Asymmetric Attention*

Alexandre N. Kohlhas   Ansgar Walther

September 2020

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: alexandre.kohlhas@iies.su.se). Walther: Imperial College London (email: a.walther@imperial.ac.uk). We are indebted to George-Marios Angeletos, Ryan Chahrour, Martin Ellison, Nicola Gennaioli, Per Krusell, Kristoffer Nimark, Robert Ulbricht, David Stromberg, Laura Veldkamp, 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.[1] 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 aggregate 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 range of survey data.[2] The participants of standard surveys, reporting their expectations about future output and inflation, not only extrapolate from recent conditions, but also underreact to aggregate information (as measured by average forecast revisions).

We show that a popular class of 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; Maćkowiak and Wiederholt, 2009), 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, extrapo-

[1]See, for example, Barberis *et al.* (2016), Bordalo *et al.* (2017), and the references therein.

[2]Specifically, 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.

lation bias or overconfidence generates overreactions. We show that, on balance, when agents' process signals of the forecasted variable, only one of these forces can dominate. In addition, we find the same result extends to several influential models with a richer information structure (e.g., Lucas, 1973; Lorenzoni, 2009; Angeletos *et al.*, 2018). 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 simultaneously explain extrapolation and underreaction to aggregate information, 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 spirit of Angeletos *et al.* (2016). 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 productivity component*, which reflects their own productivity; and (*ii*) firm beliefs about *an aggregate supply 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 aggregate supply 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.[3] 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 areas 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 recent outcomes of the forecasted variable with underreactions in *average* forecast revisions. In contemporaneous and closely related work, Bordalo *et al.* (2018) propose a behavioral model that can reconcile similar underreactions to *average* forecast revisions with overreactions to *individual* forecast revisions. 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.[4]

[3] 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.

[4] We show in Section 3.4 that the evidence for overreactions to individual-level revisions is mixed in our

Second, in common with a vast literature in macroeconomics since Lucas (1972), we emphasize 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 full information and 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 popular set of existing models and argue that these models cannot account for this observation.

data. We also provide an extension of our model with asymmetric attention that can fit the stylized facts in both papers. Crucially, this would not be possible in a model with symmetric attention choices.

2.1 Simultaneous Over- and Underreactions

We start by considering forecasts of US output growth from the *Survey of Professional Forecasters* (SPF).[5] 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, expectations about future output play a central role in the economy as determinants of consumption, inflation, and asset prices. Second, data on output forecasts are available for a longer time-span than forecasts of most other variables. We later explore the robustness of our empirical estimates by considering forecasts about future inflation, as well as alternative survey datasets for the US and the Euro Area.

We let y_{t+k} denote year-on-year output growth at time $t+k$. Consider a survey with respondents indexed by $i \in \{1, 2, ..., I\}$, and let $f_{it}y_{t+k}$ denote the forecast of y_{t+k} reported by survey respondent i at time t. The respondent's *forecast error* is $y_{t+k} - f_{it}y_{t+k}$. 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 individual forecast errors should be unpredictable. Under FIRE, no variable that is observable at time t should correlate with $y_{t+k} - f_{it}y_{t+k}$. We rely on two common tests of this prediction.

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

$$y_{t+k} - f_{it}y_{t+k} = \alpha_i + \gamma y_t + \xi_{it}, \tag{1}$$

where α_i is a constant, which also capture individual fixed effects, and ξ_{it} is an error term. The second test is a regression of forecast errors on average forecast revisions,

$$y_{t+k} - f_{it}y_{t+k} = \alpha_i + \delta\left(\bar{f}_t y_{t+k} - \bar{f}_{t-1}y_{t+k}\right) + \xi_{it}. \tag{2}$$

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 respondents' forecasts when they 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. This specification closely follows the test proposed by Coibion and Gorodnichenko (2015).[6]

[5] 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).

[6] Coibion and Gorodnichenko (2015) use *average* forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the dependent variable in (2). We prefer the individual-level regressions because it is easier to compare to candidate theories of individual expectation formation, and also because it allows for respondent-level fixed effects and assigns equal weight to all individual forecasts in an unbalanced panel. For completeness, we report both average- and individual-level estimates throughout the paper and the online appendix.

The prediction of the FIRE benchmark is that the coefficients γ and δ in (1) and (2) should both be zero, because both current output growth and the latest forecast revision are observable at time t.[7] It is useful to note that both (1) and (2) are tests of the *joint hypothesis* of full information and rational expectations. A rejection of the FIRE prediction reveals *either* that forecasters are reporting irrational expectations, *or* that they have imperfect information about current output (for $\gamma \neq 0$) or average forecast revisions (for $\delta \neq 0$).

The raw data already hints at deviations from the FIRE benchmark. Figures 1 and 2 plot average one-year-ahead forecast errors (the average left-hand side of (1) and (2) across respondents, with $k = 4$) over time and compares them, respectively, to current realizations of output growth (the right-hand side of (1)) and average one-quarter revisions (the right-hand side of (2)). In Figure 1, forecasts are frequently over-optimistic, with associated negative forecast errors when current output growth is high, and vice versa when current growth is low. This suggests that respondents 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 average 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 after negative revisions.

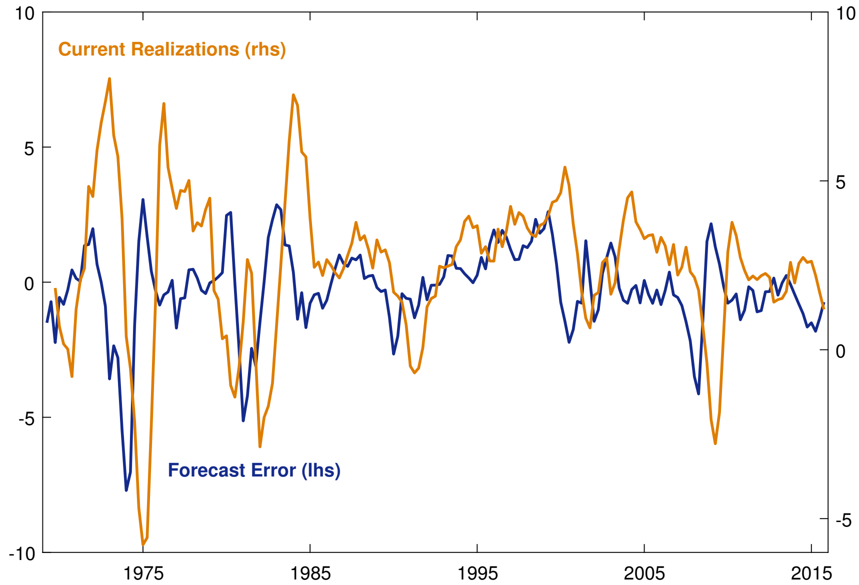
Table I confirms these impressions and reports estimates of (1) and (2) using the SPF data on one-year-ahead forecasts ($k = 4$). In the first column, we estimate (1) and find that γ is negative and statistically significant. This once more suggests extrapolation, or *overreactions* to recent realizations of output growth. In the second column, we estimate (2) using one-quarter average revisions. We find that δ 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. The multivariate estimates are similar to those in the univariate case. This suggests that the univariate results are not biased by correlation between output realizations and forecast revisions.

Taken individually, the over- and underreactions documented in Table I are in line with previous estimates. Bordalo *et al.* (2017), for example, report evidence on extrapolation based on the average-level version of regression (1). For regression (2), our estimates update those reported by Coibion and Gorodnichenko (2015, Figure I). Our results demonstrate that, in addition, extrapolation and underreactions occur *simultaneously* in the SPF data.

In contemporaneous and related work, Bordalo *et al.* (2018) analyze a different type of "overreactions" in survey expectations to that documented in Table I. Specifically, Bordalo *et al.* (2018) analyze overreactions to individual forecast revisions. By contrast, we use re-

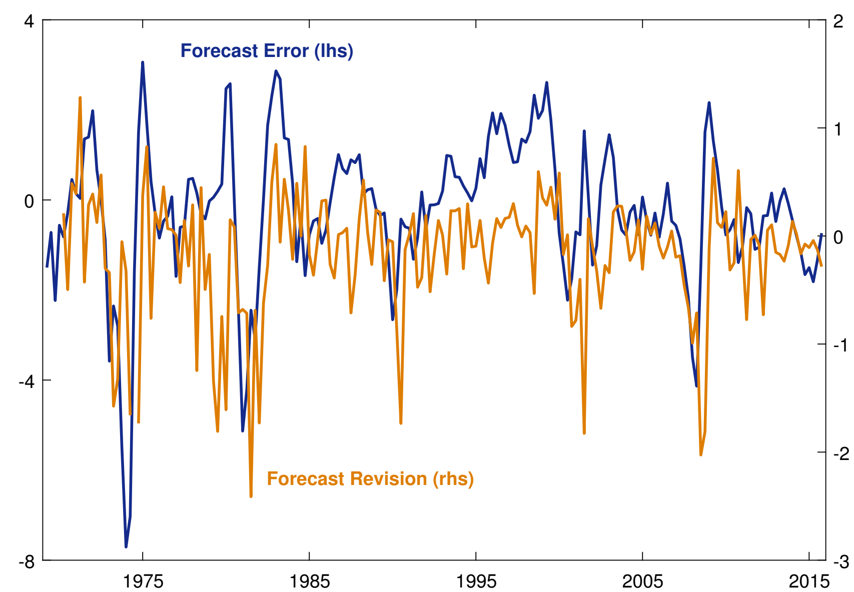
[7] We use real-time data to measure current realizations of output growth to precisely capture the definition of the output variable being forecasted (Croushore, 1998).

Figure 1: Overreactions in Output Growth Forecasts



Note: Mean one-year ahead forecast error of output growth from the *Survey of Professional Forecasters* on the left vertical axis, and the current realization on the right axis. Both scales are in percent year-on-year.

Figure 2: Underreactions in Output Growth Forecasts



Note: Mean one-year ahead forecast error of output growth from the *Survey of Professional Forecasters* on the left vertical axis, and the one-quarter revisions on the right axis. Both scales are in percent year-over-year.

Table I: Estimated Over- and Underreactions in the *SPF*

	Panel a: individual forecast error		
	(1)	(2)	(3)
Current Realization	-0.14***	–	-0.14***
	(0.05)		(0.05)
Average Revision	–	0.67***	0.66***
		(0.20)	(0.18)
Observations	6,821	6,782	6,725
F	244.8	411.8	354.3
R^2	0.04	0.06	0.10
	Panel b: average forecast error		
	(1)	(2)	(3)
Constant	0.07	-0.10	0.25*
	(0.18)	(0.10)	(0.15)
Current Realization	-0.12**	–	-0.14***
	(0.05)		(0.05)
Average Revision	–	0.77***	0.80***
		(0.26)	(0.25)
Observations	188	187	186
F	5.57	15.4	11.9
R^2	0.03	0.08	0.12

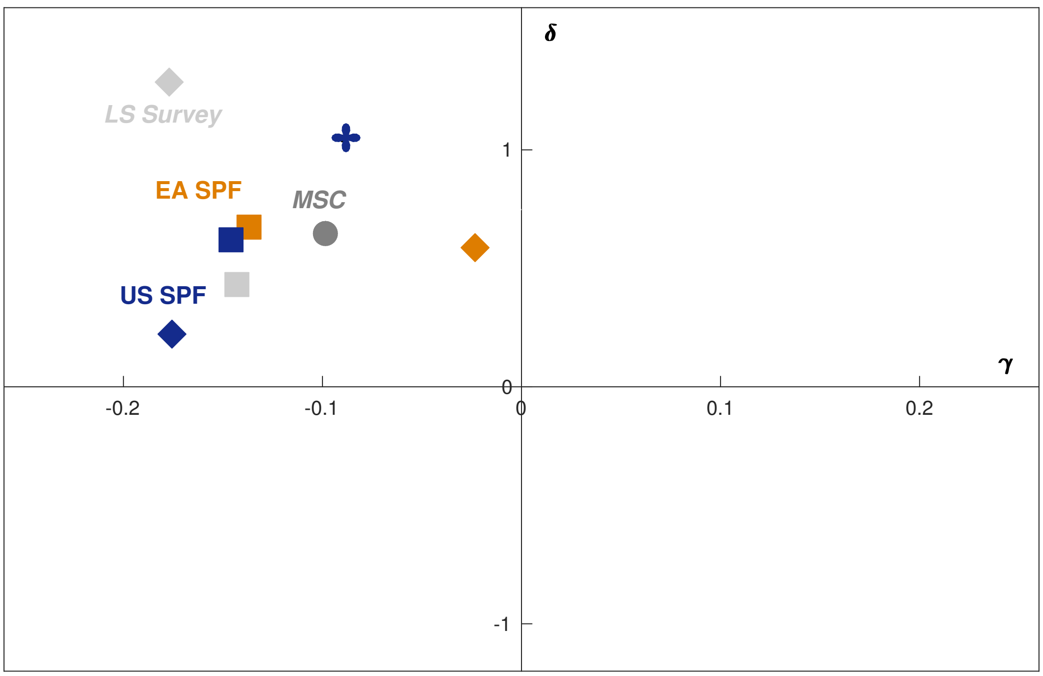
Note: Panel a: estimates of regressions (1) and (2) with individual (respondent) fixed effects. The top and bottom one percent of forecast errors and revisions have been removed. Table C.1 in the online appendix shows similar results without removing outliers. Double-clustered robust standard errors in parentheses. Panel b: estimates of regressions (1) and (2) with average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable. Robust standard errors in parentheses. Sample: 1970Q1-17Q4. Significance levels *=10%, **=5%, ***=1%.

gression (1) to emphasize that there are overreactions to recent realizations of the forecasted variable. For now, we continue to focus on our regression (1). In Section 3.4. we provide a detailed discussion of these distinct notions of overreaction.

We obtain similar estimates beyond forecasts of output growth in the US SPF. Figure 3 summarizes estimates of (1) and (2) for output and inflation 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*.[8]

[8]The *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

Figure 3: Estimated Over- and Underreactions Across Surveys



Note: Estimates of γ and δ from (1) and (2) using individual forecast errors $y_{t+k} - f_{it}y_{t+k}$ as the dependent variable. *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 lastly *MSC* the *Michigan Survey of Consumers.* $\square$ = GDP forecasts, $\diamond$ = CPI Inflation forecasts, $\star$ = GDP deflator inflation forecasts, and $\circ$ = MSC CPI 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). Figures C.1 and C.2 in the online appendix illustrate the robustness of the above estimates to alternative sample assumptions and the use of average forecast errors as the dependent variable.

We plot the coefficient γ on current realizations in (1) on the horizontal axis in Figure 3, and the coefficient δ on average forecast revisions in (2) on the vertical axis.[9] 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). Table C.6 in the online appendix contains the associated regression results. Specifically, with the exception of the Euro Area and Livingstone CPI inflation forecasts, and the GDP deflator forecasts from the US SPF, all overreaction coefficients in Figure 3 are statistically significant at the five percent level.

Tables C.2-8 in the online appendix contains further robustness checks. We show that simultaneous over- and underreactions extend to multivariate versions of (1) and (2), to the use of average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the dependent variable, and to different forecast horizons,[10] timing conventions, and assumptions about trends in the data. We also split the sample and find similar patterns in the post-1992 sample (to account for any potential structural break in the inflation series), as well as both pre- and post-Great Moderation.

