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 average new information that have been pointed out by Coibion and Gorodnichenko (2015). Our main contribution is to propose a unified model of expectation formation based on noisy rational expectations that resolves the friction between theory and data, and to explore its business cycle implications.

To empirically motivate our work, we demonstrate simultaneous overreactions and underreactions in a 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 average 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), as well as models that combine extrapolation bias or overconfidence with the presence of noisy information (e.g., Daniel *et al.*, 1998; Bordalo *et al.*, 2018). Intuitively, noisy information (or inattention) generates underreactions to new information, because individuals shrink their forecasts towards prior beliefs when the signals they observe are noisy. By contrast, extrapolation bias or overconfidence generates

[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.

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 that the same result extends to several influential models with a richer information structure (e.g., Lucas, 1973; Lorenzoni, 2009; Maćkowiak and Wiederholt, 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 observes 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 components that are either particularly volatile or 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 (2020), 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 firms' 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 productivity 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 their own productivity, because this component is substantially 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 a specific public signal (recent outcomes of the forecasted variable) with underreactions to *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 (see also Fuhrer, 2017 and Broer and Kohlhas, 2019). However, as we demonstrate in Section 2, simple versions of their framework cannot account for the simultaneous over- and underreactions of expectations that we document in the data. We therefore view these two papers as related and complementary steps towards a unified model of expectations that is consistent with over- and underreactions to new information.[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 discuss the relationship between our work and that of Bordalo *et al.* (2018) in detail in Section 3.4.

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 to 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 to 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.

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, because expectations about future output play a central role in the economy as determinants of consumption, inflation, and asset prices. Second, because data on output forecasts are available for a longer time-span than forecasts of most other variables. We later explore the robustness of our empirical estimates by considering forecasts of 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 captures 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 regression because it is easier to compare its results 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 such as ours. 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. 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 $\delta \neq 0$).

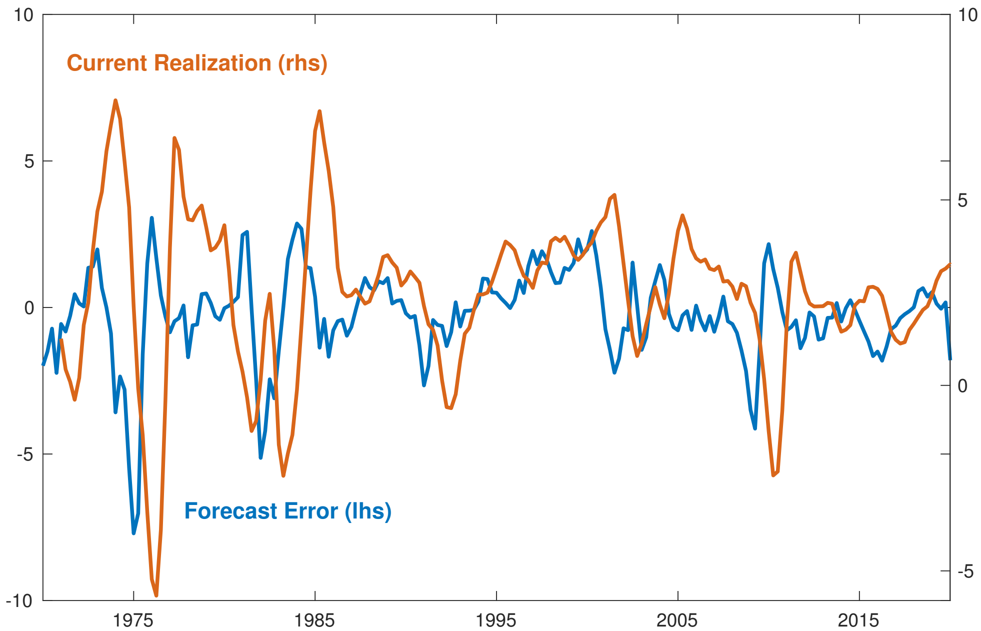
The raw data already hints at deviations from the FIRE benchmark. Figures 1 and 2 plot average one-year-ahead forecast errors (the average left-hand side of (1) and (2) across respondents, with $k = 4$) over time and compare them, respectively, to current realizations of output growth (the right-hand side of (1)) and average one-quarter revisions (the right-hand side of (2)).[7] 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 average forecast revisions are positively correlated within our sample. All else equal, this indicates that agents underreact to new information on average, as they are too pessimistic after positive forecast revisions, and vice versa after negative revisions.

Table I confirms these impressions and reports estimates of (1) and (2) using the SPF data on one-year-ahead forecasts ($k = 4$). In the first column, we estimate (1) and find that 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.

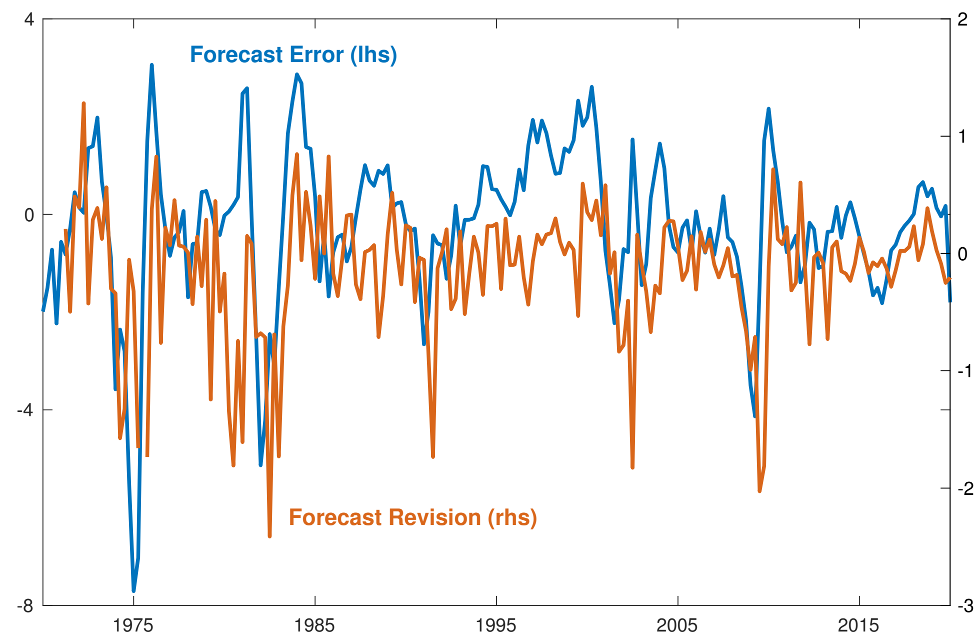
[7] We use real-time data to measure current realizations of output growth. Because the response deadline for the SPF is only one-week from the BEA's first release of output growth, we on the right-hand side of (2) average this release's value with its previous quarter's realization. This is to precisely capture the current conditions at the time the respondent institutions determine their published forecast (e.g., Croushore and Stark, 2019; Bordalo *et al.*, 2017). We do not make this adjustment for other variables and data sets that we consider below, as for these there is time to include information into published forecasts. Table C.6 in the online appendix shows that our results are similar using either of the two quarter's output growth values.

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. Current realizations are measured as the average of the BEA's first release value and its previous quarter's realization. This is to account for the timing of the SPF survey (see footnote 7 for further discussion).

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*

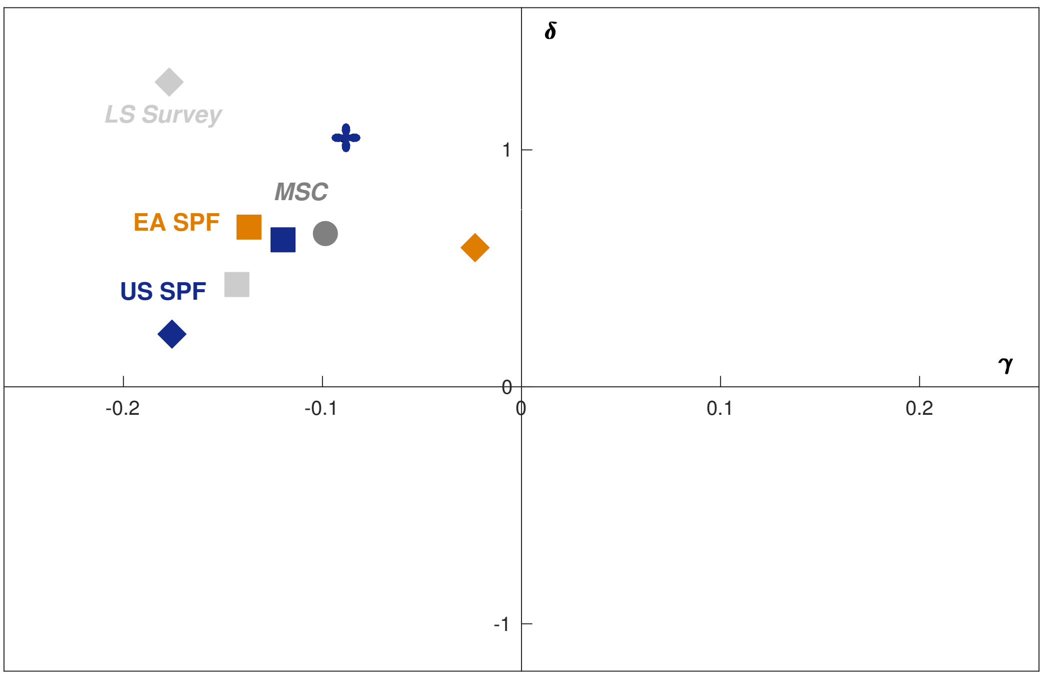
	Panel a: individual forecast error		
	(1)	(2)	(3)
Current Realization	-0.12**	–	-0.14***
	(0.05)		(0.04)
Average Revision	–	0.66***	0.69***
		(0.19)	(0.18)
Observations	7,190	7,151	7,094
F	185.0	423.5	360.9
R^2	0.03	0.06	0.10
	Panel b: average forecast error		
	(1)	(2)	(3)
Constant	0.01	-0.12	0.24
	(0.19)	(0.10)	(0.15)
Current Realization	-0.10**	–	-0.14***
	(0.05)		(0.05)
Average Revision	–	0.77***	0.84***
		(0.26)	(0.24)
Observations	198	197	196
F	3.53	16.0	11.7
R^2	0.02	0.08	0.11

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: 1970Q1-19Q4. Significance levels *=10%, **=5%, ***=1%.

In contemporaneous and closely 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 regression (1) to emphasize 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 to those in Table I 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,

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.

non-financial firms, and government agencies), and the *Michigan Survey of Consumers*.[8]

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.7 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.2-9 in the online appendix contain 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)[11], 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 average 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 systematic overreactions to recent realizations of the forecasted variable (i.e. extrapolation), but *simultaneous* underreactions to

[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 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, 2015).

