Asymmetric Attention*

Alexandre N. Kohlhas  Ansgar Walther

September, 2019

Abstract

We document that the expectations of households, firms, and professional forecasters in standard surveys simultaneously extrapolate from recent events and underreact to new information. Existing models of expectation formation, whether behavioral or rational, cannot easily account for these observations. We develop a rational theory of extrapolation based on agents’ limited attention, which is consistent with this evidence. In particular, we show that limited, asymmetric attention to different structural variables can explain the co-existence of extrapolation and underreactions. Extrapolation arises when agents choose to pay less attention to countercyclical variables. We illustrate these mechanisms in a microfounded macroeconomic model, which generates expectations that are in line with the survey data, and show that asymmetric attention increases the persistence and volatility of business cycles.

JEL codes: C53, D83, D84, E32  Keywords: Expectations, information, fluctuations

1 Introduction

Given the central role of people’s expectations in economics, it is important to have a theory of expectations formation that is consistent with the data. There is reason to believe that such a theory needs to be richer than the benchmark model of full information and rational expectations. Indeed, the original proponents of rational expectations were aware of this

*First draft: January, 2018. Kohlhas: Institute for International Economic Studies (email: andre.kohlhas@iies.su.se). Walther: Imperial College London (email: a.walther@imperial.ac.uk). We are indebted to George-Marios Angeletos, Ryan Chahrou, Martin Ellison, Nicola Gennaioli, Per Krusell, Kristoffer Nimark, Laura Veldkamp, Robert Ulbricht, Venky Venkateswaran, Mirko Wiederholt, and conference and seminar participants at the AEA Annual Meeting 2019, Barcelona Summer Forum, Cambridge University, Cornell University, CEPR Summer Symposium (Gerzensee) 2017, HKUST, Mannheim University, NORMAC 2017, Oxford University, Salento Summer Meetings, and SED 2018 for their comments. This research received generous financial support from the Lamfalussy Fellowship and Ragnar Soderberg Stiftelsen.
prospect. Muth (1961) allowed for “under-discounting” in his theory, noting that people may extrapolate from current events. Lucas (1972) studied agents who observe imperfect, noisy information, and later argued that “for most agents [...] there is no reason to specialize their information systems for diagnosing general movements correctly” (Lucas, 1977, p.21).

Many recent advances in the theory of expectations formation fall into one of two frameworks. On one hand, the noisy rational expectations approach proposed by Lucas has returned to popularity following the work of Woodford (2002) and Sims (2003). On the other hand, a common view is that such rational models cannot account for people’s pervasive tendency to extrapolate from recent events, which has been documented in the survey data. The latter view favors behavioral models of expectation formation that are consistent with extrapolation. The tension between these two frameworks is important, because the outcomes and dynamics of models with behavioral biases may differ from those with noisy rational expectations. Despite the obvious importance of this issue, no consensus has been reached.

In this paper, we argue that many existing models of expectation formation, whether behavioral or rational, cannot easily account for the survey evidence. This is because they cannot account for the fact that overreactions to recent events (i.e., extrapolation) often coincide with the type of underreactions to average new information that have been pointed out by Coibion and Gorodnichenko (2015). Our main contribution is to propose a unified model of expectation formation based on noisy rational expectations that resolves the friction between theory and data, and to explore its business cycle implications.

To empirically motivate our work, we demonstrate simultaneous overreactions and underreactions in a wide range of survey data. The participants of standard surveys, reporting their expectations about future output and inflation, not only extrapolate from recent conditions, but also make smaller average forecast revisions than they would under full information.

We show that a large class of existing models, in which agents process signals of a forecasted variable (output, for concreteness), are inconsistent with such simultaneous over- and underreactions. This class includes standard behavioral models of extrapolation bias (e.g., Cutler et al., 1990; Barberis et al., 2016), simple models of noisy rational expectations as derived from models of rational inattention (e.g., Sims, 2003), as well as models that combine extrapolation bias or overconfidence with the presence of noisy information (e.g., Daniel et al., 1998; Bordalo et al., 2018). Intuitively, noisy information (or inattention) generates underreactions to new information, because individuals shrink their forecasts towards prior beliefs when the signals they observe are noisy. By contrast, extrapolation bias or overconfidence

---

1See, for example, Barberis et al. (2016), Bordalo et al. (2017), and the references therein.
2Specifically, in Section 2, we consider output and inflation forecasts from four of the most commonly used surveys on expectations: the ASA-NBER Survey of Professional Forecasters, the ECB’s Survey of Professional Forecasters, the Michigan Survey of Consumers, and the Livingstone Survey.
generates overreactions. We show that, on balance, when agents’ process signals directly of the forecasted variable, only one of these forces can dominate. This is inconsistent with the simultaneous over- and underreactions that we find in the survey data.

Our core contribution is to develop a theory of extrapolation that is based on rational updating. We consider a model of forecasters who observe noisy information due to their limited attention. The distinguishing feature of our model is that forecasters observe noisy information of the various, structural components that comprise output, instead of observing signals directly of output itself. The combination of rational updating and noisy information implies that our theory remains consistent with observed underreactions.

In our model, output is the sum of several components. For example, these components could represent different inputs into the economy’s production function, different sectors of the economy, or different variables in the economy’s dynamic Euler equation for output. A population of forecasters observe a vector of noisy signals, where each signal contains information about a particular component. We think of attention to each component as the precision of the associated signal. Importantly, attention can be higher for some components than for others. We say that attention is asymmetric if agents receive a relatively more precise signal about some components. In this environment, we derive two main results.

The first main result is that asymmetric attention can explain the co-existence of extrapolation and underreactions, as long as attention centers on procyclical components. Consider an economy in which output is driven by only two components, which differ in their behavior over the business cycle. The first component is procyclical, while the second is countercyclical. Suppose that agents pay more attention to the procyclical component. Then, compared to the full-information benchmark, agents become more optimistic in booms and more pessimistic in busts, even though they adhere to Bayes’ rule. As a result, the measured overreactions to recent output in the survey data can be viewed as an outcome of underreactions to countercyclical components. In addition, as long as agents’ attention to the procyclical component remains imperfect, they still exhibit underreaction to new information on average, due to their rationally muted responses to noisy information. We extend this reasoning to a canonical forecasting problem with an arbitrary number of components. An auxiliary proposition generalizes our results to a comprehensive class of linear models.

Our second main result concerns the possible sources of asymmetric attention. In principle, asymmetric attention could arise from behavioral heuristics or salience effects (Gabaix, 2017). Notwithstanding such alternatives, we show that asymmetric attention arises naturally in a rational framework, in which agents optimally choose how to allocate costly attention. With standard attention cost functions, agents in our framework find it optimal to pay asymmetric attention to the components of economy-wide output that are either particularly volatile, or
particularly important for their decision-making. For example, consider a firm who reports its expectation about future output. In line with the conclusions in Lucas (1977), this firm has an incentive to focus its attention on the components of output that correlate closely with its own local conditions, especially if these components are also particularly volatile. Coibion et al. (2018) provide direct evidence of these incentives at work, using detailed firm-level data to show that firms indeed pay asymmetric attention to volatile variables that are also more important for their decision-making.

Combining our two results, we conclude that a rational model of limited attention can explain the simultaneous extrapolation and underreaction of survey expectations, as long as the volatile or important components of output that attract attention are also procyclical. This connects our results to those of Woodford (2002), Nimark (2008), and Angeletos and Huo (2019), among others, who argue that limited attention can account for the myopia and anchoring to past outcomes often documented in macroeconomics. We demonstrate that models of limited attention also have the potential to be consistent with extrapolation.

We show that an additional testable implication of our explanation, in terms of the aggregate data, is that expectations should be more precise than pure time series forecasts (e.g., forecasts from ARIMA models). Consistent with this prediction, we update estimates from Stark (2010) to show that forecasters’ survey expectations of output growth consistently outperform simple time series models, especially at short horizons.

To explore the implications of our framework, and to provide an example of the sources of asymmetric attention, we apply our framework to a standard macroeconomic model with flexible prices. In the model, firms choose output under imperfect information about productivity. We show that, in equilibrium, firms’ output choices can be split into two components: (i) firm beliefs about a local component, which reflects their own productivity; and (ii) firm beliefs about an economy-wide component, which summarizes the equilibrium effect of other agents’ choices on individual firm output. Maćkowiak and Wiederholt (2009) propose a closely related decomposition. When we sum across firms, aggregate output thus becomes the simple sum of the two components.

We show that, for standard parameter values, two key conditions are satisfied: First, the local component is procyclical, while the economy-wide component is countercyclical. The latter follows because economy-wide expansions tend to increase firms’ costs, leading each individual firm to reduce its output relative to its partial equilibrium choice. Second, if attention is costly, firms optimally choose to pay asymmetric attention to local conditions, because the local component is more volatile. As a result of these two conditions, and in line with our two main results, firms’ expectations of future aggregate output exhibit both extrapolation and underreactions to recent forecast revisions, relative to the full information
benchmark. This is qualitatively consistent with the survey evidence. The model also fits the empirical size of these effects well.

We use the macroeconomic model to explore the business cycle implications of firms’ asymmetric attention choices. We show that asymmetric attention to local components leads to more persistence and volatility in aggregate output than an equivalent model with symmetric attention. We further document that the calibrated model can match the observed increase in extrapolation post-Great Moderation, and argue that firms’ optimal attention choices may have contributed to the increased persistence of output growth during this period.

Finally, two wider implications of our analysis are worth noting. First, in the tradition of Lucas (1977), our macroeconomic model focuses on a lack of attention to equilibrium effects as the driver of extrapolation. As such, our results speak to a literature in behavioral finance, which models the neglect of equilibrium effects as fundamental behavior, and uses this to account for investment patterns (e.g. Greenwood and Hanson, 2014).

Second, motivated by the survey evidence, we focus on a setting in which agents’ forecasts appear to overreact to a particular type of public information (i.e., recent realizations of the forecasted variable). However, as we illustrate, a model of asymmetric attention may be equally consistent with underreactions to other types of public information, depending on how this information correlates with the variables to which agents pay attention.\footnote{Underreactions to public information are documented, for example, in Barberis et al. (1998), Daniel et al. (1998). Eyster et al. (2019) review further related evidence.} We therefore view this paper, more generally, as taking a first step towards integrating observed over- and underreactions to new information into a unified, rational framework.

**Related literature:** In addition to the literature cited above, this paper relates to four strands of research. We review these in reverse chronological order, starting with the most recent and ending with the long history of thought on extrapolative and adaptive expectations.

First, our paper reconciles overreactions to a specific public signal (recent outcomes of the forecasted variable) with underreactions in survey expectations. We define these phenomena in terms of average and individual expectations. In contemporaneous work, Bordalo et al. (2018) propose a behavioral model that can reconcile underreactions in aggregate forecast revisions with overreactions in individual forecast revisions (see also Broer and Kohlhas, 2019). However, as we demonstrate in Section 2, simple versions of their framework cannot account for the simultaneous over- and underreactions of expectations that we document in the data. We therefore view these two papers as related and complementary steps towards a unified model of expectations that is consistent with over- and underreactions to new information.

Second, in common with a vast literature in macroeconomics since Lucas (1972), we em-
phasize the importance of imperfect information for business cycle dynamics. Prominent studies, among many others, are Woodford (2002), Mankiw and Reis (2002), Lorenzoni (2009), Blanchard et al. (2013), Angeletos and La’O (2013), Maćkowiak and Wiederholt (2015), and Chahrour and Ulbricht (2018). We emphasize the role of agents who optimally choose how to allocate their scarce attention, and we build on the complementary literatures on “optimal information choice” (e.g. Veldkamp, 2011; Hellwig et al., 2012) and “rational inattention” (e.g., Sims, 2003; Maćkowiak and Wiederholt, 2009; Wiederholt, 2010). The contribution of our paper, in this context, is to highlight that models of imperfect information can be consistent with the observed overreactions in the survey data.

Third, we leverage the existing evidence on survey expectations. Pesaran (1987) summarizes the early evidence on deviations from full information and rational expectations, and Zarnowitz (1985) shows that survey data is consistent with models of noisy, private (instead of common, perfect) information. Relatedly, Ehrbeck and Waldmann (1996) explore the sources of bias in professional forecasts and conclude that these are unlikely to derive from agency-based considerations. More recently, Coibion and Gorodnichenko (2012; 2015) demonstrate underreactions in average forecast revisions (see also Andrade and Le Bihan, 2013, and Fuhrer, 2017), which form part of the motivation for this paper.

Finally, our focus on overreactions to recent outcomes connects this paper to the literature on adaptive and extrapolative beliefs. This includes the early work of Goodwin (1947), Cagan (1956) and Muth (1961), the experimental work on the psychology of subjective probabilities as explored by Kahneman and Tversky (1972) and Andreassen and Kraus (1988), and the modern treatments of extrapolation by DeLong et al. (1990), Cutler et al. (1990), Fuster et al. (2012), Greenwood and Shleifer (2014), Barberis et al. (2016), and Bordalo et al. (2017). This paper is the first, to our knowledge, to combine the empirical insights of this literature with a model that can also generate underreactions in aggregate expectations.

2 Motivating Evidence and Existing Theory

In this section, we revisit two simple tests of rational expectations. We document a new stylized fact: Participants’ expectations in standard surveys simultaneously overreact to recent realizations of the forecasted variable (i.e., extrapolate from recent events), but underreact in their forecast revisions. We then derive the predictions of a large class of existing models and argue that these cannot easily account for this observation.
2.1 Simultaneous Over- and Underreactions

We start by considering forecasts of US output growth from the Survey of Professional Forecasters (SPF). The SPF is a survey of between 20-100 professional forecasters and is conducted quarterly by the Federal Reserve Bank of Philadelphia. Real GDP/GNP growth estimates are available from 1968:Q4 at a quarterly frequency. We focus on output forecasts for two reasons. First, because expectations about future output play a central role in the economy as determinants of consumption, inflation, and asset prices. Second, because data on output forecasts are available for a longer time-span than forecasts of most other variables. We later show how our empirical results extend to other countries, variables, and data sources.

We let $y_{t+k}$ denote year-on-year output growth at time $t+k$, and let $\bar{f}_t y_{t+k}$ denote the average forecast from a sample of respondents of $y_{t+k}$ at time $t$. The average forecast error is $y_{t+k} - \bar{f}_t y_{t+k}$. A negative forecast error thus corresponds to an over-estimate of $y_{t+k}$. A well-known implication of full information and rational expectations (FIRE) is that forecast errors should be unpredictable. Under FIRE, no variable that is observable at time $t$ should correlate with $y_{t+k} - \bar{f}_t y_{t+k}$. We rely on two common tests of this prediction.

The first test is a regression of forecast errors on current output growth,

$$y_{t+k} - \bar{f}_t y_{t+k} = \alpha + \gamma y_t + \xi_t,$$

where $\alpha$ is constant and $\xi_t$ is an error term. The second test, proposed recently by Coibion and Gorodnichenko (2015), is a regression of forecast errors on past forecast revisions,

$$y_{t+k} - \bar{f}_t y_{t+k} = \alpha + \delta (\bar{f}_t y_{t+k} - \bar{f}_{t-1} y_{t+k}) + \xi_t.$$

The term $\bar{f}_t y_{t+k} - \bar{f}_{t-1} y_{t+k}$ on the right-hand side is the average change in forecasts when respondents are asked twice (at dates $t-1$ and $t$) to forecast the same future realization $y_{t+k}$. A positive revision arises when good news about future output arrives between $t-1$ and $t$.

The prediction of the FIRE benchmark is that the coefficients $\gamma$ and $\delta$ in (1) and (2) should both be zero, because both current output growth and the latest forecast revision are observable at time $t$. The raw data already hints at deviations from this benchmark. Figures 1 and 2 plot one-year-ahead forecast errors (the left-hand side of (1) and (2), with $k = 4$) over time, and compares them, respectively, to current realizations of output growth.

---

4 The SPF is the oldest quarterly survey of individual macroeconomic forecasts in the US, dating back to 1968. The SPF was initiated under the leadership of Arnold Zarnowitz at the American Statistical Association and the National Bureau of Economic Research, which is why it is also still often referred to as the ASA-NBER Quarterly Economic Outlook Survey (Croushore, 1993).

5 We use real-time data to measure current realizations of output growth to precisely capture the definition of the output variable being forecasted (Croushore, 1998).
In Figure 1, forecasts are frequently over-optimistic, with associated negative forecast errors when the current level of output growth is high, and vice versa when current growth is low. This suggests that forecasters extrapolate from recent events; agents are systematically too optimistic in booms and too pessimistic in busts. Figure 2, by contrast, suggests that forecast errors and current forecast revisions are positively correlated within our sample. All else equal, this indicates that agents underreact to new information on average, because they are too pessimistic after positive forecast revisions, and vice versa after negative revisions.

Table I confirms these impressions and reports estimates of (1) and (2) using the SPF data on one-year-ahead forecasts \((k = 4)\). In the first column, we estimate (1) and find that \(\gamma\) is negative and statistically significant. This once more suggests extrapolation, or overreactions to recent realizations of the forecasted variable. In the second column, we estimate (2) using one-quarter forecast revisions. We find that \(\delta\) is positive and significant, which is consistent with average forecast revisions underreacting to overall new information received within the period. The third column confirms these results in a multiple regression.

Taken individually, the over- and underreactions documented in Table I are in line with previous estimates (see, for example, Bordalo et al., 2017 for evidence on extrapolation, and Coibion and Gorodnichenko, 2015 for evidence on underreactions). Our results demonstrate
that extrapolation and underreactions occur simultaneously in the SPF data.

These patterns are stable across countries, across surveys of agents who are not professional forecasters, and across macroeconomic variables other than output. Figure 3 summarizes estimates of (1) and (2) for output forecasts from the Euro Area SPF, the Livingstone Survey (which covers academic institutions, investment banks, non-financial firms, and government agencies), and the Michigan Survey of Consumers. We also include estimates for inflation forecasts. We plot the coefficient $\gamma$ on current realizations in (1) on the horizontal axis in Figure 3, and the coefficient $\delta$ on forecast revisions in (2) on the vertical axis.

All of our estimates fall into the upper-left quadrant of the figure, where we simultaneously find that $\gamma < 0$ (overreaction) and that $\delta > 0$ (underreaction). Appendix A.2 and A.3 contain the associated regression results and further robustness checks. Specifically, with the exception of the Euro Area and Livingstone inflation forecasts, all coefficients in Figure 3 are statistically significant at the five percent level. We further show that such simultaneous over- and under-

---

6The *Livingstone Survey* is a semi-annual survey that started in 1946 (Croushore, 1997). The *Michigan Survey of Consumers* contains consumers’ inflation forecasts. A drawback of the monthly Michigan Survey of Consumers is that only one-year ahead forecasts of consumer price inflation are available. Revisions to forecasts at a fixed horizon cannot be constructed. To estimate (2), we therefore follow Coibion and Gorodnichenko (2015) and replace *ex-ante forecast revisions* with the quarterly *ex-ante forecast changes* and instrument this variable with the (log) oil price change. This approach provides an asymptotically consistent estimate. The *Euro Area’s Survey of Professional Forecasters* collects the same information as the SPF for the US.
Table I: Estimated Over- and Underreactions in the SPF

<table>
<thead>
<tr>
<th></th>
<th>Forecast Error</th>
<th>Forecast Error</th>
<th>Forecast Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.07</td>
<td>-0.10</td>
<td>0.27*</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.11)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>Current Realization</td>
<td>-0.12**</td>
<td>–</td>
<td>-0.14***</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
</tr>
<tr>
<td>Forecast Revision</td>
<td></td>
<td>0.77***</td>
<td>0.80***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.27)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>$F^2$</td>
<td>5.50</td>
<td>15.2</td>
<td>11.8</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.03</td>
<td>0.08</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Note: HAC standard errors in parentheses. Significance levels *=10%, **=5%, ***=1%.