Finally, we also consider two alternative tests from the literature to confirm the robustness of our results. First, following Coibion and Gorodnichenko (2015), we report estimates of the unconstrained version of (2) with potentially different coefficients on $\bar{f}_t y_{t+k}$ and $\bar{f}_{t-1} y_{t+k}$ (Table C.5). We fail to reject the null hypothesis that the coefficients sum to zero, validating the specification in (2). Second, in Online Appendix D, we consider the projection of forecast errors and current output growth on identified productivity shocks, as in Coibion and Gorodnichenko (2012). Consistent with underreactions, we find a positive correlation between the conditional response of forecast errors and the response of output growth.

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 to average forecast revisions. This clearly constitutes a rejection of the joint hypothesis of full information and rational expectations. In the next subsection, we consider a range of existing models that relax either full information or rational expectations. We argue that one can also use our stylized facts to determine whether existing alternative theories of expectation

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 Forecasts* collects the same information as the SPF for the US.

[9] Some of our estimates of (2) are direct updates of estimates reported by Coibion and Gorodnichenko (2015) using average forecast errors as the dependent variable. In particular, Coibion and Gorodnichenko (2015) also report estimates of (2) using CPI inflation forecasts from the Livingstone Survey and the Michigan Survey of Consumers, GDP deflator inflation forecasts from the US SPF, as well as inflation forecasts from the Euro Area (although from the Consensus Economic Survey and not the Euro Area SPF). All of these estimates are comparable to ours. Relative to their work, we focus on *simultaneous* estimates of (2) and (1), and cover a wider range of data sources for output growth forecasts, which are the focus of our analysis.

[10] The point estimates with shorter forecast horizons decline in magnitude and significance. This is consistent with a greater importance of noise in shorter horizon forecasts (see also Coibion and Gorodnichenko).

formation are consistent with the data.

2.2 Existing Theories of Expectation Formation

In this subsection, we compare our empirical evidence with several popular theories of expectation formation. On one hand, we consider *rational* forecasts in a simple model in which agents observe output with noise, as well as a collection of richer models in the existing literature. We show these models cannot explain overreactions to current output (i.e. our estimates of $\gamma < 0$ in regression (1)). On the other hand, we show that several popular *behavioral* alternatives, which are able to generate $\gamma < 0$, cannot simultaneously generate underreactions to average information (i.e., $\delta > 0$ in regression (2)).

Consider 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:

$$y_t = \rho y_{t-1} + u_t, \quad u_t \sim \mathcal{N}\left(0, \sigma_u^2\right), \tag{3}$$

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,

$$z_{it} = y_t + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right), \tag{4}$$

where the noise in agents' signals ϵ_{it} is independent of u_t at all horizons, with $\mathbb{C}\mathrm{ov}\left[\epsilon_{it}, \epsilon_{js}\right] = 0$ for all $i \neq j$ and $t \neq s$. We write $\Omega_{it} = \{z_{is}\}_{s \leq t}$ for agent i's information set at date t.[11]

Rational forecasts of output in this environment are determined by the Kalman filter:

$$f_{it} y_t = f_{it-1} y_t + g\left(z_{it} - f_{it-1} y_t\right) \tag{5}$$

$$f_{it} y_{t+k} = \rho^k f_{it} y_t \tag{6}$$

where $g = \frac{\mathbb{V}\mathrm{ar}[y_t|\Omega_{it-1}]}{\mathbb{V}\mathrm{ar}[y_t|\Omega_{it-1}] + \sigma_\epsilon^2} \in (0,1)$ is the Kalman gain. In terms of forecasts about output k periods ahead, the Kalman filter implies the following *Recursive Forecast Equation:*

$$f_{it} y_{t+k} = f_{it-1} y_{t+k} + g_k\left(z_{it} - f_{it-1} z_{it}\right), \tag{7}$$

where $g_k = \rho^k g \in \left(0, \rho^k\right)$ denotes sensitivity of forecasts about y_{t+k} to information at time t. We also consider existing behavioral models of expectations that are based on modifications of this recursive equation. Concretely, we characterize forecasts in the following cases:

[11] We allow agents to observe an infinite history of signals, so that their signal extraction problem is initialized in steady state at date 0. This assumption follows the convention in e.g. Maćkowiak *et al.* (2018).

- *Noisy Rational expectations (Limited Attention):* Agents forecasts follow (7) with $g_k \leq \rho^k$. This specification is identical to those from models with noisy rational expectations (Woodford, 2002) or rational inattention (Sims, 2003).[12] The special case in which agents observe output without noise ($\sigma_\epsilon^2 = 0$) corresponds to the case of full information and rational expectations (FIRE), and implies that $g_k = \rho^k > 0$.

- *Overconfidence* (e.g., Daniel *et al.*, 1998; Hirshleifer *et al.*, 2011): Agents overestimate the precision of their own information. They believe that the variance of the noise in their signals of output is $\hat{\sigma}_\epsilon^2 < \sigma_\epsilon^2$.[13] Agents forecasts follow the Recurse Forecast Equation (7) with a sensitivity parameter g_k that exceeds its rational value.

- *Extrapolation* (e.g., De Long *et al.*, 1990; Fuster *et al.*, 2012): Agents overestiamte the extent to which current output predicts future realizations. They observe output without noise ($\sigma_\epsilon^2 = 0$), but believe that the persistence parameter in the output equation (3) is $\hat{\rho} > \rho$. Agents' forecasts satisfy (7) with a sensitivity parameter $g_k = \hat{\rho}^k > \rho^k$.[14]

- *Diagnostic Expectations* (Bordalo *et al.*, 2017, 2018): The model is related to the overconfidence case, but the effect of overconfidence is temporary and does not affect forecasts at future dates. Equation (7) is replaced by $f_{it}y_{t+k} = \mathbb{E}_{it-1}y_{t+k} + g_k\left(z_{it} - \mathbb{E}_{it-1}y_t\right)$, where $\mathbb{E}_{it}\left[\cdot\right]$ denotes the conditional expectation operator, and g_k exceeds its rational value.

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) and (4).

Proposition 1. *Consider forecasts defined by (i) noisy rational expectations or limited attention, (ii) overconfidence, (iii) extrapolation, or (iv) diagnostic expectations. Then, the coefficients γ in (1) and δ in (2) both have the same sign as $\rho^k - g_k$.*

This proposition demonstrates that the rational model and the three behavioral models we have considered all imply either underreactions in both of our main regressions ($\gamma > 0$ and $\delta > 0$), or overreactions in both regressions ($\gamma < 0$ and $\delta < 0$). This prediction is at odds with our empirical estimates of simultaneous overreactions to recent output and underreactions to average forecast revisions. One can see this discrepancy clearly in terms of Figure 3. Proposition 1 shows that the econometrician's estimates in the alternative models we have considered will fall either into the upper-right quadrant or the lower-left quadrant of the figure. This is inconsistent with our empirical estimates, which fall into the upper-left quadrant.

[12] A 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).

[13] For further analysis of overconfidence, see Broer and Kohlhas (2019) and the references therein.

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

Proposition 1 reveals a fundamental link between rational and behavioral theories of expectation formation. Recall that agents' gain parameter in the FIRE case, where they perfectly observes current output, is equal to $g_k = \rho^k$. 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. In all cases, deviations from the FIRE benchmark are characterized by the sufficient statistic $\rho^k - g_k$, which measures the responsiveness of forecasts to current signals relative to FIRE benchmark.

Two counteracting effects determine the size of g_k. First, noise in agents' signals reduces the rational responsiveness of forecasts, and all else equal pushes g_k below its FIRE value. This effect generates underreactions. Intuitively, an econometrician running (1) and (2) has more information than agents in the model because he observes output and average revisions perfectly. Therefore, forecast errors are predictable. Furthermore, because agents respond to noisy information in a muted way, this predictability takes the shape of measured underreactions. Second, the behavioral biases we have considered heighten agents' responsiveness to current information, and push g_k above its FIRE value. This effect generates overreactions.

In the rational model with noisy information, only the first effect exists, so that we observe underreactions in both regressions (1) and (2). In the behavioral models characterized by Proposition 1, overreactions and underreactions are possible, but only one of these forces can come to dominate the sufficient statistic $\rho^k - g_k$. Hence, in all of the above cases, agents either over- or underreact, but do not over- and underreact *simultaneously*.[15]

It follows that one must consider models with more than one sufficient statistic for belief formation to explain the survey data. In the next section, we achieve this aim by proposing a noisy rational expectation model in which agents pay limited but asymmetric attention to different structural components of the forecasted variable. Before turning to our model, we briefly consider more sophisticated existing models of rational expectations.[16]

We focus on richer models from two influential strands of literature. First, the literature on rational inattention includes more sophisticated models following Maćkowiak and Wiederholt (2009), in which agents rationally allocate their attention between aggregate and individual-specific conditions. Individual-specific conditions, and the signals that agents obtain about them, are uncorrelated with aggregate output by assumption. Hence, forecasts about output

[15]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.

[16]In addition, Online Appendix E characterizes the more sophisticated behavioral model in Angeletos *et al.* (2018), who introduce a small deviation from rational expectations into a model of dispersed information. Intuitively, agents in their model adjust their expectations in proportion to exogenous confidence shocks. We show that this model predicts *overreactions* to both output and average revisions (i.e., $\gamma < 0$, $\delta < 0$).

behave *as if* agents obtained only a noisy signal of output. Indeed, Online Appendix E.1 shows that (7) exactly describes output expectations in Maćkowiak and Wiederholt (2009).

Second, we consider models with dispersed information, in which agents observe local economic conditions (on "islands") accurately, but observe economy-wide conditions only with noise (e.g. Lucas, 1973; Lorenzoni, 2009). In Online Appendix E.2-3, we explicitly solve the models in Lucas (1973) and Lorenzoni (2009), and show that these models also generate underreactions to current output ($\gamma > 0$). The intuition in these models is similar to the simple model with noisy observations of output: Agents have less information about aggregates than the econometrician, and they respond to this information in a muted fashion, which generates measured underreactions. Indeed, we show that one can directly use (7) to obtain an analytical expression for the underreactions in Lucas (1973).

To summarize, it is instructive to view the results in this section in terms of our empirical findings using (1) and (2). Our estimates show that $\gamma < 0$ and $\delta > 0$, and reject the FIRE benchmark. This reveals that *either* the assumption of full information *or* the assumption of rationality are violated. However, our analysis of existing models establishes that it is not obvious how to match the data by relaxing either assumption. Although the list of models we have considered is clearly not exhaustive, we are unaware of a pre-existing model that can explain our results. This motivates the development of our model in the next section.

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.

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},

$$y_t = x_{1t} + x_{2t} + \ldots + x_{Nt}. \tag{8}$$

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

$$x_{jt} = a_j \theta_t + b_j u_{jt}, \quad u_{jt} \sim \mathcal{N}\left(0,\, 1\right), \tag{9}$$

where θ_t denotes a latent factor that follows the autoregressive process

$$\theta_t = \rho \theta_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}\left(0,\, \tau_\eta^{-1}\right), \tag{10}$$

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

The output response to a positive fundamental shock $d\theta_t > 0$ is $\frac{dy_t}{d\theta_t} = \sum_j a_j$. We assume that $\sum_j a_j > 0$ without loss of generality, so that output correlates positively with θ_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 latent factor. 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

$$z_{ijt} = x_{jt} + q_j \epsilon_{ijt}, \quad \epsilon_{ijt} \sim \mathcal{N}\left(0,\, 1\right), \quad j = \left\{1, 2, \ldots, N\right\}, \tag{11}$$

where q_j parameterizes the amount of noise (or inattention) in agents' signals about the jth structural component, and ϵ_{ijt} is an idiosyncratic error term. Agent i's information set at time t is the infinite history of her past signals $\Omega_{it} = \{z_{i1s}, \ldots, z_{iNs}\}_{s \leq t}$.[17] Agents thus infer information about the latent factor θ_t from noisy signals of x_{jt} that may covary either positively ($a_j > 0$) or negatively ($a_j < 0$) with the latent factor.

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 case (i) in Proposition 1). First, output is determined by several underlying components. Second, agents learn about these components separately: The information structure in (11) 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

[17] This assumption follows the convention in e.g. Maćkowiak *et al.* (2018). By allowing agents to observe an infinite history of signals, we ensure that their signal extraction problem is initialized in steady state.

embedded in (11), combined with a component-based structure in (8), 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 (11) into the normalized parameters

$$m_j \equiv \frac{\mathbb{V}\text{ar}(x_{jt}|\theta_t)}{\mathbb{V}\text{ar}(z_{ijt}|\theta_t)} = \frac{b_j^2}{b_j^2 + q_j^2} \in (0,1)\,. \tag{12}$$

These parameters measure the sensitivity of agents' beliefs to new information about the jth structural component. Suppose that agent i knows θ_t, and is then asked to predict component x_{jt} based on her own noisy signal z_{ijt}. Her estimate will be:[18]

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

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 $\mathbb{E}[x_{jt}|\theta_t]$ 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 jth component. We therefore call m_j the *attention* dedicated to the jth component.

While we have motivated our definition of m_j in the hypothetical case where agents observe the latent factor θ_t, these quantities also determine agents' expectations about θ_t.

Lemma 1. *For each agent $i \in [0,1]$, expectations about the latent factor θ_t satisfy*

$$\mathbb{E}_{it}[\theta_t] \;=\; \mathbb{E}_{it-1}[\theta_t] + \sum_j g_j \left(z_{ijt} - \mathbb{E}_{it-1}[z_{ijt}]\right), \tag{13}$$

where $g_j = \mathbb{V}[\theta_t|\,\Omega_{it}]\,\frac{a_j}{b_j^2} m_j$ denotes the weight placed on signal z_{ijt}.

The lemma confirms that attention coefficients m_j drive the each agent's Bayesian updates about θ_t. The agent responds to each of her signals at date t in proportion to the Kalman gain g_j. This gain is the product of the steady state variance of θ_t and a measure of the precision of signal z_{ijt}, which is in turn proportional to attention m_j.[19]

[18] We assume that all individuals choose the same attention allocation $\{m_j\}$. This is true in our model of optimal attention choice in Section 4 and 5. It is also a standard assumption in the information choice literature (see, for example, Veldkamp, 2011).

[19] To see why g_j captures the precision of z_{ijt}, consider the normalized signal $\hat{z}_{ijt} = z_{ijt}/a_j = \theta_t + \xi_{ijt}$, with

3.3 Attention, Overreactions, and Underreactions

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

$$\gamma = \mathbb{C}\text{ov}\left[y_{t+k} - \mathbb{E}_{it}y_{t+k}, y_t\right]\mathbb{V}\text{ar}\left[y_t\right]^{-1} = d_0\mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, y_t\right], \tag{14}$$

where $d_0 = \left(\rho^k \sum_j a_j\right)\mathbb{V}\text{ar}\left[y_t\right]^{-1} > 0$, and $\mathbb{E}_{it}y_{t+k} = f_{it}y_{t+k}$ denotes the k-period ahead forecast of output. Since agents are rational, their forecasts are equal to their conditional expectations. The equality in (14) follows because y_{t+k} depends only on θ_t and on shocks that are uncorrelated with date-t information. We note that the sign of γ is determined only by the covariance between the tracking error $\theta_t - \mathbb{E}_{it}\theta_t$ and current output.