[11]The Federal Reserve Bank of Philadelphia took over ownership of the SPF in 1990Q2.

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 formation are consistent with the data.

2.2 Existing Theories of Expectation Formation

We compare our estimates to a parsimonious framework, where agents observe noisy signals of the forecasted variable, which captures several popular models of expectation formation. On the one hand, we show that *rational* forecasts are inconsistent with overreactions to current output (i.e. $\gamma < 0$ in (1)), and that this extends to a collection of richer models. 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 (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 $\mathbb{C}\text{ov}\left[\epsilon_{it}, \epsilon_{js}\right] = 0$ 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.[12]

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

$$f_{it}y_{t+k} = \lambda f_{it-1}y_{t+k} + g_k\left(z_{it} - \lambda f_{it-1}z_{it}\right), \tag{5}$$

where $f_{it}y_{t+k} = \rho^k f_{it}y_t$. As with the textbook Kalman filter, the agent starts with her forecast of output at time $t-1$, and updates it in proportion to the new information in her signal at time t. Departing from the standard filter, we allow $g_k \geq 0$ to be an arbitrary gain parameter

[12] 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).

that measures agents' responsiveness to new information.[13] We also allow the prior update parameter $\lambda \in [0, 1]$ to be less than one. Despite its simplicity, the formulation in (5) nests a wide range of existing models of expectation formation. We demonstrate this through a series of examples, which we delineate into rational and behavioral theories:

Noisy Rational Expectations: Agents forecasts equal their conditional expectation $f_{it}y_{t+k} = \mathbb{E}\left[y_{t+k} \mid \Omega_{it}\right]$ and follow (5) with a gain parameter $g_k = \mathbb{C}\text{ov}_{t-1}\left[y_{t+k}, z_{it}\right] / \mathbb{V}\text{ar}_{t-1}\left[z_{it}\right] < \rho^k$ while $\lambda = 1$. This specification is identical to those from models with noisy rational expectations (Woodford, 2002) or rational inattention (Sims, 2003).[14] The special case in which agents observe output without noise ($\sigma_\epsilon = 0$) corresponds to the case of full information and rational expectations (FIRE), and implies that $f_{it}y_{t+k} = \rho^k y_t$ with $g_k = \rho^k > 0$.

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

- *Overconfidence*: Agents overestimate the precision of new information. They believe that the variance of the noise in their signals is $\hat{\sigma}_\epsilon^2 < \sigma_\epsilon^2$ (e.g. Daniel *et al.*, 1998; Hirshleifer *et al.*, 2011).[15] Agents forecasts follow the Recursive Forecast Equation in (5) with a sensitivity parameter $g_k \in (0, 1)$ that exceeds its rational value and $\lambda = 1$.

- *Extrapolation:* Agents overestimate the extent to which current output predicts future realizations. They observe output without noise ($\sigma_\epsilon^2 = 0$), but believe that the persistence parameter for output is $\hat{\rho} > \rho$ (e.g. DeLong *et al.*, 1990; Fuster *et al.*, 2012). Agents' forecasts satisfy (5) with a sensitivity parameter $g_k = \hat{\rho}^k > \rho^k$ and $\lambda = 0$.[16]

- *Diagnostic Expectations:* The model in Bordalo *et al.* (2017) and Bordalo *et al.* (2018) corresponds to the overconfidence case, but the effect of overconfidence is temporary and does not effect forecasts at future dates. Equation (5) 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 g_k exceeds its rational value. Despite the non-recursivity of forecasts, we include the model in this list because the properties of its forecast errors $y_{t+k} - f_{it}y_{t+k}$ depend only on $(\rho^k - g_k)(y_{t+k} - \mathbb{E}\left[y_{t+k} \mid \Omega_{it}\right])$, and thus exclusively on those from the noisy rational expectations case and (5) (Corollary 1 in Appendix A.1).

[13] To ensure that forecasts in (5) are well-defined, we impose that $g_k = \rho^k g_0$ has $g_0 \in (0, 2)$.

[14] 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).

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

[16] Thus, $f_{it}y_{t+k} = \hat{\rho}^k y_t$. The introduction contains a further list of references using such forecasts.

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

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

Proposition 1 demonstrates that models described by the Recursive Forecast Equation, such as the rational and two behavioral models above, all imply either underreactions in both of our main regressions ($\gamma > 0$ and $\delta > 0$), or overreactions ($\gamma < 0$ and $\delta < 0$). Indeed, Proposition 1 also implies that the coefficients γ and δ in the model of diagnostic expectations also both have the same sign as $\rho^k - g_k$ (Corollary 1 in Appendix A.1). This is at odds with our empirical estimates of *simultaneous* over- and underrreactions. One can see this discrepancy clearly in terms of Figure 3. Proposition 1 shows that an econometrician's estimates will fall either into the upper-right quadrant or the lower-left quadrant of the figure. This is inconsistent with our empirical estimates that center on the upper-left quadrant.

To interpret Proposition 1 further, recall that agents' gain parameter in the FIRE case, in which they perfectly observes current output, is equal to $g_k = \rho^k$ (since $f_{it}y_{t+k} = \rho^k y_t$). Proposition 1 states that there are two possible parametric regions, corresponding to systematic underreactions or overreactions, depending on whether agents' responsiveness to new information g_k is smaller or greater than in the FIRE benchmark.

Two counteracting effects determine the size of g_k. First, the presence of noise in agents' signals dampens agents' responsiveness to new information, which rationally pushes g_k below its FIRE value. This effect, all else equal, creates measured underreactions: An econometrician estimating (1) and (2) has access to more information than agents in the model, because he observes current output and average forecast revisions perfectly. As a result, forecast errors are predictable. And because agents respond to noisy information in a muted fashion, this predictability takes the shape of measured underreactions. Second, behavioral biases, such as overconfidence or extrapolation, heighten agents' responsiveness to new information, which in turn increases the gain coefficient. However, as Proposition 1 shows, only one of these forces can come to dominate the sufficient statistic g_k. Hence, in all of the above cases, agents either over- or underreact, but do not over- and underreact *simultaneously*.[17]

We conclude that a popular class of models, in which agents form Bayesian or non-Bayesian expectations based on noisy signals of the forecasted variable, is inconsistent with simultaneous over- and underreactions. In particular, it is clear that to explain the survey data, we must

[17]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.

consider a model with more than one sufficient statistic for belief formation. 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 however briefly consider more sophisticated existing models of noisy rational expectations.[18]

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 of future aggregate output behave *as if* agents obtained only a noisy signal of output itself. Indeed, Online Appendix E.1 shows that the above noisy rational expectations case, where $\gamma > 0$, 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 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 is similar to that in 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 creates underreactions. Indeed, we show that one can directly use (5) and the noisy rational expectations case above 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 is 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 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.

[18]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., $\gamma < 0$, $\delta < 0$).

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},

$$y_t = x_{1t} + x_{2t} + \ldots + x_{Nt}. \tag{6}$$

These components could, for example, represent different inputs into the economy's production function, different sectors of the economy, or different variables in firms' optimal production plans. We discuss one such example at length in Section 5. Each component x_{jt} is determined by the linear relationship

$$x_{jt} = a_j \theta_t + b_j u_{jt}, \quad u_{jt} \sim \mathcal{N}(0,\, 1), \tag{7}$$

where θ_t denotes a latent factor that follows the autoregressive process

$$\theta_t = \rho\theta_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}\left(0,\, \tau_\eta^{-1}\right), \tag{8}$$

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

$$z_{ijt} = x_{jt} + q_j \epsilon_{ijt}, \quad \epsilon_{ijt} \sim \mathcal{N}(0, 1), \quad j = \{1, 2, \ldots, N\}, \tag{9}$$

where q_j parameterizes the noise (or inattention) in agents' signals about the jth component, and ϵ_{ijt} is an idiosyncratic error term. Agent i's information set at time t is $\Omega_{it} = \{z_{i1s}, \ldots, z_{iNs}\}_{s \leq t}$.[19] Agents thus infer information about the latent factor θ_t from 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 (9) restricts 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 embedded in (9), combined with a component-based structure in (6), is a simple and common way to model this idea (see e.g. Maćkowiak and Wiederholt, 2009). We discuss the role of these restrictions in more detail in Section 4, where we also consider an alternative setup with fully flexible information design.

3.2 Definition of Attention

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

$$m_j \equiv \frac{\mathbb{V}\mathrm{ar}(x_{jt}|\theta_t)}{\mathbb{V}\mathrm{ar}(z_{ijt}|\theta_t)} = \frac{b_j^2}{b_j^2 + q_j^2} \in (0, 1). \tag{10}$$

These parameters measure the sensitivity of agents' expectations 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:[20]

[19]This assumption follows the convention in the literature. By allowing agents to observe an infinite history of signals, we ensure that their signal extraction problem is initialized in steady state.

[20]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 and the references therein).

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

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 condition on 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*

$$\mathbb{E}_{it}[\theta_t] \quad = \quad \mathbb{E}_{it-1}[\theta_t] + \sum_j g_j \left(z_{ijt} - \mathbb{E}_{it-1}[z_{ijt}]\right), \tag{11}$$

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

The lemma confirms that attention coefficients m_j drive agents' responses to new information. 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.[21]

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:

$$\gamma = \mathbb{C}\text{ov}\left[y_{t+k} - \mathbb{E}_{it} y_{t+k}, y_t\right] \mathbb{V}\text{ar}\left[y_t\right]^{-1} = d_0 \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, y_t\right], \tag{12}$$

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 (12) 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.

[21] 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 $\xi_{ijt} = (b_j u_{jt} + q_j \epsilon_{ijt})/a_j$. 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 $\frac{a_j^2}{b_j^2} m_j$.

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

$$\begin{aligned}\delta &= \mathbb{C}\text{ov}\left[y_{t+k} - \mathbb{E}_{it}y_{t+k}, \bar{\mathbb{E}}_t y_{t+k} - \bar{\mathbb{E}}_{t-1}y_{t+k}\right] \mathbb{V}\text{ar}\left[\bar{\mathbb{E}}_t y_{t+k} - \bar{\mathbb{E}}_{t-1}y_{t+k}\right]^{-1} \\ &= d_1 \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, \bar{\mathbb{E}}_t\theta_t - \bar{\mathbb{E}}_{t-1}\theta_t\right], \end{aligned} \tag{13}$$

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 (12) and (13) 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 (12) 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. $\square$

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

The economic intuition behind this fact, which captures one of the main 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 $\bar{\mathbb{E}}_t\theta_t - \bar{\mathbb{E}}_{t-1}\theta_t = \int_0^1 (\mathbb{E}_{jt}\theta_t - \mathbb{E}_{jt-1}\theta_t)\, dj$, the linearity of the covariance operator and the symmetry of attention choices 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 $\theta_t - \mathbb{E}_{jt}\theta_t$, 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.[22] The intuition is as 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 demonstrate how attention choices map into the over- and underreaction coefficients γ and δ, respectively. Specifically, they show 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 $j \in \{1, \ldots, N\}$ such that $0 < m_j < 1$.*

[22] 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 $\delta > 0$.