Figure 3: Estimated Over- and Underreactions Across Surveys

Note: Estimates of $\gamma$ and $\delta$ from (1) and (2). US SPF represents the estimates for the US Survey of Professional Forecasters, EA SPF the ECB’s Survey of Professional Forecasters, LS Survey the Livingstone Survey, and last MSC the Michigan Survey of Consumers. □ = GDP forecasts, ◆ = Inflation forecasts, and ◇ = MSC inflation forecasts that have been instrumented. All estimates are for one-year ahead forecasts, and estimates of (2) use semi-annual revisions (Livingstone Survey) or one-quarter revisions (all others). A HP-trend ($\lambda = 1,600$) has been deducted from recent inflation and output in this figure to account for potential structural changes.
reactions extend to other forecast horizons than one-year-ahead, that they are present both pre- and post-Great Moderation, and that they are robust to different assumptions about the potential for structural trends in the data. Lastly, while our results focus on average forecasts $f_t y_{t+k}$, we show that extrapolation is also prominent at the individual level. As we will discuss below, our proposed theoretical framework is also consistent with this observation.

In summary, the results in Table I and Figure 3 document pervasive overreactions to recent realizations of the forecasted variable (i.e. extrapolation), but simultaneous underreactions in forecast revisions. This clearly constitutes a rejection of the FIRE benchmark. However, as we now argue, we can also use these stylized facts to determine whether existing alternative theories of expectation formation are consistent with the data.

2.2 Existing Theories of Expectation Formation

We consider a parsimonious framework that captures several alternative models of expectation formation. There is a continuum of measure one of agents who make forecasts of future output $y_{t+k}$. We assume that output $y_t$ follows the autoregressive process:

$$y_t = \rho y_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2)$$

(3)

where $\rho \in (0, 1)$ and $u_t$ is serially uncorrelated. At the start of each period, each agent $i \in [0, 1]$ observes a noisy signal of current output,

$$z_{it} = y_t + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0, \sigma_{it}^2)$$

(4)

where $\sigma_{it}^2$ denotes the variance of the type-specific signal and $\epsilon_{it}$ is independent of $u_t$ at all horizons with Cov $[\epsilon_{it}, \epsilon_{js}] = 0$ for all $i \neq j$ and $t \neq s$. We write $v^t = \{v_s\}_{s \leq t}$ for the history of any stochastic process $v_t$ up to date $t$. Agent $i$’s information set at time $t$ is $\Omega_{it} = \{z_{it}\}$.\textsuperscript{7}

We assume that agents’ forecasts follow a Generalized Kalman Filter. Let $f_{it} y_{t+k}$ and $f_{it-1} y_{t+k}$ denote agent $i$’s forecasts of future output at dates $t$ and $t - 1$, respectively, and let $f_{it-1} z_{it}$ be her forecast of her own signal one period ahead. Agent $i$’s output forecast then follows the updating equation:

$$f_{it} y_{t+k} = f_{it-1} y_{t+k} + g_k (z_{it} - f_{it-1} z_{it})$$

(5)

As with the textbook Kalman Filter, the agent starts with her forecast of output at time $t - 1$, and updates it in proportion to the unexpected innovation $z_{it} - f_{it-1} z_{it}$ in her signal at time $t$.

\textsuperscript{7}We assume that all agents receive a signal $z_{0i}$ in period zero that initializes their beliefs in the steady state of their respective filter. Maćkowiak \textit{et al.} (2018), for example, employ a similar approach.
Departing from the standard filter, we allow $g_k \in (0, 1)$ to be an arbitrary gain parameter that measures agents’ responsiveness to new information. Despite its simplicity, the formulation in (5) nests a wide range of models that have been considered in the literature. We demonstrate this through a series of examples, which we delineate into rational and behavioral theories.

**Rational Expectations:** Suppose that agents form their expectations in accordance with Bayes’ rule. In this case, $f_{it}y_{t+k}$ equals the conditional expectation of $y_{t+k}$ based on $\Omega_{it}$, $f_{it}y_{t+k} = \mathbb{E}[y_{t+k} \mid \Omega_{it}]$, and the gain parameter satisfies $g_k = \text{Cov}_{t-1}[y_{t+k}, z_{it}] / \mathbb{V}_{t-1}[z_{it}]$. This captures two important sub-cases:

1. **FIRE:** When agents perfectly observe output ($\sigma_e = 0$), their forecasts reduce to the FIRE benchmark with $g_k = \rho^k$. Agents’ forecasts collapse to $f_{it}y_{t+k} = \mathbb{E}[y_{t+k} \mid y^t] = \rho^k y_t$.

2. **Noisy RE (Limited Attention):** By contrast, when agents observe a noisy signal of output ($\sigma_e > 0$), their forecasts $f_{it}y_{t+k}$ become identical to those from models with noisy rational expectations (Woodford, 2002), or rational inattention (Sims, 2003). The gain coefficient in (5) now satisfies $g_k < \rho^k$, as agents rationally shrink their forecasts towards prior beliefs when their information is noisy.

**Behavioral Expectations:** A common way to model behavioral biases is to assume agents perceive the data-generating process to be different from its true parametrization, but then update correctly under this wrong model. The formulation in (5) nests several of such cases:

3. **Extrapolation Bias:** Agents perfectly observe current output ($\sigma_e = 0$), but perceive its persistence parameter to be $\hat{\rho} > \rho$. As a result, individuals are in effect overconfident in the extent to which current output predicts future realizations. Agents forecasts in (5) reduce to $f_{it}y_{t+k} = g_k y_t$, where $g_k = \hat{\rho}^k > \rho^k$.

4. **Overconfidence:** Equation (5) also captures a related model of overconfidence, in which agents correctly perceive the persistence of output but believe the noise in their information to be $\hat{\sigma_e} < \sigma_e$ (Daniel et al., 1998; Hirshleifer et al., 2011). Agents forecasts are then effectively based on a Kalman Filter with a gain $g_k$ that exceeds its rational value.

---

8A more comprehensive list of papers in this tradition is in the introduction. The Gaussian signal $z_{it}$ we have specified is optimal in a rational inattention setting if agents minimize their squared forecast errors and their cost of processing information is based on the reduction in entropy (see Maćkowiak et al., 2018).

9See Fuster et al. (2012), for example. The introduction contains a list of references on extrapolation.

10This follows from: $f_{it}y_{t+k} = \hat{\rho}^{1+k}y_{t-1} + \hat{\rho}^k (y_t - \hat{\rho} y_{t-1}) = \hat{\rho}^k y_t$.

11For further analysis of overconfidence, see Broer and Kohlhas (2019) and the references therein.
5. Diagnostic Expectations: The model in Bordalo et al. (2017) and Bordalo et al. (2018) corresponds to the overconfidence case, but where the prior forecasts $f_{it-1}y_{t+k}$ and $f_{it-1}z_{it}$ in (5) are reset to their rational, full information values in each period.\footnote{That is, $f_{it-1}y_{t+k}$ and $f_{it-1}z_{it}$ are set to $E[y_{t+k} | y^{t-1}]$ and $E[z_{it} | y^{t-1}]$, respectively, in each period.}

We now characterize the results that an econometrician would obtain in the above cases when running (1) and (2), assuming that the true data-generating process satisfies (3) to (5).

**Proposition 1.** If agents form their expectations according to (5), based on signals in (4). Then, the coefficient $\gamma$ in (1) and $\delta$ in (2) both have the same sign as $(\rho^k - g_k)$.

To interpret Proposition 1, recall that agents’ gain parameter in the FIRE case, where they perfectly observes current output, is equal to $g_k = \rho^k$ (since $f_{it}y_{t+k} = \rho^k y_t$). Proposition 1 states that there are two possible parametric regions, depending on whether agents’ beliefs are more or less responsive to current output than under FIRE.

First, if the gain parameter on new information is lower than in the FIRE benchmark, with $g_k < \rho^k$, then the econometrician will measure $\gamma > 0$ and $\delta > 0$. His estimates will fall into the upper-right quadrant of Figure 3. Individuals in this case underreact to new information compared to the FIRE benchmark. For instance, consider the rational model with limited attention (Example 2 above). In this example, the measured underreactions are driven by two forces. On one hand, rational agents respond to new information in a muted fashion because they realize that their signals are noisy. On the other hand, the econometrician has access to strictly better information than any individual agent: In (1), he controls for a noiseless version of current output, which no individual agent observes. In (2), he controls for the average forecast revision, which is more informative than the individual revisions observed by agents, because the noise in agents’ signals cancels on average. This superiority of the econometrician’s information implies that forecast errors are predictable, while agents’ muted responses to new information imply that this predictability takes the form of measured underreactions. We therefore obtain $\gamma > 0$ in (1) and $\delta > 0$ in (2).

Second, by contrast, if the responsiveness to new information is higher than under FIRE, with $g_k > \rho^k$, then the econometrician will measure $\gamma < 0$ and $\delta < 0$. His estimates will fall into the lower-left quadrant of Figure 3. This corresponds to the case in which agents overreact to both recent data and with their forecast revisions. We obtain such a situation, for example, in the case with extrapolation bias (Example 3 above).

In all of the above cases, agents’ deviations from full information and rational expectations enter their beliefs through a single sufficient statistic, namely, the gain $g_k$ in the Generalized Kalman Filter. Two counteracting effects determine the size of $g_k$. First, the presence of noisy information dampens agents’ responses to new information, even if they are rational.
This decreases the gain coefficient below its FIRE value. Second, behavioral biases, such as extrapolation or overconfidence, heighten agents’ responsiveness, which in turn increases the gain coefficient. However, as Proposition 1 shows, only one of these forces can come to dominate the sufficient statistic \( g_k \). Hence, in all of the above cases, agents either over- or underreact, but do not over- and underreact simultaneously. This is inconsistent with the evidence summarized in Figure 3, where all of the estimates fall in the upper-left quadrant.

Furthermore, and for the same reason, a simple model with heterogeneous expectation formation among agents is also inconsistent with our estimates. In an economy with heterogeneous types of forecasters, who have different degrees of behavioral biases or limited attention, the Generalized Kalman gain \( g_k \) in our formulation can be reinterpreted as the weighted average of each type’s response to new information. Hence, average forecasts will either over- or underreact, but cannot do so at the same time.

In conclusion, a large class of existing models are inconsistent with the simultaneous over- and underreactions that we find in the data. This motivates the development of our model in the next section. In particular, it is clear that to explain the survey data, we must consider a model with more than one sufficient statistic for belief formation. We achieve this by proposing a model in which output is the sum of various components, and by allowing attention to be asymmetric across these.

3 Asymmetric Attention

In this section, we consider a rational model of limited attention. The central difference to the standard model from the previous section is that we view output as comprised of a set of structural components. We show that the over- and underreactions that we have documented can be rationalized if agents pay more attention to some components than others; that is if agents’ attention is asymmetric. Our approach in this section is to take attention choices as given and derive conditions under which the model can account for our empirical results. In the next section, we then examine the possible sources of asymmetric attention. In Appendix B.2, we show how our results generalize to a canonical class of linear models.

3.1 Environment

A continuum of measure one of agents are asked to forecast future output \( y_{t+k} \). Aggregate output \( y_t \) is driven by the sum of \( N \) structural components \( x_{jt} \)

\[
y_t = x_{1t} + x_{2t} + \ldots + x_{Nt}.
\]
These components could, for example, represent different inputs into the economy’s production function, different sectors of the economy, or different variables in firms’ optimal production plans. We discuss one such example at length in Section 5. Each component $x_{jt}$ is determined by the linear relationship

$$x_{jt} = a_j \theta_t + b_j u_{jt}, \quad u_{jt} \sim \mathcal{N}(0, 1),$$

(7)

where $\theta_t$ denotes a persistent fundamental that follows the autoregressive process

$$\theta_t = \rho \theta_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, \tau_{\eta}^{-1}),$$

(8)

with $\rho \in (0, 1)$. The error terms $u_{jt}$ and $\eta_t$ are serially uncorrelated, mutually independent, and it is common knowledge that $\theta_1 \sim \mathcal{N}(0, \tau_{\theta}^{-1})$. As a result, each component depends both on a common fundamental $\theta_t$ and on a transitory, component-specific shock $u_{jt}$.

The output response to a positive fundamental shock $d\theta_t > 0$ is

$$\frac{dy_t}{d\theta_t} = \sum_j a_j.$$ \[15\]

We assume that $\sum_j a_j > 0$ without loss of generality, so that output correlates positively with $\theta_t$. The contribution of component $x_{jt}$ to this output response is $a_j$. We refer to a component $x_{jt}$ as procyclical if $a_j > 0$, so that $x_{jt}$ reinforces the response of output to the persistent fundamental. Analogously, we say that $x_{jt}$ is countercyclical if it dampens the response with $a_j < 0$.

Output and its components are not directly observable to agents, because of their limited attention. Instead, each agent $i \in [0, 1]$ observes the history of $N$ noisy signals

$$z_{ijt} = x_{jt} + q_j \epsilon_{ijt}, \quad \epsilon_{ijt} \sim \mathcal{N}(0, 1), \quad j = \{1, 2, \ldots, N\},$$

(9)

where $q_j$ parameterizes the amount of noise (or inattention) in agents’ signals about the $j$th structural component, and $\epsilon_{ijt}$ is an idiosyncratic error term.\[13\] Agent $i$’s information set at time $t$ is thus $\Omega_{it} = \{z_{i1t}, \ldots, z_{iNt}\}$.

Notice that there are two key differences between this environment and that in Section 2, which also nested a rational case with noisy signals (see Example 2 in that section). First, output is determined by several underlying components. Second, agents learn about these components separately: The information structure in (9) restrict agents to observing conditionally independent signals of each component. This formalizes the idea that paying attention to one component is a separate activity from paying attention to another. Combined, these features capture the notion that, to form expectations, individuals first need to pay attention to information about the various components of the forecasted variable, and then combine these different pieces of information into a single prediction. The conditional independence

\[13\] We set $\tau_\theta$ so that agents’ signal extraction problem is initialized in steady state (Maćkowiak et al., 2018).
embedded in (9), combined with a component-based structure in (6), is a simple and common
way to model this idea (see e.g. Maćkowiak and Wiederholt, 2009). We discuss the role of
these restrictions in more detail in Section 4, where we also consider an alternative setup with
fully flexible information design.

3.2 Definition of Attention

To characterize agents’ attention to the various structural components, we transform the noise
parameters \( q_j \) in (9) into the normalized parameters

\[
m_j \equiv \frac{\text{Var}(x_{jt}|\theta_t)}{\text{Var}(z_{ijt}|\theta_t)} = \frac{b_j^2}{b_j^2 + q_j^2} \in (0, 1).
\]

These parameters measure the sensitivity of agents’ beliefs to new information about the
\( j \)th structural component. Suppose that agent \( i \) knows \( \theta_t \), and is then asked to predict component
\( x_{jt} \) based on her own noisy signal \( z_{ijt} \). Her estimate will be:

\[
E[x_{jt}|z_{ijt}, \theta_t] = m_j z_{ijt} + (1 - m_j) E[x_{jt}|\theta_t].
\]

If \( m_j = 0 \) (i.e., if the noise parameter \( q_j \to \infty \)), then the agent has no new information about
\( x_{jt} \) and sticks to her prior \( E[x_{jt}|\theta_t] \) when observing \( z_{ijt} \). By contrast, if \( m_j = 1 \) (i.e., if the
noise parameter \( q_j = 0 \)), then the agent perfectly observes \( x_{jt} \) and ignores her own prior in
her expectation of \( x_{jt} \). In this sense, \( m_j \) captures how much information agents obtain about
the \( j \)th component. We therefore call \( m_j \) the attention dedicated to the \( j \)th component.

3.3 Attention, Overreactions, and Underreactions

We now derive the coefficients for extrapolation in (1) and underreaction in (2) that an econo-
metrician would estimate for this economy. The coefficient on current output in (1) satisfies:

\[
\gamma = \text{Cov} \left[ y_{t+k} - \bar{E}_t y_{t+k}, y_t \right] \text{Var}[y_t]^{-1} = d_0 \text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, y_t \right],
\]

where \( d_0 = (\rho^k \sum_j a_j) \text{Var}[y_t]^{-1} > 0 \), and \( \bar{E}_t y_{t+k} = \bar{f}_t y_{t+k} \) denotes the average \( k \)-period ahead
forecast of output. Since agents are rational, this forecast is equal to the average conditional
expectation. The equality in (11) follows because \( y_{t+k} \) depends only on \( \theta_t \) and on shocks that
are uncorrelated with date-\( t \) information. We note that the sign of \( \gamma \) is determined only by
the covariance between the average tracking error \( \theta_t - \bar{E}_t \theta_t \) and current output.

\[\text{We assume that all individuals make the same attention allocations } m_j. \text{ This is true in our model of}
\text{optimal attention choice in Section 4 and 5. It is also a standard assumption in the information choice}
\text{literature (see, for example, Veldkamp, 2011).} \]
Meanwhile, the coefficient $\delta$ on the latest forecast revision in (2) is:

$$
\delta = \text{Cov} \left[ y_{t+k} - \bar{E}_t y_{t+k}, \bar{E}_t y_{t+k} - \bar{E}_{t-1} y_{t+k} \right] \text{Var} \left[ \bar{E}_t y_{t+k} - \bar{E}_{t-1} y_{t+k} \right]^{-1} \\
= d_1 \text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, \bar{E}_t \theta_t - \bar{E}_{t-1} \theta_t \right],
$$

(12)

where $d_1 = \left( \rho^k \sum_j a_j \right)^2 \text{Var} \left[ \bar{E}_t y_{t+k} - \bar{E}_{t-1} y_{t+k} \right]^{-1} > 0$. Hence, the sign of $\delta$ is determined only by the covariance between the average tracking error of $\theta_t$ and the latest forecast revision.

We start with two stark examples that demonstrate how the two covariances in (11) and (12) depend on individuals’ attention choices. This, in turn, allows us to provide a simple illustration of the mechanisms behind our main results.