Meanwhile, the coefficient δ on the average forecast revision in (2) is:

$$\begin{aligned}\delta &= \mathbb{C}\text{ov}\left[y_{t+k} - \mathbb{E}_{it}y_{t+k}, \bar{\mathbb{E}}_t y_{t+k} - \bar{\mathbb{E}}_{t-1}y_{t+k}\right]\mathbb{V}\text{ar}\left[\bar{\mathbb{E}}_t y_{t+k} - \bar{\mathbb{E}}_{t-1}y_{t+k}\right]^{-1}\\ &= d_1\mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, \bar{\mathbb{E}}_t\theta_t - \bar{\mathbb{E}}_{t-1}\theta_t\right],\end{aligned} \tag{15}$$

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

We start with two stark examples that demonstrate how the two covariances in (14) and (15) 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 (14) becomes

$$\begin{aligned}\gamma &= d_0\mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, x_{1t} + x_{2t}\right]\\ &= d_0\mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, x_{2t}\right] = d_0\mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, a_2\theta_t\right] = a_2 d_0\mathbb{V}\text{ar}\left[\theta_t|\Omega_{it}\right],\end{aligned}$$

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

$\xi_{ijt} = (b_j u_{jt} + q_j\epsilon_{ijt})/a_j$. The standard Gaussian updating formula implies that the gain on $\hat{z}_{ijt}$ is proportional to the precision (inverse variance) of ξ_{ijt}. The proof of Lemma 1 shows that this precision equals $\frac{a_j^2}{b_j^2}m_j$.

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 represents good news about the latent factor θ_t. However, the countercyclical component x_{2t}, on average, also tends to be large, which dampens any good news about the latent factor. 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: $0 < m_1 < 1$. Since the average revision is $\bar{\mathbb{E}}_t\theta_t - \bar{\mathbb{E}}_{t-1}\theta_t = \int_0^1 \left(\mathbb{E}_{jt}\theta_t - \mathbb{E}_{jt-1}\theta_t\right) dj$, the linearity of the covariance operator and the symmetry of attention choices across agents imply that

$$\begin{aligned}\delta &= d_1 \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, \bar{\mathbb{E}}_t\theta_t - \bar{\mathbb{E}}_{t-1}\theta_t\right] \\ &= d_1 \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, \mathbb{E}_{jt}\theta_t - \mathbb{E}_{jt-1}\theta_t\right] = d_1 \mathbb{C}\text{ov}\left[\mathbb{E}_{jt}\theta_t - \mathbb{E}_{it}\theta_t, \mathbb{E}_{jt}\theta_t - \mathbb{E}_{jt-1}\theta_t\right],\end{aligned}$$

where the third equality follows by adding and subtracting agent j's forecast error $\theta - \mathbb{E}_{jt}\theta_t$, and noting that it is uncorrelated with j's forecast revision. We conclude that $\delta > 0$ if, for all i and $j \neq i$, $\mathbb{C}\text{ov}\left[\mathbb{E}_{jt}\theta_t, \mathbb{E}_{jt}\theta_t - \mathbb{E}_{jt-1}\theta_t\right] > \mathbb{C}\text{ov}\left[\mathbb{E}_{it}\theta_t, \mathbb{E}_{jt}\theta_t - \mathbb{E}_{jt-1}\theta_t\right]$. This always holds in our example. Intuitively, when $m_1 < 1$, agent i and j observe *different* signals, which makes agent j's forecast revision more strongly correlated with her own expectation. □

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 extends the results in Coibion and Gorodnichenko (2015) to our case.[20] The intuition is as

[20]The baseline model in Coibion and Gorodnichenko (2015) assumes uncorrelated noise terms across agents. In an extension, Coibion and Gorodnichenko (2015, Online Appendix A) note that the coefficient δ measured by an econometrician will be attenuated by the presence of common noise terms u_{jt}. A novel result in this example and Proposition 2 that follows is that, despite this effect, we always have $\delta > 0$.

discussed above. As long as information is dispersed, rational individuals respond less strongly to *average* new information than agents in the fully-informed rational benchmark. This leads to underreactions of expectations similar to those documented in the survey data.

Combined, the above examples have shown how attention choices map into the over- and underreaction coefficients γ and δ, respectively. Specifically, they have shown how limited, asymmetric attention to a procyclical component can explain the simultaneous over- and underractions 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 $j \in \{1, \dots, N\}$ 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 (i.e. extrapolation) whenever this statistic is negative. This is clearly the case when agents are inattentive ($m_j \simeq 0$) to components that are countercyclical, which covary negatively with the latent factor ($a_j < 0$), and are more attentive to procyclical components ($a_j > 0$). 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 (case (i) in Proposition 1), 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.

3.4 Summary and Extensions

In summary, our model is able to match the stylized facts whenever attention is both *limited* and *asymmetric*. We close this section by discussing two important extensions.

First, we have presented a latent factor model with several components of output. This classical structure conveys our main contribution and leads naturally to our macroeconomic example in Section 5. However, the model in this section is not the only possible parametrization in which asymmetric attention explains the patterns that we find in the data. In particular, Proposition 6 in Appendix B fully characterizes the coefficients in (1) and (2) for a larger class of linear models, in which we allow for (*i*) the direct effects of several, latent factors on output, (*ii*) for the correlation between component-specific shocks, and (*iii*) for the explicit observation of (and dependence on) lagged outcomes. This extension, which encompasses most linear macroeconomic models, delivers necessary and sufficient conditions for over- and underreactions based on asymmetric attention more generally.

Second, we have focused our discussion of forecast revisions on (2), which is the regression of forecast errors on *average* forecast revisions proposed by Coibion and Gorodnichenko (2015). By contrast, in contemporaneous and closely related work, Bordalo *et al.* (2018) consider the regression of forecast errors on *individual* forecast revisions:

$$y_{t+k} - f_{it}y_{t+k} = \alpha + \delta^{ind}\left(f_{it}y_{t+k} - f_{it-1}y_{t+k}\right) + \xi_{it}. \tag{16}$$

Using a range of survey data, Bordalo *et al.* (2018) estimate that $\delta^{ind} < 0$, which is inconsistent with the predictions of our baseline model, and also with other models with rational expectations in which agents recall their own forecast revisions. Table C.1 in the online appendix reports estimates of (16) for output forecasts in the US SPF. We estimate overreactions to individual revisions ($\delta^{ind} < 0$), but unlike our estimates of (1) and (2), which motivate our analysis, this result appears sensitive to outliers.[21] Online Appendix F considers an extension of our framework, which allows for both asymmetric attention and irrational overconfidence. We show that, when one introduces a small bias, the extended model can account not only for the stylized facts that we have emphasized ($\gamma < 0$ in (1) and $\delta > 0$ in (2)), but also for overreactions to individual revisions ($\delta^{ind} < 0$ in (16)). Crucially, the extended model can fit these stylized facts *only if* one introduces asymmetric attention. As discussed in Section 2.2, the baseline model in Bordalo *et al.* (2018) predicts that γ and δ have the same sign. Thus, regardless of whether there are overreactions to individual revisions, asymmetric attention is *necessary* to reconcile the varied survey evidence within the class of models considered.

So far, we have considered reduced-form economies. 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 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

[21] Indeed, we cannot reject that $\delta^{ind} = 0$ once we remove outliers in the top one percent of forecast errors and revisions. This is in contrast to our estimates of (1) and (2).

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 to incorporate attention choice. To do so, we assume the following timing of events: At the start of each period, each agent chooses her attention allocation m_j to the different components x_{jt} of output (or equivalently, the noise terms q_j). She makes this choice *ex ante*, before she observes the realization of her signal vector z_i^t. Then, the agent observes her signals and chooses an action a_{it}.

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

$$\mathcal{U}_{it} = -\left(a_t^\star - a_{it}\right)^2 - K(m). \tag{17}$$

The first term in (17) is a quadratic loss that the individual incurs when she deviates from her ideal action $a_t^\star$. 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 full information about all stochastic disturbances, can depend both on the unobserved latent factor and on the structural components:

$$a_t^\star = w_\theta \theta_t + \sum w_{xj} x_{jt}, \tag{18}$$

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 Ω_{it} in the last stage at date t is to set $a_{it} = \mathbb{E}\left[a_t^\star | \Omega_{it}\right]$.

Equations (17) and (18) 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 $w_{xj} = 0$, $a_t^\star$ becomes the full-information mean squared optimal forecast of y_{t+k}, which is $\mathbb{E}_t^{FIRE}\left[y_{t+k}\right] = \rho^k \sum_j a_j \theta_t$. However, (17) and (18) 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' attention choices. To do so, it is instructive to first derive agents' expected utility at the start of period t, before they observe the realization of their signals.

Lemma 2. *Each agent's expected utility at the start of period t equals*

$$\begin{aligned}
\mathbb{E}\left[\mathcal{U}_{it}\right] &= -\mathbb{V}\left[a_t^\star \mid \Omega_{it}\right] - K(m) &(19)\\
&= -\sum_j w_{xj}^2 b_j^2 (1-m_j) - \mathbb{V}ar_t\left[\theta_t\right]\left[w_\theta + \sum_j w_{xj} a_j\left(1-m_j\right)\right]^2 - K(m). &(20)
\end{aligned}$$

Lemma 2 first provides a natural characterization of an agent's expected utility at the beginning of period t (i.e., before she observes her signals z_i^t). Intuitively, for every realization of her signals at date t, the agent will set $a_{it} = \mathbb{E}\left[a_t^\star | \Omega_{it}\right]$. Hence, her maximized utility depends on the expected squared deviation of $\mathbb{E}\left[a_t^\star | \Omega_{it}\right]$ from $a_t^\star$, which reduces to the conditional variance in (19). Lemma 2 then derives an explicit expression for the conditional variance, using the law of total variance:

$$\mathbb{V}\text{ar}\left[a_t^\star \mid \Omega_{it}\right] = \mathbb{V}\text{ar}\left[a_t^\star \mid \Omega_{it}, \theta_t\right] + \mathbb{V}\text{ar}\left[\mathbb{E}\left[a_t^\star \mid \Omega_{it}, \theta_t\right] \mid \Omega_{it}\right].$$

Accordingly, the first term in (20) reflects the uncertainty about the optimal action conditional on the latent factor. It equals the sum of the conditional variances $\mathbb{V}\text{ar}\left[x_{jt} | \Omega_{it}, \theta_t\right]$ 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 in (20) measures the residual uncertainty $\mathbb{V}\text{ar}\left[\theta_t | \Omega_{it}\right] \equiv \mathbb{V}\text{ar}_t\left[\theta_t\right]$, scaled by the uncertainty about the ideal action $a_t^\star = w_\theta \theta_t + \sum_j w_j x_j$ that is attributable to θ_t (i.e., by the term in square brackets). We provide a brief derivation of $\mathbb{V}\text{ar}_t\left[\theta_t\right]$, to show how it depends on agents' attention choices. In turn, combined with (17) and (20), this will then allow us to derive an expression for agents' optimal attention choices.

Recall that the effective precision of signal z_{ijt} about θ_t is $\tau_j = \frac{a_j^2}{b_j^2 + q_j^2}$, and let

$$\tau(m) = \sum_j \tau_j \tag{21}$$

denote the total precision of date t signals. Starting at date t, the conditional variance about next period's fundamental is $\mathbb{V}\text{ar}_t\left[\theta_{t+1}\right] = \rho^2 \mathbb{V}ar_t\left[\theta_t\right] + \sigma_\theta^2$. After updating based on date $t+1$ signals, this variance satisfies the linear precision rule $\mathbb{V}\text{ar}_{t+1}\left[\theta_{t+1}\right]^{-1} = \mathbb{V}\text{ar}_t\left[\theta_{t+1}\right]^{-1} + \tau(m)$. Solving for a steady state where $\mathbb{V}\text{ar}_t\left[\theta_t\right] = \mathbb{V}\text{ar}_{t+1}\left[\theta_{t+1}\right] = V$ 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 τ of an agent's signals is a sufficient statistic for her uncertainty about the persistent fundamental, and we can write

$$\mathbb{V}\mathrm{ar}_t\left[\theta_t\right] = V\left[\tau(m)\right], \tag{22}$$

where $V'(\tau) < 0$ and $\partial\tau/\partial m_j > 0$ from (21). Combined, (17), (20), and (22) allow us to characterize agents' attention choices. Proposition 3 summarizes the results.

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

$$w_{xj}^2 b_j^2 + \mu_\tau a_j^2 b_j^{-2} + \mu_\alpha w_{xj} a_j = \frac{\partial K(m)}{\partial m_j}, \tag{23}$$

where $\mu_\tau > 0$ and $\mu_\alpha > 0$ 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 (23), consists of three terms. The first term is the benefit of resolving uncertainty about the optimal action conditional on θ_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 latent factor θ_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 τ of agents' signals about θ_t. The multiplier μ_τ 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 are highly volatile (high b_j^2). The third term measures an adjustment to this effect, namely, that information about θ_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 μ_α is the shadow value of reducing the residual uncertainty about $a_t^\star$ that is attributable to θ_t.

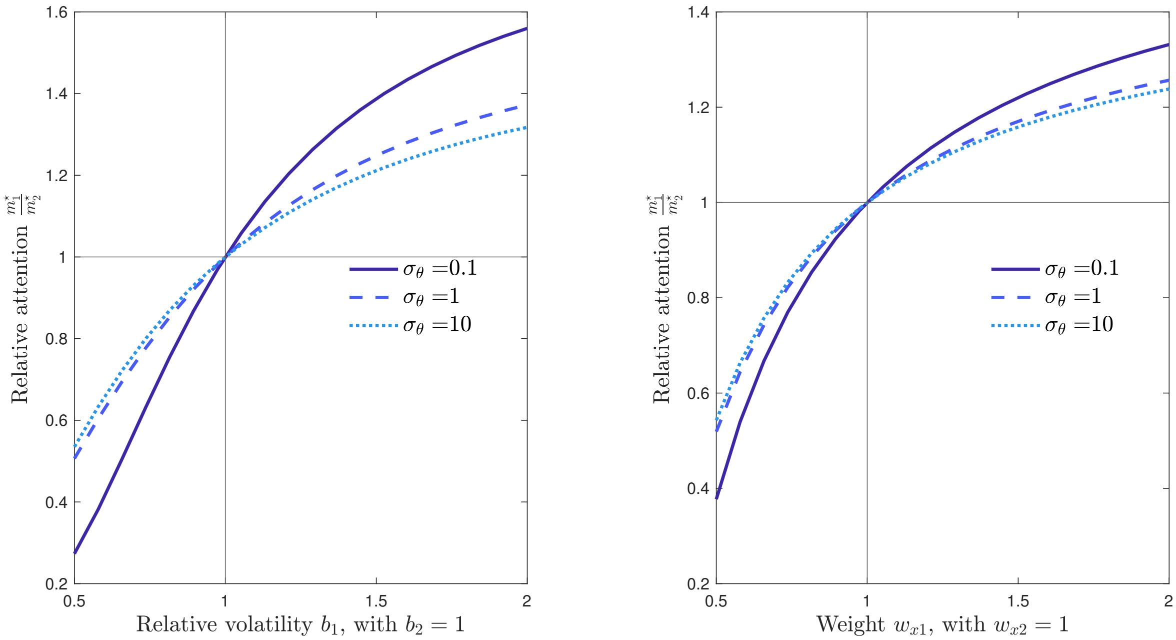
While these effects are subtle, the underlying intuition is clear. On one hand, agents are more likely to pay attention to components that are important for their utility, those with large weights w_{xj} in (18). On the other hand, agents also prefer to pay attention to volatile components (with a high idiosyncratic variance b_j^2), as long as learning spillovers are not too strong. 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.