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 enviroment 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 limited, 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:[23]

$$y_{t+k} - f_{it}y_{t+k} = \alpha + \delta^{\text{ind}} \left(f_{it}y_{t+k} - f_{it-1}y_{t+k} \right) + \xi_{it}. \tag{14}$$

[23] See also Fuhrer (2017) and Broer and Kohlhas (2019) for related results using inflation forecasts.

Using a range of survey data, Bordalo *et al.* (2018) estimate that $\delta^{\text{ind}} < 0$, 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 (14) for output forecasts in the US SPF. We estimate overreactions to individual revisions ($\delta^{\text{ind}} < 0$), but unlike our estimates of (1) and (2), which motivate our analysis, this result appears sensitive to outliers.[24] Online Appendix F considers an extension of our framework, which allows for both asymmetric attention and irrational overconfidence (e.g. Moore and Healy, 2008 and Broer and Kohlhas, 2019). 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 ($\delta^{\text{ind}} < 0$ in (14)). Crucially, the extended model can fit these empirical patterns *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 examined.

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 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 in her signals). 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:

[24] Indeed, we cannot reject that $\delta^{\text{ind}} = 0$ 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). See also Angeletos and Huo (2020) for similar empirical results using inflation forecasts.

$$\mathcal{U}_{it} = -\left(a_t^{\star} - a_{it}\right)^2 - K(m). \tag{15}$$

The first term in (15) 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:

$$a_t^{\star} = w_\theta \theta_t + \sum w_{xj} x_{jt}, \tag{16}$$

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 (15) and (16) 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, (15) and (16) 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*

$$\begin{aligned} \mathbb{E}\left[\mathcal{U}_{it}\right] &= -\mathbb{V}ar\left[a_t^{\star} \mid \Omega_{it}\right] - K(m) && (17) \\ &= -\sum_j w_{xj}^2 b_j^2 (1-m_j) - \mathbb{V}ar_t\left[\theta_t\right]\left[w_\theta + \sum_j w_{xj} a_j \left(1-m_j\right)\right]^2 - K(m). && (18) \end{aligned}$$

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 (17). 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 (18) 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 (18) 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 (16) and (18), 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

$$\tau(m) = \sum_j \tau_j \tag{19}$$

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 latent factor, and we can write

$$\mathbb{V}\text{ar}_t\left[\theta_t\right] = V\left[\tau(m)\right], \tag{20}$$

where $V'(\tau) < 0$ and $\partial\tau/\partial m_j > 0$ from (19). Combined, (18) and (20) allow us to characterize agents' attention choices. Proposition 3 summarizes the results.

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

$$w_{xj}^2 b_j^2 + \mu_\tau a_j^2 b_j^{-2} + \mu_\alpha w_{xj} a_j = \frac{\partial K(m)}{\partial m_j}, \tag{21}$$

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 (21) 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).

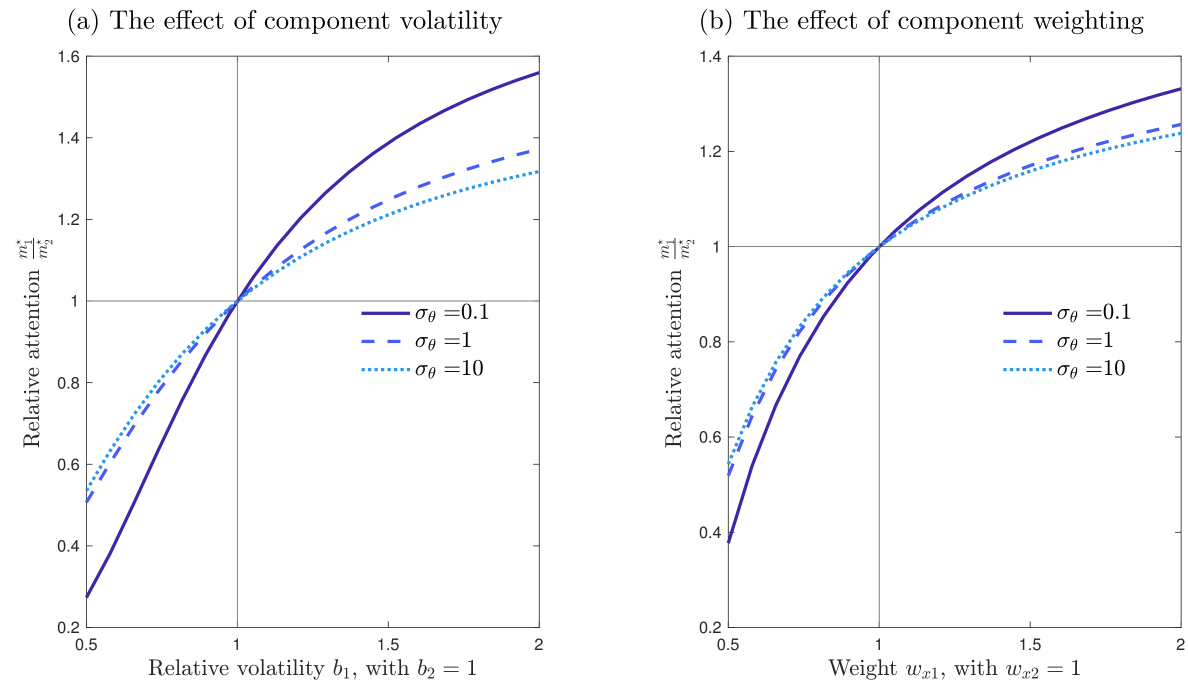
The second and third terms capture a more nuanced effect: By learning about x_{jt}, the agent also acquires information about the latent factor θ_t, which generates *learning spillovers* by resolving uncertainty about x_{kt} for $k \neq j$. The second term measures the effect of attention m_j on the 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 (16). 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 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 (21) explicitly for an entropy-based cost function, and shows that the main comparative statics remain the same. In addition, we show

Figure 4: Optimal Attention: Numerical Example



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.

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 (9) restricts agents to acquire N separate, conditionally independent signals z_{ijt} about the components x_{jt} of output. This is one of two popular approaches. An alternative approach is to instead allow agents full flexibility when designing the conditional distribution of their signals given the state of the economy (e.g., Sims, 2003). The choice between the two approaches is typically made based on the problem at hand, and on tractability. In the context of our analysis, it is interesting to compare the predictions of each approach.

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:[25]

$$s_{it}^{\star} \quad = \quad a_t^{\star} + h' v_t + q^{\star} \epsilon_{it}, \tag{22}$$

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

[25] Heuristically, we derive this result in two steps. We first express $a_t^{\star}$ as an ARMA process in reduced form. In particular, substituting (7) and (8) into (16) shows that

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

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 (22).

$K(m)$. Equation (22) 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 (15) is equivalent to the mean-squared error of next period's output forecast ($w_\theta = \rho \sum_j a_j$ and $w_{xj} = 0$, as discussed above). In this case, it follows that $h = 0$ in (22).[26] As a result, the fully optimal signal boils down to $s_{it}^\star = (\rho \sum_j a_j)\theta_t + q^\star \epsilon_{it}$, 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.[27] 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 $s_{it}^\star$ 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 $s_{it}^\star = x_{1t} + q^\star \epsilon_{it}$. 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}. □

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

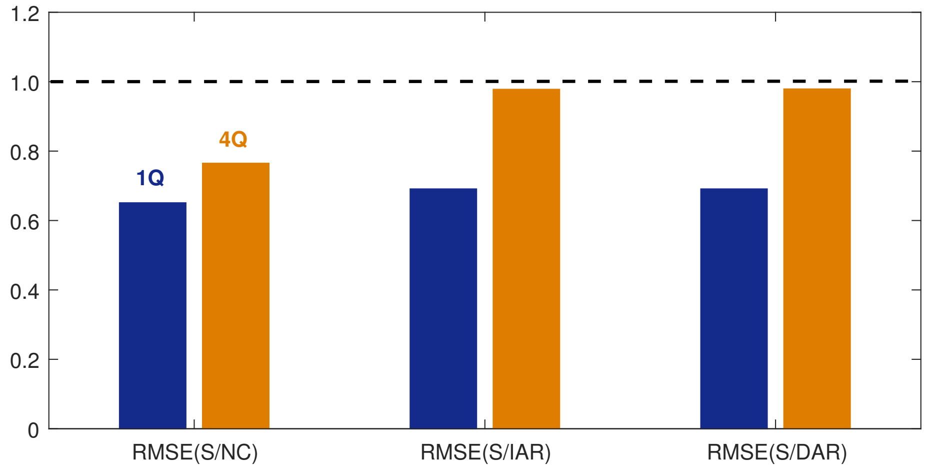
[26] See Corollary H.2 in the online appendix, or Cover and Thomas (2012) for the standard result in which the optimal action $a_t^\star$ is proportional to a simple AR(1) process.

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

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

where we have also used that $y_t = \sum_j a_j \theta_t + \sum_j b_j u_{jt}$ and that $\sum_j a_j > 0$.

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.

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.

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.[28] 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

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

derive forecasts from a misspecified ARMA model. Recent behavioral theory, such as Bordalo *et al.* (2018), is more nuanced, and further work would be needed to test whether forecasts in the data are more or less accurate than such theories predict. Hence, we interpret the supplementary evidence as a sanity check, which implies that our theory is consistent with moments of the data beyond the motivating evidence in Section 2.

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

5 A Macroeconomic Example

In this section, we illustrate the sources and effects of asymmetric attention in a 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 firm's own productivity; and second, an *aggregate supply component*, which captures the effects of other agents' behavior on an individual firm's 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 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

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log C_t - \xi_t N_t\right], \quad \xi_t > 0, \tag{23}$$

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

$$C_t = \left[\int_0^1 C_{it}^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}, \quad P_t = \left[\int_0^1 P_{it}^{\frac{1}{\sigma-1}} di\right]^{\sigma-1}, \tag{24}$$

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

$$\int_0^1 P_{it} C_{it} di + B_{t+1} \leq \int_0^1 \Pi_{it} di + W_t N_t + (1 + R_t) B_{t+1} + T_t^h, \tag{25}$$

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 (23) subject to (25).