**Example 1. Asymmetric Attention and Extrapolation:** Suppose that output has two components with $y_t = x_{1t} + x_{2t}$, and that the first component is procyclical with $a_1 > 0$. Agents pay full attention to the first component and none to the second ($m_1 = 1$, $m_2 = 0$). Then, the extrapolation coefficient in (11) becomes

$$
\gamma = d_0 \text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, x_{1t} + x_{2t} \right] = d_0 \text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, x_{2t} \right] = d_0 \text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, a_2 \theta_t \right] = a_2 d_0 \text{Var} \left[ \theta_t | \Omega_{it} \right],
$$

where the first equality follows from $\text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, x_{1t} \right] = 0$ for all agents $i \in [0, 1]$, because each agent is fully rational and observes $x_{1t}$ perfectly. The second equality follows from $\text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, x_{2t} \right] = a_2 \text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, \theta_t \right]$, while the third one is due to individual rationality implying $\text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, \theta_t \right] = \text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, \theta_t - \bar{E}_t \theta_t \right]$. We conclude that $\gamma = a_2 d_0 \text{Var} \left[ \theta_t | \Omega_{it} \right]$, and thus that the extrapolation coefficient $\gamma$ has the same sign as $a_2$. $\square$

In this example, the econometrician will find extrapolation, i.e. overreactions to current output ($\gamma < 0$), if and only if $a_2 < 0$; that is, if and only if the component $x_{2t}$, to which agents pay no attention, is countercyclical. This highlights how our rational model can generate overreactions. In effect, the example shows that the overreaction to recent output documented in the survey data can be interpreted as an underreaction to countercyclical components.

The economic intuition behind this fact, which captures one of the key ideas of this paper, is as follows: When output $y_t$ is high, the procyclical component $x_{1t}$, all else equal, also tends to be high, which is good news about the persistent fundamental $\theta_t$. However, the countercyclical component $x_{2t}$, on average, also tends to be large, which dampens any good news about $\theta_t$. When agents pay relatively less attention to countercyclical components, their posteriors place only a small weight on this dampening effect. As a result, when output is high, agents tend to be more optimistic than the econometrician (who controls for total output) about the future.
This leads to a seeming extrapolation, which manifests itself in a negative correlation between future forecast errors and current output.

Our second example shows that our environment, despite such overreactions, remains consistent with the underreactions documented in Section 2.

Example 2. Limited Attention and Underreactions: Consider the setting in Example 1, but now suppose that agents’ attention to the first component of output is also limited: $0 < m_1 < 1$. Agent $i$’s forecast revision in (12) obeys the Kalman Filter: $\mathbb{E}_{it}\theta_t - \mathbb{E}_{it-1}\theta_t = g(z_{1it} - \mathbb{E}_{it-1}z_{1it})$, where $g = \text{Cov}_{t-1}\theta_t, z_{1it}] / \text{Var}_{t-1}\theta_t$ denotes the rational Kalman Gain. For comparison, let $g^* = \text{Cov}_{t-1}\theta_t, x_{1it}] / \text{Var}_{t-1}\theta_t$ denote the rational Kalman Gain of an agent who perfectly observes $x_{1it}$. It is easy to see that $0 < g < g^*$. The first inequality follows because $x_{1it}$ is procyclical, while the second follows from $m_1 < 1$ (an agent with limited attention accounts for the noise in her signal and hence becomes less responsive to it). Inserting individuals’ forecast revision about $\theta_t$ into (12), summing across $i$, and combining terms then provides us with:

$$\delta = (g^* - g)d_1\text{Var} [\theta_t - \bar{f}_{t-1}\theta_t] g > 0.$$

This second example shows that the econometrician will estimate underreactions to average forecast revisions ($\delta > 0$) when agents’ attention to at least one component is limited. This is familiar from Coibion and Gorodnichenko (2015) and Proposition 1, and the intuition is as discussed above. As long as information is imperfect, rational individuals attach some weight to their prior beliefs. Consequently, they respond less strongly to new information than agents in the fully-informed rational benchmark, which leads to the underreaction of aggregate expectations documented in the survey data.

Combined, the above examples have shown how attention choices map into the over- and underreaction coefficients $\gamma$ and $\delta$, respectively. Specifically, they have shown how limited, asymmetric attention to a procyclical component can explain the simultaneous over- and underreactions of survey expectations ($\gamma < 0$ and $\delta > 0$). Using similar steps, Proposition 2 extends our results to the general case with $N$ components and arbitrary attention choices.

**Proposition 2.** Output forecasts overreact to current output ($\gamma < 0$ in (1)) if and only if agents pay asymmetric attention to procyclical components, so that $\sum_j a_j(1 - m_j) < 0$. Output forecasts underreact to new information on average ($\delta > 0$ in (2)) if and only if attention is limited, i.e. if there exists $j \in \{1, \ldots, N\}$ such that $0 < m_j < 1$.

The first part of the proposition states the key sufficient statistic: $\sum_j a_j(1 - m_j)$. Our model is consistent with overreactions to current output (extrapolation) whenever this statistic is
negative. This is clearly the case when agents are inattentive \((m_j \approx 0)\) to components that are countercyclical \((a_j < 0)\), and are more attentive to procyclical components. Thus, asymmetric attention to procyclical components is a sufficient condition for extrapolation \((\gamma < 0)\).

The proposition further implies that asymmetric attention is also a necessary condition for extrapolation. If attention were symmetric with \(m_j \equiv \bar{m}\) for all \(j\), then we would have \(\sum_j a_j (1 - \bar{m}) \geq 0\), since \(\sum_j a_j > 0\), and hence, \(\gamma \geq 0\). Intuitively, the symmetric case is similar to the rational benchmark with noisy information about output (Example 1 in Section 2), where rational updating induces underreactions in both \((1)\) and \((2)\). Hence, the symmetric case is inconsistent with the large body of evidence documenting extrapolation.

The second part of the proposition extends the results of Coibion and Gorodnichenko (2015) to our framework. We find that underreactions to new information occur whenever attention is limited for at least one component.

In Appendix B.2, we further show that the underlying logic behind Proposition 2 extends to a large class of models, which encompasses most linear macroeconomic models. In that extension (see Proposition 5), the condition for underreactions remains the same, while the condition for extrapolation is adjusted to allow for \((i)\) the direct effects of several, persistent fundamentals on output, \((ii)\) for the correlation between component-specific shocks, and \((iii)\) for the explicit observation of lagged outcomes. We use some of these extensions in Section 5.

### 3.4 Summary and Further Implications

In summary, our model is able to match the stylized facts whenever attention is both limited and asymmetric. We close this section with two additional observations.

First, while we have focused on model predictions about average forecasts, our model generates extrapolative expectations at the individual level. Indeed, Example 1 and the proof of Proposition 2 (Appendix B) compute the correlation between individual forecast errors and current output, and show that this correlation is negative under the same condition that we have derived for \(\gamma < 0\). Intuitively, and unlike the coefficient \(\delta\) measuring underreactions, extrapolation is not driven by the averaging of forecasts across agents, but rather by the fact that each individual agent focuses on procyclical variables. The same result would remain true if agents were heterogeneous, as long as a critical mass of them pay asymmetric attention.

Second, while we have focused on extrapolation, our framework delivers a broader message: General patterns of predictability of forecast errors can arise from individuals’ asymmetric attention choices. For example, suppose that instead of regression \((1)\), we directly regress forecast errors onto one component \(x_{jt}\) of output. The slope coefficient from this regression would be proportional to \(a_j (1 - m_j)\), which could be either positive (representing an underreaction to \(x_{jt}\)) or negative (representing an overreaction), depending on the cyclicality of \(x_{jt}\) (the sign of
In principle, the conditions in Proposition 2 could therefore be extended and used to account for the much broader patterns of predictability documented, for example, by Pesaran and Weale (2006) and Fuhrer (2017). The conditions could, for instance, be used to account for the presence of both over- and underreactions to different public signals. In this sense, we view our model as a step towards a unified, rational explanation for the co-existence of over- and underreactions to new information (e.g., Bordalo et al., 2018; Eyster et al., 2019).

So far, we have considered a reduced-form economy. In deriving our results, we have taken agents’ attention choices, as summarized by the set of $m_j$, as given. We now move on to studying the potential sources of limited and asymmetric attention.

## 4 Attention Choices

In this section, we consider agents’ attention choices. We show that attention gravitates towards volatile components that are important to decision-makers. Combined with our previous results, this demonstrates that a rational theory of limited attention can match the survey evidence when procyclical components are either more volatile or more important.

### 4.1 A Model with Attention Choice

We augment our environment from Section 3 to incorporate attention choice. To do so, we assume the following timing of events: At the start of each period, each individual chooses her attention allocation $m_j$ to the different components $x_{jt}$ of output (or equivalently, the noise $q_j$ contained in her signals). She makes this choice before she observes the realization of her signals. Then, the agent observes her signals and chooses an action $a_{it}$.

The agent’s realized utility at the end of the period is:

$$U = -(a_{it} - a_t^*)^2 - K(m).$$

The first term in (13) is a quadratic loss that the individual incurs when she deviates from her ideal action $a_t^*$. The second term reflects the cost of attention $K(m)$. We assume that $K(\cdot)$ is positive, increasing in all $m_j$, and convex. We further assume that the ideal action, which the agent would take under perfect information (i.e., if attention $m_j = 1$ for all $j$), can depend both on the unobserved fundamental and on the structural components:

$$a_t^* = w_0 \theta_t + \sum w_{xj}x_{jt},$$

---

15 This condition follows from analogous steps to the proof of Proposition 2.

16 See, for example, Landier et al. (2017), Broer and Kohilas (2019), and the references therein.
where \( w_\theta \in \mathbb{R} \) and \( w_{xj} \in \mathbb{R} \) for all \( j \). With these preferences, the optimal choice of an agent who has information \( \Omega_{it} \) in the last stage at date \( t \) is to set 
\[
a_{it} = \mathbb{E}[a^*_t | \Omega_{it}].
\]

Equations (13) and (14) nest the benchmark case in which agents care only about forecasting future output as accurately as possible: When \( w_\theta = \rho^k \sum_j a_j \) and \( w_{xj} = 0 \), \( a^*_t \) becomes the full information mean-squared optimal forecast of \( y_{t+k} \), which is 
\[
\mathbb{E}^{FIRE}_t[y_{t+k}] = \rho^k \sum_j a_j \theta_t.
\]
However, (13) and (14) also allow us to capture more general cases in which agents’ ideal choice depends differently on the various structural components of output. This allows us to account for cases in which agents do not necessarily design their attention choices with the objective of predicting future output as accurately as possible. Instead, agents can also skew their attention choices towards the components of output that are the most important for their own specific decision problems. A firm, for example, might choose to pay more attention to its own sector than the economy as a whole (see Section 5 for a related example).

### 4.2 Optimal Attention to Important and Volatile Variables

We now derive agents’ optimal attention choices. To do so, it is instructive to first derive agents’ expected loss at the start of period \( t \), before they observe the realization of their signals. This expected loss equals (Appendix C.1):
\[
\mathbb{E}\left[(a_{it} - a^*_t)^2\right] = \text{Var}[a^*_t | \Omega_{it}, \theta_t] + \text{Var}[\mathbb{E}[a^*_t | \Omega_{it}, \theta_t] | \Omega_{it}]
\]
\[
= \sum_j w_{xj}^2 b_j^2 (1 - m_j) + \text{Var}_t[\theta_t] \left[w_\theta + \sum_j w_{xj} a_j (1 - m_j)\right]^2.
\] (15)

The first term in (15) reflects the uncertainty about the optimal action conditional on the fundamental. It equals the sum of the conditional variances \( \text{Var}[x_{jt} | \Omega_{it}, \theta_t] \) across the components \( x_{jt} \), weighted by their importance \( w_{xj} \) in agents’ utility. The uncertainty about each component naturally increases in its volatility \( b_j^2 \) but decreases in agents’ attention \( m_j \).

The second term measures the residual uncertainty \( \text{Var}[\theta_t | \Omega_{it}] \equiv \text{Var}_t[\theta_t] \), scaled by the uncertainty about the ideal action \( a^*_t = w_\theta \theta_t + \sum_j w_j x_j \) that is attributable to \( \theta_t \) (i.e., by the term in square brackets). We provide a brief derivation of \( \text{Var}_t[\theta_t] \), to show how it depends on agents’ attention choices. In turn, combined with (13) and (15), this will then allow us to derive an expression for agents’ optimal choices.
The effective precision of signal $z_{ijt}$ about $\theta_t$ is $\tau_j = \frac{a_j^2}{b_j^2 + \sigma_j^2}$,\(^{17}\) and we let

$$
\tau(m) = \sum_j \tau_j = \sum_j \frac{a_j^2}{b_j^2} m_j
$$

(16)
denote the total precision of date $t$ signals. Starting at date $t$, the conditional variance about next period’s fundamental is $\text{Var}_t [\theta_{t+1}] = \rho^2 \text{Var}_t [\theta_t] + \sigma_\theta^2$. After updating based on date $t+1$ signals, this variance satisfies the linear precision rule $\text{Var}_{t+1} [\theta_{t+1}]^{-1} = \text{Var}_t [\theta_{t+1}]^{-1} + \tau(m)$. Solving for a steady state where $\text{Var}_t [\theta_t] = \text{Var}_{t+1} [\theta_{t+1}] = V$ then delivers:

$$
\sigma_\theta^2 = V \left[ 1 - \rho^2 + \tau(m) \sigma_\theta^2 \right] + V^2 \tau(m) \rho^2.
$$

Thus, the total precision $\tau$ of an agent’s signals is a sufficient statistic for her uncertainty about the persistent fundamental, and we can write

$$
\text{Var}_t [\theta_t] = V [\tau(m)],
$$

(17)
where $V'(\tau) < 0$ and $\partial \tau / \partial m_j > 0$ from (16). Combined, (13), (15), and (17) allow us to characterize agents’ attention choices. Proposition 3 summarizes the results.

**Proposition 3.** Agents’ optimal attention choices satisfy, for all $j$ such that $0 < m_j < 1$,

$$
w_j^2 b_j^2 + \mu \tau_j^2 b_j^{-2} + \mu \lambda w_j a_j = \frac{\partial K(m)}{\partial m_j},
$$

(18)
where $\mu \lambda > 0$ and $\mu \tau > 0$ denote Lagrange multipliers.

Proposition 3 uses the fact that optimal (interior) attention choices equate the marginal benefit of paying more attention to each component to its marginal cost. The marginal benefit, on the left-hand side of (18), consists of three terms. The first term is the benefit of resolving uncertainty about the optimal action conditional on $\theta_t$. This benefit is higher for components that are more important for the optimal action (high $w_j$) and more volatile (high $b_j$).

The second and third terms capture a more nuanced effect: By learning about $x_{jt}$, the agent also acquires information about the fundamental $\theta_t$, which generates *learning spillovers* by resolving uncertainty about $x_{kt}$ for $k \neq j$. The second term measures the effect of attention $m_j$ on the effective precision $\tau$ of agents’ signals about $\theta_t$. The multiplier $\mu \tau$ is the shadow value of increasing this precision. This benefit of attention is larger for components that are highly correlated with the fundamental (high $a_j^2$), but spillovers are attenuated for components that

\(^{17}\)To see this, note that $z_{ijt}$ conveys the same information as its scaled version $\hat{z}_{ijt} = z_{ijt} / a_j$, which satisfies $\hat{z}_{ijt} = \theta_t + \xi_t$, where $\xi_t$ is Gaussian with mean zero and variance $\tau_j^{-1}$.
are highly volatile (high $b_j^2$). The third term measures an adjustment to this effect, namely, that information about $\theta_t$ becomes less valuable to an agent if she already has precise information about the structural components $x_{jt}$, and hence about her optimal action. The multiplier $\mu_\lambda$ is the shadow value of reducing the residual uncertainty about $a_t^*$ that is attributable to $\theta_t$.

While these effects are subtle, the underlying intuition is clear. On one hand, agents are more likely to pay attention to components that are important for their utility, those with large weights $w_{xj}$ in (14). On the other hand, agents also prefer to pay attention to volatile components (with a high idiosyncratic variance $b_j^2$), as long as learning spillovers are not too weak. This tendency for attention to gravitate towards important and volatile variables is familiar from much of the literature on information choice (Veldkamp, 2011), and has recently received additional empirical support in micro-level firm data (Coibion et al., 2018). Proposition 3 confirms that this intuition carries over to our component-based model.

We have so far kept the functional form of the attention cost function $K(m)$ general. In Appendix C.3, we derive the first-order condition (18) explicitly for an entropy-based cost function, and show that the main comparative statics remain the same. In addition, we show that an entropy-based cost function naturally yields limited attention choices $m_j < 1$, because it implies that the marginal cost of full attention is infinite ($\lim_{m_j \to 1} \frac{\partial K(m)}{\partial m_j} = \infty$).

In sum, asymmetric attention arises naturally from costly attention choice if some components are either more volatile, or more important to decision-makers. Combined with the insights of the previous section, we can therefore conclude that a rational theory of limited attention can match the survey evidence when procyclical components are either more volatile or more important. In the next section, we apply this reasoning to a simple macroeconomic model and show that, for reasonable parameters, attention gravitates to procyclical variables.

Before moving on to the application, we however consider two more points. First, we explore an alternative model of information choice in which agents have full flexibility in their information design. Second, we revisit the data and show that the survey evidence is consistent with an additional prediction of our framework.

### 4.3 Fully Flexible Information Choice

Proposition 3 characterizes the solution to a constrained information choice problem. Equation (9) restricts agents to acquire $N$ separate, conditionally independent signals $z_{ijt}$ about the components $x_{jt}$ of output. This is one of two popular approaches. An alternative approach is to instead allow agents full flexibility when designing the conditional distribution of their signals given the state of the economy (e.g., Sims, 2003). The choice between the two approaches is typically made based on the problem at hand, and on tractability. In the context of our analysis, it is interesting to compare the predictions of each approach.
With full flexibility and an entropy-based cost function, agents in our model would optimally choose to receive a single signal of the optimal action (see Maćkowiak et al., 2018):

\[ s_{it}^* = a_t^* + h u_t + q^* \epsilon_{it}, \]

(19)

where \( h \) is proportional to \( \tilde{w}_x = [ w_{x1} b_1 \ldots w_{xN} b_N ]' \), we stack the component-specific shocks into a vector \( u_t = [ u_1 \ldots u_N ]' \), and \( q^* \) is a scalar that depends only on the cost \( K(m) \) of attention.\(^\text{18}\) Equation (19) shows that the asymmetry of attention now depends only on the weights \( w_{xj} \) in agents’ optimal action, which influences both \( a_t^* \) and \( h \) in the optimal signal. This has important empirical implications.

For example, consider the benchmark case in which agents’ utility in (13) is equivalent to the mean-squared error of next period’s output forecast (\( w_\theta = \rho \sum_j a_j \) and \( w_{xj} = 0 \), as discussed above). In this case, it follows that \( h = 0 \) in (19). As a result, the fully optimal signal boils down to \( s_{it}^* = (\rho \sum_j a_j) \theta_t + q^* \epsilon_{it} \), which is a simple noisy signal of \( \theta_t \). Similar to the results in Proposition 1, and due to the symmetry of underlying preferences, such a signal is inconsistent with extrapolation.\(^\text{19}\) Indeed, in this case, agents systematically underreact to new information about output, yielding \( \gamma > 0 \) in (1).