Figure 4 provides a numerical example, which illustrates the effects of component volatility and utility weighting on agents' optimal attention choices. To demonstrate the role of

Figure 4: Optimal Attention: Numerical Example

(a) The effect of component volatility

(b) The effect of component weighting



The charts show the properties of optimal attention choices as a function of component volatilities b_j, utility weighting w_{xj}, and the variance σ_θ^2 of the latent factor, in a numerical example with two components. The parameters not detailed in the figure are set at $a_1 = a_2 = 1$, $\rho = 0.9$, $w_\theta = 0$. The cost function $K(m)$ is set to the reduction in entropy, as derived in Proposition G.1 in the online appendix.

learning spillovers, the figure considers three scenarios for the variance σ_θ^2 of the latent factor. Intuitively, spillovers are minimized when the variance of the latent factor θ_t is small. The two panels confirm the main points in our discussion: The relative attention $m_1^\star/m_2^\star$ paid to component 1 increases as this component becomes more volatile ($\uparrow b_1$ in panel (a)) and more important in agents' objective function ($\uparrow w_{x1}$ in panel (b)). In both cases, the rate of increase is smaller when there are strong spillovers (high σ_θ^2). This reflects the intuition that strong learning spillovers incentivize an agent to push on all margins to learn more about the latent factor, which in turn leads her to respond less strongly to component-specific features.

We have so far kept the functional form of the attention cost function $K(m)$ general. Online Appendix G derives the first-order condition (23) explicitly for an entropy-based cost function, and shows 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 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 (11) 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.

Building closely on recent work by Maćkowiak *et al.* (2018), Proposition F.1 in the online appendix shows that agents in our model, when equipped with an entropy-based cost function, would optimally choose to receive a single signal of the optimal action:[22]

$$s_{it}^{\star} \quad = \quad a_t^{\star} + h' v_t + q^{\star} \epsilon_{it}, \tag{24}$$

where h depends on the utility weights w_θ and w_{xj}, we stack the common shocks into a vector $v_t = \begin{bmatrix} \eta_t & u_{1t} & \dots & u_{Nt} \end{bmatrix}'$, and $q^{\star}$ denotes a scalar that depends only on the cost of attention $K(m)$. Equation (24) shows that the asymmetry of attention now depends on the weights w_{xj} in agents' optimal action both through their influence on $a_t^{\star}$ and the vector h in the optimal signal. This has important empirical implications.

For example, consider the benchmark case in which agents' utility in (17) 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

[22]Heuristically, we derive this result in two steps. We first express $a_t^{\star}$ as an ARMA process in reduced form. In particular, substituting (9) and (10) into (18) shows that

$$\begin{aligned} a_t^{\star} = \underbrace{\left(w_\theta + \sum w_{xj} a_j\right)}_{\equiv \bar{w}_\theta} \theta_t + \sum \underbrace{w_{xj} b_j}_{\equiv \bar{w}_{xj}} u_{jt} \quad &= \quad \rho a_{t-1}^{\star} + \bar{w}_\theta \eta_t + \bar{w}_x' u_t - \rho \bar{w}_x' u_{t-1}. \\ &\equiv \quad \rho a_{t-1}^{\star} + c_0' v_t + c_1' v_{t-1}. \end{aligned}$$

Hence, $a_t^{\star}$ is an ARMA process whose vector of innovations is $v_t = [\eta_t \;\; u_t]'$. We then modify the results in Maćkowiak *et al.* (2018), which apply to ARMA processes with scalar-valued innovations, to arrive at (24).

discussed above). In this case, it follows that $h = 0$ in (24).[23] As a result, the fully optimal signal boils down to $s^\star_{it} = (\rho \sum_j a_j)\theta_t + q^\star \epsilon_{it}$, which is a simple noisy signal of θ_t. Similar to the results in Proposition 1, and due to the symmetry of underlying preferences, such a signal is inconsistent with extrapolation.[24] Indeed, in this case, agents systematically underreact to new information about current 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. In this case, agents' forecasts of future output given $s^\star_{it}$ can exhibit extrapolation. Similar to the results in Proposition 2, this occurs when the weights w_{xj} are tilted towards procyclical components. This is easiest to see in the following example, which extends our previous Example 1 to flexible information choice:

Example 3. Asymmetric Attention and Extrapolation (cont.): As in Example 1, suppose that output has two components, where $a_1 > 0$ and $a_2 < 0$. Agents' ideal action depends only on the first component ($w_{x1} > 0$ while $w_{x2} = w_\theta = 0$). Corollary H.2 in the online appendix shows that, if the costs of attention are sufficiently small, the optimal signal tends to $s^\star_{it} = x_{1t} + q^\star \epsilon_{it}$. Hence, the information structure is identical to that in Example 1 and 2, where $0 < m_1 < 1$ and $m_2 = 0$. The arguments in Example 1 and 2 now imply that $\gamma < 0$ and $\delta > 0$. By continuity, the model with flexible information choice generates $\gamma < 0$ and $\delta > 0$ as long as the weight w_{x1} is sufficiently large relative to w_θ and w_{x2}. $\square$

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 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 underrreactions documented in the data, so long as there are differences in the volatility of the underlying components.

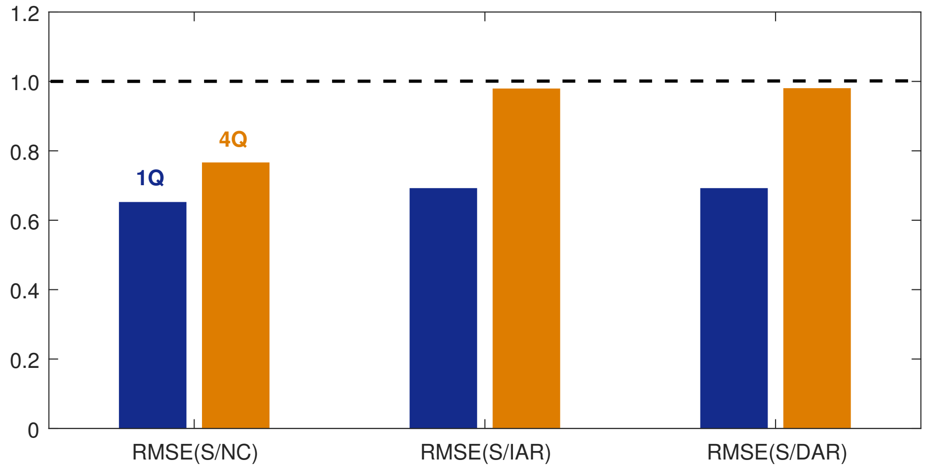
[23]See Corollary H.2 in the online appendix, or Cover and Thomas (2012) for the standard case in which the optimal action $a^\star_t$ is proportional to a simple AR(1) process.

[24]Consider the extrapolation coefficient in (1) based on $s^\star_{it} = (\rho \sum_j a_j)\theta_t + q^\star \epsilon_{it}$. It follows that

$$\begin{aligned}
\gamma &= \mathbb{C}\mathrm{ov}\left(y_{t+k} - \mathbb{E}_{it} y_{t+k}, y_t\right) \mathbb{V}\mathrm{ar}\left[y_t\right]^{-1} \\
&= d_0 \mathbb{C}\mathrm{ov}\left(\theta_t - \mathbb{E}_{it}\theta, y_t\right) = d_0 \sum_j a_j \mathbb{V}\mathrm{ar}\left[\theta_t \mid s_i^{\star,t}\right] > 0,
\end{aligned}$$

where we have also used that $y_t = \sum_j a_j \theta_t + \sum_j b_j u_{jt}$.

Figure 5: Forecast Precision Relative to Time Series Models



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.

4.4 Are Attention Choices Optimal? Supplementary Evidence

We briefly return to the data to compare the quality of agents' expectations to that of standard time series models. Figure 5 shows updated values from Stark (2010), available from the *Federal Reserve Bank of Philadelphia's* website.[25] 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, in which agents pay attention to underlying, structural components of the forecasted variable, but inconsistent with a model where agents consider only the past time series of output (see, for instance, 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

[25] https://www.philadelphiafed.org/research-and-data/real-time-center.html

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 flexible-price business cycle model. We analyze an environment in which firms choose output under imperfect information. We show that firms' output choices can be decomposed into two components: First, a *productivity component*, which summarizes the effects of a firms' own productivity; and second, an *aggregate supply component*, which captures the effects of other agents' behavior on an individual firms' output choice. We document that, for standard parameters, the productivity component is procyclical, while the aggregate supply component is countercyclical. In accordance with the evidence in Coibion *et al.* (2018), we show that firms' attention choices are asymmetric and tend to abstract from the aggregate supply component. As a result, and in line with the above analysis, we find that firms' expectations of 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 $i \in [0, 1]$, which specialize in the production of differentiated goods.

Households: The representative household has lifetime utility

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log\left(C_t\right) - \xi_t N_t\right], \quad \xi_t > 0, \tag{25}$$

where β denotes the time discount factor, C_t the consumption index at time t, N_t the number of hours worked by the household, and ξ_t a shock to the disutility of labor. The consumption index C_t and associated welfare-based price index P_t are

$$C_t = \left[\int_0^1 C_{it}^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}, \quad P_t = \left[\int_0^1 P_{it}^{\frac{1}{\sigma-1}} di\right]^{\sigma-1}, \tag{26}$$

where C_{it} is the amount the household consumes of goods produced by firm i at price P_{it}, and $\sigma > 1$. The household's per-period budget constraint is

$$\int_0^1 P_{it} C_{it} di + B_{t+1} \leq \int_0^1 \Pi_{it} di + W_t N_t + (1 + R_t) B_{t+1} + T_t^h, \tag{27}$$

where Π_{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 (25) subject to (27).

Firms: A representative firm $i \in [0, 1]$ chooses its output of Y_{it} to maximize its own expectation of the household's valuation of its profits, using the stochastic discount factor $(P_t C_t)^{-1}$. The expected valuation of profits at time t is equal to

$$\mathcal{V}_{it} = \mathbb{E}_{it} \left[\frac{1}{P_t C_t} \Pi_{it} \right], \ \Pi_{it} = P_{it} Y_{it} - W_t N_{it} \tag{28}$$

where the inverse-demand for a firm's product is consistent with household optimality: $P_{it} = P_t \left(Y_{it} / Y_t \right)^{-\frac{1}{\sigma}}$. Firm output is produced in accordance with the production function

$$Y_{it} = A_{it} N_{it}^{\alpha}, \quad \alpha \in (0, 1), \tag{29}$$

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

$$a_{it} = \theta_t + u_t^x + \epsilon_{it}^a, \tag{30}$$

where the persistent, common component θ_t follows an AR(1) process,

$$\theta_t = \rho \theta_{t-1} + u_t^{\theta}, \quad u_t^{\theta} \sim \mathcal{N} \left(0, \sigma_{\theta}^2 \right), \tag{31}$$

while the transitory and firm-specific components are distributed as $u_t^x \sim \mathcal{N}(0, \sigma_x^2)$ and $\epsilon_{it}^a \sim \mathcal{N}(0, \sigma_a^2)$, 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

$$\log \xi_t = \bar{\xi} + u_t^n, \quad u_t^n \sim \mathcal{N} \left(0, \sigma_n^2 \right), \tag{32}$$

where $\bar{\xi} \in \mathbb{R}$. We show below that the labor supply shock in (32) introduces a component-specific innovation to aggregate output. In effect, u_t^n will play the role of one of the component-specific disturbances u_{jt} discussed in Section 3. We assume that the innovations u_t^x, u_t^θ , u_t^n, and ϵ_{it}^a are independent of each other, across time, and across firms.

Timeline: In each period, nature determines the realization of the innovations u_t^x, u_t^θ , u_t^n, and ϵ_{it}^a. The economy then proceeds through three stages. In the first stage, firms choose how much attention to devote to the various components of output, which we define below, and commit to their output choices. After output choices are sunk, the economy transitions to the second stage, in which the labor market opens. Each firm observes its own productivity a_{it} and hires the amount of labor $n_{it} = \alpha^{-1}(y_{it} - a_{it})$ that is necessary to implement its previous output choice y_{it}. The representative household observes its marginal disutility $\xi_t = \bar{\xi} + u_t^n$ of labor and the permanent productivity component θ_t, and then makes its labor supply choice.[26] The real wage adjusts to clear the labor market. In the third and final stage, goods markets open, goods prices adjust to clear them, and the household consumes.

Information Structure: To complete the description of the economy, it is necessary to specify the information structure, and firms' associated attention choice problem. Our assumptions are based on the following decomposition of firms' expected profits:

Proposition 4. *A second-order approximation of firm i's expected discounted profits satisfies*

$$v_{it} \simeq -\frac{1}{2}\mathbb{E}_{it}\left[(y_{it} - y_{it}^\star)^2\right], \tag{33}$$

where the firm's ideal output under full information $y_{it}^\star$ can be decomposed into

$$y_{it}^\star = x_{i1t} + x_{2t} \tag{34}$$

with

$$x_{i1t} = ra_{it}, \qquad x_{2t} = \alpha r\left(\sigma^{-1}y_t - \omega_t\right), \tag{35}$$

and where $\omega_t = w_t - p_t$ denotes the real wage, $y_t = \int_0^1 y_{it}di$, and $r = \frac{\sigma}{\sigma+\alpha(1-\sigma)} > 1$.

In the spirit of Lucas (1977) and Maćkowiak and Wiederholt (2009), condition (34) and (35) decompose each firm's ideal output choice into two components: We refer to x_{i1t} as the

[26] Because the household does not observe the realization of u_t^x in the second stage, output will respond differently to innovations in θ_t and u_t^x. This friction creates a meaningful distinction between these two shocks. Without this friction, only shocks to the sum $\int_0^1 a_{it}di = \theta_t + u_t^x$ 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).

productivity 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 *aggregate supply component*, which encapsulates the general equilibrium effects of other agents' behavior on an individual firm's optimal output choice. The aggregate supply component, in turn, is comprised of two terms: On one hand, firms produce more when aggregate demand in the economy y_t is high. On the other hand, a firm also chooses to produce less when the real wage ω_t is high. Both effects are captured in (35).[27]

Given this decomposition, our assumptions about firms' information sets and attention choices mirror those in our baseline model. Specifically, we assume that firm i's information set consists of the infinite history of component-based signals:

$$\Omega_{it} = \left\{ z_{i1}^t, z_{i2}^t \right\}, \tag{36}$$

where

$$z_{i1t} = x_{i1t} + q_1 \epsilon_{i1t}, \quad z_{i2t} = x_{2t} + q_2 \epsilon_{i2t}, \tag{37}$$

and $\epsilon_{ijt} \sim \mathcal{N}(0,1)$ is independently distributed across time and firms for $j = \{1,2\}$. Furthermore, as in our reduced-form framework, we also assume that at the start of each period each firm chooses normalized attention parameters $m_j = \frac{\mathbb{V}ar(x_{jt}|\theta_t)}{\mathbb{V}ar(x_{jt}|\theta_t)+q_j^2}$ at a cost $K(m)$.