Firms: A representative firm $i \in [0, 1]$ chooses its output 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

$$\mathcal{V}_{it} = \mathbb{E}_{it} \left[\frac{1}{P_t C_t} \Pi_{it} \right], \ \Pi_{it} = P_{it} Y_{it} - W_t N_{it}, \tag{26}$$

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

$$Y_{it} = A_{it} N_{it}^{\alpha}, \quad \alpha \in (0, 1), \tag{27}$$

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

$$a_{it} = \theta_t + u_t^x + \epsilon_{it}^a, \tag{28}$$

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

$$\theta_t = \rho \theta_{t-1} + u_t^{\theta}, \quad u_t^{\theta} \sim \mathcal{N}\left(0, \sigma_{\theta}^2\right), \tag{29}$$

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

$$\log \xi_t = \bar{\xi} + u_t^n, \quad u_t^n \sim \mathcal{N}\left(0, \sigma_n^2\right), \tag{30}$$

where $\bar{\xi} \in \mathbb{R}$. We show below that the labor supply shock 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 persistent productivity component θ_t, and then makes its labor supply choice.[29] 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*

$$v_{it} \simeq -\frac{1}{2}\mathbb{E}_{it}\left[(y_{it} - y_{it}^\star)^2\right], \tag{31}$$

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

$$y_{it}^\star = x_{i1t} + x_{2t} \tag{32}$$

with

$$x_{i1t} = ra_{it}, \qquad x_{2t} = \alpha r\left(\sigma^{-1}y_t - \omega_t\right), \tag{33}$$

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

In the spirit of Lucas (1977) and Maćkowiak and Wiederholt (2009), equation (32) and (33) decompose each firm's ideal output choice into two components: We refer to x_{i1t} as the

[29] 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, since it depends on a firm's own productivity a_{it}. 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 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 it faces ω_t is high. Both effects are captured in (33).[30]

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:

$$\Omega_{it} = \{z_{i1s}, z_{i2s}\}_{s \leq t}, \tag{34}$$

where

$$z_{i1t} = x_{i1t} + q_1 \epsilon_{i1t}, \quad z_{i2t} = x_{2t} + q_2 \epsilon_{i2t}, \tag{35}$$

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 in 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

$$y_t = \int_0^1 y_{it} di = x_{1t} + x_{2t}, \tag{36}$$

[30] 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 B.

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 (33). As for the productivity component x_{1t}, we have

$$x_{1t} = r\theta_t + ru_t^x, \tag{37}$$

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 $\omega_t = \mathbb{E}_{ht} y_t + u_t^n$, where $\mathbb{E}_{ht}[\cdot]$ denotes household expectations. Thus, we conclude from (33) and (36) that

$$x_{2t} = \alpha r \left(\frac{1-\sigma}{\sigma} y_t + \left(\frac{1}{1-\alpha} - r \right) u_t^x - u_t^n \right) = (1-r)\theta_t + \left(\frac{1}{1-\alpha} - r \right) u_t^x - \alpha u_t^n. \tag{38}$$

The first equality in (38) 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 (33). By contrast, with perfect competition ($\sigma = 1$), firms are price-takers and act independently of one another. The second equality in (38) 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 aggregate 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 relevant features of the closely related 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 3). Hence, while qualitative explorations of models of strategic complementarity between firms' choices have yielded important theoretical 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 type of model.

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

$$\gamma_i = \mathbb{C}ov\left(y_{t+1} - \mathbb{E}_{it} y_{t+1}, y_t\right) \mathbb{V}\text{ar}\left[y_t\right]^{-1} = -\rho \frac{\alpha}{1-\alpha} \frac{\mathbb{V}\text{ar}_t\left[\theta_t\right]}{\mathbb{V}\text{ar}\left[y_t\right]} < 0, \tag{39}$$

so that firm i appears to extrapolate.[31]

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 their corresponding expected profits:

Proposition 5. *An individual firm'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, discounted profits are* $v_{it}^{\star} \simeq -\frac{1}{2}\mathbb{V}ar[y_{it}^{\star} \mid \Omega_{it}]$.

The characterization in Proposition 5 follows directly from 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

$$\max_{\{m_1, m_2\} \in [0,1]^2} -\frac{1}{2} \mathbb{V}\text{ar}\left[y_{it}^{\star} \mid \Omega_{it}\right] - K\left(m\right). \tag{40}$$

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

$$y_{it} = \mathbb{E}\left[y_{it}^{\star} \mid z_{i1}^{t}, z_{i2}^{t}\right] = \mathbb{E}_{it}\left[x_{i1t} + x_{2t}\right], \tag{41}$$

where x_{2t} depends upon $y_t = \int_0^1 y_{it} di$. Notice that the problem in (40) and (41) 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.[32]

[31] This follows from

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

[32] 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 the full-attention version of the model, the equilibrium dynamics of output can no longer be derived analytically when firms pay limited attention. Instead, we solve the model numerically, looking for linear equilibria in which the law of motion for the components and the latent factor take the form of an infinite dimensional vector,

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + Bu_t, \quad u_t = \begin{bmatrix} u_t^\theta & u_t^x & u_t^n \end{bmatrix}', \tag{42}$$

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 (28) and (33).

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

$$\begin{aligned} y_t &= \bar{\mathbb{E}}_t \left[x_{1t} + x_{2t} \right] \\ &= \begin{bmatrix} 1 & 1 & \mathbf{0} \end{bmatrix} \bar{\mathbb{E}}_t \left[\boldsymbol{x}_t \right] = \begin{bmatrix} 1 & 1 & \mathbf{0} \end{bmatrix} \Xi \boldsymbol{x}_t, \end{aligned} \tag{43}$$

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 $m = \begin{bmatrix} m_1 & m_2 \end{bmatrix}$, which 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 (40) 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 (40). 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 (1) and (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).[33] 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 revisions. 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. Clearly, survey respondents may 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.[34] Following the literature, we view the estimates from professional forecasters as providing a useful lower-bound on deviations from full-information rationality.[35]

[33] 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.

[34] 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.

[35] See, for example, Lorenzoni (2009), Nimark (2014), and Angeletos and Huo (2020). We note that the SPF includes forecasts from large industrial firms, in addition to those from financial and government institutions, and forecasting agencies. The bi-annual Livingstone survey estimates reported in Section 2, which resemble

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 aggregate component 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 problem. 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.[36] 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.2 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

those from the SPF, includes a broader range of non-financial firms.

[36] 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).

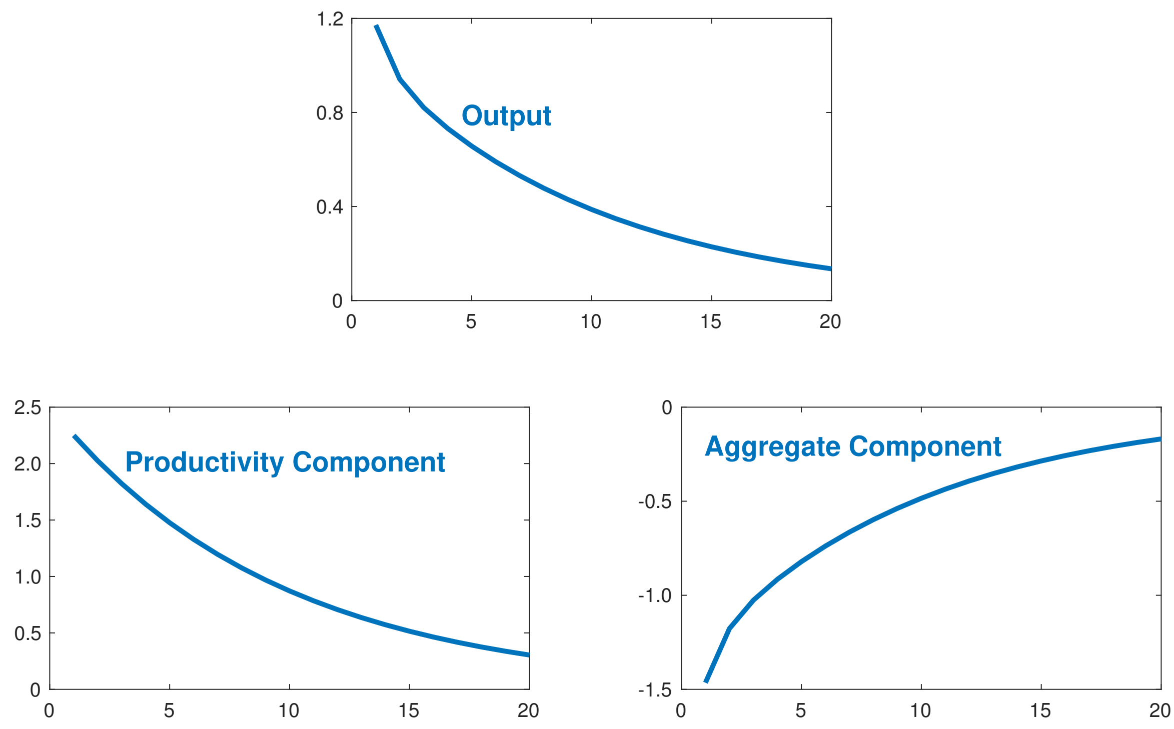
Table II: Attention Choices in Equilibrium

Component	*Variance*	*Weight*	q	m
Productivity component (x_{1t})	3.73	1.00	1.29	0.75
Economy-wide component (x_{2t})	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		−0.05	
	(–)		(0.06)	
Average Revision		0.39		0.39***
		(–)		(0.16)
Sample	(–)	(–)	70Q1:19:Q4	70Q1:19:Q4
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		−0.12**	
	(–)		(0.05)	
Average Revision		0.28		0.66**
		(–)		(0.19)
Sample	(–)	(–)	70Q1:19Q4	70Q1:19Q4
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.[37]

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 their optimal output:

$$s_{it} = y_t^{\star} + q\epsilon_{it} = x_{1t} + x_{2t} + q_{\star}\epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0, 1), \tag{44}$$

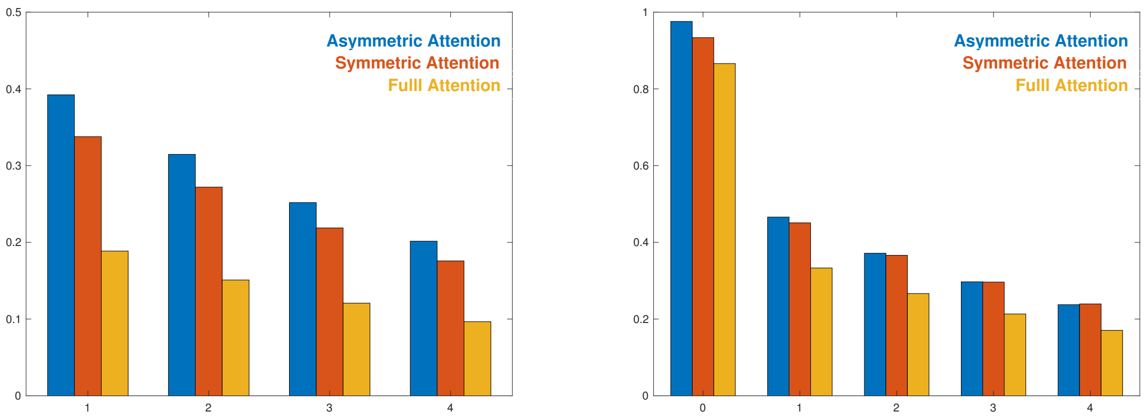
where the noise parameter q (or corresponding attention parameter m) is again calibrated to match the one-quarter-ahead estimate of δ from the SPF.