Consider now instead the case in which the weights \( w_{xj} \) in agents’ optimal action are asymmetric across the structural components. It is then easy to show that agents’ forecasts of future output given \( s_{it}^* \) can exhibit extrapolation. Similar to the results in Proposition 2, this occurs when the weights \( w_{xj} \) are tilted towards procyclical components (Appendix C.4).

Combined, these examples show that we cannot test, based on survey data alone, whether the asymmetry of attention is driven by conditionally independent signals or by a flexibly designed, skewed signal. We only know that the fully flexible case is rejected by the data if agents

\(^{18}\)To see why (19) arises, notice that

\[
\begin{align*}
a_t^* &= w_\theta \theta_t + w_x' x_t = \tilde{w}_\theta \theta_t + \tilde{w}_x' u_t \\
&= \rho a_{t-1}^* + (\tilde{w}_\theta \sigma_\theta u_\theta + \tilde{w}_x' u_t - \rho \tilde{w}_x' u_{t-1},
\end{align*}
\]

in which \( \tilde{w}_\theta = w_\theta + \sum w_{xj} a_j \). Maćkowiak et al. (2018) show that the optimal choice of information (when \( a_t^* \) follows an ARMA(1,1) and attention costs are entropy-based) consists of a single noisy signal of the form:

\[ z_{it}^* = \omega a_t^* + (1-\omega)(\rho a_{t-1}^* - \rho \tilde{w}_x' u_t) + q^* \epsilon_{it}, \]

(20)

where \( \omega > 0 \) and \( q^* > 0 \). The signal in (20) is proportional and hence equivalent to that in (19).

\(^{19}\)Consider the extrapolation coefficient in (1) based on \( s_{it}^* = (\rho \sum_j a_j) \theta_t + q^* \epsilon_{it} \). It follows that

\[
\gamma = \text{Cov}(y_{t+k} - \tilde{E}_t y_{t+k}, y_t) \text{Var}[y_t]^{-1} = d_0 \text{Cov}(\theta_t - E_t \theta, y_t) = d_0 \sum_j a_j \text{Var}[\theta_t | s_{it}^*] > 0,
\]

where we have also used that \( y_t = \sum_j a_j \theta_t + \sum_j b_j u_{jt} \).
The chart shows updated values from Stark (2010), available from the Federal Reserve Bank of Philadelphia’s website. The chart illustrates the relative root mean-squared error of one-quarter and four-quarter ahead forecasts of output growth from the US Survey of Professional Forecasters (S) relative to three time series models: NC denotes a Random Walk forecast, IAR forecasts from an ARMA model chosen to minimize one-quarter ahead forecast errors, and DAR forecasts from ARIMA models chosen to minimize forecast errors at each forecast horizon. The sample period is 1985Q1:2015Q2. A RRMSE ratio below unity indicates that the SPF consensus forecast is more accurate. The sample period is 1985Q1:2015Q2.

care exclusively about the mean-squared error of output forecasts. By contrast, Proposition 3 shows that the “conditionally-independent signals” structure, even in the mean-squared error case, can be consistent with the simultaneous over- and underreactions documented in the data, so long as there are differences in the volatility of the underlying components.

4.4 Are Attention Choices Optimal? Supplementary Evidence

We briefly return to the data to compare the quality of forecasters’ expectations to that of standard time series models. Figure 4 shows updated values from Stark (2010), available from the Federal Reserve Bank of Philadelphia’s website. The chart illustrates the relative root mean-squared error (RRMSE) of one-quarter and four-quarter ahead forecasts of output growth from US SPF relative to three optimally-chosen time series models. A RRMSE ratio below unity indicates that the SPF consensus forecast is more accurate. All time series models fall short of survey forecasts at the one-quarter horizon, while the more sophisticated ARMA models achieve a close match with the SPF at the four-quarter horizon.

This supplementary evidence suggests that forecasters do better than simple time series models at forecasting output. This is consistent with our model, where agents pay attention to underlying, structural components of the forecasted variable, but inconsistent with a model

---

20https://www.philadelphiafed.org/research-and-data/real-time-center.html
where agents consider only the past time series of output (see, for instance, the results in Proposition 11.2 in Lütkepohl, 2007). In addition, this evidence rejects a simple behavioral story where agents derive forecasts from a misspecified ARMA model. Recent behavioral theory, such as Bordalo et al. (2018), is more nuanced, and further work would be needed to test whether forecasts in the data are more or less accurate than such theories predict. Hence, we interpret the supplementary evidence as a sanity check, which implies that our theory is consistent with moments of the data beyond the motivating evidence in Section 2.

We now turn to an application of our ideas to a standard macroeconomic model.

5 A Macroeconomic Example

In this section, we illustrate the sources and effects of asymmetric attention in a simple flexible-price business cycle model. We analyze an environment in which firms determine output under imperfect information. We show that firms’ output choices can be decomposed into two components: First, a local component, which summarizes the effects of a firm’s own local conditions, such as its own productivity; and second, an economy-wide component, which captures the effects of other agents’ behavior on an individual firm’s output choice. We document that, for standard parameters, the local component is procyclical, while the economy-wide component is countercyclical. In accordance with the evidence in Coibion et al. (2018), we show that firms’ attention choices are asymmetric and tend to focus on the local component. As a result, and in line with the above analysis, we find that firms’ expectations about output qualitatively and quantitatively match the estimated extrapolation and underreactions from the survey data. Finally, we show that asymmetric attention leads to more volatility and persistence in output.

5.1 Model Setup

The economy consists of a representative household and a continuum of monopolistically competitive firms, which specialize in the production of differentiated goods.

**Households:** A representative household has lifetime utility

\[ U = E_0 \sum_{t=0}^{\infty} \beta^t \left[ \log (C_t) - \xi_t N_t \right], \quad \xi_t > 0, \]  

(21)

where \( \beta \) denotes the time discount factor, \( C_t \) the consumption index at time \( t \), \( N_t \) the number of hours worked by the household, and \( \xi_t \) a shock to the disutility of labor. The consumption
index $C_t$ and associated welfare-based price index $P_t$ are given by

$$C_t = \left[ \int_0^1 C_{it}^{\sigma_t} di \right]^{\frac{\sigma_t}{\sigma_t - 1}}, \quad P_t = \left[ \int_0^1 P_{it}^{\frac{1}{\sigma_t - 1}} di \right]^{\sigma_t - 1}, \quad \text{(22)}$$

where $C_{it}$ is the amount the household consumes of goods produced by firm $i \in [0, 1]$ at price $P_{it}$ and $\sigma > 1$. Because the household receives all profits and labor income, its per-period budget constraint is

$$\int_0^1 P_{it}C_{it} di + B_{t+1} \leq \int_0^1 \Pi_{it} di + W_tN_t + (1 + R_t)B_{t+1} + T_t^h, \quad \text{(23)}$$

where $\Pi_{it}$ denotes the profits of firm $i$, $W_t$ the nominal wage, $R_t$ the nominal rate of return on riskless bonds, $B_t$ its holdings of riskless bonds, and $T_t^h$ lump-sum nominal transfers. The representative household’s objective is to maximize its utility (21) subject to (23).

**Firms:** There is a continuum of firms $i \in [0, 1]$ who specialize in the production of differentiated goods, also indexed by $i$. A representative firm $i$ chooses its output to maximize its own expectation of the household’s valuation of its profits, using the stochastic discount factor $(P_tC_t)^{-1}$. Profits at time $t$ are equal to

$$\Pi_{it} = (P_tC_t)^{-1} [P_{it}Y_{it} - W_tN_{it}], \quad \text{(24)}$$

where the demand for a firm’s product is consistent with the consumption index in (22). Firm output is produced using the production function

$$Y_{it} = A_{it}N_{it}^\alpha, \quad \alpha \in (0, 1), \quad \text{(25)}$$

where $N_{it}$ denotes the amount of labor input used and $A_{it}$ firm-specific productivity.

**Shocks:** We let lower-case letters denote natural logarithms of their upper-case counterparts. Firm-specific productivity $a_{it} = \log A_{it}$ is

$$a_{it} = \theta_t + u_{it}^\theta + \epsilon_{it}^a, \quad \text{(26)}$$

where the persistent, common component $\theta_t$ follows the AR(1) process

$$\theta_t = \rho\theta_{t-1} + u_{it}^\theta, \quad u_{it}^\theta \sim \mathcal{N} \left(0, \sigma_\theta^2 \right), \quad \text{(27)}$$
while the transitory and firm-specific components are distributed as $u^x_t \sim N(0, \sigma^2_x)$ and $\epsilon^a_{it} \sim N(0, \sigma^2_a)$, respectively. This is similar to the decomposition used in Kydland and Prescott (1982). The household’s disutility of labor is subject to a transitory shock with

$$\log \xi_t = \bar{\xi} + u^n_t, \quad u^n_t \sim N \left(0, \sigma^2_n\right)$$

where $\bar{\xi} \in \mathbb{R}$. We show below that the labor supply shock in (28) introduces a component-specific innovation to aggregate output. In effect, $u^n_t$ will play the role of one of the component-specific disturbances $u_{jt}$ discussed in Section 3. We assume that the innovations $u^x_t$, $u^\theta_t$, $u^n_t$, and $\epsilon^a_{it}$ are independent of each other, across time, and across firms.

**Timeline:** Each period consists of three stages. In the first stage, firms choose how much attention to devote to local and economy-wide conditions (defined below) and commit to their output choices. After output choices are sunk, the economy transitions to the second stage. The representative household now meets with firms, who hire labor to implement their previous output choices, and the shocks $\theta_t$ and $u^n_t$ are realized. The real wage adjusts to clear the labor market. In the final stage, good markets open, the transitory productivity shocks $u^x_t$ and $\epsilon^a_{it}$ are realized, good prices adjust, and the household consumes.\(^{21}\)

### 5.2 Firm Profits and Information

We focus on log-linear approximations of households’ and firms’ decision rules around the full information, non-stochastic steady state. Consistent with this approach, we analyze a log-quadratic approximation of an individual firms’ problem. Appendix D.1 shows that a second-order approximation of (24) yields, up to a constant term:

$$\pi_{it} \approx -\frac{1}{2} E_{it} \left[y_{it} - y^*_{it}\right]^2.$$  \hspace{1cm} (29)

The ideal, full information output choice $y^*_{it}$ is

$$y^*_{it} = ra_{it} + \alpha r \left(\sigma^{-1} y_t - E_{ht} [y_t] - u^n_t\right) \equiv x_{i1t} + x_{2t},$$  \hspace{1cm} (30)

\(^{21}\)Because the household does not observe the realization of $u^x_t$ in the second stage, output will respond differently to innovations in $\theta_t$ and $u^x_t$. This friction creates a meaningful distinction between these two shocks. Without this friction, only shocks to the sum $\int_0^1 a_idi = \theta_t + u^x_t$ would matter for output. An equivalent way to create distinct dynamics would be to study a model in which one of the factors of production, such as capital, is pre-determined before the realization of some of the shocks (see, for example, Angeletos et al., 2016).
where \( y_t = \int_0^1 y_t \, di \), we have defined \( r = \frac{\sigma}{\sigma + \alpha(1-\sigma)} > 1 \), and

\[
x_{1t} = r (\theta_t + u_t^x + \epsilon_{i1t}), \quad x_{2t} = \alpha r \left( \sigma^{-1} y_t - \mathbb{E}_{ht} [y_t] - u_t^n \right).
\] (31)

In the spirit of Lucas (1977), (30) and (31) decompose each firm’s ideal output choice into two components: We refer to \( x_{1t} \) as the local component, which depends on the firm’s own productivity. Clearly, each firm produces more when it is more productive. We refer to the second component, \( x_{2t} \), as the economy-wide component, which encapsulates the general equilibrium effects of other agents’ behavior on an individual firm’s optimal output choice. The economy-wide component, in turn, is comprised of two terms: On one hand, firms produce more when aggregate demand \( y_t \) is high. On the other hand, a firm also chooses to produce less when the real wage, equal to \( \omega_t = \mathbb{E}_{ht} [y_t] + u_t^n \), is high, where \( \mathbb{E}_{ht} [\cdot] \) denotes household expectations (Appendix D.1). Both effects are captured in (31).

Unlike the similar decomposition used in Maćkowiak and Wiederholt (2009), the local and economy-wide components are correlated in this application. For example, a shock to \( \theta_t \) will affect both components. Furthermore, in contrast to the baseline model from Section 3, the error terms in the two components are also correlated, since both depend on the transitory productivity shock \( u_t^x \). Hence, in order to characterize the properties of firms’ expectations, we will use the more general results listed in Proposition 5 in Appendix B.2.

As in our reduced-form framework, we assume that firms’ information set consists of the infinite history of component-based signals:

\[
\Omega_{it} = \{ z_{i1t}^l, z_{i2t}^l \},
\] (32)

where

\[
z_{i1t} = x_{1t} + q_1 \epsilon_{i1t}, \quad z_{i2t} = x_{2t} + q_2 \epsilon_{i2t},
\] (33)

and \( \epsilon_{ijt} \sim \mathcal{N}(0, 1) \) is independently distributed across time and firms for \( j = \{1, 2\} \). As in Section 3, we work with normalized parameters that capture attention:

\[
m_j = \frac{\text{Var}(x_{jiti})}{\text{Var}(x_{jiti}(\theta_t) + q_j^x)}.
\]

5.3 Optimal Firm Choices

An individual firm’s problem is as follows: At the start of the first stage of each period, the firm chooses attention coefficients \( m_1 \) and \( m_2 \) to maximize

\[
\max_{\{m_1, m_2\} \in [0, 1]^2} -\frac{1}{2} \mathbb{E} [y_{it} - y_{it}^*]^2 - K(m),
\] (34)
while anticipating that its optimal output choice will later be

\[ y_{it} = E \left[ y_{it}^* \mid z_{i1}^t, z_{i2}^t \right] = E_{it} \left[ x_{i1t} + x_{2t} \right]. \]  

(35)

The firm thus chooses ex ante how much attention to devote to local and economy-wide conditions, so as to maximize expected discounted profits. Notice that the problem in (34) and (35) is an application of the problem we studied in Section 4. There are \( N = 2 \) components of output, which determine the firm’s ideal action \( y_{it}^* \). The weight on each component \( x_{jt} \) is one (\( w_j = 1 \)). A small modification is that, due to firm-specific shocks, the ideal output \( y_{it}^* \) is now firm-specific.\(^{22}\) We summarize firms’ optimal output choices in Proposition 4.

**Proposition 4.** An individual firm’s output choice satisfies the fixed-point relationship

\[ y_{it} = E_{it} \left[ x_{i1t} + x_{2t} \right] \]

(36)

\[ = E_{it} \left[ r \left( \theta_t + u_t^x + \epsilon_{it}^0 \right) + \alpha r \left( \sigma^{-1} y_t - E_{ht} \left[ y_t \right] - u_t^a \right) \right], \]

where \( \sigma > 1 \), \( r = \frac{\sigma}{\sigma + \alpha (1 - \sigma)} > 1 \), and \( y_t = \int_0^t y_i \, di \).

Proposition 4 connects firms’ output choices to the optimal actions studied in the large class of beauty contest models analyzed in, for example, Morris and Shin (2002) and Angeletos and Pavan (2007). Similar to those models, output choices are a weighted average of firms’ expectations about unobserved fundamentals and firms’ expectations about the choices of others in the economy (see also, e.g., Angeletos et al., 2016). Unlike in these papers, however, our model includes a persistent unobserved fundamental \( \theta_t \). Proposition 4 thus closely relates to the price-setting models that follow in the tradition of Woodford (2002), in which firm choices are static but underlying fundamentals are dynamic.

### 5.4 Equilibrium Characterization

We now characterize equilibrium output choices in the economy.

#### 5.4.1 Equilibrium with Full Attention

We start with the case in which firms pay full attention to both components of output (i.e., \( m_j = 1 \) for \( j = 1, 2 \)) and there are no firm-specific productivity shocks (\( \sigma_a = 0 \)). This special case illustrates some important findings, which will carry over to our numerical solution of the full model with limited attention.

\(^{22}\)Nevertheless, from a firm’s perspective, firm-specific shocks are equivalent to an increase in the volatility of component-specific disturbances. Hence, the same conditions as in Section 4 apply here.
In this special case, each firm sets $y_t = x_{1t} + x_{2t}$, such that

$$y_t = \int_0^1 y_{it} \, di = x_{1t} + x_{2t},$$

(37)

precisely as in the baseline model from Section 3. The components $x_{jt}$ of output can now further be characterized directly from (31). As for the local component $x_{1t}$, we have

$$x_{1t} = r \theta_t + ru_t^x,$$

(38)

which is procyclical, since it places a positive weight $r > 0$ on the persistent fundamental $\theta_t$. Turning to the economy-wide component $x_{2t}$, we obtain from (31) and (37) that

$$x_{2t} = \alpha r \left( \frac{1 - \sigma}{\sigma} y_t + (1 - \alpha) u_t^n \right) = (1 - r) \theta_t + \left( \frac{1}{1 - \alpha} - r \right) u_t^x - \alpha u_t^n.$$  

(39)

The first equality in (39) shows that output choices are strategic substitutes: When other firms raise their output $y_t$, each individual firm’s output choice responds negatively (since $\sigma > 1$). Indeed, the increase in the real wage when output is high dominates the increase in demand in (31). By contrast, with perfect competition ($\sigma = 1$), firms are price-takers and act independently of one another. The second equality in (39) expresses the same relationship in equilibrium, in terms of the fundamental $\theta_t$ and other primitive shocks. We conclude that, due to strategic substitutability, the economy-wide component is countercyclical, since it places a negative weight $(1 - r) < 0$ on the persistent fundamental. This type of strategic substitutability (or “general equilibrium offset”) arises commonly in flexible-price business cycle models that generate realistic amount of volatility in hours worked (Hansen, 1985; Rogerson, 1988), because increases in other firms’ output tend to drive up production costs.

These properties, along with our results in Proposition 2 and 3, suggest that firms’ expectations about future output will match the survey data when firms pay imperfect, asymmetric attention to the procyclical, local component. For example, consider the hypothetical case in which all firms except firm $i$ pay full attention to both components, while firm $i$ pays full attention to $x_{1t}$ but none to $x_{2t}$. Then, it immediately follows that the slope coefficient in a regression of firm $i$’s forecast errors on recent output (that is, similar to (1)) becomes

$$\gamma_i = \text{Cov} (y_{t+1} - \mathbb{E} it y_{t+1}, y_t) \text{Var} [y_t]^{-1} = -\rho \frac{\alpha}{1 - \alpha} \frac{\text{Var}_t [\theta_i]}{\text{Var} [y_t]} < 0,$$

(40)
so that firm \( i \) appears to extrapolate.\(^{23}\)

### 5.4.2 Equilibrium with Limited Attention

We now return to the full model in which firms pay limited attention. When firms pay limited attention, the equilibrium dynamics of output can no longer be derived analytically. Instead, we solve the model numerically, looking for linear equilibria in which the law of motion for the local and economy-wide component, as well as the persistent fundamental, take the form of an infinite dimensional vector

\[
x_t = Ax_{t-1} + Bu_t, \quad u_t = \begin{bmatrix} u_t^0 & u_t^x & u_t^n \end{bmatrix}',
\]

where \( x_t = \begin{bmatrix} \bar{x}_t'_{t-1} & \bar{x}_t'_{t-2} & \ldots \end{bmatrix}' \) with \( \bar{x}_t = \begin{bmatrix} x_{1t} & x_{2t} & \theta_t \end{bmatrix}' \) and \( x_{1t} = \int_0^1 x_{it} di \). The appropriate rows of \( A \) and \( B \) conform with (31) and the law of motion for \( \theta_t \) in (26).