5.2 Equilibrium Characterization

We now proceed to characterize equilibrium output for the economy.

5.2.1 Equilibrium with Full Attention

We start with the case in which firms pay full attention to both components (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. In this special case, Proposition 4 directly implies that each firm sets $y_{it} = y_{it}^\star = x_{i1t} + x_{2t}$, so that

$$y_t = \int_0^1 y_{it} di = x_{1t} + x_{2t}, \tag{38}$$

[27] Unlike the similar decomposition used in Maćkowiak and Wiederholt (2009), the two components x_{i1t} and x_{2t} are correlated in this application. For example, a shock to θ_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 6 in Appendix A.5.

with $x_t = \int_0^1 x_{it} di$. Thus, output has the same component-based structure as in the baseline model from Section 3. The components x_{jt} of output can now further be characterized directly from (35). As for the productivity component x_{1t}, we have

$$x_{1t} = r\theta_t + ru_t^x, \tag{39}$$

This component is *procyclical*, since it places a positive weight $r > 0$ on the latent factor θ_t. Turning to the aggregate supply component x_{2t}, the real wage in equilibrium is $\omega_t = y_t + (1-\alpha)u_t^n$. Thus, we conclude from (35) and (38) that

$$x_{2t} = \alpha r \left(\frac{1-\sigma}{\sigma} y_t + (1-\alpha)\, u_t^n \right) = (1-r)\theta_t + \left(\frac{1}{1-\alpha} - r \right) u_t^x - \alpha u_t^n. \tag{40}$$

The first equality in (40) 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 (35). By contrast, with perfect competition ($\sigma = 1$), firms are price-takers and act independently of one another. The second equality in (40) expresses the same relationship in equilibrium, in terms of the latent factor θ_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 latent factor. This type of strategic substitutability (or "general equilibrium offset") arises commonly in flexible-price business cycle models, especially those that generate realistic amounts of volatility in hours worked (Hansen, 1985; Rogerson, 1988), because increases in other firms' output tend to drive up production costs.

In Online Appendix I, we consider a model that nests both our example and the model in Angeletos and La'O (2010) and Angeletos *et al.* (2016). In this extension, among other additional parameters, households have a flexible coefficient ψ of relative risk aversion (our model fixes $\psi = 1$). We show that output choices are strategic substitutes if and only if $\sigma\psi > 1$. Common values in macroeconomics for σ and ψ are $\sigma \geq 4$ and $\psi \geq 1$ (e.g. Gali, 2008, Chapter 2). Hence, while qualitative explorations of models of strategic complementarity have yielded important insights (e.g. Angeletos *et al.*, 2016), we view the case in which output choices are strategic substitutes as a quantitatively relevant one for this class of models.

The above 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 first component x_{1t}. 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

$$\gamma_i = \mathbb{C}ov\left(y_{t+1} - \mathbb{E}_{it} y_{t+1}, y_t\right) \mathbb{V}\text{ar}\left[y_t\right]^{-1} = -\rho \frac{\alpha}{1-\alpha} \frac{\mathbb{V}\text{ar}_t\left[\theta_t\right]}{\mathbb{V}\text{ar}\left[y_t\right]} < 0, \tag{41}$$

so that firm i appears to extrapolate.[28]

5.2.2 Equilibrium with Limited Attention

We now return to the full model with limited attention. We start by describing firms' optimal output choices under limited attention, and the corresponding expected profits:

Proposition 5. *An individual firm i's output choice under limited attention satisfies* $y_{it} = \mathbb{E}_{it}\left[y_{it}^{\star}\right] = \mathbb{E}_{it}\left[x_{i1t} + x_{2t}\right]$, *and the associated expected, maximized profits are* $v_{it}^{\star} \simeq -\frac{1}{2}\mathbb{V}ar[y_{it}^{\star} \mid \Omega_{it}]$.

The characterization in Proposition 5 follows from that Proposition 4. It allows us to state an individual firm's attention choice problem as follows: At the start of the first stage of each period, the firm chooses attention coefficients m_1 and m_2 to maximize

$$\max_{\{m_1, m_2\} \in [0,1]^2} -\frac{1}{2} \mathbb{V}\text{ar}\left[y_{it}^{\star} \mid \Omega_{it}\right] - K\left(m\right). \tag{42}$$

while anticipating that its optimal output choice in the subsequent stage be

$$y_{it} = \mathbb{E}\left[y_{it}^{\star} \mid z_{i1}^t, z_{i2}^t\right] = \mathbb{E}_{it}\left[x_{i1t} + x_{2t}\right], \tag{43}$$

where x_{2t} depends upon $y_t = \int_0^1 y_{it} di$. Notice that the problem in (42) and (43) 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 $y_{it}^{\star}$. 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 $y_{it}^{\star}$ is now firm-specific.[29]

[28] This follows from

$$\begin{aligned}
\gamma_i = \mathbb{C}ov\left(y_{t+1} - \mathbb{E}_{it} y_{t+1}, y_t\right) \mathbb{V}\text{ar}\left[y_t\right]^{-1} &= \mathbb{C}ov\left[y_{t+1} - \mathbb{E}_{it} y_{t+1}, x_{2t} \pm \frac{1}{r}\left(\frac{1}{1-\alpha} - r\right) x_{1t}\right] \mathbb{V}\text{ar}\left[y_t\right]^{-1} \\
&= \rho \mathbb{C}ov\left[\theta_t - \mathbb{E}_{it}\theta_t, (1-r)\theta_t - \left(\frac{1}{1-\alpha} - r\right)\theta_t\right] \mathbb{V}\text{ar}\left[y_t\right]^{-1} \\
&= -\rho \frac{\alpha}{1-\alpha} \mathbb{V}\text{ar}_t\left[\theta_t\right] \mathbb{V}\text{ar}\left[y_t\right]^{-1} < 0.
\end{aligned}$$

[29] 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.

5.2.3 Numerical Solution Method

Unlike in the full-attention version of the model, 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 components and the latent factor take the form of an infinite dimensional vector

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + Bu_t, \quad u_t = \begin{bmatrix} u_t^\theta & u_t^x & u_t^n \end{bmatrix}', \tag{44}$$

where $\boldsymbol{x}_t = \begin{bmatrix} \bar{x}'_{t-1} & \bar{x}'_{t-2} & \dots \end{bmatrix}'$ with $\bar{x}_t = \begin{bmatrix} x_{1t} & x_{2t} & \theta_t \end{bmatrix}'$ and $x_{1t} = \int_0^1 x_{i1t} di$, and where A and B are matrices of undetermined coefficients whose rows conform with (35) and (30).

To solve for the rational expectations equilibrium, we further conjecture that

$$\begin{aligned} y_t &= \bar{\mathbb{E}}_t \left[x_{1t} + x_{2t}\right] \\ &= \begin{bmatrix} 1 & 1 & \mathbf{0} \end{bmatrix} \bar{\mathbb{E}}_t \left[\boldsymbol{x}_t\right] = \begin{bmatrix} 1 & 1 & \mathbf{0} \end{bmatrix} \Xi \boldsymbol{x}_t \end{aligned} \tag{45}$$

where Ξ is another matrix of undetermined coefficients.

Solving the model requires finding values for the matrices A, B, and Ξ, as well as firms' attention choices $m = \begin{bmatrix} m_1 & m_2 \end{bmatrix}$, that are consistent with firm optimality, Bayesian updating of expectations, and market-clearing. We do so by first truncating the infinite-dimensional vector $\boldsymbol{x}_t$. In accordance, with Hellwig and Venkateswaran (2009) and Lorenzoni (2009), we truncate it at $\bar{x}_{t-T}$ where $T = 50$, but our numerical results are already stable from around $T = 10$. We then iterate on the following two steps until convergence.

First, we hold attention choices m fixed and derive new matrices A, B, and Ξ implied by Bayesian updating and firm optimality. Specifically, we solve firms' signal extraction problem using the Kalman filter, which implies a new matrix Ξ, characterizing average expectations about $\boldsymbol{x}_t$. This matrix, along with firms' optimality conditions, implies new matrices A and B characterizing the law of motion for $\boldsymbol{x}_t$, which in turn implies a new matrix Ξ. We iterate on these updates until the coefficients in A, B, and Ξ converge in the sense of absolute difference.

Second, we hold coefficients in A, B, and Ξ fixed and derive new values m for firms' optimal attention choices. We derive an expression for firms' profits in (28) as a function of attention choices, which closely resembles the expression in Lemma 2. We then find new optimal choices m by solving the problem in (42). We halt the iteration between these two steps when attention choices m have converged in the sense of absolute difference. Online Appendix J contains further details about the solution method and its implementation.

5.3 A Quantitative Exploration

We now explore the quantitative implications of the model. We address 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 γ in (1) and δ 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 latent factor θ_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 functional form $K(q) = \mu \sum_j q_j^{-2}$; that is, a marginal cost μ multiplied by the sum of signal precisions $1/q_j^2$ across the components of output (Veldkamp, 2011).[30] The free parameter is the marginal cost μ, which determines the overall imperfection in firms' information. 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 δ in (2) that measures underreactions in average expectations. Hence, we calibrate μ to match estimated underreactions. Concretely, we solve the model repeatedly, varying μ, until the estimate of $\hat{\delta}$ obtained from the model's output matches the empirical estimate obtained from one-quarter-ahead forecasts in the SPF. This approach yields $\mu = 1.30$. This calibration implicitly assumes that forecasts reported by respondents in the SPF are similar to the expectations of firms in our model, who aim to maximize the accuracy of their forecasts in (42). Clearly, survey participants might instead be motivated by career concerns, a desire to attract publicity, or other biased incentives (e.g. Ehrbeck and Waldmann, 1996; Lamont, 2002; Ottaviani and Sørensen, 2006). The related empirical evidence is mixed.[31] Following the recent macroeconomic literature, we view the estimates from professional forecasters as providing a useful lower-bound on deviations from full-information rationality.[32]

[30] 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.

[31] For example, Lamont (2002) finds evidence for strategic forecasts in the non-anonymized *Business Week Survey,* but Stark (1997) argues that the same hypothesis is rejected in the anonymized SPF. Ehrbeck and Waldmann (1996) reject a model of strategically biased forecasts in T-bill forecasts from the Blue-Chip Survey.

[32] See, for example, Lorenzoni (2009), Nimark (2014), and Angeletos and Huo (2019). We note that the SPF includes forecasts from large industrial firms, in addition to those from financial and government institutions,

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 6 and Table II illustrate these mechanisms in general equilibrium.

Figure 6 shows that, as in the full information case, the productivity component is procyclical, while the economy-wide one is countercyclical in equilibrium. Output as a whole is procyclical. The first two columns in Table II show the significance of the productivity and aggregate supply component in firms' decision problems. While both components have a utility weight of one in firms' ideal output choice (Proposition 4), the productivity component is much more volatile for baseline parameters.[33] 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 productivity component x_{1t} because of its larger volatility. In particular, firms optimally choose to pay around three times more attention to x_{1t}. This is consistent with the conclusions from Lucas (1977) (also cited in the introduction) that for most firms there is little reason to pay particularly close attention to aggregate conditions. Coibion *et al.* (2018) provide evidence in favor of this supposition. 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 (Table C.10 in the online appendix). The underreaction coefficient δ at the one-quarter frequency was a targeted moment. Due to firms' asymmetric attention to the procyclical component of output, the coefficient γ 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

and forecasting agencies. The bi-annual Livingstone survey estimates reported in Section 2, which resembles those from the SPF, includes a broader range of non-financial firms.

[33] 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).

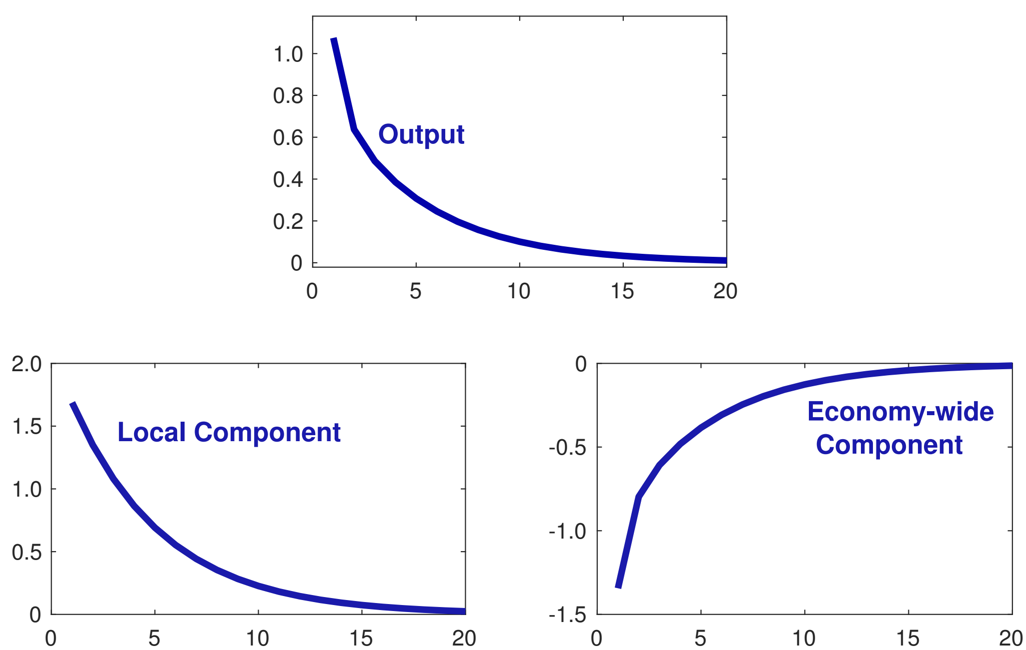
Table II: Attention Choices in Equilibrium

Component	*Variance*	*Weight*	q	m
Productivity component (x_1)	3.73	1.00	1.29	0.75
Economy-wide component (x_2)	1.08	1.00	2.10	0.19

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

Figure 6: Cyclicality of Structural Components and Output:

Impulse Response to a One Unit Standard Deviation Shock to θ_t



Note: The chart depicts the impulse response function to a unit standard deviation shock to θ_t on the vertical axis. Time is measured in quarters on the horizontal axis.

Table III: Over- and Underreactions

(a) One-quarter Ahead Output Growth

	Model Estimates		Data Estimates	
	Forecast Error	*Forecast Error*	*Forecast Error*	*Forecast Error*
Current Realization	-0.09		-0.06	
	(–)		(0.05)	
Average Revision		0.41		0.42***
		(–)		(0.14)
Sample	(–)	(–)	01/70:10/17	01/70:10/17
Relative RMSE	0.88			

(b) Four-Quarter Ahead Output Growth

	Model Estimates		Data Estimates	
	Forecast Error	*Forecast Error*	*Forecast Error*	*Forecast Error*
Current Realization	-0.07		-0.14***	
	(–)		(0.05)	
Average Revision		0.28		0.67***
		(–)		(0.20)
Sample	(–)	(–)	01/70:01/17	01/70:01/17
Relative RMSE	0.85			

Note: Double-clustered robust standard errors in parentheses. The top/bottom one percent of forecast errors and revisions has been trimmed pre-estimation. Significance levels *=10%, **=5%, ***=1%.
Relative RMSE denotes the root mean-squared-error of individual forecasts relative to an estimated AR(1).

targeted moments in the calibration.[34]

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.