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.

[37] 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 than 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. Column (2) is calibrated. The equilibrium noise in signals about the 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 40-50 percent after 1985, while the volatility of firm-specific shocks 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 45 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, underreactions become weaker while extrapolation becomes somewhat stronger. This is because the decrease in the volatility of common 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 q_1), and 77 percent less attention to the countercylical component.

The results in this subsection have highlighted two 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 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 (2020), 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.[38] 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.

[38] 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: The proof proceeds in three steps. We first derive the MA-form for the nowcast $f_{it}y_t$. We then use this result to derive slope coefficients in (1) and (2).

Step (i): MA-form of nowcast. Solving (5) backwards for $k = 0$, we find that

$$f_{it}y_t = g_0 z_{it} + \lambda\rho\,(1-g_0)\,f_{it-1}y_{t-1} = g_0 \sum_{h=0}^{\infty} \lambda^h \rho^h (1-g_0)^h z_{it-h}, \tag{A1}$$

where we have also used that $f_{it-1}z_{it} = f_{it-1}y_t = \rho f_{it-1}y_{t-1}$.

Step (ii): Slope coefficient γ in (1). The overreaction coefficient γ equals

$$\begin{aligned}
\gamma &= \mathbb{C}\text{ov}\left[y_{t+k} - f_{it}y_{t+k}, y_t\right]\mathbb{V}\text{ar}\left[y_t\right]^{-1} \\
&= \rho^k \mathbb{C}\text{ov}\left[y_t - f_{it}y_t, y_t\right]\mathbb{V}\text{ar}\left[y_t\right]^{-1} = \rho^k \left(1 - \mathbb{C}\text{ov}\left[f_{it}y_t, y_t\right]\mathbb{V}\text{ar}\left[y_t\right]^{-1}\right),
\end{aligned}$$

because $f_{it}y_{t+k} = \rho^k f_{it}y_t$, and where (A1) shows that

$$\mathbb{C}\text{ov}\left[f_{it}y_t, y_t\right] = g_0 \sum_{h=0}^{\infty} \lambda^h \rho^h (1-g_0)^h \rho^h \mathbb{V}\text{ar}\left[y_t\right] = g_0 \frac{1}{1-\lambda\rho^2(1-g_0)}\mathbb{V}\text{ar}\left[y_t\right].$$

Hence,

$$\gamma = \rho^k \left(1 - \frac{g_0}{1-\lambda\rho^2(1-g_0)}\right) = \rho^k (1-g_0)\frac{1-\lambda\rho^2}{1-\lambda\rho^2(1-g_0)}.$$

We conclude the sign of γ depends only the sign of $\rho^k(1-g_0) = \rho^k - g_k$, since the responsiveness coefficient g_k satisfies $g_k = g_0\rho^k$ from (5) and $f_{it}y_{t+k} = \rho^k f_{it}y_t$.

Step (iii): Slope coefficient δ in (2). Averaging (5) across i for $k = 0$, using that $\bar{f}_t y_{t+k} = \rho^k \bar{f}_t y_t$, and rearranging terms as in Coibion and Gorodnichenko (2015) shows that:

$$y_{t+k} - \bar{f}_t y_{t+k} = \frac{1-g_0}{g_0}\rho^k \left(\bar{f}_t y_t - \lambda \bar{f}_{t-1} y_t\right) + u_{t,t+k},$$

where $u_{t,t+k}$ denotes a linear combination of future shocks $(u_{t+s})_{0<s\leq k}$ to output.

Thus, the underreaction coefficient δ equals

$$\begin{aligned}
\delta &= \mathbb{C}\text{ov}\left[y_{t+k} - f_{it}y_{t+k}, \bar{f}_t y_{t+k} - \bar{f}_{t-1} y_{t+k}\right]\mathbb{V}\text{ar}\left[\chi_t\right]^{-1} \\
&= \mathbb{C}\text{ov}\left[y_{t+k} - \bar{f}_t y_{t+k}, \bar{f}_t y_{t+k} - \bar{f}_{t-1} y_{t+k}\right]\mathbb{V}\text{ar}\left[\chi_t\right]^{-1} \\
&= \rho^k \frac{1-g_0}{g_0}\rho^k \mathbb{C}\text{ov}\left[\bar{f}_t y_t - \lambda \bar{f}_{t-1} y_t, \bar{f}_t y_t - \bar{f}_{t-1} y_t\right]\mathbb{V}\text{ar}\left[\chi_t\right]^{-1},
\end{aligned}$$

where $\chi_t \equiv \bar{f}_t y_{t+k} - \bar{f}_{t-1} y_{t+k}$, and the second equality follows from the linearity of the covariance operator, and because the signals in (4) have the same steady-state distribution for all i. We have used that $\bar{f}_t y_{t+k} = \rho^k \bar{f}_t y_t$ for the third equality. Finally, because g_k satisfies $g_k = g_0 \rho^k$, all that remains to show is that $\mathbb{C}\text{ov}\left[\bar{f}_t y_t - \lambda \bar{f}_{t-1} y_t, \bar{f}_t y_t - \bar{f}_{t-1} y_t\right] > 0$.

Multiplying out terms, and using the stationarity of forecasts, we find that

$$\begin{aligned}\mathbb{C}\text{ov}\left[\bar{f}_t y_t - \lambda \bar{f}_{t-1} y_t, \bar{f}_t y_t - \bar{f}_{t-1} y_t\right] &= \left(1 + \lambda \rho^2\right) \mathbb{V}\text{ar}\left[\bar{f}_t y_t\right] - \rho \left(1 + \lambda\right) \mathbb{C}\text{ov}\left[\bar{f}_t y_t, \bar{f}_{t-1} y_{t-1}\right] \\ &\geq (1-\rho)(1-\lambda\rho) \mathbb{V}\text{ar}\left[\bar{f}_t y_t\right] > 0\end{aligned}$$

since $\bar{f}_{t-1} y_t = \rho \bar{f}_{t-1} y_{t-1}$ and $\mathbb{V}\text{ar}\left[\bar{f}_t y_t\right] \geq \mathbb{C}\text{ov}\left[\bar{f}_t y_t, \bar{f}_{t-1} y_{t-1}\right] > 0$.[39] We conclude the sign of δ depends only the sign of the sufficient statistic $\rho^k - g_k$. This completes the proof. □

Corollary 1. *Consider the diagnostic expectations model:* $f_{it} y_{t+k} = \mathbb{E}_{it-1} y_{t+k} + g_k \left(z_{it} - \mathbb{E}_{it-1} y_t\right).$ *Then, the coefficients* γ *in (1) and* δ *in (2) both have the same sign as* $\rho^k - g_k$.

Proof of Corollary 1: The proof follows from Proposition 1. To see this implication, first notice that the diagnostic nowcast error at time t equals

$$\begin{aligned} y_t - f_{it} y_t &= (1 - g_0)\left(y_t - \mathbb{E}_{it-1} y_t\right) - g_0 \epsilon_{it} \\ &= (1 - g_0) \left(y_t - \frac{1}{1 - g_0^\star} \mathbb{E}_{it} y_t + \frac{g_0^\star}{1 - g_0^\star} z_{it} \right) - g_0 \epsilon_{it} \\ &= (1 - g_0)\left(1 - g_0^\star\right)^{-1} \left(y_t - \mathbb{E}_{it} y_t\right) + \left[g_0^\star (1 - g_0)\left(1 - g_0^\star\right)^{-1} - g_0\right] \epsilon_{it}, \end{aligned}$$

where the second equality exploits (5) in the rational case, and we let $g_0^\star \in (0, 1)$ denote the noisy rational expectation gain on z_{it}. It now follows from $y_{t+k} - f_{it} y_{t+k} = \rho^k \left(y_t - f_t y_t\right) + u_{t,t+k}$, where $u_{t,t+k}$ denotes a linear combination of future shocks $\left(u_{t+s}\right)_{0<s\leq k}$ to output, that

$$y_{t+k} - f_{it} y_{t+k} = (1 - g_0)\left(1 - g_0^\star\right)^{-1} \left(y_{t+k} - \mathbb{E}_{it} y_{t+k}\right) + \text{t.u.w},$$

where t.u.w denotes *terms uncorrelated with* y_t or $\bar{f}_t y_{t+k} - \bar{f}_{t-1} y_{t+k}$, and we have used (3) and $\mathbb{E}_{it} y_{t+k} = \rho^k \mathbb{E}_{it} y_t$. We conclude $\gamma = (1 - g_0)\left(1 - g_0^\star\right)^{-1} \gamma_{NRE}$ and $\delta = (1 - g_0)\left(1 - g_0^\star\right)^{-1} \delta_{NRE}$, where γ_{NRE} and δ_{NRE} denote the over- and underreaction coefficients, respectively, in the noisy

[39] It follows from (5) that

$$\bar{f}_t y_t = \rho \left[1 + \lambda(1 - g_0)\right] \bar{f}_{t-1} y_{t-1} - \lambda \rho^2 \left(1 - g_0\right) \bar{f}_{t-2} y_{t-2} + g_0 u_t.$$

Thus,

$$\mathbb{C}\text{ov}\left(\bar{f}_t y_t, \bar{f}_{t-1} y_{t-1}\right) = \rho \frac{1 + \lambda(1 - g_0)}{1 + \lambda\rho(\rho - g_0 \rho)} \mathbb{V}\text{ar}\left[\bar{f}_t y_t\right] > 0.$$

rational expectation case. Proposition 1 implies $\gamma_{NRE} > 0$ and $\delta_{NRE} > 0$. Thus, the sign of γ and δ depend only $1 - g_0 = (\rho^k - g_k)\rho^{-k}$, which depends only on $\rho^k - g_k$. $\square$

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 signal z_{ijt} is informationally equivalent to the signal

$$\hat{z}_{ijt} \equiv \frac{z_{ijt}}{a_j} = \theta_t + \frac{1}{a_j}\left(b_j u_{jt} + q_j \epsilon_{ijt}\right) \equiv \theta_t + \xi_{ijt},$$

which has precision $\tau_j \equiv \mathbb{V}\text{ar}\left[\hat{z}_{ijt} \mid \theta_t\right]^{-1}$ equal to