To solve for the rational expectations equilibrium, we conjecture and verify that

\[
y_t = \int_0^1 y_{it} di = \bar{E}_t [x_{1t} + x_{2t}] = \psi x_t,
\]

where the beliefs follow from the Kalman Filter

\[
\bar{E}_t [x_t] = A\bar{E}_{t-1} [x_{t-1}] + G \left( \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} - \bar{E}_{t-1} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \right),
\]

in which \( G \) denotes a matrix of Kalman Gains and \( \int_0^1 \begin{bmatrix} z_{1it} & z_{2it} \end{bmatrix}' di = \begin{bmatrix} x_{1t} & x_{2t} \end{bmatrix}' \).

The solution of the model requires finding coefficients \( A, B, \) and \( \psi \) that are consistent with firm optimality, Bayesian updating of expectations, and market-clearing. Computing the equilibrium requires truncating the infinite-dimensional vector \( x_t \). In accordance, with Hellwig and Venkateswaran (2009) and Lorenzoni (2009), we truncate it at \( \bar{x}_{t-T} \) where \( T = 50 \), but the numerical results are stable from already around \( T = 10 \). In Appendix D.2, we describe the algorithm that we use to find the approximate equilibrium solution.

\(^{23}\)This follows from

\[
\gamma_i = \text{Cov} (y_{t+1} - \text{E}_t y_{t+1}, y_t) \text{Var} [y_t]^{-1} = \text{Cov} \left( y_{t+1} - \text{E}_t y_{t+1}, x_{2t} \pm \frac{1}{r} \left( \frac{1}{1-\alpha} - r \right) x_{1t} \right) \text{Var} [y_t]^{-1} = \rho \text{Cov} \left( \theta_t - \text{E}_t \theta_t, (1-r) \theta_t - \left( \frac{1}{1-\alpha} - r \right) \theta_t \right) \text{Var} [y_t]^{-1} = -\rho \frac{\alpha}{1-\alpha} \text{Var}_t [\theta_t] \text{Var} [y_t]^{-1} < 0.
\]
5.5 A Quantitative Exploration

We now explore the quantitative implications of the model. The model is too stylized for a full quantitative investigation, so we only take a first pass at two basic questions: First, can the model match the extrapolation and underreaction from the survey data? Second, if so, what are the implications for the dynamics of output? To tackle these questions, we parameterize the model and compare estimates of $\gamma$ in (1) and $\delta$ in (2) to those from the data.

**Calibration:** We set the labor share $\alpha = 2/3$ and elasticity of substitution $\sigma = 6$. The persistence of the unobserved fundamental $\theta_t$ is set to $\rho = 0.90$ and the standard deviation of the shock to $\sigma_\theta = 1$. The standard deviation of the transitory component of productivity is likewise set to $\sigma_x = 1$, while the standard deviation of the labor supply shock is set to $\sigma_n = 0.1$. These values are all within the range used in standard DSGE models with monopolistic competition. Our baseline calibration eliminates firm-specific productivity shocks by setting $\sigma_a = 0$, to cleanly illustrate the effect of attention choices without exogenous noise in firms’ information. We later explore the robustness of our results towards this assumption.

For the attention cost function, we use the simple functional form $K(m) = \mu \sum_j q_j(m)^{-2}$; that is, a marginal cost $\mu$ multiplied by the sum of signal precisions $1/q_j^2$ across the components of output (Veldkamp, 2011). The free parameter is the marginal cost $\mu$, which determines the overall imperfection in firms’ information in equilibrium. For example, if $\mu = 0$, then we obtain the full information benchmark, because firms can obtain infinitely precise signals at no cost. As Coibion and Gorodnichenko (2015) point out, information frictions relate directly to the observable coefficient $\delta$ in (2) that measures underreactions in average expectations. Hence, we calibrate $\mu$ to match our estimated coefficient $\hat{\delta}$ from one-quarter-ahead forecasts in the SPF, consistent with a quarterly calibration. This yields $\mu = 1.30$.

**Components of Output and Attention Choices:** Recall from Proposition 2 and 3 that (i) asymmetric attention to procyclical variables can rationalize apparent extrapolation and underreactions, and that (ii) these patterns are consistent with optimal attention choices if procyclical variables are either more volatile or more important for agents’ decision-making. Figure 5 and Table II illustrate these mechanisms in general equilibrium.

Figure 5 shows that, as in the full information case, the local component of output is procyclical in equilibrium, while the economy-wide one is countercyclical. Output as a whole is procyclical. The first two columns in Table II show the significance of the local and economy-wide components in firms’ decision problems. While both components have a utility weight

---

24 In equilibrium, there is a one-to-one mapping between the precision parameters $q_j$ and the attention parameters $m_j$. Similar conclusions as those presented in Table II arise with an entropy-based cost function.
Table II: Attention Choices in Equilibrium

<table>
<thead>
<tr>
<th>Component</th>
<th>Variance</th>
<th>Weight</th>
<th>q</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local component ($x_1$)</td>
<td>3.72</td>
<td>1.00</td>
<td>1.28</td>
<td>0.93</td>
</tr>
<tr>
<td>Economy-wide component ($x_2$)</td>
<td>1.03</td>
<td>1.00</td>
<td>2.10</td>
<td>0.29</td>
</tr>
</tbody>
</table>

(i) Note: Variances have been scaled by the variance of output.

Figure 5: Cyclicality of Structural Components and Output:
Impulse Response to a One Unit Standard Deviation Shock to $\theta_t$

Note: The chart depicts the impulse response function to a unit standard deviation shock to $\theta_t$ on the vertical axis. Time is measured in quarters on the horizontal axis.
of one in firms’ ideal output choice (see (35)), the local component is much more volatile for baseline parameters. The third and fourth columns in Table II show firms’ optimal attention choices \((m_j)\), or equivalently noise choices \((q_j)\), for both output components. As expected, attention gravitates towards the local component because of its larger volatility. In particular, firms optimally choose to pay around three times more attention to the local component. This is consistent with the conclusions from Lucas (1977) (also cited in the introduction) that for most firms there is no reason to pay particularly close attention to aggregate conditions. It is also in accordance with the recent evidence in Coibion et al. (2018), which shows that firms’ attention choices are asymmetric and tend to focus on local variables. We now explore the implications of these asymmetries for firms’ expectations in equilibrium.

**Over- and Underreactions:** The first two columns in Table IIIa show the results of estimating the extrapolation regression (1) and the underreactions regression (2) on firms’ simulated expectations of one-quarter ahead output in equilibrium. The third and fourth columns compare these estimates to those obtained in the survey data at the one-quarter horizon (see Appendix A.2). The underreaction coefficient \(\delta\) at the one-quarter frequency was a targeted moment. Due to firms’ asymmetric attention to the procyclical, local component of output, the coefficient \(\gamma\) on current output in (1) is negative, generating apparent overreactions in expectations that are qualitatively and quantitatively close to those in the data. As a result, firms’ expectations are simultaneously consistent with extrapolation and underreactions.

Table IIIb shows the implied estimates at the four-quarter horizon, which mirror the specification in Table I. The model does not match the full magnitude of these coefficients, largely because the stationarity of the model implies that the estimates of (1) and (2) should decline with the time horizon. However, despite its simplicity, the model still accounts for a sizable proportion of the empirical estimates at the four-quarter horizon, neither of which were targeted moments in the calibration.

The last row in Table III shows that firms in the simulated model make better forecasts (in a root-mean square error sense) than they would achieve using a simple time series model. This is consistent with our empirical results in Section 4.

---

25 Notice that, because firms have imperfect information about both components, the variance of each component in Table II can exceed that of output itself (which is the expectation of the sum).

26 An alternative approach is to calibrate the model by targeting the four-quarter \(\delta\) estimate in Table I. In this case, we arrive at estimates for \(\gamma\) which are close to their empirical counterparts. The implied one-quarter ahead estimates, however, suggest slightly more extrapolation that what we see in the data.
Table III: Over- and Underreactions

(a) One-quarter Ahead Output Growth

<table>
<thead>
<tr>
<th></th>
<th>Model Estimates</th>
<th>Data Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast Error</td>
<td>Forecast Error</td>
</tr>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td>(–)</td>
</tr>
<tr>
<td>Current Realization</td>
<td>−0.09</td>
<td>−0.06</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td></td>
</tr>
<tr>
<td>Forecast Revision</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>(–)</td>
<td></td>
</tr>
<tr>
<td>Relative RMSE</td>
<td>0.88</td>
<td></td>
</tr>
</tbody>
</table>

(b) Four-Quarter Ahead Output Growth

<table>
<thead>
<tr>
<th></th>
<th>Model Estimates</th>
<th>Data Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast Error</td>
<td>Forecast Error</td>
</tr>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td>(–)</td>
</tr>
<tr>
<td>Current Realization</td>
<td>−0.07</td>
<td>−0.12**</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td></td>
</tr>
<tr>
<td>Forecast Revision</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>(–)</td>
<td></td>
</tr>
<tr>
<td>Relative RMSE</td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>

Note: HAC standard errors in parentheses. Significance levels *=10%, **=5%, ***=1%.

Relative RMSE denotes the root mean-squared-error of individual forecasts relative to an estimated AR(1).
5.6 Further Implications of Asymmetric Attention

We leverage our calibrated model to illustrate two wider implications of asymmetric attention. First, we show that asymmetric attention causes the equilibrium dynamics of output to be more persistent and more volatile. Second, we show that our model is also consistent with the increased responsiveness to new information, and the increased extrapolation, that emerges in the data after the Great Moderation (as we also document empirically in Appendix A.2).

5.6.1 Asymmetric Attention and Output Dynamics

We compare the dynamics of output in our model with those that arise in an equivalent model where attention is limited but \textit{symmetric}. In this symmetric case, firms observe only one noisy signal of optimal output:

\[ s_{it} = y_t + q\epsilon_{it} = x_{1t} + x_{2t} + q\epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0, 1), \]  

where the noise parameter \( q \) (or corresponding attention parameter \( m \)) is again calibrated to match the one-quarter-ahead estimate of \( \delta \).

Figure 6 summarizes the results. The left panel shows that the model with asymmetric attention results in more persistence in output (larger autocorrelation). This is intuitive: When firms focus their attention on the local component, their beliefs and actions become more persistent because the local component directly tracks the dynamics of the fundamental \( \theta_t \). This increase in output persistence occurs even though all input choices happen within period. An additional, pre-determined factor of production, such as capital, would amplify these effects by allowing firms’ extrapolative expectations to directly affect future output.

Relatedly, the right panel in Figure 6 shows that output responses are also more correlated with the unobserved fundamental itself when there is asymmetric attention. The bottom panel, in turn, shows that asymmetric attention also causes the unconditional variance of output to increase. For the same overall information friction (as measured by the coefficient \( \delta \) in (2)), the asymmetry of attention increases the volatility of output, and pushes it closer to its full information value.

Lastly, in line with our results from Section 3, we note that the model with symmetric attention produces a positive estimate of \( \gamma \) (\( \gamma = 0.08 \)), which is inconsistent with the data.

5.6.2 Asymmetric Attention and the Great Moderation

One manifestation of the Great Moderation was a reduction in the size of aggregate versus firm-specific shocks. As discussed in, for example, Arias \textit{et al.} (2007) and Galí and Gambetti (2009), the standard deviation of aggregate productivity shocks declined by around 50 percent.
Figure 6: Asymmetric Attention and Output Dynamics

(a) Autocorrelation of Output  
(b) Correlation of Output with Productivity  
(c) Variance Relative to Full Information Benchmark

<table>
<thead>
<tr>
<th></th>
<th>Asymmetric Attention</th>
<th>Symmetric Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Variance</td>
<td>0.52</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Note: The left panel shows the autocorrelation of output on the vertical axis, with the lags of output up to four quarters on the horizontal axis. The right panel shows the correlation of output with total factor productivity $a_t = \theta_t + u^2_t$ once more up to a four-quarter lag. We depict these both for the calibrated asymmetric attention model, the symmetric attention model, as well as the full information case. The bottom panel illustrates the variance of output in the asymmetric and symmetric case relative to the full information benchmark.
Table IV: Model Estimates Pre/Post-Great Moderation

<table>
<thead>
<tr>
<th></th>
<th>Pre-Great Moderation</th>
<th>Post-Great Moderation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast Error</td>
<td>Forecast Error</td>
</tr>
<tr>
<td>Constant</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td>(–)</td>
</tr>
<tr>
<td>Realization</td>
<td>-0.16</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td></td>
</tr>
<tr>
<td>Forecast Revision</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(–)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The equilibrium noise in the signal about the local component is $q_1 = 0.92$; by contrast, the noise in the signal about the economy-wide component is $q_2 = 3.15$.

after 1985, while the volatility of firm-specific innovations appears mostly unchanged (Comin and Philippon, 2005). We explore the implications of a similar structural shift in our model.

Following Arias et al. (2007), we assume that all of the decrease in the volatility of aggregate productivity is due to a decrease in the common, persistent component $\sigma_\theta$. To model the economy before the Great Moderation, we use our baseline calibration above, but re-introduce firm-specific productivity shocks $\sigma_a > 0$. This parameter is calibrated to match the level of information frictions before the Great Moderation, which we estimate by running regression (2) for one-year ahead forecasts on a sample until 1985Q1. To model the economy after the Great Moderation, we then reduce the volatility of $\sigma_\theta$ by 50 percent.

Table IV shows the resulting estimates of (1) and (2) on model-generated data before and after the Great Moderation. As in the equivalent regressions on the actual survey data, which we report in Appendix A.2, extrapolation becomes stronger while underreactions become weaker after the Great Moderation. This is because the decrease in the volatility of persistent productivity shocks causes firms to choose more asymmetric attention. Indeed, compared to the full sample calibration in Table II, our solution shows that post-Great Moderation firms pay 28 percent more attention to the local component (as measured by the noise in firms’ information $q_1$), and 50 percent less attention to the economy-wide component.

The results in this subsection have highlighted two important implications of asymmetric attention. First, asymmetric attention not only affects the properties of expectations, but also heightens the persistence and volatility of output fluctuations in general equilibrium. Second, an exploration of the Great Moderation provides further validation of our example framework. A simple model based on asymmetric attention to firms’ local component of output can qualitatively match the empirical observation that extrapolation strengthened while underreactions subsided at a time when aggregate productivity became less volatile.
6 Conclusion

In this paper, we have contributed to a research agenda that seeks to find a data-consistent model of expectation formation. The framework we have considered relies on minimal frictions relative to the classical benchmark. The only primitive deviation from full information and rational expectations is limited attention. Previous work by Woodford (2002), Sims (2003), Angeletos and Huo (2019), and others, have demonstrated that limited attention offers an explanation for the myopia and anchoring to past outcomes commonly documented in macroeconomics. Our results show that extrapolation, and more generally overreactions to public information, can also be explained by this framework.

We have documented that households’, firms’, and professional forecasters’ expectations simultaneously overreact to recent outcomes of the forecasted variable but underreact to new information on average. These facts are inconsistent with standard behavioral models of extrapolation, as well as with models that combine the overconfidence inherent to extrapolation with noisy information. To resolve this friction, we have proposed a simple, rational model of limited attention in which people internalize that a forecasted variable is comprised of several components. We characterized the conditions under which this model is consistent with the data. In doing so, we have developed a rational theory of extrapolation that is also consistent with observed underreactions. This theory is based on individuals’ asymmetric attention to procyclical variables. Through the lens of this model, the overreactions to recent outcomes documented in survey data can be viewed as underreactions to countercyclical components.

To illustrate our results, we embedded our analysis in a workhorse macroeconomic model. For reasonable parameters, we showed that firms’ expectations exhibit extrapolation and underreactions, similar to their empirical counterparts. This application also allowed us to study the implications of asymmetric attention for the dynamics of output, and to validate the model further by studying its implications for structural changes around the Great Moderation.

Beyond the analysis in this paper, our results suggest that models of limited, asymmetric attention can account for flexible patterns of predictability in people’s forecast errors. We see important scope for extending our results to account for the more general under- and overreactions to public information documented in the literature. Another avenue for future research is to combine models of optimal information choice with insights from behavioral economics, such as those discussed recently by Bordalo et al. (2018). The latter approach would allow for an empirical estimate of the relative contribution of each component to the predictability of forecast errors. Overall, we view the research in this paper as a useful step towards a unified, data-consistent model of expectations based on a minimal set of frictions.
Appendix A: Motivating Evidence

This Appendix details the proof of Proposition 1 and the robustness of the empirical results.

Appendix A.1: Existing Theories

Proof of Proposition 1: The proof proceeds in three steps. We first derive the slope coefficients in (1) and (2) for all cases with the exception of the extrapolation model and that of diagnostic expectations. We then derive the slope coefficients for each of these in turn.

Step 1: The Benchmark Case: We start with the extrapolation coefficient $\gamma$ in (1). The Average Generalized Kalman Filter is from (5):

$$\tilde{f}_i y_{t+k} = \tilde{f}_{i-1} y_{t+k} + g_k \left( y_t - \tilde{f}_{i-1} y_t \right), \quad g_k \in (0, 1). \tag{A1}$$

Using (A1), we then find that

$$\gamma = \text{Cov} \left[ y_{t+k} - \tilde{f}_{i} y_{t+k}, y_t \right] \mathbb{V}[y_t]^{-1}$$

$$= \text{Cov} \left[ y_{t+k} - \tilde{f}_{i-1} y_{t+k} - g_k \left( y_t - \tilde{f}_{i-1} y_t \right), y_t \right] \mathbb{V}[y_t]^{-1}$$

$$= \text{Cov} \left[ y_{t+k} - \tilde{f}_{i-1} y_{t+k} - \rho^k \left( y_t - \tilde{f}_{i-1} y_t \right) + \left( \rho^k - g_k \right) \left( y_t - \tilde{f}_{i-1} y_t \right), y_t \right] \mathbb{V}[y_t]^{-1}$$

$$= \text{Cov} \left[ y_{t+k} - \rho^k y_t + \left( \rho^k - g_k \right) \left( y_t - \tilde{f}_{i-1} y_t \right), y_t \right] \mathbb{V}[y_t]^{-1}$$

$$= \left( \rho^k - g_k \right) \text{Cov} \left[ y_t - \tilde{f}_{i-1}, y_t \right] \mathbb{V}[y_t]^{-1}, \tag{A2}$$

where the third equality follows from $\tilde{f}_{i-1} y_{t+k} = \rho^k \tilde{f}_{i-1} y_t$, while the fourth equality follows from the orthogonality of FIRE forecast errors to observed information.