5.4 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 increased responsiveness to new information, and increased extrapolation, after the onset of the Great Moderation (as we also document empirically in the online appendix).

5.4.1 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 *symmetric*. In this symmetric case, firms observe only one noisy signal of optimal output:

$$s_{it} = y_t^\star + q\epsilon_{it} = x_{1t} + x_{2t} + q_\star \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0,1), \tag{46}$$

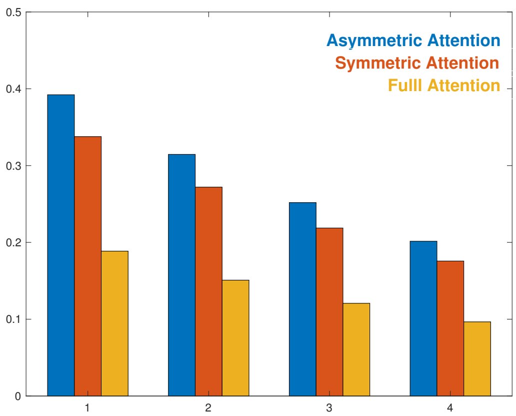
where the noise parameter q (or corresponding attention parameter m) is again calibrated to match the one-quarter-ahead estimate of δ.

Figure 7 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 procyclical, productivity component their beliefs and actions become more persistent because this component directly tracks the dynamics of the latent factor. This increase in 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.

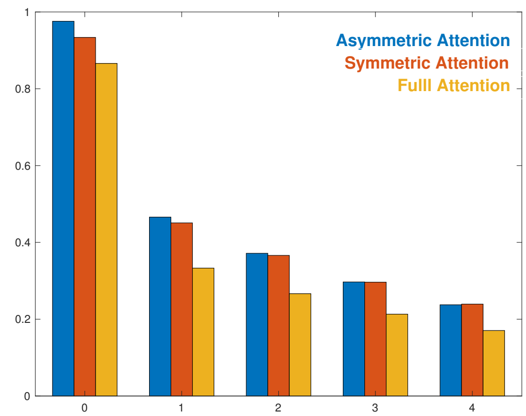
Relatedly, the right panel in Figure 7 shows that output responses are also more correlated with the latent factor itself when there is asymmetric attention. The bottom panel, in turn, shows that asymmetric attention also causes the unconditional variance of output to increase. For the same overall information friction (as measured by δ in (2)), the asymmetry of attention increases the volatility of output, and pushes it closer to its full information value.

[34] An alternative approach is to calibrate the model by targeting the four-quarter δ estimate in Table I. In this case, we arrive at estimates for γ 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.

Figure 7: Asymmetric Attention and Output Dynamics



(a) Autocorrelation of Output



(b) Correlation of Output with Productivity

(c) Variance Relative to Full Information Benchmark

	Asymmetric Attention	Symmetric Attention
Relative Variance	0.51	0.47

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 $a_t = \theta_t + u_t^x$ 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

	Pre-Great Moderation		Post-Great Moderation	
	Forecast Error	*Forecast Error*	*Forecast Error*	*Forecast Error*
Current Realization	-0.09	–	-0.13	
	(–)		(–)	
Average Revision		0.56		0.42
		(–)		(–)

Note: Columns (1) and (3) report estimates using one-year ahead forecast, while columns (2) and (4) employ one-quarter ahead forecasts. The equilibrium noise in the signals about the two components is pre-Great Moderation $q_1 = 1.31$ and $q_2 = 2.57$, and post-Great Moderation $q_1 = 1.29$ and $q_2 = 4.56$.

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

5.4.2 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 *et al.* (2007) and Galí and Gambetti (2009), the standard deviation of aggregate productivity shocks declined by around 50 percent 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 σ_θ. 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-quarter ahead forecasts on a sample until 1985Q1. To model the economy after the Great Moderation, we then reduce the volatility of σ_θ 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 are in Table C.10 in the online appendix, 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 pre-Great Moderation values, our solution shows that post-Great Moderation firms pay two percent more attention to the procyclical component (as measured by the noise in firms' information q_1), and 77 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 a procyclical, 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.[35] 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.

[35] Consider, for example, our baseline model from Section 3, and suppose that instead of regression (1) we 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 a_j). In principle, we therefore conjecture that the conditions in Proposition 2 could be extended and used to account for the much broader patterns of predictability documented, for example, by Pesaran and Weale (2006) and Fuhrer (2017).

A Proofs and Derivations

A.2 Alternative Models

Proof of Proposition 1: We prove the proposition separately for each of the cases that we analyze. In all cases, the slope coefficient in (1) and (2) equal, respectively:

$$\gamma = \mathbb{C}ov\left[y_{t+k} - f_{it}y_{t+k}, y_t\right]\mathbb{V}\text{ar}\left[y_t\right]^{-1} = \mathbb{C}ov\left[y_{t+k} - \bar{f}_t y_{t+k}, y_t\right]\mathbb{V}\text{ar}\left[y_t\right]^{-1} \tag{A1}$$

and

$$\delta = \frac{\mathbb{C}ov\left[y_{t+k} - f_{it}y_{t+k}, \bar{f}_t y_{t+k} - \bar{f}_{t-1}y_{t+k}\right]}{\mathbb{V}\text{ar}\left[\bar{f}_t y_{t+k} - \bar{f}_{t-1}y_{t+k}\right]} = \frac{\mathbb{C}ov\left[y_{t+k} - \bar{f}_t y_{t+k}, \bar{f}_t y_{t+k} - \bar{f}_{t-1}y_{t+k}\right]}{\mathbb{V}\text{ar}\left[\bar{f}_t y_{t+k} - \bar{f}_{t-1}y_{t+k}\right]}, \tag{A2}$$

where the second equality in (A1) and (A2) follows by integrating across $i \in [0, 1]$.

Case (i): Rational Expectations. Let the responsiveness coefficient g_k satisfy $g_k = g\rho^k$, where imperfect information implies that $g \in (0, 1)$. We need to show that δ and γ have the same sign as $\rho^k - g_k = \rho^k(1 - g)$. Averaging (7) across i, using that $f_{it-1}y_{t+k} = \rho^k f_{it-1}y_t$, and rearranging terms as in (Coibion and Gorodnichenko, 2015, Equation (2)), we find that:

$$y_{t+k} - \bar{f}_t y_{t+k} = \frac{1-g}{g}\rho^k\left(\bar{f}_t y_t - \bar{f}_{t-1}y_t\right) + u_{t,t+h},$$

where $u_{t,t+h}$ denotes a linear combination of future shocks $(u_{t+s})_{0<s\leq h}$ to output. It follows that $\delta = \frac{1-g}{g\rho^k}\rho^k$ and $\gamma = \frac{1-g}{g}\rho^k\mathbb{C}ov\left[\bar{f}_t y_t - \bar{f}_{t-1}y_t, y_t\right]\mathbb{V}\left[y_t\right]^{-1}$. It remains to show that $\mathbb{C}ov\left[\bar{f}_t y_t - \bar{f}_{t-1}y_t, y_t\right] > 0$. Solving (7) backwards shows that $f_{it}y_t = g\sum_{h=0}^{\infty}\lambda^h z_{it-h}$, where $\lambda \equiv (1-g)\rho$. Using this characterization, we obtain that $\mathbb{C}ov\left[\bar{f}_t y_t - \bar{f}_{t-1}y_t, y_t\right]/\mathbb{V}\text{ar}\left[y_t\right] = g\frac{1-\rho^2}{1-\lambda\rho} > 0$. We conclude that both δ and γ have the required sign.

Case (ii): Overconfidence. Let the responsiveness coefficient g_k satisfy $g_k = \hat{g}\rho^k$. Overconfidence in the precision of the noisy signal z_{it} ($\hat{\sigma}_\epsilon^{-2} > \sigma_\epsilon^{-2}$) implies that $\hat{g} = \frac{\mathbb{V}ar[y_t|\Omega_{it}]}{\mathbb{V}ar[y_t|\Omega_{it}]+\hat{\sigma}_\epsilon^2} > g$, where g denotes the rational Kalman gain from case (i). The proof of this case now follows by repeating the same steps as in case (i).

Case (iii): Extrapolation. Individual forecasts follow $f_{it}y_{t+k} = \hat{\rho}^k y_t = \bar{f}_t y_{t+k}$, so that the responsiveness coefficient g_k equals $\hat{\rho}^k > \rho^k$. We need to show that both δ and γ have the same sign as $\rho^k - g_k = \rho^k - \hat{\rho}^k$. To do so, notice that average forecast errors equal

$$y_{t+k} - \bar{f}_t y_{t+k} = \left(\rho^k - \hat{\rho}^k\right)y_t + u_{t,t+h}.$$

It follows that $\gamma = \rho^k - \hat{\rho}^k$ and $\delta \propto \left(\rho^k - \hat{\rho}^k\right)\hat{\rho}^k\mathbb{C}ov\left[\bar{f}_t y_t - \bar{f}_{t-1}y_t, y_t\right]$. By a parallel argument to case (i), we have that $\mathbb{C}ov\left[\bar{f}_t y_t - \bar{f}_{t-1} y_t, y_t\right] = \left(1 - \rho\hat{\rho}^k\right) \mathbb{V}ar\left[y_t\right] > 0$. Thus, both coefficients have the same sign as $\rho^k - g_k = \rho^k - \hat{\rho}^k$.

Case (iv): Diagnostic expectations. Individual forecasts equal $f_{it} y_{t+k} = \mathbb{E}_{it-1} y_{t+k} + g_k \left(z_{it} - \mathbb{E}_{it-1} z_{it}\right)$, where the responsiveness coefficient satisfies $g_k = \hat{g}\rho^k$ with $\hat{g} > g$, and where g once more is the rational Kalman gain from case (i). We need to show that δ and γ have the same sign as $\rho^k - g_k = \rho^k(1 - \hat{g})$. Averaging individual forecasts across agents and rearranging terms shows that $\bar{f}_t y_t = \bar{\mathbb{E}}_t y_t + \frac{\hat{g}-g}{g}\left(\bar{\mathbb{E}}_t y_t - \bar{\mathbb{E}}_{t-1} y_t\right)$. Thus,

$$
\begin{aligned}
\gamma &= \rho^k \mathbb{C}ov\left[y_t - \bar{f}_t y_t, y_t\right] \mathbb{V}\left[y_t\right]^{-1} = \rho^k \left(\frac{1-g}{g} - \frac{\hat{g}-g}{g}\right) \mathbb{C}ov\left[\bar{\mathbb{E}}_t y_t - \bar{\mathbb{E}}_{t-1} y_t, y_t\right] \mathbb{V}\left[y_t\right]^{-1} \\
&= \rho^k \frac{1-\hat{g}}{g} \mathbb{C}ov\left[\bar{\mathbb{E}}_t y_t - \bar{\mathbb{E}}_{t-1} y_t, y_t\right] \mathbb{V}\left[y_t\right]^{-1},
\end{aligned}
$$

where we have used the expression for γ from case (i) to evaluate the rational forecasts $\bar{\mathbb{E}}_t\left[\cdot\right]$. Since $\mathbb{C}ov\left[\bar{\mathbb{E}}_t y_t - \bar{\mathbb{E}}_{t-1} y_t, y_t\right] > 0$ (see case (i)), γ has the same sign as $\rho^k - g_k$. Finally, Proposition 2 in Bordalo *et al.* (2018) shows that δ has the same sign as $1 - \hat{g}$ (their notation sets $\hat{g} = g(1+\theta)$, where $\theta > 0$), which is proportional to $\rho^k - \hat{g}$. This completes the proof. □

A.3 Asymmetric Attention

Proof of Lemma 1: The proof follows directly from the derivation of the Kalman gain g_j. At date t, agent i's signals z_{ijt} are informationally equivalent to

$$
\hat{z}_{ijt} \equiv \frac{z_{ijt}}{a_j} = \theta_t + \theta_t + \xi_{ijt},
$$

where the noise terms equal $\xi_{ijt} \equiv \frac{1}{a_j}\left[b_j u_{jt} + q_j \epsilon_{ijt}\right]$, and have precision (inverse variance)

$$
\tau_j = \frac{a_j^2}{b_j^2 + q_j^2} = \frac{a_j^2}{b_j^2} m_j.
$$

The standard formula for Gaussian updating implies that

$$
\mathbb{E}_{it}\left[\theta_t\right] = \mathbb{E}_{it-1}\left[\theta_t\right] + \sum_j \left(\frac{\tau_j}{\bar{\tau} + \sum_k \tau_k}\right) \left(\hat{z}_{ijt} - \mathbb{E}_{it-1}\left[\hat{z}_{ijt}\right]\right), \tag{A3}
$$

where $\bar{\tau} \equiv \mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it-1}\right]^{-1}$, while the posterior precision satisfies $\mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it}\right]^{-1} = \bar{\tau} + \sum_k \tau_k$.

Combining, and inserting the definition of $\hat{z}_{ijt}$ into (A3), we obtain that

$$\mathbb{E}_{it}\left[\theta_t\right] = \mathbb{E}_{it-1}\left[\theta_t\right] + \sum_j \mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it}\right] \frac{a_j}{b_j^2} m_j \left(z_{ijt} - \mathbb{E}_{t-1} z_{ijt}\right).$$

Equating $g_j = \mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it}\right] \frac{a_j}{b_j^2} m_j$ then completes the proof. $\square$

Proof of Proposition 2: We start with the characterization of the extrapolation coefficient γ in (1). Equation (14) shows that the sign of γ is determined by

$$\begin{aligned}
\gamma \propto \sum_j \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, x_{jt}\right] &= \sum_j \left(a_j \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, \theta_t\right] + b_j \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, u_{jt}\right]\right) \\
&= \sum_j \left(a_j \mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it}\right] - b_j \mathbb{C}\text{ov}\left[\mathbb{E}_{it}\theta_t, u_{jt}\right]\right),
\end{aligned} \tag{A4}$$

since $\mathbb{C}\text{ov}\left(\theta_t, u_{jt}\right) = 0$ and $\mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, \theta_t\right] = \mathbb{E}\left[\left(\theta_t - \mathbb{E}_{it}\theta_t\right)^2\right] = \mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it}\right]$.

Lemma 1 now implies that

$$\mathbb{C}\text{ov}\left[\mathbb{E}_{it}\theta_t, u_{jt}\right] = \mathbb{C}\text{ov}\left[g_j z_{ijt}, u_{jt}\right] = g_j b_j = \mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it}\right] \frac{a_j}{b_j} m_j.$$

Substituting this expression into (A4), we conclude that

$$\gamma \propto \sum_j \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, x_{jt}\right] = \mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it}\right] \sum_j a_j (1 - m_j).$$

This completes the first step of the proposition.

Turning to the characterization of the underreaction coefficient δ in (2), we start by solving the Kalman Filter in (13) backwards to obtain

$$\mathbb{E}_{it}\left[\theta_t\right] = \sum_{h=0}^{\infty} \lambda^h \hat{z}_{it-h}, \tag{A5}$$

where we define the precision-weighted signal $\hat{z}_{it} \equiv \sum_j g_j z_{ijt}$, and let $\lambda \equiv \left(1 - \sum g_j a_j\right)\rho$. The average precision-weighted signal is $\int_0^1 \hat{z}_{it} di = \hat{z}_{it} - \hat{\epsilon}_{it}$ for all $i \in [0, 1]$, with $\hat{\epsilon}_{it} = \sum_j g_j q_j \epsilon_{ijt}$.