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

The standard formula for Gaussian updating now implies that

$$\mathbb{E}_{it}\left[\theta_t\right] = \mathbb{E}_{it-1}\left[\theta_t\right] + \sum_j \left(\frac{\tau_j}{\bar{\tau} + \sum_k \tau_k}\right)\left(\hat{z}_{ijt} - \mathbb{E}_{it-1}\left[\hat{z}_{ijt}\right]\right), \tag{A2}$$

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 terms, and inserting the definition of $\hat{z}_{ijt}$ into (A2), 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 (12) shows that the sign of γ is determined by

$$\begin{aligned}
\gamma \propto \sum_j \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, x_{jt}\right] &= \sum_j \left(a_j \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, \theta_t\right] + b_j \mathbb{C}\text{ov}\left[\theta_t - \mathbb{E}_{it}\theta_t, u_{jt}\right]\right) \\
&= \sum_j \left(a_j \mathbb{V}\text{ar}\left[\theta_t \mid \Omega_{it}\right] - b_j \mathbb{C}\text{ov}\left[\mathbb{E}_{it}\theta_t, u_{jt}\right]\right),
\end{aligned} \tag{A3}$$

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 (A3), 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 (11) backwards to obtain

$$\mathbb{E}_{it}\left[\theta_t\right] = \sum_{h=0}^{\infty} \lambda^h \hat{z}_{it-h}, \tag{A4}$$

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 $\hat{\epsilon}_{it} \equiv \sum_j g_j q_j \epsilon_{ijt}$.

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 (13) yields:

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

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

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

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

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

A.4 Optimal Attention Choice

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

$$L_t^\star \equiv \mathbb{E}\left\{\min_{a_{it}} \mathbb{E}\left[(a_{it}-a_t^\star)^2\middle|\,\Omega_{it}\right]\right\}. \tag{A5}$$

The minimizer to this problem is

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

Substituting this expression into (A5) shows that

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

Now, using the law of total variance, we can decompose $L_t^\star$ into

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

To complete the proof, we need to derive expressions for the two components of (A6).

To do so, we first note that

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

Agent i's information set Ω_{it} contains the unbiased signal z_{ijt} of x_{jt}, defined in (9), which has precision q_j^{-2}. All other elements of Ω_{it} are independent of x_{jt} conditional on θ_t.

We can therefore use Bayes' law for Gaussian variables to show that

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

and

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

We are now ready to compute the two components of (A6).

Computing the first term in (A6):

$$\begin{aligned}\mathbb{V}\text{ar}\left[a_t^\star \mid \Omega_{it},\theta_t\right] &= \mathbb{V}\text{ar}\left[w_\theta\theta_t+\sum_j w_{xj}x_{jt}\middle|\,\Omega_{it},\theta_t\right] = \mathbb{V}\text{ar}\left[\sum_j w_{xj}x_{jt}\middle|\,\Omega_{it},\theta_t\right]\\ &= \sum_j w_{xj}^2\mathbb{V}\text{ar}\left[x_{jt}\middle|\,\Omega_{it},\theta_t\right]+\sum_j\sum_{k\neq j}\underbrace{\mathbb{C}\text{ov}\left[x_{jt},x_{kt}\middle|\,\Omega_{it},\theta_t\right]}_{=0}\\ &= \sum_j w_{xj}^2 b_j^2(1-m_j). \end{aligned}\tag{A7}$$

Computing the second term in (A6):

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

so that

$$\begin{aligned}\mathbb{V}\text{ar}\left[\mathbb{E}\left[a_t^\star \mid \Omega_{it},\theta_t\right]\mid\Omega_{it}\right] &= \mathbb{V}\text{ar}\left[w_\theta\theta_t+\sum_j w_{xj}\left((1-m_j)a_j\theta_t+m_j z_{ijt}\right)\middle|\,\Omega_{it}\right]\\ &= \mathbb{V}\text{ar}\left[\left(w_\theta+\sum_j w_{xj}(1-m_j)a_j\right)\theta_t\middle|\,\Omega_{it}\right]\\ &= \left[w_\theta+\sum_j w_{xj}(1-m_j)a_j\right]^2\mathbb{V}\text{ar}\left[\theta_t\mid\Omega_{it}\right].\end{aligned}\tag{A8}$$

Substituting (A7) and (A8) into (A6) then yields the desired expression. □

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

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

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

A.5 A Macroeconomic Example

Proof of Proposition 4: We start with a firm's output choice,[40]

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

Thus,

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

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

$$v(y_i, y, a_i, \omega) \approx v_1 y_i + \frac{v_{11}}{2} y_i^2 + v_{12} y_i y + v_{13} y_i a_i + v_{14} y_i \omega + t.i.a, \tag{A9}$$

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

As a result of (A9), a firm's optimal, full-information choice of output is

$$y_i^\star = \frac{v_{12}}{\mid v_{11} \mid} y + \frac{v_{13}}{\mid v_{11} \mid} a_i + \frac{v_{14}}{\mid v_{11} \mid} \omega, \tag{A10}$$

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

$$y_i = \mathbb{E}_i [y_i^\star]. \tag{A11}$$

It remains to derive the optimal output choice under full information in (A10). A few simple but tedious derivations combine to show that

$$y_i^\star = r a_i + \alpha r \left(\sigma^{-1} y - \omega\right) \equiv x_{i1} + x_2. \tag{A12}$$

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

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

Finally, we can use (A10) and (A11) 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 $v_i^\star$:

$$\begin{aligned} v_i - v_i^\star &= \frac{v_{11}}{2} y_i^2 - \frac{v_{11}}{2} y_i^{\star 2} + (v_{12} y + v_{13} a_i + v_{14}\omega)(y_i - y_i^\star) \\ &= \frac{v_{11}}{2} y_i^2 - \frac{v_{11}}{2} y_i^{\star 2} - v_{11} y_i^\star (y_i - y_i^\star) = \frac{v_{11}}{2}(y_i - y_i^\star)^2, \end{aligned} \tag{A13}$$

where we have used the first-order condition for optimal output in (A9). □

Proof of Proposition 5: Follows immediately from (A11) and (A13). □

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 latent factors, 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 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

$$y_t = D\theta_t + Ex_t + Fu_t \tag{A14}$$

$$x_t = A\theta_t + Bx_{t-1} + Cu_t, \tag{A15}$$

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),

$$\theta_t = M\theta_{t-1} + Nu_t, \tag{A16}$$

where M and N are conformable matrices.

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

$$z_{it} = x_t + Q\epsilon_{it}, \quad Q = diag(q), \tag{A17}$$

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 (A14) to (A16), as

$$y_t = \alpha\bar{\theta}_t + \beta u_t, \tag{A18}$$

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

$$\bar{\theta}_t = \bar{M}\bar{\theta}_{t-1} + \bar{N}u_t, \tag{A19}$$

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 (A17) as

$$z_{it} = L_0\bar{\theta}_t + L_1\bar{\theta}_{t-1} + Ru_t + Q\epsilon_{it}, \tag{A20}$$

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 (A14)-(A17), then the population coefficients in the regression equations (1) and (2) satisfy:*

$$\gamma < 0 \iff \alpha\bar{M}^k\left(GQQ'E' + \Sigma_{\theta\bar{\theta}}D' + \Omega\right) < 0 \tag{A21}$$

$$\delta > 0 \iff \exists q_j \in (0\,,\infty)\,, \tag{A22}$$

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 underreactive 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 the 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 latent factors can have on output itself ($D \neq \underline{0}$),[41] (ii) for any cross-correlation in errors between the signal

[41] As an unnamed referee has pointed out to us, our central insight about asymmetric attention can also be seen in a reductionist manner in the case of several, independent latent factors. Suppose θ_{1t} and θ_{2t} follow independent AR(1) processes with persistence parameters ρ_j, in which $\rho_1 > 0$ and $\rho_2 < 0$. We further assume

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). Finally, we also use our results to 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 in (A20) have the same steady-state distribution for all i. In the remainder of the proof, we therefore use individual and average 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

$$G_k = \mathbb{C}ov\left(\bar{\theta}_{t+k} - \mathbb{E}_{it-1}\bar{\theta}_{t+k},\, z_{it} - \mathbb{E}_{it-1} z_t\right) \mathbb{V}\left[z_{it} - \mathbb{E}_{it-1} z_t\right]^{-1}. \tag{A23}$$

We note that

$$\bar{\mathbb{E}}_t\left[y_{t+k}\right] = \bar{\mathbb{E}}_{t-1}\left[y_{t+k}\right] + \alpha G_k\left(x_t - \bar{\mathbb{E}}_{t-1}\left[x_t\right]\right). \tag{A24}$$

We can now use (A24) to show that

$$\begin{aligned} \bar{\mathbb{E}}_t\left[y_{t+k}\right] - \bar{\mathbb{E}}_{t-1}\left[y_{t+k}\right] &= \alpha G_k\left(x_t - \bar{\mathbb{E}}_{t-1}\left[x_t\right]\right) && \text{(A25)} \\ y_{t+k} - \bar{\mathbb{E}}_t\left[y_{t+k}\right] &= \alpha\left(\bar{\theta}_{t+k} - \bar{\mathbb{E}}_t\left[\bar{\theta}_{t+k}\right]\right) + F u_{t+k}. && \text{(A26)} \end{aligned}$$

This completes the first step.

that $D = A = I_{2\times 2}$, $E = B = C = F = 0_{2\times 2}$, and that agents pay full attention to their first signal but none to their second ($q_1 \to 0$, $q_2 \to \infty$), as in Example 1. Then, condition (A21) shows that $\gamma < 0$ because $\rho_2 \mathbb{V}\text{ar}\left[\theta_{2t}\right] < 0$. Thus, as in the body of this paper, the *overreaction* to recent output documented in the survey data can be interpreted as an *underreaction* to countercyclical components ($\rho_2 < 0$).

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 $G_k = \bar{M}^k G$. 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 (A25), (A26) 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 $Q = 0$). We conclude that

$$\begin{aligned} \delta &\propto \alpha\left(\bar{G}_k - G_k\right)\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right]G_k'\alpha' \\ &= \left(\bar{d}_k - d_k\right)\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right]d_k', \end{aligned} \tag{A27}$$

where $\bar{d}_k \equiv \alpha\bar{G}_k$ and $d_k \equiv \alpha G_k$. 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 that $\mid \bar{d}_k \mid > \mid d_k \mid$. Combined, it now follows from (A27) that, because $\mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1}x_t\right]$ is positive semi-definite, $\delta > 0$ (Abadir and Magnus, 2005; Chpt.8). □

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:

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

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

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

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. (2020). 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. and Stark, T. (2019). Fifty years of the survey of professional forecasters. *Economic Insights*, **4** (4), 1–11.