It remains to show that $\text{Cov} \left[ y_t - \tilde{f}_{i-1}, y_t \right] > 0$ in (A2). Notice that from (A1):

$$\tilde{f}_{i-1} y_t = \tilde{f}_{i-2} y_t + g_1 \left( y_{t-1} - \tilde{f}_{i-1} y_{t-1} \right)$$

$$= g_1 y_{t-1} + (\rho - g_1) \tilde{f}_{i-2} y_{t-1} = g_1 \sum_{i=1}^{\infty} (\rho - g_1)^{i-1} y_{t-i},$$

where the sum converges since $| \rho - g_1 | < 1$, and we have used that $\tilde{f}_{i-2} y_t = \rho \tilde{f}_{i-2} y_{t-1}$.

It therefore follows that

$$\text{Cov} \left[ \tilde{f}_{i-1} y_t, y_t \right] = g_1 \sum_{i=1}^{\infty} (\rho - g_1)^{i-1} \text{Cov} \left[ y_{t-i}, y_t \right] = g_1 \sum_{i=1}^{\infty} (\rho - g_1)^{i-1} \rho^i \mathbb{V}[y_t].$$
Thus, we have from (A2) that
\[
\gamma = (\rho^k - g_k) \left[ 1 - g_1 \sum_{i=1}^{\infty} (\rho - g_1)^{i-1} \rho^i \right] = (\rho^k - g_k) \left[ 1 - \frac{\rho g_1}{1 - \rho^2 + g_1 \rho} \right]. \tag{A3}
\]

Equation (A3) shows that the sign of \( \gamma \) depends only on the sign of \( \rho^k - g_k \) since \( \frac{\rho g_1}{1 - \rho^2 + g_1 \rho} < 1 \).

Let us now turn to the \( \delta \) coefficient. Using (A1) once more, we have that
\[
\delta = \text{Cov} \left[ y_{t+k} - \tilde{f}_t y_{t+k}, \tilde{f}_t y_{t+k} - \tilde{f}_{t-1} y_{t+k} \right] \mathbb{V} \left[ \tilde{f}_t y_{t+k} - \tilde{f}_{t-1} y_{t+k} \right]^{-1} \\
\propto \text{Cov} \left[ y_{t+k} - \rho^k y_t, \tilde{f}_t y_{t+k} - \tilde{f}_{t-1} y_{t+k} \right] \\
= (\rho^k - g_k) \text{Cov} \left[ y_t - \tilde{f}_{t-1} y_t, \tilde{f}_t y_{t+k} - \tilde{f}_{t-1} y_{t+k} \right], \tag{A4}
\]

where " \( \propto \) " denotes "positively proportional to", and we have used the same steps as those that resulted in (A2).\(^{27}\) The definition of the average Generalized Kalman Filter then shows that
\[
\delta \propto (\rho^k - g_k) g_k \mathbb{V} \left[ y_t - \tilde{f}_{t-1} y_t \right]. \tag{A5}
\]

Equation (A5) demonstrates that the sign of \( \delta \) is once more tied to the sign of \( \rho^k - g_k \).

**Step 2: Extrapolation Model:** The above derivations several times relied on \( \tilde{f}_{t-1} y_{t+k} = \rho^k \tilde{f}_{t-1} y_t \). In the classical extrapolation case, however, \( \tilde{f}_{t-1} y_{t+k} = g_k \tilde{f}_{t-1} y_t \), where \( g_k = g_k^1 > \rho^k \) since agents think that output is more persistent than it actually is. However, we can still in this case link the sign of \( \gamma \) and \( \delta \) to the sign of \( \rho^k - g_k \). Consider the \( \gamma \) coefficient
\[
\gamma = \text{Cov} \left[ y_{t+k} - \tilde{f}_t y_{t+k}, y_t \right] \mathbb{V}[y_t]^{-1} = \text{Cov} \left[ y_{t+k} - g_k y_t, y_t \right] \mathbb{V}[y_t]^{-1} \\
= \text{Cov} \left[ y_{t+k} - \rho^k y_t + (\rho^k - g_k) y_t, y_t \right] \mathbb{V}[y_t]^{-1} = \rho^k - g_k.
\]

Consider next the \( \delta \) coefficient
\[
\delta = \text{Cov} \left[ y_{t+k} - \tilde{f}_t y_{t+k}, \tilde{f}_t y_{t+k} - \tilde{f}_{t-1} y_{t+k} \right] \mathbb{V} \left[ \tilde{f}_t y_{t+k} - \tilde{f}_{t-1} y_{t+k} \right]^{-1} \\
\propto \text{Cov} \left[ y_{t+1} - g_k y_t, g_k y_t - g_1 y_{t-1} \right] \\
= \text{Cov} \left[ (\rho^k - g_k) y_t, g_k y_t - g_1^{1+k} y_{t-1} \right] = (\rho^k - g_k) g_k (1 - \rho g_1) \mathbb{V}[y_t]
\]

where we have once more added and subtracted the FIRE forecast \( \rho^k y_t \) in the third line; we have also used that \( g_k = g_k^1 \) in the third line of the relationship.

\(^{27}\)Notice that FIRE forecast errors are also orthogonal to any linear combinations of past outcomes of \( y_t \) (Granger, 1974), and hence also to \( \tilde{f}_t y_{t+k} - \tilde{f}_{t-1} y_{t+k} \).
**Step 3: Diagnostic Expectations Model:** The diagnostic expectations model is not recursive; the terms \( \tilde{f}_{t-1}y_{t+k} \) and \( \tilde{f}_{t-1}y_t \) in the average Generalized Kalman Filter are replaced by their FIRE counterparts. This, however, does not complicate the derivations connecting the sign of \( \gamma \) and \( \delta \) to the sign of \( \rho^k - g_k \). The derivations leading up to (A2) now show that

\[
\gamma = (\rho^k - g_k) \text{Cov} \left[ y_t - \mathbb{E} \left[ y_t \mid y_{t-1} \right], y_t \right] \mathbb{V}[y_t]^{-1}
\]
\[
\gamma = (\rho^k - g_k) \text{Cov} \left[ y_t - \rho y_{t-1}, y_t \right] \mathbb{V}[y_t]^{-1} = (\rho^k - g_k) \mathbb{V} [u_t] \mathbb{V}[y_t]^{-1}.
\]

Likewise, with the \( \delta \) coefficient. Consider (A4). We now have that

\[
\delta \propto (\rho^k - g_k) \text{Cov} \left[ y_t - \mathbb{E} \left[ y_t \mid y_{t-1} \right], \tilde{f}_t y_{t+k} - \mathbb{E} \left[ y_{t+k} \mid y_{t-1} \right] \right]
\]

because of the rational reset in the diagnostic expectations model. Thus,

\[
\delta \propto (\rho^k - g_k) g_k \mathbb{V} \left[ y_t - \mathbb{E} \left[ y_t \mid y_{t-1} \right] \right].
\]

This completes the proof. \( \square \)

**Appendix A.2: Robustness of Evidence**

This appendix details additional robustness checks of the empirical results. We consider whether the simultaneous over- and underreactions documented in the paper extend (i) to shorter horizons than one-year ahead forecasts; (ii) to both pre- and post-Great Moderation; (iii) across different means of de-trending the data; and last (iv) to the individual-level.

In Table V, the first two columns show our baseline estimates of (1) and (2) with a one-year forecast horizon (the same as reported in Table I), while the third and fourth column show the equivalent results with a one-quarter forecast horizon. The point estimates continue to suggest simultaneous over- and underreactions, but the estimated coefficients decline in magnitude and significance. This is consistent with a greater importance of noise in shorter horizon forecasts, as also documented elsewhere (e.g. Coibion and Gorodnichenko, 2015).

Table VIII also replicates the baseline case in the first two columns. The third and fourth columns contain the same estimates for a sample that ends before the Great Moderation (1985Q1), while the final two columns correspond to the post-Great Moderation sample (1986Q1-2006Q4). After the Great Moderation, the overreaction increases, while underresponsiveness becomes somewhat weaker.

Recall that our estimates of (1) have current output growth on the right-hand side. As a result, they may require an appropriate method to remove structural trends. This is to account for any potential, unanticipated structural shifts in the level of output growth, and
especially, in the level of inflation, which could otherwise create serial correlation in forecast errors. Our baseline estimates, without any de-trending, are reiterated in the first column of Table VI. The second column shows the estimates that arise from de-trending output growth with a HP-filter where $\lambda = 1,600$. A Baxter-King filter yields almost indistinguishable results, which we do not report. The final column uses a linear trend, which results in very similar estimates as those in our baseline.

Lastly, Table VII shows that extrapolation is, if anything, more pronounced at the individual-level compared to the average-level that we use as the baseline for our results.

Table V: GDP Forecasts SPF (Alternative Forecast Horizon)

<table>
<thead>
<tr>
<th></th>
<th>Four-Quarter Ahead</th>
<th></th>
<th>One-Quarter Ahead</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast Error</td>
<td>Forecast Error</td>
<td>Forecast Error</td>
<td>Forecast Error</td>
</tr>
<tr>
<td>Constant</td>
<td>0.07</td>
<td>-0.10</td>
<td>0.28</td>
<td>0.25**</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.11)</td>
<td>(0.23)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Current Realization</td>
<td>-0.12**</td>
<td>-</td>
<td>-0.06</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>Forecast Revision</td>
<td>-</td>
<td>0.77***</td>
<td>-</td>
<td>0.42***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.27)</td>
<td></td>
<td>(0.14)</td>
</tr>
</tbody>
</table>

Table VI: GDP Forecasts SPF (Alternative Means to De-trend)

<table>
<thead>
<tr>
<th></th>
<th>Benchmark</th>
<th>HP Detrend</th>
<th>Linear Detrend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.07</td>
<td>-0.22**</td>
<td>-0.22**</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.12)</td>
<td>(0.12)</td>
</tr>
<tr>
<td>Current Realization</td>
<td>-0.12**</td>
<td>-0.15**</td>
<td>-0.12**</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
</tbody>
</table>

Note: HAC standard errors in parentheses. Significance levels *=10%, **=5%, ***=1%.
Table VII: GDP Forecasts SPF (Individual-level)

<table>
<thead>
<tr>
<th>Four-Quarter Ahead</th>
<th>Avg. Forecast Error</th>
<th>Ind. Forecast Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.59)</td>
</tr>
<tr>
<td>Current Realization</td>
<td>-0.12**</td>
<td>-0.17***</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Sample</td>
<td>1970Q1-2017Q4</td>
<td>1970Q1-2017Q4</td>
</tr>
<tr>
<td>Observations</td>
<td>188</td>
<td>6,850</td>
</tr>
</tbody>
</table>

Note: HAC standard errors (average level); White double-clustered standard errors (individual level). Individual level estimates include forecaster fixed effects. Significance levels *=10%, **=5%, ***=1%.
Table VIII: GDP Forecasts SPF (Pre/Post-Great Moderation)

<table>
<thead>
<tr>
<th></th>
<th>Full Sample</th>
<th>Pre-Great Moderation</th>
<th>Post-Great Moderation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forecast Error</td>
<td>Forecast Error</td>
<td>Forecast Error</td>
</tr>
<tr>
<td>Constant</td>
<td>0.07</td>
<td>-0.10</td>
<td>-0.45</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.11)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>Current Realization</td>
<td>-0.12**</td>
<td>–</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.08)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>Forecast Revision</td>
<td>–</td>
<td>0.77***</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.38)</td>
<td>(0.30)</td>
</tr>
<tr>
<td>$F$</td>
<td>5.50</td>
<td>15.2</td>
<td>1.35</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.03</td>
<td>0.08</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Note: HAC standard errors in parentheses. Significance levels * = 10%, ** = 5%, *** = 1%.
## Appendix A.3: Output and Inflation Estimates

### Table IX: Estimated Over- and Underreactions across Surveys

<table>
<thead>
<tr>
<th></th>
<th>US SPF</th>
<th></th>
<th>EA SPF</th>
<th></th>
<th>LS Survey</th>
<th></th>
<th>MSC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Output</td>
<td>Inflation</td>
<td>Output</td>
<td>Inflation</td>
<td>Output</td>
<td>Inflation</td>
<td>Inflation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-0.22</td>
<td>-0.03</td>
<td>-0.09</td>
<td>-0.05</td>
<td>0.09</td>
<td>0.11</td>
<td>0.09</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.02)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.16)</td>
<td>(0.15)</td>
<td>(0.09)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Current Realization</td>
<td>-0.15**</td>
<td>-0.21**</td>
<td>-0.23***</td>
<td>-0.31*</td>
<td>-0.19***</td>
<td>-0.07</td>
<td>-0.47***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.10)</td>
<td>(0.08)</td>
<td>(0.16)</td>
<td>(0.09)</td>
<td>(0.16)</td>
<td>(0.16)</td>
<td></td>
</tr>
<tr>
<td>Forecast Revision</td>
<td>0.77***</td>
<td>1.14***</td>
<td>0.68***</td>
<td>0.70*</td>
<td>0.45***</td>
<td>1.91***</td>
<td>0.71***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.26)</td>
<td>(0.28)</td>
<td>(0.25)</td>
<td>(0.40)</td>
<td>(0.16)</td>
<td>(0.24)</td>
<td>(0.26)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>6.50</td>
<td>15.24</td>
<td>7.20</td>
<td>42.6</td>
</tr>
<tr>
<td>R²</td>
<td>0.04</td>
<td>0.08</td>
<td>0.04</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Note: HAC standard errors in parentheses. Significance levels *=10%, **=5%, ***=1%. A HP trend ($\lambda = 1,600$) has been deducted from current output and inflation to account for potential structural changes.
Appendix B: Asymmetric Attention

This Appendix derives the proofs of the propositions mentioned in Section 3.

Appendix B.1: Baseline Model

In this Appendix, we detail the proof of Proposition 2. We proceed in two steps: First, we derive a simple result that relates individuals’ attention choices \( m_j \) to the Kalman Gain in their expectations about the unobserved fundamental. Second, we then use this results to prove Proposition 2. For the first step, notice that from the Kalman Filter:

\[
 f_{it} \theta_t = \mathbb{E}_{it} [\theta_t] = \mathbb{E}_{it-1} [\theta_t] + \sum_j g_j (z_{ijt} - \mathbb{E}_{it-1} [z_{ijt}]), \tag{A6}
\]

where \( g_j \) now denotes the Kalman Gain on the \( j \)th signal \( z_{ijt} \) about the fundamental \( \theta_t \).

Lemma 1 relates these gain coefficients to individuals’ attention choices \( m_j \).

Lemma 1. The update coefficients in (A6) equal \( g_j = \bar{g} m_j a_j b_j^{-2} \), in which \( \bar{g} \in (0, 1) \), and \( \bar{g} \) is strictly increasing in individuals’ attention choices \( m_j \) for \( j = \{1, \ldots, N\} \).

Proof of Lemma 1: The Kalman Gain in (A6) is equal to:\(^{28}\)

\[
g = P1' [11'P + BB' + QQ']^{-1}, \tag{A7}
\]

where \( g = \begin{bmatrix} g_1 & \cdots & g_N \end{bmatrix} \), \( Q = \text{diag}(q_j/a_j) \), \( B = \text{diag}(b_j/a_j) \), and \( P = \mathbb{V} [\theta_{t+1} | \Omega_{it}] \) solves the Riccati equation:

\[
P = \rho^2 \left[ P - P^2 1' (11'P + BB' + QQ')^{-1} 1 \right] + \sigma_\theta^2. \tag{A8}
\]

We now derive expressions for terms in (A7) and (A8).

1. \( BB' + QQ' \): It follows that

\[
 BB' + QQ' = \text{diag} \left( \frac{b_j^2 + q_j^2}{a_j^2} \right) = \text{diag} \left( \frac{b_j^2}{a_j^2 m_j} \right) = \text{diag} \left( \frac{1}{n_j} \right), \quad n_j = m_j a_j^2 b_j^2.
\]

2. \( (11'P + BB' + QQ')^{-1} \): Sherman-Morrison’s formula provides us with

\[
 (11'P + BB' + QQ')^{-1} = \left( \sqrt{P} 11' \sqrt{P} + BB' + QQ' \right)^{-1} = \text{diag} (n_j) - \frac{\text{diag} (n_j) \sqrt{P} 11' \sqrt{P} \text{diag} (n_j)}{1 + 1' \sqrt{P} \text{diag} (n_j) \sqrt{P} 1}.
\]

\(^{28}\)See, for example, Anderson and Moore (2012).
Now, notice that
\[
\text{diag}(n_j) \sqrt{P} 11' \sqrt{P} \text{diag}(n_j) = P \begin{bmatrix}
  n_1^2 & n_1 n_2 & \ldots & n_1 n_N \\
  n_1 n_2 & n_2^2 & \ldots & n_2 n_N \\
  \vdots & \vdots & \ddots & \vdots \\
  n_1 n_N & n_2 n_N & \ldots & n_N^2
\end{bmatrix}
\]

while
\[
1 + 1' \sqrt{P} \text{diag}(n_j) \sqrt{P} 1 = 1 + P \sum n_j.
\]
Thus,
\[
(11'P + BB' + QQ')^{-1} = \begin{bmatrix}
  n_1 & \ldots & 0 \\
  \vdots & \ddots & \vdots \\
  0 & \ldots & n_N
\end{bmatrix} - \frac{P}{1 + P \sum n_j} \begin{bmatrix}
  n_1^2 & n_1 n_2 & \ldots & n_1 n_N \\
  n_1 n_2 & n_2^2 & \ldots & n_2 n_N \\
  \vdots & \vdots & \ddots & \vdots \\
  n_1 n_N & n_2 n_N & \ldots & n_N^2
\end{bmatrix}.
\]

Multiplying (A9) with unit vectors and combining terms then shows that (A8) condenses to
\[
\sum n_j p^2 + \left(1 - \rho^2 - \sigma^2 \sum n_j \right) P - \sigma^2 = 0, \quad n_j = m_j a_j^2 / b_j^2
\]
while (A7) becomes
\[
g = \frac{P (\sum n_j)}{1 + P (\sum n_j)} \begin{bmatrix}
  n_1 & \ldots & n_N
\end{bmatrix}.
\]
This completes the proof. □

**Proof of Proposition 2:** We start with the slope coefficient in (1). The estimated coefficient on the current level of output \( y_t \) is, due to the linearity of the covariance operator,\(^{29}\)
\[
\gamma = \text{Cov} \left[ y_{t+k} - \tilde{E}_t \left[ y_{t+k} \right], y_t \right] \text{Var} \left[ y_t \right]^{-1} = \int_0^1 \text{Cov} \left[ y_{t+k} - \tilde{E}_t y_{t+k}, y_t \right] \text{Var} \left[ y_t \right]^{-1} \, di
\]
\[
= \text{Cov} \left[ y_{t+k} - \tilde{E}_t y_{t+k}, y_t \right] \text{Var} \left[ y_t \right]^{-1}.
\]
The covariance in this expression equals
\[
\text{Cov} \left[ y_t - \tilde{E}_t y_t, y_t \right] = (\Sigma_j a_j) \rho^k \text{Cov} \left[ \theta_t - \tilde{E}_t \theta_t, y_t \right].
\]