We thus find that the average forecast revision equals

$$\bar{\mathbb{E}}_t\theta_t - \bar{\mathbb{E}}_{t-1}\theta_t = \bar{\mathbb{E}}_t\theta_t - \rho\bar{\mathbb{E}}_{t-1}\theta_{t-1} = \sum_{h=0}^{\infty} \lambda^h \left(\hat{z}_{it-h} - \hat{\epsilon}_{it-h}\right) - \rho \sum_{h=1}^{\infty} \lambda^{h-1} \left(\hat{z}_{it-h} - \hat{\epsilon}_{it-h}\right).$$

By the projection theorem, agent i's forecast error $\theta_t - \mathbb{E}_{it}\theta_t$ is uncorrelated with $\hat{z}_{it-h}$ for all $h \geq 0$. Thus, the characterization of δ in (15) yields:

$$
\begin{aligned}
\delta &\propto \mathbb{C}\mathrm{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, \bar{\mathbb{E}}_t\theta_t - \bar{\mathbb{E}}_{t-1}\theta_t\right] \\
&= \mathbb{C}\mathrm{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, -\sum_{h=0}^{\infty}\lambda^h\hat{\epsilon}_{it-h} + \rho\sum_{h=1}^{\infty}\lambda^{h-1}\hat{\epsilon}_{it-h}\right] \\
&= \mathbb{C}\mathrm{ov}\left[\sum_{h=0}^{\infty}\lambda^h\hat{z}_{it-h}, \sum_{h=0}^{\infty}\lambda^h\hat{\epsilon}_{it} - \rho\sum_{h=1}^{\infty}\lambda^{h-1}\hat{\epsilon}_{it-h}\right] \\
&= \mathbb{V}\mathrm{ar}\left[\epsilon_{it}\right]\left[1 + \sum_{h=1}^{\infty}\lambda^h\left(\lambda^h - \rho\lambda^{h-1}\right)\right] = \mathbb{V}\mathrm{ar}\left[\epsilon_{it}\right]\frac{1-\lambda\rho}{1-\lambda^2},
\end{aligned}
$$

where the third and fourth equality use $\mathbb{C}\mathrm{ov}\left[\theta_t, \hat{\epsilon}_{it-h}\right] = 0$ and $\mathbb{C}\mathrm{ov}\left[\hat{z}_{it-\ell}, \hat{\epsilon}_{it-h}\right] = 1_{\ell=h}\mathbb{V}\left[\hat{\epsilon}_{it}\right]$.

Since $\lambda < \rho \leq 1$, we conclude that

$$
\delta \propto \mathbb{V}\mathrm{ar}\left[\epsilon_{it}\right] = \sum_j g_j^2 q_j^2 = \mathbb{V}\mathrm{ar}\left[\theta_t \mid \Omega_{it}\right]\sum_j \frac{a_j^2}{b_j^2}m_j(1-m_j)
$$

This expression is positive whenever $0 < m_j < 1$ for at least one j. □

A.4 Optimal Attention Choice

Proof of Lemma 2: We consider the minimized expected loss at the start of period t:

$$
L_t^\star \equiv \mathbb{E}\left\{\min_{a_{it}}\mathbb{E}\left[\left(a_{it} - a_t^\star\right)^2\middle|\,\Omega_{it}\right]\right\}
$$

The minimizer in this problem is

$$
a_{it} = \mathbb{E}\left[a_t^\star|\Omega_{it}\right]
$$

Substituting:

$$
\begin{aligned}
L_t^\star = \mathbb{E}\left[\left(a_t^\star - \mathbb{E}\left[a_t^\star|\Omega_{it}\right]\right)^2\right] \quad &= \quad \mathbb{E}\left[\mathbb{E}\left[\left(a_t^\star - \mathbb{E}\left[a_t^\star|\Omega_{it}\right]\right)^2 \mid \Omega_{it}\right]\right] \\
&= \quad \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[a_t^\star|\,\Omega_{it}\right]\right] = \mathbb{V}\mathrm{ar}\left[a_t^\star|\,\Omega_{it}\right].
\end{aligned}
$$

Using the law of total variance:

$$
L_t^\star = \mathbb{V}\mathrm{ar}\left[a_t^\star \mid \Omega_{it}, \theta_t\right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[a_t^\star \mid \Omega_{it}, \theta_t\right] \mid \Omega_{it}\right]. \tag{A6}
$$

Note that:

$$
x_{jt}|\theta_t \sim N\left(a_j\theta_t, b_j^2\right)
$$

Agent i's information Ω_{it} contains the unbiased signal z_{ijt} of x_{jt}, defined in (11), which has precision q_j^{-2}. All other elements of Ω_{it} are independent of x_{jt} conditional on θ_t.

Using Bayes' law for Gaussian variables:

$$\begin{aligned}
\mathbb{E}[x_{jt}|z_{ijt},\theta_t] &= \mathbb{E}[x_{jt}|\theta_t] + \frac{\mathbb{C}\text{ov}\left[x_{jt}, z_{ijt}|\,\theta_t\right]}{\mathbb{V}\text{ar}\left[z_{ijt}|\,\theta_t\right]}\left(z_{ijt} - \mathbb{E}[x_{jt}|\theta_t]\right) \\
&= a_j\theta_t + \underbrace{\frac{b_j^2}{b_j^2 + q_j^2}}_{\equiv m_j}\left(z_{ijt} - a_j\theta_t\right) = (1-m_j)a_j\theta_t + m_j a_{ijt}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}\text{ar}\left[x_{jt} \mid \Omega_{it},\theta_t\right] &= \mathbb{V}\text{ar}\left[x_{jt} \mid \theta_t\right] - \frac{\mathbb{C}\text{ov}\left[x_{jt}, z_{ijt}|\,\theta_t\right]^2}{\mathbb{V}\text{ar}\left[z_{ijt} \mid \theta_t\right]} \\
&= b_j^2 - \frac{b_j^4}{b_j^2+q_j^2} = b_j^2\left(1 - \frac{b_j^2}{b_j^2+q_j^2}\right) = b_j^2\left(1-m_j\right).
\end{aligned}$$

Computing the first term in (A6)

$$\begin{aligned}
\mathbb{V}\text{ar}\left[a_t^\star \mid \Omega_{it},\theta_t\right] &= \mathbb{V}\text{ar}\left[w_\theta\theta_t + \sum_j w_{xj}x_{jt}\,\middle|\,\Omega_{it},\theta_t\right] = \mathbb{V}\text{ar}\left[\sum_j w_{xj}x_{jt}\,\middle|\,\Omega_{it},\theta_t\right] \\
&= \sum_j w_{xj}^2\mathbb{V}\text{ar}\left[x_{jt}|\,\Omega_{it},\theta_t\right] + \sum_j\sum_{k\neq j}\underbrace{\mathbb{C}\text{ov}\left[x_{jt},x_{kt}|\,\Omega_{it},\theta_t\right]}_{=0} = \sum_j w_{xj}^2 b_j^2(1-m_j).
\end{aligned}$$

Computing the second term in (A6)

$$\begin{aligned}
\mathbb{E}\left[a_t^\star \mid \Omega_{it},\theta_t\right] &= \mathbb{E}\left[w_\theta\theta_t + \sum_j w_{xj}x_{jt} \mid \Omega_{it},\theta_t\right] \\
&= w_\theta\theta_t + \sum_j w_{xj}\mathbb{E}\left[x_{jt} \mid \Omega_{it},\theta_t\right] \\
&= w_\theta\theta_t + \sum_j w_{xj}\left((1-m_j)a_j\theta_t + m_j z_{ijt}\right)
\end{aligned}$$

and

$$
\begin{aligned}
\mathbb{V}\text{ar}\left[\mathbb{E}\left[a_t^\star \mid \Omega_{it}, \theta_t\right] \mid \Omega_{it}\right] &= \mathbb{V}\text{ar}\left[w_\theta \theta_t + \sum_j w_{xj}\left((1-m_j)a_j\theta_t + m_j z_{ijt}\right) \middle| \Omega_{it}\right] \\
&= \mathbb{V}\text{ar}\left[\left(w_\theta + \sum_j w_{xj}(1-m_j)a_j\right)\theta_t \middle| \Omega_{it}\right] \\
&= \left[w_\theta + \sum_j w_{xj}(1-m_j)a_j\right]^2 \mathbb{V}\text{ar}_t\left[\theta_t\right].
\end{aligned}
$$

Substituting into (A6) yields the relevant expression in Lemma 2 in the text. □

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

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

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

The Lagrangian for this problem is

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

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

A.5 A Macroeconomic Example

Proof of Proposition 4: We start with a firm's output choice,[36]

$$
\begin{aligned}
Y_i = argmax \mathcal{V}_i &= \mathbb{E}_i\left[\frac{1}{PY}\left(PY^{\frac{1}{\sigma}}Y_i^{1-\frac{1}{\sigma}} - WN_i\right)\right] \\
&= \mathbb{E}_i\left[\left(\frac{Y_i}{Y}\right)^{1-\frac{1}{\sigma}} - \frac{W}{PY}\left(\frac{Y_i}{A_i}\right)^{\frac{1}{\alpha}}\right].
\end{aligned}
$$

Thus,

$$
\mathcal{V}_i = \mathcal{V}\left(Y_i, Y, A_i, \frac{W}{P}\right).
$$

[36]Since all actions are taken within period, we remove time subscripts, to economize on notation.

A second-order log-linear approximation of $\mathcal{V}$ then results in

$$v(y_i, y, a_i, \omega) \approx v_1 y_i + \frac{v_{11}}{2} y_i^2 + v_{12} y_i y + v_{13} y_i a_i + v_{14} y_i \omega + t.i.a, \tag{A7}$$

where $\omega = w - p$ and $t.i.a$ stands for *terms independent* of the firm's *action* y_i.

As a result of (A7), a firm's optimal, full-information choice of output is

$$y_i^\star = \frac{v_{12}}{|\, v_{11} \,|} y + \frac{v_{13}}{|\, v_{11} \,|} a_i + \frac{v_{14}}{|\, v_{11} \,|} \omega, \tag{A8}$$

while a firm's optimal choice under imperfect information is, because of certainty-equivalence,

$$y_i = \mathbb{E}_i \left[y_i^\star \right]. \tag{A9}$$

It remains to derive the optimal output choice under full information in (A8). A few simple but tedious derivations combine to show that

$$y_i^\star = r a_i + \alpha r \left(\sigma^{-1} y - \omega \right) \equiv x_{i1} + x_2. \tag{A10}$$

We note for later use that the equilibrium expression for the real wage in (A10) is $\omega = \mathbb{E}_h y + u^n$.

Finally, we can use (A8) and (A9) to derive the difference between a firm's valuation of its profits $v_i = v(y_i, y, a_i, \omega)$ and those that would have arisen under full information $v_i^\star$:

$$\begin{aligned} v_i - v_i^\star &= \frac{v_{11}}{2} y_i^2 - \frac{v_{11}}{2} y_i^{\star 2} + (v_{12} y + v_{13} a_i + v_{14} \omega)(y_i - y_i^\star) \\ &= \frac{v_{11}}{2} y_i^2 - \frac{v_{11}}{2} y_i^{\star 2} - v_{11} y_i^\star (y_i - y_i^\star) = \frac{v_{11}}{2} (y_i - y_i^\star)^2, \end{aligned} \tag{A11}$$

where we have used the first-order condition for optimal output in (A7). □

Proof of Proposition 5: Follows immediately from (A9) and (A11). □

B Over- and Underreactions in a General Linear Model

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

$$y_t = D\theta_t + Ex_t + Fu_t \quad \text{(A12)}$$
$$x_t = A\theta_t + Bx_{t-1} + Cu_t, \quad \text{(A13)}$$

where y_t is a scalar variable, θ_t is an $n_\theta \times 1$ vector of fundamental states, x_t is an $n_x \times 1$ vector of structural components, and lastly 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),

$$\theta_t = M\theta_{t-1} + Nu_t, \quad \text{(A14)}$$

where M and N are conformable matrices.

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

$$z_{it} = x_t + Q\epsilon_{it}, \quad Q = diag\,(q)\,, \quad \text{(A15)}$$

where ϵ_{it} is an $n_x \times 1$ vector of *i.i.d.* standard normal random variables.

It is useful to re-write the system, comprised of (A12) to (A14), as

$$y_t = \alpha\bar{\theta}_t + \beta u_t, \quad \text{(A16)}$$

where $\alpha = \begin{bmatrix} D & E \end{bmatrix}$, $\bar{\theta}_t = \begin{bmatrix} \theta_t' & x_t' \end{bmatrix}'$ and $\beta = F$. We further have that

$$\bar{\theta}_t = \bar{M}\bar{\theta}_{t-1} + \bar{N}u_t, \quad \text{(A17)}$$

where

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

We can now also re-write (A15) as

$$z_{it} = L_0\bar{\theta}_t + L_1\bar{\theta}_{t-1} + Ru_t + Q\epsilon_{it}, \quad \text{(A18)}$$

where L_0, L_1 and R are implicitly defined.

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

Proposition 6. *If the economy evolves according to (A12)-(A15), then the population coefficients in the regression equations (1) and (2) satisfy:*

$$\gamma < 0 \iff \alpha \bar{M}^k \left(GQQ'E' + \Sigma_{\theta\bar{\theta}} D' + \Omega\right) < 0 \tag{A19}$$

$$\delta > 0 \iff \exists q_j \in (0\,,\infty)\,, \tag{A20}$$

where G is the Kalman Gain on z_{it} when forming expectations about $\bar{\theta}_t$, $\Sigma_{\theta\bar{\theta}}$, denotes the covariance term $\Sigma_{\theta\bar{\theta}} = \mathbb{C}ov\left(\theta_t,\, \bar{\theta}_t\right)$, and $\Omega = \left[\bar{N} - G\left(L_0\bar{N} + R\right)\right] F'$.

Similar to the results in Proposition 2, expectations are generically underresponsive in Proposition 6; $\delta > 0$ whenever agents pay limited attention to structural components. Furthermore, limited attention to countercyclical components (that is, those that are assigned a negative weight in G, or directly have a negative element in E) once more tend to push expectations towards measured overreactions to recent outcomes ($\gamma < 0$). This generalizes our key insight from the body of this paper. In deriving this proposition, we have in effect "adjusted" the $\gamma-$condition in Proposition 2 for (i) the direct impact that several, persistent fundamentals can have on output itself ($D \neq \underline{0}$), (ii) for any cross-correlation in errors between the signal vector and output ($\Omega \neq \underline{0}$); and lastly ($iii$) for any effects that lagged components may have on output (see the expression for $\bar{M}$). The business cycle model in Section 5 provides an example of a model in which the second extension is relevant.

Proof of Proposition 6: The proof proceeds in three steps: First, we derive an expression for one-period ahead forecast errors and the corresponding one-period ahead forecast revision. Then, we compute the extrapolation coefficient γ in (1). Lastly, we use our results to also calculate the underreaction coefficient δ in (2).

As a preliminary step, we note that for any random variable Z, the covariance of individual forecast errors with Z equals the covariance of average forecast errors with Z:

$$\mathbb{C}ov\left(y_{t+1} - \mathbb{E}_{it} y_{t+k}, Z\right) = \mathbb{C}ov\left(y_{t+1} - \bar{\mathbb{E}}_t y_{t+k}, Z\right).$$

This follows because the right-hand side is the integral of the left-hand side across individuals, and because the signals defined in (A18) have the same steady-state distribution for all individuals. In the remainder of the proof, we therefore use individual and average forecast errors interchangeably.