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.

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.

Moore, D. A. and Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, **115** (2), 502.

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.

Online Appendix to:

“Asymmetric Attention”

Alexandre N. Kohlhas and Ansgar Walther

IIES, Stockholm University and Imperial College London

September, 2020

Contents of Online Appendix:

This *Online Appendix* details additional extensions of the models analyzed in the paper.

C Additional Empirical Results

C.1 Robustness of Evidence

Table C.1: Regression of forecast errors on individual forecast revisions

	All Observations			*Excluding Outliers*		
	(1)	(2)	(3)	(4)	(5)	(6)
Current Realization	-0.13^{**}			-0.12^{**}		
	(0.06)			(0.05)		
Average Revision		0.72^{***}			0.66^{***}	
		(0.24)			(0.19)	
Individual Revision			-0.19^{***}			-0.03
			(0.06)			(0.08)
Observations	7,417	7,377	5,532	7,190	7,151	5,357
R^2	0.02	0.05	0.02	0.03	0.06	0.00

Note: Estimates of regressions (1), (2), and (14) with individual (respondent) fixed effects. Columns (4) to (6) remove the top and bottom one percent of forecast errors and revisions. Double-clustered robust standard errors in parentheses. Sample: 1970Q1-2019Q4. Significance levels *=10%, **=5%, ***=1%.

Table C.2: Estimates using one quarter ahead forecasts

	Panel a: individual forecast error					
	Four-quarter ahead			One-quarter ahead		
	(1)	(2)	(3)	(1)	(2)	(3)
Current Realization	-0.12**	–	-0.14***	-0.05	–	-0.05
	(0.05)		(0.04)	(0.06)		(0.06)
Average Revision	–	0.66***	0.69***	–	0.38**	0.40***
		(0.19)	(0.18)		(0.15)	(0.15)
Observations	7,190	7,151	7,094	7,546	7,546	7,546
F	185.0	423.5	360.9	33.46	308.5	164.1
R^2	0.03	0.06	0.10	0.01	0.04	0.04
	Panel b: average forecast error					
	Four-quarter ahead			One-quarter ahead		
	(1)	(2)	(3)	(1)	(2)	(3)
Constant	0.01	-0.12	0.24	0.27	0.24*	0.44**
	(0.19)	(0.10)	(0.15)	(0.21)	(0.14)	(0.21)
Current Realization	-0.10**	–	-0.14***	-0.06	-	-0.08
	(0.05)		(0.05)	(0.06)		(0.06)
Average Revision	–	0.78***	0.84***		0.42***	0.46***
		(0.26)	(0.24)		(0.14)	(0.14)
Observations	198	197	196	188	188	188
F	3.53	16.0	11.7	2.13	16.1	9.27
R^2	0.02	0.08	0.11	0.01	0.08	0.09

Note: Estimates of regressions (1) and (2) using one-quarter ahead forecasts of output growth. Panel a: Estimates with individual (respondent) fixed effects. Panel b: Estimates with average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable. Robust standard errors (double clustered in Panel a) in parentheses. The top and bottom one percent of forecast errors and revisions have been removed in Panel a. Sample: 1970Q1-2019Q4. Significance levels *=10%, **=5%, ***=1%.

Table C.3: Estimates after removing trends in output growth

	Panel a: individual forecast error		
	Benchmark	*Level detrend*	*Linear detrend*
Current Realization	-0.12**	-0.14***	-0.12***
	(0.05)	(0.05)	(0.05)
Observations	7,190	7,190	7,190
F	185.0	251.8	185.0
R^2	0.03	0.04	0.03
	Panel b: average forecast error		
	Benchmark	*Level detrend*	*Linear detrend*
Constant	0.01	0.08	0.01
	(0.19)	(0.18)	(0.19)
Current Realization	-0.10**	-0.14***	-0.10**
	(0.05)	(0.05)	(0.05)
Observations	198	198	198
F	3.53	6.74	3.53
R^2	0.02	0.03	0.02

Note: Estimates of regressions (1) using different methods for detrending output growth. Column (1): No detrending, as in the baseline specification. Column (2): adjusting for the structural (level) increase in output growth between 1995 and 2000 (e.g. Jacobson and Occhino, 2012). Column (3): Linear detrending. Panel a: Estimates with individual (respondent) fixed effects. Panel b: Estimates with average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable. Robust standard errors (double clustered in Panel a) in parentheses. The top and bottom one percent of forecast errors and revisions have been removed in Panel a pre-estimation. Sample: 1970Q1-2019Q4. Significance levels *=10%, **=5%, ***=1%.

Table C.4: Estimates before and after Great Moderation

	Panel a: individual forecast error			
	Pre-Great Moderation		Post-Great Moderation	
	(1)	(2)	(1)	(2)
Current Realization	-0.13**	–	-0.19**	–
	(0.06)	(–)	(0.08)	(–)
Average Revision	–	0.76***	–	0.53*
	(–)	(0.24)	(–)	(0.32)
Observations	2,284	2,245	4,574	4,574
F	92.0	185.6	159.4	159.6
R^2	0.04	0.08	0.04	0.04

	Panel b: average forecast error			
	Pre-Great Moderation		Post-Great Moderation	
	(1)	(2)	(1)	(2)
Current Realization	-0.15**	–	-0.14*	–
	(0.07)		(0.08)	
Average Revision	–	0.93**	–	0.56*
		(0.37)		(0.34)
Observations	60	59	122	122
F	2.79	6.56	2.44	5.33
R^2	0.05	0.10	0.02	0.04

Note: Estimates of regressions (1) before and after the Great Moderation. Panel a: Estimates with individual (respondent) fixed effects. Panel b: Estimates with average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable. Robust standard errors (double clustered in Panel a) in parentheses. Sample: 1970Q1-2019Q4 (split into 1970Q1-1985Q1 and 1990Q1-2019Q4; Stock and Watson, 2002; Table I). We adjust for the structural increase in output between 1995 and 2000 (Jacobson and Occhino, 2012). The top and bottom one percent of forecast errors and revisions have been removed in Panel a pre-estimation. Significance levels *=10%, **=5%, ***=1%.

Table C.5: Estimates of unconstrained version of regression (2)

	(1) *Individual errors*	(2) *Average errors*
Constant	–	0.18
		(0.40)
Avr. Forecast from Time t (δ_0)	0.70***	0.82***
	(0.20)	(0.26)
Avr. Forecast from Time $t-1$ (δ_1)	−0.64***	−0.92***
	(0.28)	(0.31)
Observations	7,151	197
F Statistic	248.3	8.420
R^2	0.07	0.08

Model	Df.	χ^2	$Pr(>\chi^2)$
(1) Individual Forecast Errors	1	0.15	0.70
(2) Average Forecast Errors	1	0.59	0.44

Note: *Upper table*: Estimates of the unconstrained regression $y_{t+k} - f_{it}y_{t+k} = \alpha_i + \delta_0 \bar{f}_t y_{t+k} + \delta_1 \bar{f}_{t-1} y_{t+k} + \epsilon_{it}$. Column (1): Estimates with individual (respondent) fixed effects. Column (2): Estimates with average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable. Robust standard errors (double clustered in column (1)) in parentheses. The top and bottom one percent of forecast errors and revisions have been removed in column (1) pre-estimation. Sample: 1970Q1-2019Q4. Significance levels *=10%, **=5%, ***=1%. *Lower table*: Hypothesis tests of $\delta_0 + \delta_1 = 0$, which is imposed by regression (2) in the paper.

Table C.6: Estimates of concurrent version of regression (1)

	Baseline	Level		Recent	Detrend
	(1)	(2)	(3)	(4)	(5)
Current Realization	-0.12**	-0.09*	-0.13***	-0.25***	-0.11**
	(0.05)	(0.05)	(0.04)	(0.09)	(0.05)
Average Revision	–	–	0.72***	–	–
			(0.17)		
Observations	7,190	7.247	7,151	3,276	7,247
R^2	0.03	0.02	0.09	0.07	0.02
F	185.0	97.2	325.1	219.6	144.1

Note: Estimates of regression (1) with individual (respondent) fixed effects. Column (1): baseline specification. Columns (2-5) use only the BEA's first release of output growth as the right-hand side variable in regression (1). Column (4) considers the post-2000 sample. Column (5) adjusts for the structural (level) increase in output growth between 1995 and 2000 (e.g. Jacobson and Occhino, 2012). The top and bottom one percent of forecast errors and revisions have been removed pre-estimation. Double-clustered robust standard errors in parentheses. Sample: 1970Q1-2019Q4. Significance levels *=10%, **=5%, ***=1%.

Table C.7: Estimates in different surveys

Panel a: individual forecast error

	US SPF						EA SPF				LS Survey				MSC†	
	Output		Inflation		Deflator		Output		Inflation		Output		Inflation		Inflation	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Current Realization	-0.12**	–	-0.18***	–	-0.09	–	-0.13***	–	-0.01	–	-0.14**	–	-0.18*	–	-0.10	–
	(0.05)		(0.07)		(0.06)		(0.05)		(0.12)		(0.07)		(0.11)		(0.10)	
Average Revision	–	0.66***	–	0.19	–	1.04***	–	0.73***	–	0.42	–	0.41***	–	1.31***	–	0.66***
		(0.19)		(0.23)		(0.22)		(0.23)		(0.36)		(0.12)		(0.31)		(0.33)
Observations	7,190	7,151	3,995	4,067	5,441	5,468	3,972	4,017	4,260	4,090	4,092	4,266	1,912	1,912	147	151
Sample	1970Q1:2020Q1		1982Q2:2020Q1		1970Q1:2020Q1		1999Q4:2020Q1		1999Q4:2020Q1		1970Q1:2020Q1		1992Q1:2020Q1		1982Q1:2020Q1	

Panel b: average forecast error

	US SPF						EA SPF				LS Survey				MSC	
	Output		Inflation		Deflator		Output		Inflation		Output		Inflation		Inflation	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Current Realization	-0.10**	–	-0.19***	–	0.08*	–	-0.12**	–	-0.04	–	-0.13**	–	-0.16*	–	-0.10	–
	(0.05)		(0.07)		(0.05)		(0.05)		(0.13)		(0.06)		(0.10)		(0.10)	
Average Revision	–	0.78***	–	0.27	–	1.13***	–	0.62***	–	0.74*	–	0.43***	–	1.50***	–	0.66**
		(0.26)		(0.23)		(0.29)		(0.23)		(0.40)		(0.11)		(0.43)		(0.33)
Observations	198	197	147	151	198	197	83	79	83	79	94	97	57	57	147	151
Sample	1970Q1:2020Q1		1982Q2:2020Q1		1970Q1:2020Q1		1999Q4:2020Q1		1999Q4:2020Q1		1970Q1:2020Q1		1992Q1:2020Q1		1982Q1:2020Q1	

Note: Estimates of (1) and (2) across different surveys: US SPF, Euro Area SPF, Livingstone Survey, and Michigan Survey of Consumers. *Inflation* is the percentage change in the CPI index; *Deflator* is the percentage change in the GDP deflator. Panel a: Estimates with individual (respondent) fixed effects. Panel b: Estimates with average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable. For the Michigan Survey: Regression (2) is estimated using instrumental variables (see footnote 7 in the paper), and because respondent fixed effects are not feasible without a repeated panel, Panel a contains copies of the estimates in Panel b. Robust standard errors (double clustered in Panel a) in parentheses. The top and bottom one percent of forecast errors and revisions have been removed in each survey in Panel a pre-estimation. All regressions in Panel b include a constant term. Significance levels *=10%, **=5%, ***=1%.