\(^{29}\)We assume that individuals make the same attention choices \( m_j \). This is true in our model of optimal attention allocation in Section 4 and 5. It is also the workhorse assumption in the information choice literature (see, for example, Veldkamp, 2011).
To derive an expression for $\text{Cov}[\theta_t - E_{it}\theta_t, y_t]$, notice that we can always write output as

$$y_t = \sum_j x_{jt} = \sum_j m_j z_{ijt} + (x_{jt} - m_j z_{ijt})$$

$$= \sum_j m_j z_{ijt} + (1 - m_j) a_j \theta_t + (1 - m_j) b_j u_{jt} - m_j q_j \epsilon_{ijt}. \quad \text{(A13)}$$

As a result of (A13), $\text{Cov}[\theta_t - E_{it}\theta_t, y_t]$ can be expanded into

$$\text{Cov}[\theta_t - E_{it}\theta_t, y_t] = \text{Cov}(\theta_t - E_{it}\theta_t, y_t)$$

$$= \sum_j (1 - m_j) a_j \text{Var}[\theta_t | \Omega_{it}] \quad \text{(A14)}$$

$$- \text{Cov} \left[ E_{it}\theta_t, \sum_j (1 - m_j) b_j u_{jt} - m_j q_j \epsilon_{ijt} \right] \quad \text{(A15)}$$

because $\text{Cov}(\theta_t - E_{it}\theta_t, z_{ijt}) = 0$ and $\text{Var}[\theta_t | \Omega_{it}] = E[\theta_t - E_{it}\theta_t]^2 = \text{Cov}(\theta_t - E_{it}\theta_t, \theta_t)$. Furthermore, because it follows from (A6) that

$$(1 - m_j) b_j = \frac{q^2_j b_j}{q^2_j + b^2_j} \quad m_j q_j = \frac{b^2_j q_j}{q^2_j + b^2_j},$$

we find that (A15) becomes

$$\text{Cov} \left[ E_{it}\theta_t, (1 - m_j) b_j u_{jt} - m_j q_j \epsilon_{ijt} \right] = k_j \frac{q^2_j b^2_j}{q^2_j + b^2_j} - k_j \frac{q^2_j b^2_j}{q^2_j + b^2_j} = 0. \quad \text{(A16)}$$

Thus, from (A11), (A14), (A15), and (A16) we conclude that

$$\gamma = \bar{a}_k \sum_j a_j (1 - m_j), \quad \text{(A17)}$$

where $\bar{a}_k = p^k \frac{\text{Var}[\theta_t | \Omega_{it}]}{\text{Var}[y_t]} \sum_j a_j > 0. \quad \text{This completes the first step of the proof.}$

We now turn to the estimated coefficient in (2) on the average forecast revisions:

$$\delta = \text{Cov} \left[ y_{t+1} - E_t y_{t+1}, \bar{E}_{t-1} y_{t+1} - \bar{E}_{t-1} y_{t+1} \right] \text{Var} \left[ \bar{E}_{t-1} y_{t+1} - \bar{E}_{t-1} y_{t+1} \right]^{-1}, \quad \text{(A18)}$$

in which the covariance term reduces to

$$\text{Cov} \left[ y_{t+k} - \bar{E}_t y_{t+k}, \bar{E}_t y_{t+k} - \bar{E}_{t-1} y_{t+k} \right] = (\Sigma_j a_j)^2 \rho^{2k} \text{Cov} \left[ \theta_t - \bar{E}_t \theta_t, \bar{E}_t \theta_t - \bar{E}_{t-1} \theta_t \right]. \quad \text{(A19)}$$

\[\text{Notice that because of the symmetry of agents' attention choice Var}[\theta_t | \Omega_{it}] \text{ is the same across all } i \in [0, 1].\]
Hence, the key statistic for (A18) is the covariance between the nowcast error of the unobserved fundamental $\theta_t$ and its previous revision. To derive this covariance, we make use of Lemma 1. Specifically, because of Lemma 1, we can write the average expectation of $\theta_t$ in (A6) as

$$\bar{E}_t \theta_t = \bar{E}_{t-1} \theta_t + \tilde{g} \left( \bar{x}_t - \bar{E}_{t-1} \bar{x}_t \right),$$  

(A20)

where $\bar{x}_t = \sum_j \tilde{g}_{j1} x_t$. Furthermore, it follows from Lemma 1 that $\tilde{g}$ in (A20) is always less than its full attention value when at least one $m_j < 1$. Inserting (A20) into (A19), and using the same steps as those that lead to Proposition 1, then shows that $\delta > 0$ iff agents pay limited attention ($0 < m_j < 1$) to at least one component. □

Appendix B.2: General Linear Model

We extend the results from Section 3 to economies in which output is driven by several fundamentals, correlated disturbances, and where the structural components themselves can depend on their own history. This allows us to encapsulate most linearized macroeconomic models, including those with imperfect information.

Setup: We once more consider a discrete-time economy with a continuum of agents $i \in [0, 1]$. Output $y_t$ and its components $x_t$ are defined as

$$y_t = D\theta_t + Ex_t + Fu_t, \quad (A21)$$

$$x_t = A\theta_t + Bx_{t-1} + Cu_t, \quad (A22)$$

where $y_t$ is a scalar variable, $\theta_t$ is an $n_\theta \times 1$ vector of fundamental states, $x_t$ is an $n_x \times 1$ vector of structural components, and last $u_t$ is a $n_u \times 1$ vector of i.i.d. standard normal random variables. Most linear DSGE models can be written in this form (Fernández-Villaverde et al., 2007). The vector of fundamentals follows a simple VAR(1),

$$\theta_t = M\theta_{t-1} + Nu_t. \quad (A23)$$

Each agent $i \in [0, 1]$ observes the vector of signals

$$z_{it} = x_t + Q\epsilon_{it}, \quad Q = diag (q), \quad (A24)$$
where $\epsilon_{it}$ is an $n_x \times 1$ vector of i.i.d. standard normal random variables, which are uncorrelated with $u_t$. It is useful to re-state the system comprised of (A21) to (A23) as

$$y_t = \alpha \bar{\theta}_t + \beta u_t,$$

where $\alpha = \begin{bmatrix} D & E \end{bmatrix}$, $\bar{\theta}_t = \begin{bmatrix} \theta_t' & x_t' \end{bmatrix}'$, $\beta = F$, and in which

$$\bar{\theta}_t = \begin{bmatrix} \theta_t' & x_t' \end{bmatrix}' = \tilde{M}\bar{\theta}_{t-1} + \tilde{N}u_t,$$

where

$$\tilde{M} = \begin{bmatrix} M & 0 \\ AM & B \end{bmatrix}, \quad \tilde{N} = \begin{bmatrix} N \\ AN + C \end{bmatrix}.$$

We can now also re-state (A24) as

$$z_{it} = L_0\bar{\theta}_t + L_1\bar{\theta}_{t-1} + Ru_t + Q\epsilon_{it},$$

where $L_0$, $L_1$ and $R$ are implicitly defined.

**General Result:** We can now extend Proposition 2 to this more general case:

**Proposition 5.** If the economy evolves according to (A21)-(A24), then the population coefficients in the regression equations (1) and (2) satisfy:

$$\gamma < 0 \iff \alpha \tilde{M}^k (KQQ'F' + \Sigma_{\bar{\theta}\bar{\theta}} D' + \Omega) < 0 \quad (A25)$$

$$\delta > 0 \iff \exists q_j \in (0, \infty), \quad (A26)$$

where $\Sigma_{\bar{\theta}\bar{\theta}}$, denotes the covariance term $\Sigma_{\bar{\theta}\bar{\theta}} = \text{Cov} (\theta_t, \bar{\theta}_t)$ $K$ is the Kalman Gain on $z_{it}$ when forming expectations about $\bar{\theta}_t$, and $\Omega = \tilde{N} - K (L_0\tilde{N} + R)$.

Similar to the results in the main text, expectations are generically underresponsive in Proposition 5; $\delta > 0$ whenever agents pay limited attention to structural components. Likewise, informative, countercyclical components (that is, those that are assigned a large negative weight in $K$ or negative elements in $E$) also push expectations towards measured overreactions with $\gamma < 0$. In deriving this proposition, we have adjusted the latter condition for $(i)$ the direct impact of the persistent fundamental on output itself ($D \neq 0$), and $(ii)$ for the cross-correlation in errors between the signal vector and output ($\Omega \neq 0$). The second and third component of (A25) correspond to these additional effects.
Proof of Proposition 5: The proof proceeds in three steps: First, we derive an expression for the one-period ahead forecast error and the corresponding one-period ahead forecast revision. Then, we compute the extrapolation coefficient $\gamma$. Last, we use our results to also calculate the underreaction coefficient $\delta$.

To start, we use the Kalman Filter for systems with lagged states in the measurement equation (Nimark, 2015). This directly provides us with

$$
\mathbb{E}_{it} [y_{t+k}] = \alpha \mathbb{E}_{it} [\hat{\theta}_{t+k}] = \alpha \left\{ \mathbb{E}_{it-1} [\hat{\theta}_{t+k}] + G_k (z_{it} - \mathbb{E}_{it-1} [z_{it}]) \right\} 
= \mathbb{E}_{it-1} [y_{t+k}] + \alpha G_k (z_{it} - \mathbb{E}_{it-1} [z_{it}]),
$$

where $G_k$ is equal to

$$
G_k = \text{Cov} (\hat{\theta}_{t+k} - \mathbb{E}_{it-1} \hat{\theta}_{t+k}, z_{it} - \mathbb{E}_{it-1} z_t) \mathbb{V} [z_{it} - \mathbb{E}_{it-1} z_t]^{-1},
$$

and denotes the vector of Kalman Gains, such that

$$
\mathbb{E}_t [y_{t+k}] = \mathbb{E}_{t-1} [y_{t+k}] + \alpha G_k \left( x_t - \mathbb{E}_{t-1} [x_t] \right).
$$

We can then use (A28) to show that

$$
\begin{align*}
\mathbb{E}_t [y_{t+k}] - \mathbb{E}_{t-1} [y_{t+k}] &= \alpha G_k \left( x_t - \mathbb{E}_{t-1} [x_t] \right) \quad \text{(A29)} \\
y_{t+k} - \mathbb{E}_t [y_{t+k}] &= \alpha \left( \hat{\theta}_{t+k} - \mathbb{E}_t \hat{\theta}_{t+k} \right) + Fu_{t+k}. \quad \text{(A30)}
\end{align*}
$$

This completes the first step.

We are now ready to derive the overreaction coefficient $\gamma$. Combined, (A30) and the fact that $x_t = z_{it} - Q \epsilon_{it}$ and $y_{t+k} - \mathbb{E}_t [y_{t+k}] = \int_0^1 y_{t+k} - \mathbb{E}_t [y_{t+k}] \, di$ yields that

$$
\begin{align*}
\gamma &\propto \text{Cov} \left( \int_0^1 y_{t+k} - \mathbb{E}_t [y_{t+k}] \, di, \ y_t \right) = \int_0^1 \text{Cov} \left[ y_{t+k} - \mathbb{E}_t [y_{t+k}], \ E(z_{it} - Q \epsilon_{it}) + D \theta_t + Fu_t \right] \, di \\
&= \text{Cov} \left[ \alpha \left( \hat{\theta}_{t+k} - \mathbb{E}_t \hat{\theta}_{t+k} \right), -EQ \epsilon_{it} + D \theta_t + Fu_t \right] \\
&= \alpha M^k \left\{ \text{Cov} \left( \hat{\theta}_t - \mathbb{E}_t \hat{\theta}_t, -\epsilon_{it} \right) Q'E'x \right\} + \text{Cov} \left( \hat{\theta}_t - \mathbb{E}_t \hat{\theta}_t, \theta_t \right) D' + \text{Cov} \left( \hat{\theta}_t - \mathbb{E}_t \hat{\theta}_t, u_t \right) F'. \quad \text{(A31)}
\end{align*}
$$
But since
\[ \text{Cov} \left( \bar{\theta}_t - \text{E}_{it} \bar{\theta}_t, -\epsilon_{it} \right) = KQ \]
\[ \text{Cov} \left( \bar{\theta}_t - \text{E}_{it} \bar{\theta}_t, \theta_t \right) = \text{Cov} \left( \bar{\theta}_t - \text{E}_{it} \bar{\theta}_t, \theta_t - \text{E}_{it} \theta_t \right) = \Sigma_{\theta \theta} \]
\[ \text{Cov} \left( \bar{\theta}_t - \text{E}_{it} \bar{\theta}_t, u_t \right) = \bar{N} - K \left( L_0 \bar{N} + R \right) \]

because
\[ \text{E}_{it} \left[ \theta_t \right] = \text{E}_{it-1} \left[ \bar{\theta}_t \right] + K \left( z_{it} - \text{E}_{it-1} \left[ z_{it} \right] \right), \]
where \( G_k = \bar{M}^k K \) from the Kalman Filter, we can rewrite the condition in (A31) as
\[ \gamma \propto \alpha \bar{M}^k \left\{ KQ' E' + \Sigma_{\theta \theta} D' + \bar{N} - K \left( L_0 \bar{N} + R \right) \right\}. \]

This completes the second step of the proof.

Last, we compute the underreaction coefficient \( \delta \). Equation (A29) and (A30) show that
\[ \delta \propto \text{Cov} \left( y_{t+k} - \bar{E}_t [ y_{t+k} ], \bar{E}_t [ y_{t+k} ] - \bar{E}_{t-1} [ y_{t+k} ] \right) \]
can be rewritten as
\[ \delta \propto \alpha \text{Cov} \left( \bar{\theta}_{t+k} - \bar{E}_t \bar{\theta}_{t+k}, x_t - \bar{E}_{t-1} x_t \right) G_k' \alpha' \]
\[ = \alpha \text{Cov} \left( \bar{\theta}_{t+k} - \bar{E}_t \bar{\theta}_{t+k} - G_k \left( x_t - \bar{E}_{t-1} \left[ x_t \right] \right), x_t - \bar{E}_{t-1} x_t \right) G_k' \alpha' \]
\[ = \alpha \left\{ \bar{G}_k \text{V} \left[ x_t - \bar{E}_{t-1} x_t \right] - G_k \text{V} \left[ x_t - \bar{E}_{t-1} x_t \right] \right\} G_k' \alpha', \]

where
\[ \bar{G}_k = \text{Cov} \left( \bar{\theta}_{t+k} - \bar{E}_{t-1} \bar{\theta}_{t+k}, x_t - \bar{E}_{t-1} x_t \right) \text{V} \left[ x_t - \bar{E}_{t-1} x_t \right]^{-1}. \]

Thus,
\[ \delta \propto \alpha \left( \bar{G}_k - G_k \right) \text{V} \left[ x_t - \bar{E}_{t-1} x_t \right] G_k' \alpha', \]
\[ = \left( \bar{d}_k - d \right) \text{V} \left[ x_t - \bar{E}_{t-1} x_t \right] d'_k \quad (A32) \]

where \( \bar{d}_k = \alpha \bar{G}_k \) and \( d_k = \alpha G_k \). We note that the sign of \( \bar{d}_k \) is the same as that for \( d_k \), because \(| \bar{G}_{j,k} | > | G_{j,k} | \) (due to the noise in private signals) and \( \text{sign}(\bar{G}_{j,k}) = \text{sign}(G_{j,k}) \). We also note for the same reasons \(| \bar{d}_k | > | d_k | \). Combined, it now follows from (A32) that, because \( \text{V} \left[ x_t - \bar{E}_{t-1} x_t \right] \) is also positive semi-definite, \( \delta > 0 \) (Abadir and Magnus, 2005). \( \Box \)

**Alternative Proof of Proposition 2:** The model in Section 3 is a special case of the general structural model in the previous subsection. In particular, we obtain the model in Section 3
by setting:

\[ D = F = B = 0, \quad E = 1_{1 \times N} \]

\[ A = \begin{bmatrix} 0_{N \times 1}, \quad \text{diag}(a_1, \ldots, a_N) \end{bmatrix}, \quad C = \begin{bmatrix} 0_{N \times 1}, \quad \text{diag}(b_1, \ldots, b_N) \end{bmatrix} \]

\[ M = \rho, \quad N = \begin{bmatrix} \sigma_\theta, \quad 0_{1 \times N} \end{bmatrix} \]

An application of Proposition 5, with \( K \) evaluated according to the standard expression for Kalman Gains (Anderson and Moore, 2012), then also establishes Proposition 2.

**Appendix C: Optimal Attention Choice**

This Appendix derives agents’ ex-ante utility in (13) and states the proof of Proposition 3.

**Appendix C.1: Mean-squared Error Preferences**

We start by writing the optimal actions as

\[ a^*_t = \left( w_\theta + \sum_j w_{xj} a_j \right) \theta_t + \sum_j w_{xj} b_j u_{jt}. \] (A33)

From the Law of Total Variance, it follows that

\[ \mathbb{E} [a_t - a^*_t]^2 = \text{Var} [a^*_t | \Omega_{it}, \theta_t] + \text{Var} [\mathbb{E} [a^*_t | \Omega_{it}, \theta_t] | \Omega_{it}], \]

where because of (A33)

\[ \mathbb{E} [a^*_t | \Omega_{it}, \theta_t] = \left( w_\theta + \sum_j w_{xj} a_j \right) \theta_t + \sum_j w_{xj} b_j \mathbb{E} [u_{jt} | \Omega_{it}]. \]

Thus,

\[ \mathbb{E} [a_t - a^*_t]^2 = \text{Var} [a^*_t | \Omega_{it}, \theta_t] + \text{Var} [\mathbb{E} [a^*_t | \Omega_{it}, \theta_t] | \Omega_{it}], \]

\[ = \sum_j w_{xj}^2 b_j^2 (1 - m_j) + \text{Var} [\theta_t] \left( w_\theta + \sum_j w_{xj} a_j (1 - m_j) \right), \]

which provides us with the desired expression in (13).
Appendix C.2: Attention Choices

Proof of Proposition 3: An individual agent $i$’s attention choice problem can be written as

$$\max_{(m_j), V, \alpha, \tau} \left( \sum_j w_{xj}^2 b_j^2 (1 - m_j) - V \alpha - K(m) \right)$$

s.t. $V \geq V(\tau), \quad \alpha \geq w_\theta + \sum_j w_{xj} a_j (1 - m_j) \quad \tau \leq \sum_j a_j^2 b_j^2 m_j$

The Lagrangian for this problem is

$$\mathcal{L} = -\sum_j w_{xj}^2 b_j^2 (1 - m_j) - V \alpha - K(m) + \mu_V [V - V(\tau)] + \mu_\alpha \left[ \alpha - w_\theta - \sum_j w_{xj} a_j (1 - m_j) \right] + \mu_\tau \left[ \sum_j a_j^2 b_j^2 m_j - \tau \right]$$

The desired first-order condition is now obtained by rearranging $\frac{\partial \mathcal{L}}{\partial m_j} = 0$. □

Appendix C.3: Attention Choices with Entropy Costs

Suppose that the costs of attention are equal to the reduction in relative entropy:

$$I = \mu \lim_{T \to \infty} \frac{1}{T} \left\{ H(\theta^T, x^T) - H(\theta^T, x^T | z_i^T) \right\}, \tag{A34}$$

where $H(x|y)$ denotes the conditional entropy of $x$ given $y$, and $x^T$ denotes the history of the process $\{x_i\}_{i=-\infty}^T$. In this appendix, we first show that $I = K(m)$ for a well-defined cost function $K(\cdot)$, so that the reduction in entropy is merely a special case of our analysis in Proposition 3. We then derive the comparable first-order condition to (18).