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

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

where G_k is equal to

$$G_k=\mathbb{C}ov\left(\bar{\theta}_{t+k}-\mathbb{E}_{it-1}\bar{\theta}_{t+k},\, z_{it}-\mathbb{E}_{it-1}z_t\right)\mathbb{V}\left[z_{it}-\mathbb{E}_{it-1}z_t\right]^{-1}. \tag{A21}$$

We note that

$$\bar{\mathbb{E}}_t\left[y_{t+k}\right]=\bar{\mathbb{E}}_{t-1}\left[y_{t+k}\right]+\alpha G_k\left(x_t-\bar{\mathbb{E}}_{t-1}\left[x_t\right]\right). \tag{A22}$$

We can now use (A22) to show that

$$\bar{\mathbb{E}}_t\left[y_{t+k}\right]-\bar{\mathbb{E}}_{t-1}\left[y_{t+k}\right] = \alpha G_k\left(x_t-\bar{\mathbb{E}}_{t-1}\left[x_t\right]\right) \tag{A23}$$

$$y_{t+k}-\bar{\mathbb{E}}_t\left[y_{t+k}\right] = \alpha\left(\bar{\theta}_{t+k}-\bar{\mathbb{E}}_t\left[\bar{\theta}_{t+k}\right]\right)+Fu_{t+k}. \tag{A24}$$

This completes the first step.

We are now ready to derive the overreaction coefficient γ:

$$\begin{aligned}\gamma &\propto \mathbb{C}ov\left(y_{t+k}-\mathbb{E}_{it}\left[y_{t+k}\right],\, y_t\right)=\mathbb{C}ov\left(y_{t+k}-\mathbb{E}_{it}\left[y_{t+k}\right],\, E\left(z_{it}-Q\epsilon_{it}\right)+D\theta_t+Fu_t\right) \\ &= \mathbb{C}ov\left(\alpha\left(\bar{\theta}_{t+k}-\mathbb{E}_{it}\bar{\theta}_{t+k}\right),\, -EQ\epsilon_{it}+D\theta_t+Fu_t\right) \\ &= \alpha\bar{M}^k\left\{\mathbb{C}ov\left(\bar{\theta}_t-\mathbb{E}_{it}\bar{\theta}_t,\, -\epsilon_{it}\right)Q'E'+\mathbb{C}ov\left(\bar{\theta}_t-\mathbb{E}_{it}\bar{\theta}_t,\, \theta_t\right)D'+\mathbb{C}ov\left(\bar{\theta}_t-\mathbb{E}_{it}\bar{\theta}_t,\, u_t\right)F'\right\},\end{aligned}$$

where the second line used that $x_t=z_{it}-Q\epsilon_{it}$. But since

$$\begin{aligned}\mathbb{C}ov\left(\bar{\theta}_t-\mathbb{E}_{it}\bar{\theta}_t,\, \theta_t\right) &= \mathbb{C}ov\left(\bar{\theta}_t-\mathbb{E}_{it}\bar{\theta}_t,\, \theta_t-\mathbb{E}_{it}\theta_t\right)=\Sigma_{\bar{\theta}\theta} \\ \mathbb{C}ov\left(\bar{\theta}_t-\mathbb{E}_{it}\bar{\theta}_t,\, u_t\right) &= \bar{N}-G\left(L_0\bar{N}+R\right) \\ \mathbb{C}ov\left(\bar{\theta}_t-\mathbb{E}_{it}\bar{\theta}_t,\, -\epsilon_{it}\right) &= GQ,\end{aligned}$$

where the last two equalities follow from

$$\mathbb{E}_{it}\left[\bar{\theta}_t\right]=\mathbb{E}_{it-1}\left[\bar{\theta}_t\right]+G\left(z_{it}-\mathbb{E}_{it-1}\left[z_{it}\right]\right).$$

We note that $G_k=\bar{M}^kG$. Thus,

$$\gamma\propto\alpha\bar{M}^k\left\{GQQ'E'+\Sigma_{\theta\bar{\theta}}D'+\left[\bar{N}-G\left(L_0\bar{N}+R\right)\right]F'\right\}.$$

This completes the second step of the proof.

Lastly, we compute the underreaction coefficient δ. Equation (A23), (A24) show that $\delta \propto \mathbb{C}ov\left(y_{t+k} - \bar{\mathbb{E}}_t\left[y_{t+k}\right], \bar{\mathbb{E}}_t\left[y_{t+k}\right] - \bar{\mathbb{E}}_{t-1}\left[y_{t+k}\right]\right)$ can be rewritten as

$$\begin{aligned}
\delta &\propto \alpha\mathbb{C}ov\left(\bar{\theta}_{t+k} - \bar{\mathbb{E}}_t\bar{\theta}_{t+k}, x_t - \bar{\mathbb{E}}_{t-1}x_t\right) G_k'\alpha' \\
&= \alpha\mathbb{C}ov\left(\bar{\theta}_{t+k} - \bar{\mathbb{E}}_{t-1}\bar{\theta}_{t+k} - G_k\left(x_t - \bar{\mathbb{E}}_{t-1}\left[x_t\right]\right), x_t - \bar{\mathbb{E}}_{t-1}x_t\right) G_k'\alpha' \\
&= \alpha\left\{\bar{G}_k\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right] - G_k\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right]\right\} G_k'\alpha',
\end{aligned}$$

where we define

$$\bar{G}_k \equiv \mathbb{C}ov\left(\bar{\theta}_{t+k} - \bar{\mathbb{E}}_{t-1}\bar{\theta}_{t+k}, x_t - \bar{\mathbb{E}}_{t-1}x_t\right)\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right]^{-1}.$$

Notice that $\bar{G}_k$ corresponds to the Kalman Gain of a hypothetical agent who at time t has the prior belief that $\bar{\theta}_{t+k} \sim \mathcal{N}\left(\bar{\mathbb{E}}_{t-1}\bar{\theta}_{t+k}, P\right)$, where $P = \mathbb{V}\left[\bar{\theta}_{t+k} \mid z_i^{t-1}\right]$, but observes x_t perfectly (i.e. without noise). We conclude that

$$\begin{aligned}
\delta &\propto \alpha\left(\bar{G}_k - G_k\right)\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right] G_k'\alpha' \\
&= \left(\bar{d}_k - d_k\right)\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right] d_k', \qquad \text{(A25)}
\end{aligned}$$

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 $\mid \bar{G}_{j,k} \mid > \mid G_{j,k} \mid$ (due to the noise in private signals) and $sign(\bar{G}_{j,k}) = sign(G_{j,k})$. We also note for the same reasons $\mid \bar{d}_k \mid > \mid d_k \mid$. Combined, it now follows from (A25) that, because $\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right]$ is also positive semi-definite, $\delta > 0$ (Abadir and Magnus, 2005; Chpt.8). $\square$

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

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

$$A = \left[\begin{array}{cc} 0_{N\times 1} & \text{diag}\left(a_1, ..., a_N\right) \end{array}\right], \quad C = \left[\begin{array}{cc} 0_{N\times 1} & \text{diag}\left(b_1, ..., b_N\right) \end{array}\right]$$

$$M = \rho, \quad N = \left[\begin{array}{cc} \sigma_\theta, & 0_{1\times N} \end{array}\right]$$

An application of Proposition 6, with G evaluated according to the standard expression for Kalman Gains (Anderson and Moore, 2012), then also establishes Proposition 2.

References

Abadir, K. M. and Magnus, J. R. (2005). *Matrix algebra*, vol. 1. Cambridge University Press.

Anderson, B. D. and Moore, J. B. (2012). *Optimal filtering*. Courier Corporation.

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

Andreassen, P. and Kraus, S. (1988). Judgmental prediction by extrapolation. *Unpublished paper, Department of Psychology, Harvard University*.

Angeletos, G.-M., Collard, F. and Delad, H. (2018). Quantifying confidence. *Econometrica*.

— and Huo, Z. (2019). Myopia and anchoring.

— and La'O, J. (2010). Noisy business cycles. *NBER Macroeconomics Annual*, **24** (1), 319–378.

— and — (2013). Sentiments. *Econometrica*, **81** (2), 739–779.

—, — and Iovino, L. (2016). Real rigidity, nominal rigidity, and the social value of information. *American Economic Review*, **106(1)**, 200–227.

Arias, A., Hansen, G. D. and Ohanian, L. E. (2007). Why have business cycle fluctuations become less volatile? *Economic theory*, **32** (1), 43–58.

Barberis, N., Greenwood, R., Jin, L. and Shleifer, A. (2016). *Extrapolation and bubbles*. Tech. rep., National Bureau of Economic Research.

—, Shleifer, A. and Vishny, R. (1998). A model of investor sentiment. *Journal of financial economics*, **49** (3), 307–343.

Blanchard, O. J., L'Huillier, J.-P. and Lorenzoni, G. (2013). News, noise, and fluctuations: An empirical exploration. *American Economic Review*, **103** (7), 3045–70.

Bordalo, P., Gennaioli, N., Ma, Y. and Shleifer, A. (2018). *Overreaction in macroeconomic expectations*. Tech. rep., Working Paper.

—, — and Shleifer, A. (2017). Diagnostic expectations and credit cycles. *Journal of Finance*.

Broer, T. and Kohlhas, A. (2019). Forecaster (mis-)behavior.

Cagan, P. (1956). *The monetary dynamics of hyperinflation; In: Studies in the Quality Theory of Money (M. Friedman ed.* Tech. rep.

Chahrour, R. and Ulbricht, R. (2018). Robust predictions for dsge models with incomplete information. *Available at SSRN 3292550*.

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

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

—, — and Kumar, S. (2018). How do firms form their expectations? new survey evidence. *American Economic Review*, **108** (9), 2671–2713.

Comin, D. and Philippon, T. (2005). The rise in firm-level volatility: Causes and consequences. *NBER macroeconomics annual*, **20**, 167–201.

Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory.*

Croushore, D. (1998). Evaluating inflation forecasts.

Croushore, D. D. (1993). Introducing: the survey of professional forecasters. *Business Review-Federal Reserve Bank of Philadelphia*, **6**, 3.

— (1997). The livingston survey: still useful after all these years. *Business Review-Federal Reserve Bank of Philadelphia*, **2**, 1.

Cutler, D. M., Poterba, J. M. and Summers, L. H. (1990). Speculative dynamics and the role of feedback traders. *The American Economic Review*, **80** (2), 63–68.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, **53** (6), 1839–1885.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, **98** (4), 703–738.

DeLong, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J. (1990). Positive feedback investment strategies and destabilizing rational speculation. *the Journal of Finance*, **45** (2), 379–395.

Ehrbeck, T. and Waldmann, R. (1996). Why are professional forecasters biased? agency versus behavioral explanations. *The Quarterly Journal of Economics*, **111** (1), 21–40.

Eyster, E., Rabin, M. and Vayanos, D. (2019). Financial markets where traders neglect the informational content of prices. *The Journal of Finance*, **74** (1), 371–399.

Fernández-Villaverde, J., Rubio-Ramirez, J. and Sargent, T. (2007). The abc (and d) of strucutual vars. *American Economic Review*, **97**, 1021–1026.

Fuhrer, J. (2017). *Intrinsic Expectations Persistence: Evidence from professional and household survey expectations*. Tech. rep.

Fuster, A., Hebert, B. and Laibson, D. (2012). Natural expectations, macroeconomic dynamics, and asset pricing. *NBER Macroeconomics Annual*, **26** (1), 1–48.

Gabaix, X. (2017). Behavioral inattention. *NBER Working Paper*.

Gali, J. (2008). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.

Galí, J. and Gambetti, L. (2009). On the sources of the great moderation. *American Economic Journal: Macroeconomics*, **1** (1), 26–57.

Goodwin, R. M. (1947). Dynamical coupling with especial reference to markets having production lags. *Econometrica*, pp. 181–204.

Greenwood, R. and Hanson, S. G. (2014). Waves in Ship Prices and Investment. *The Quarterly Journal of Economics*, **130** (1), 55–109.

— and Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, **27** (3), 714–746.

Hansen, G. D. (1985). Indivisible labor and the business cycle. *Journal of monetary Economics*, **16** (3), 309–327.

Hellwig, C., Kohls, S. and Veldkamp, L. (2012). Information choice technologies. *American Economic Review*, **102** (3), 35–40.

— and Venkateswaran, V. (2009). Setting the right prices for the wrong reasons. *Journal of Monetary Economics*, **56**, 57–77.

Hirshleifer, D., Lim, S. S. and Teoh, S. H. (2011). Limited investor attention and stock market misreactions to accounting information. *The Review of Asset Pricing Studies*, **1** (1), 35–73.

Kahneman, D. and Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, **3** (3), 430–454.

Kydland, F. E. and Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pp. 1345–1370.

Lamont, O. A. (2002). Macroeconomic forecasts and microeconomic forecasters. *Journal of economic behavior & organization*, **48** (3), 265–280.

Lorenzoni, G. (2009). A theory of demand shocks. *American Economic Review*, **99** (5), 2050–84.

Lucas, R. (1977). Understanding business cycles. In *Carnegie-Rochester conference series on public policy*, North-Holland, vol. 5, pp. 7–29.

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

— (1973). Some international evidence on output-inflation tradeoffs. *The American Economic Review*, **63** (3), 326–334.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Maćkowiak, B., Matějka, F. and Wiederholt, M. (2018). The rational inattention filter. *Journal of Economic Theory*.

— and Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *The American Economic Review*, **99** (3), 769–803.

— and Wiederholt, M. (2015). Business cycle dynamics under rational inattention. *The Review of Economic Studies*, **82** (4), 1502–1532.

Mankiw, N. G. and Reis, R. (2002). Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. *The Quarterly Journal of Economics*, **117** (4), 1295–1328.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, pp. 315–335.

Nimark, K. (2008). Dynamic pricing and imperfect common knowledge. *Journal of monetary Economics*, **55** (2), 365–382.

Nimark, K. P. (2014). Man-bites-dog business cycles. *American Economic Review*, **104** (8), 2320–67.

— (2015). A low dimensional kalman filter for systems with lagged states in the measurement equation. *Economics Letters*, **127**, 10–13.

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

Pesaran, M. H. (1987). *The limits to rational expectations*.

— and Weale, M. (2006). Survey expectations. *Handbook of economic forecasting*, **1**, 715–776.

Rogerson, R. (1988). Indivisible labor, lotteries and equilibrium. *Journal of monetary Economics*, **21** (1), 3–16.

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

Stark, T. (1997). Macroeconomic forecasts and microeconomic forecasters in the survey of professional forecasters. *Federal Reserve Bank of Philadelphia Working Paper*.

— (2010). Realistic evaluation of real-time forecasts in the survey of professional forecasters. *Federal Reserve Bank of Philadelphia Research Rap, Special Report*, **1**.

Veldkamp, L. L. (2011). *Information choice in macroeconomics and finance*. Princeton University Press.

Wiederholt, M. (2010). Rational inattention. *The New Palgrave Dictionary of Economics (Online Edition ed.)*.

Woodford, M. (2002). *Imperfect Common Knowledge and the Effects of Monetary Policy*. Nber working papers, Department of Economics, Columbia University.

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