We use the following properties of conditional entropy:

**Lemma 2.** Let $X$, $Y$ and $Z$ be random vectors. Then:

1. **Symmetry of mutual information:** $H(X) - H(X|Y) = H(Y) - H(Y|X)$
2. **Chain rule of conditional entropy:** $H(X, Y) = H(X) + H(Y|X)$
3. **Conditional independence:** If $Y$ is independent of $Z$ conditional on $X$, then

$$H(Y|X, Z) = H(Y|X)$$

**Proof of Lemma 2:** See Cover and Thomas (2012). □

---

31See, for example, Maćkowiak et al. (2018).
To start, let $s = \{\theta, x\}$. Symmetry and the chain rule for entropy, then allows us to write

$$H(s^T) - H(s^T | z_i^T) = H(z_i^T) - H(z_i^T | s^T)$$

$$= \sum_{t=1}^{T} H(z_{it}|z_{i}^{t-1}) - H(z_{it}|z_{i}^{t-1}, s^T).$$

(A35)

Note that conditional on $s_t = \{\theta_t, x_t\}$, the vector of signals $z_{it} = x_t + \text{diag}(q_j)\epsilon_{it}$ is independent of $s_{t'}$ for $t' \neq t$, since $\epsilon_{it}$ is serially uncorrelated. This, in turn, implies that

$$H(z_{it}|z_{i}^{t-1}) - H(z_{it}|z_{i}^{t-1}, s^T) = H(z_{it}|z_{i}^{t-1}) - H(z_{it}|z_{i}^{t-1}, s_t)$$

$$= H(s_t|z_{i}^{t-1}) - H(s_t|z_{i}^{t})$$

$$= H(\theta_t|z_{i}^{t-1}) - H(\theta_t|z_{i}^{t}) + H(x_t|z_{i}^{t-1}, \theta_t) - H(x_t|z_{i}^{t}, \theta_t),$$

(A36)

where the second equality follows from symmetry and the third from the chain rule for entropy.

For the first term in (A36), since all variables are jointly Gaussian, we have that

$$H(\theta_t|z_{i}^{t-1}) - H(\theta_t|z_{i}^{t}) = \frac{1}{2} \log \left[ \frac{\text{Var}_{t-1}[\theta_t]}{\text{Var}_t[\theta_t]} \right].$$

Now focus on the steady state where $\text{Var}_t[\theta_t] = \text{Var}_{t-1}[\theta_{t-1}] = V(\tau)$, with $\tau$ defined in (16). Using the AR(1) dynamics of $\theta_t$, we have that

$$\text{Var}_{t-1}[\theta_t] = \rho^2 V(\tau) + \sigma_\theta^2,$$

which after substituting gives us

$$H(\theta_t|z_{i}^{t-1}) - H(\theta_t|z_{i}^{t}) = \frac{1}{2} \log \left[ \rho^2 + \frac{\sigma_\theta^2}{V(\tau)} \right] \equiv K(\tau),$$

(A37)

in which $K'(\tau) > 0$ since $V'(\tau) < 0$.

For the second term in (A36), note that $x_t$ is independent of $z^{t-1}$ conditional on $\theta_t$, so that

$$H(x_t|z_{i}^{t-1}, \theta_t) = H(x_t|\theta_t)$$

$$= \frac{1}{2} \log \left[ \frac{\text{det}(\text{Var}[x_t|\theta_t])}{\text{det}(\text{Var}[x_t|\theta_t, z_{it}])} \right] = \frac{1}{2} \log \left[ \frac{\prod_{i=1}^{m} b_i^2}{\prod_{i=1}^{m} b_i^2 (1 - m_i)} \right]$$

$$= \frac{1}{2} \log \left[ \frac{1}{\prod_j (1 - m_j)} \right] = -\frac{1}{2} \sum_{j=1}^{m} \log(1 - m_j).$$

(A38)
Substituting (A38) and (A37) into (A36) then shows that

\[ I = K(\tau) - \frac{1}{2} \sum_{j=1}^{m} \log(1 - m_j) \equiv K(m), \]

which is well-defined since \( \tau \) is a function of \( m \). Last, combining (A35) with (A34) and using stationarity, we find that our cost function satisfies \( K(m) = I \).

We can now repeat the steps from the proof of Proposition 3 to show that the first-order condition for \( m_j \) at an interior optimum satisfies:

\[ w_j^2 b_j^2 + \mu_j a_j^2 + \mu \alpha w_j a_j = \frac{1}{2} \frac{1}{1 - m_j}, \tag{A39} \]

where the multiplier measuring learning spillovers is

\[ \mu = -\alpha V'(\tau) + K'(\tau). \]

The second term in (A39) is specific to the entropy cost formulation, because entropy reductions also depend on the sufficient statistic \( \tau \). The comparative statics remain the same as in our version with a generic cost function: It is optimal to pay attention to important components (high \( w_j \)), and to volatile components (high \( b_j \)) as long as spillovers are not too strong. In addition, we see that an entropy cost function naturally yields \( m_j < 1 \) for all \( j \): Attention is always imperfect because the entropy costs of full attention \( m_j \to 1 \) are infinite.

**Appendix C.4: Rational Inattention**

Suppose agents’ utility in (13) is equivalent to the mean-squared-error of next-period’s output forecast \((w_\theta = \rho \sum_j a_j \text{ and } w_{xj} = 0)\). It then follows from Section 4 that \( h = 0 \) in (19). As a result, the rational inattention signal becomes

\[ s_{it}^* = (\rho \sum_j a_j) \theta_t + q^* \epsilon_{it}, \tag{A40} \]

which is a simple noisy signal of \( \theta_t \). This signal is inconsistent with the extrapolation documented in Section 2. Consider the extrapolation coefficient in (1) based on (A40)

\[ \gamma = \text{Cov} (y_{t+k} - \bar{E}_t y_{t+k}, y_t) \text{Var}[y_t]^{-1} = \alpha_k \text{Cov} (\theta_t - \bar{E}_{it}\theta_t, y_t) = \alpha_k \sum_j a_j \text{Var} [\theta_t | s_{i,t}^*] > 0, \]

where we have used that \( y_t = \sum_j a_j \theta_t + \sum_j b_j u_{jt} \).
Appendix D: A Macroeconomic Model

This Appendix details the derivations used in Section 5.

Appendix D.1: Firm Choices

Proof of Proposition 4: We start with a firm’s output choice,

\[ Y_i = \arg \max \mathbb{E}_i \Pi_i = \mathbb{E}_i \left[ \frac{1}{PY} \left( PY \frac{1}{\sigma} Y_i^{1-\frac{1}{\sigma}} - WN_i \right) \right] = \mathbb{E}_i \left[ \frac{Y_i^{1-\frac{1}{\sigma}}}{Y} - \frac{W}{PY} \left( \frac{Y_i}{A_i} \right) \right]. \]

Thus,

\[ \Pi_i = \pi \left( Y_i, Y, A_i, \frac{W}{P} \right). \]

A second-order log-linear approximation of \( \Pi \) then results in

\[ \pi (y_i, y, a_i, \omega) \approx \pi_{11} y_i + \frac{\pi_{11}^2}{2} y_i^2 + \pi_{12} y_i y + \pi_{13} a_i y_i + \pi_{14} a_i \omega + t.i.a, \] (A41)

where \( \omega = w - p \) and \( t.i.a \) stands for terms independent of the firm’s action \( y_i \).

As a result of (A41), a firm’s optimal, full information choice of output is

\[ y_i^* = \frac{\pi_{12}}{\pi_{11}^2} y + \frac{\pi_{13}}{\pi_{11}^2} a_i + \frac{\pi_{14}}{\pi_{11}^2} \omega, \] (A42)

while a firm’s optimal choice under imperfect information is

\[ y_i = \mathbb{E}_i [y_i^*]. \] (A43)

It remains to derive the optimal output choice under full information in (A42). A few simple but tedious derivations combine to show that

\[ y_i^* = ra_i + \alpha r \left( \sigma^{-1} y - \omega \right) \equiv x_{i1} + x_2. \] (A44)

Substituting the expression for the real wage \( \omega \) in (A44) then completes the proof. \( \square \)

Firm Profit Expression: We can use (A42) and (A43) to derive the difference between a firm’s

\[ \text{Firm Profit Expression: We can use (A42) and (A43) to derive the difference between a firm’s} \]

\[ \text{Since all actions are taken within period, we remove time subscripts, to economize on notation.} \]

\[ \text{Since all actions are taken within period, we remove time subscripts, to economize on notation.} \]
realized profits $\pi_i = \pi(y_i, y, a_i, \omega)$ and those that would have arisen under full information $\pi^*_i$:

$$\pi_i - \pi^*_i = \frac{\pi_{11}}{2} y_i^2 - \frac{\pi_{11}}{2} y^*_i^2 + (\pi_{12} y + \pi_{13} a_i + \pi_{14} \omega) (y_i - y^*_i)$$

$$= \frac{\pi_{11}}{2} y_i^2 - \frac{\pi_{11}}{2} y^*_i^2 - \pi_{11} y^*_i (y_i - y^*_i) = \frac{\pi_{11}}{2} (y_i - y^*_i)^2,$$

where we have used the first-order condition for optimal output from (A41). This documents the quadratic expression used for firm profits in Section 5.

**Appendix D.2: Equilibrium Characterization**

We solve the model backwards. First, we solve for the imperfect information equilibrium given a set of attention choices. Then, we solve for the optimal attention choices.

**Step 2: Equilibrium Given Attention Choices:** Consider the equation for aggregate output that arises from Lemma 4:

$$y_t = \int_0^1 y_{it} di = \bar{E}_t [x_{1t} + x_{2t}], \quad (A45)$$

where $x_{1t} = \int_0^1 x_{1it} di$ and

$$x_{1t} = r\theta_t + ru^x_t, \quad x_{2t} = \alpha r \sigma^{-1} y - \alpha r \left( \bar{E}_t [y_t] + u_t^\xi \right).$$

Now let $\mathbf{x}_t = \begin{bmatrix} \bar{x}_{t-1} & \bar{x}_{t-2} & \ldots \end{bmatrix}'$ where $\bar{x}_t = \begin{bmatrix} x_{1t} & x_{2t} & \theta_t \end{bmatrix}'$. We look for linear equilibria where the law of motion for the unobserved components and the fundamental takes the form of the infinite dimensional vector

$$\mathbf{x}_t = A \mathbf{x}_{t-1} + B u_t, \quad u_t = \begin{bmatrix} u^0_t & u^x_t & u^a_t \end{bmatrix}' \quad (A46)$$

where

$$A = \begin{bmatrix} 0 & 0 & r \rho \theta & 0 \\ 0 & A_p & \rho \theta & 0 \\ 0 & 0 & \rho \theta & 0 \\ 0 & 0 & \rho \theta & 0 \end{bmatrix}, \quad B = \begin{bmatrix} r & r & 0 \\ B_p & 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (A47)$$

To solve for the rational expectations equilibrium, we conjecture and verify below that

$$y_t = \psi \mathbf{x}_t, \quad x_{2t} = c_0 \mathbf{x}_t + c_1 u_t, \quad (A48)$$

The steps used to find this equilibrium are identical to those described in Lorenzoni (2009).
where \( \psi, c_0, \) and \( c_1 \) are vectors of coefficients.

**Coefficients and Conjectures:** It follows from (A45) that

\[
y_t = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \bar{E}_t [x_t] \overset{\text{should}}{=} \psi x_t,
\]

where \( \overset{\text{should}}{=} \) denotes “should equal”. We conclude from (A49) that to verify our conjecture we need to find a matrix \( \Xi \) such that

\[
\bar{E}_t [x_t] = \Xi x_t.
\]

Now since

\[
\bar{E}_t^b [y_t] = \psi \left\{ A x_{t-1} + B \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} u_t \right\} = \psi x_t - \psi B \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} u_t,
\]

it also follows that

\[
x_{2t} = \alpha r^{\sigma^{-1}} \psi x_t - \alpha r \left\{ \psi x_t - \psi B \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} u_t + c_3 u_t \right\} \overset{\text{should}}{=} c_0 x_t + c_1 u_t,
\]

where \( e_l \) denotes a row vector with a one in the \( l \)'s position but zero elsewhere.

**Individual and Average Inference:** An individual firm’s signal vector is

\[
s_{it} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} x_t + Q \epsilon_{it}, \quad Q = diag \left[ q_1 \ q_2 \right].
\]

\[
\equiv \ L x_t + Q \epsilon_{it}
\]

Thus,

\[
\bar{E}_{it} [x_t] = A \bar{E}_{it-1} [x_t] + K (s_{it} - L \bar{E}_{it-1} [z_{t-1}]) ,
\]

where the Kalman Gain \( K \) is given by the standard expression (Anderson and Moore, 2012).

Then, from (A50) and (A53) it has to hold for all \( t \) that

\[
\Xi x_t = (I - KL) A \Xi x_{t-1} + KL x_t
\]
Fixed Point: We have from (A49) and (A51) that
\[
\psi = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \Xi, \quad c_0 = \alpha r \left( \sigma^{-1} - 1 \right) \psi, \quad c_1 = \alpha r \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + c_3 \tag{A55}
\]

Equilibrium and Computation: An equilibrium is characterized by (i) a set of coefficients that describe aggregate dynamics \( \{A_p, B_p, \psi, c_0, c_1\} \), and (ii) a set of coefficients that detail the learning dynamics \( \{K, \Xi\} \). Computing the equilibrium requires truncating the infinite-dimensional vector \( \mathbf{x}_t \). Specifically, we instead consider the vector \( \mathbf{x}_t^{[T]} = [\mathbf{x}_{t-1}' \mathbf{x}_{t-2}' \ldots \mathbf{x}_{t-T}']' \).

To find the equilibrium, we apply the following algorithm: We start with some initial values for \( A_p \) and \( B_p \) (for simplicity, we use those from the corresponding full information solution). We then use these values to compute \( K \) from (A52) and (A53). This, in turn, allows us to find an expression for \( \Xi \) from (A54) since
\[
\Xi \mathbf{x}_t^{[T]} = (I - KL) A \Xi M \mathbf{x}_t^{[T]} + KL \mathbf{x}_t^{[T]},
\]
where
\[
M = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},
\]
which gives us the following relationship that we solve for \( \Xi \)
\[
\Xi = (I - KL) A \Xi M + KL. \tag{A56}
\]
We can now use (A55) to find an expression for \( \psi, c_0, \) and \( c_1 \).

Last, we use these expressions to compute new values of \( A_p \) and \( B_p \) from (A47), and then repeat these steps until convergence is achieved. The criterion used is that maximum absolute difference between the new and old elements of \( A_p \) and \( B_p \).

Step 1: Attention Choices Given Equilibrium: Given the above aggregate equilibrium, we solve a firm’s \textit{ex-ante} attention choice problem. That is, we solve
\[
\min_{m_1, m_2} \mathbb{E}_{it} [y_{it} - y_{it}^*]^2 + K(m), \quad K(m) = \mu \left( q_1^{-2} + q_2^{-2} \right), \tag{A57}
\]
where \( q_j = \text{Var} (x_{jt} | \theta_t) [m_j - \text{Var} (x_{jt} | \theta_t)]^{-1} \) for \( j = \{1, 2\} \) and we have that
\[
y_{it}^* = x_{i1t} + x_{2t},
\]
62
in which

\[ x_{i1t} = r \theta_t + ru_t^2 + re_{it} = re'_{T}x_i^T + re'_{T}r_xu_t \equiv a_1x_i^T + b_1u_t + e_{it}^2 \]

\[ x_{2t} \equiv a_2x_i^T + b_2u_t, \]

and where \( a_1 \) and \( b_1 \) are implicitly defined while \( a_2 = c_0 \) and \( b_2 = c_1 \).

To minimize (A57), we first derive an expression for the quadratic component

\[ \mathbb{E} [y_{it}^* - \mathbb{E}[y_{it}^*]]^2 = 1' \mathbb{V} [x_{it} | z_i^t] 1, \quad x_{it} = \begin{bmatrix} x_{i1t} & x_{2t} \end{bmatrix}' \]

in which

\[ \mathbb{V} [x_{it} | z_i^t] = \mathbb{V} [x_{it} | z_{it}, x_i^T] + \mathbb{V} [\mathbb{E} [x_{it} | z_i^t, x_i^T] | z_i^t] \quad (A58) \]

by the Law of Total Variance.

It now follows that the first component in (A58) is

\[ \mathbb{V} [x_{it} | z_i^t, x_i^T] = \mathbb{V} [x_{it} | z_{it}, x_i^T] = bb' + \bar{r}r' - (bb' + \bar{r}r') [bb' + QQ' + \bar{r}r']^{-1} (bb' + \bar{r}r')' = bb' + \bar{r}r' - \bar{m} (bb' + \bar{r}r')' \]

where \( b = \begin{bmatrix} b_1 & b_2 \end{bmatrix}' \), \( \bar{r} = \begin{bmatrix} r \sigma_a & 0 \end{bmatrix}' \), and \( \bar{m} = (bb' + \bar{r}r') [bb' + QQ' + \bar{r}r']^{-1} \).

To derive the second component in (A58), notice that

\[ \mathbb{E} [x_{it} | z_i^t, x_i^T] = \mathbb{E} [x_{it} | z_{it}, x_i^T] = \mathbb{E} [x_{it} | x_i^T] + \bar{m} \left( z_{it} - \mathbb{E} [z_{it} | x_i^T] \right) = (I - \bar{m}) a x_i^T + \bar{m} z_{it}, \]

where \( a = \begin{bmatrix} a_1 & a_2 \end{bmatrix}' \). Thus,

\[ \mathbb{V} \left[ \mathbb{E} [x_{it} | z_i^t, x_i^T] | z_i^t \right] = (I - \bar{m}) a \mathbb{V} [x_i^T | z_i^t] a' (I - \bar{m})', \]

in which \( \mathbb{V} [x_i^T | z_i^t] \) can be found from the Kalman Filter run in (A53).

In sum, we have that the quadratic term (A57) becomes

\[ \mathbb{E} [y_{it}^* - \mathbb{E}[y_{it}^*]]^2 = 1' \left[ bb' + \bar{r}r' - \bar{m} (bb' + \bar{r}r')' \right] 1 + 1' (I - \bar{m}) a \mathbb{V} [x_i^T | z_i^t] a' (I - \bar{m})', \]

which allows us to solve the problem in (A57).

**Equilibrium:** We iterate on two steps described in Step 1 and Step 2 until convergence. As
a convergence criteria, we use the maximum absolute difference in attention coefficients. We use the full information case in which $m_1 = m_2 = 1$ as the initial values.
References