Table C.8: Multivariate estimates in different surveys

Panel a: individual forecast error

	US SPF			EA SPF		LS Survey	
	Output	Inflation	Deflator	Output	Inflation	Output	Inflation
Current Realization	−0.14***	−0.18***	-0.12**	−0.27***	−0.17	−0.15*	−0.14
	(0.04)	(0.07)	(0.05)	(0.05)	(0.17)	(0.08)	(0.10)
Average Revision	0.69***	0.27	1.11***	0.97***	0.85*	0.43***	1.29***
	(0.17)	(0.22)	(0.23)	(0.23)	(0.45)	(0.11)	(0.32)
Observations	7,094	3,995	5,468	4,090	4,124	1,677	
Sample	70Q1:20Q1	82Q2:20Q1	70Q1:20Q1	99Q4:20Q1	99Q4:20Q1	70Q1:20Q1	92Q1:20Q1

Panel b: average forecast error

	US SPF			EA SPF		LS Survey	
	Output	Inflation	Deflator	Output	Inflation	Output	Inflation
Current Realization	−0.14***	−0.18***	0.04	−0.29***	−0.20	−0.12**	−0.11*
	(0.05)	(0.07)	(0.04)	(0.07)	(0.18)	(0.06)	(0.07)
Average Revision	0.84***	0.16	1.08***	0.88***	1.05**	0.42***	1.46***
	(0.24)	(0.22)	(0.28)	(0.23)	(0.48)	(0.10)	(0.42)
Observations	196	147	183	79	79	94	57
Sample	70Q1:20Q1	82Q2:20Q1	70Q1:20Q1	99Q4:20Q1	99Q4:20Q1	70Q1:20Q1	92Q1:20Q1

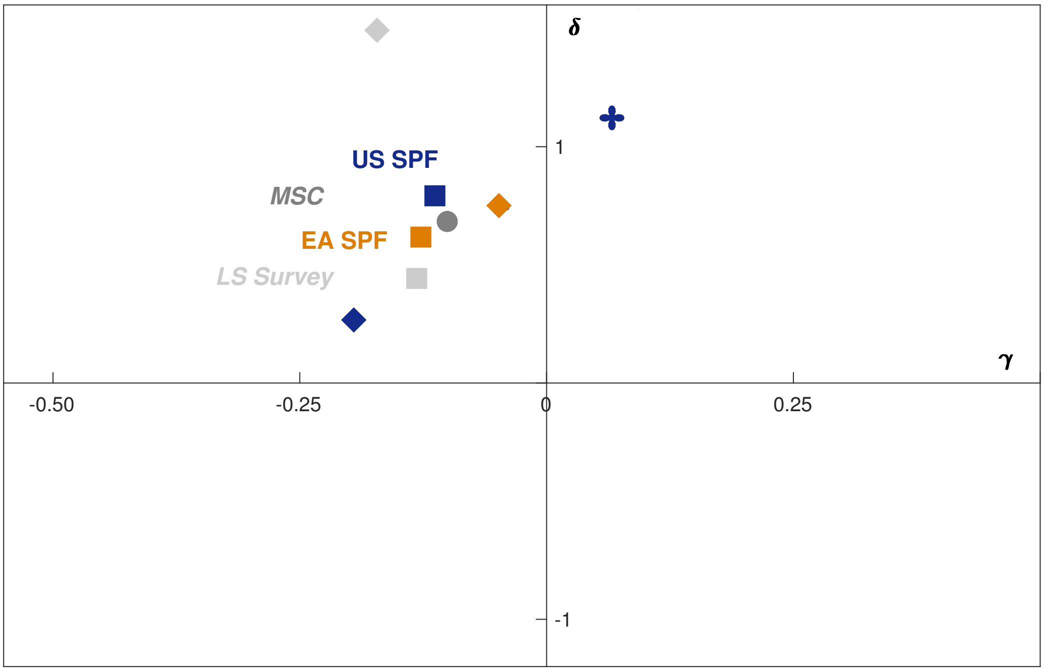
Note: Multivariate estimates of regressions (1) and (2) across surveys: the US SPF, the EA SPF, and the Livingstone Survey. *Inflation* refers to the percentage change in the CPI index; *Deflator* to the percentage change in the GDP deflator. Panel a: Estimates with individual (respondent) fixed effects. Panel b: Estimates with average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable. Robust standard errors (double clustered in Panel a) in parentheses. The top and bottom one percent of forecast errors and revisions have been removed in each survey in Panel a pre-estimation. Regressions in Panel b include a constant. Significance levels *=10%, **=5%, ***=1%.

Table C.9: Estimates in different surveys using inflation data after 1992

	US SPF						EA SPF				LS Survey				MSC	
	Output		Inflation		Deflator		Output		Inflation		Output		Inflation		Inflation	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Constant	0.01	-0.11	0.37*	-0.11	-0.02	-0.30**	0.02	0.03	0.15	0.16	0.03	-0.02	-0.05	0.18	-0.69	-1.35***
	(0.19)	(0.10)	(0.22)	(0.10)	(0.12)	(0.06)	(0.17)	(0.14)	(0.22)	(0.11)	(0.22)	(0.14)	(0.24)	(0.23)	(0.39)	(0.17)
Current Realization	-0.10**	–	-0.21**	–	-0.16**	–	-0.12***	–	-0.04	–	-0.13**	–	-0.16	–	-0.29*	–
	(0.05)		(0.07)		(0.06)		(0.05)		(0.13)		(0.06)		(0.11)		(0.18)	
Average Revision	–	0.78***	–	0.17	–	0.43	–	0.62***	–	0.74*	–	0.43***	–	1.50***	–	0.74*
		(0.26)		(0.41)		(0.27)		(0.23)		(0.40)		(0.11)		(0.43)		(0.43)
Sample	1970Q1:2020Q1		1992Q1:2020Q1		1992Q1:2020Q1		1999Q4:2020Q1		1999Q4:2020Q1		1970Q1:2020Q1		1992Q1:2020Q1		1992Q1:2020Q1	
F	3.53	16.4	6.64	0.40	4.01	2.22	1.74	11.1	0.52	4.41	4.02	29.2	1.81	23.2	5.78	13.6
R^2	0.03	0.08	0.06	0.01	0.04	0.02	0.02	0.13	0.01	0.05	0.04	0.24	0.03	0.30	0.05	0.08

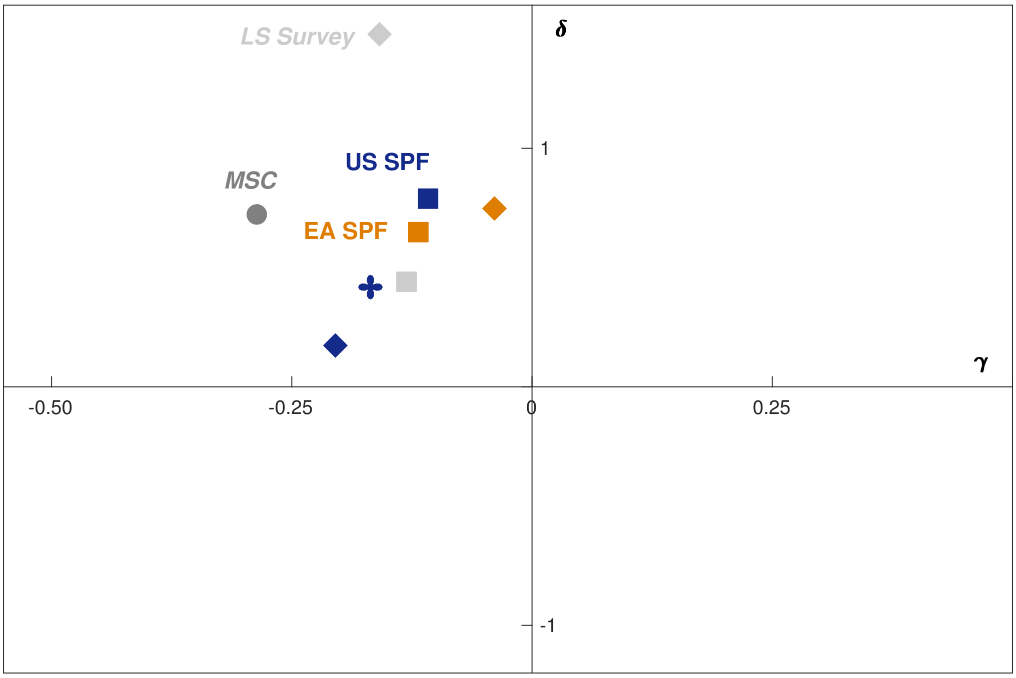
Note: Estimates of (1) and (2) across surveys using inflation data post 1992 and average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable: US SPF, Euro Area SPF, Livingstone Survey, and Michigan Survey of Consumers. *Inflation* is the percentage change in the CPI index; *Deflator* is the percentage change in the GDP deflator. For the Michigan Survey: Regression (2) is estimated using IV (see footnote 7 in the paper). Robust standard errors in parentheses. Significance levels *=10%, **=5%, ***=1%.

Figure C.1: Alternative version of Figure 3 based on Table C.6b (average errors)



Note: Estimates of γ and δ from (1) and (2) using average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the dependent variable. $\square$ = GDP forecasts, $\diamond$ = CPI inflation forecasts, $\star$ = GDP deflator inflation forecasts, and $\circ$ = MSC inflation forecasts that have been instrumented. All estimates are for one-year ahead forecasts, and estimates of (2) use semi-annual revisions (LS Survey) or one-quarter revisions (all others).

Figure C.2: Alternative version of Figure 3 based on Table C.8 (inflation data after 1992)



Note: Estimates of γ and δ from (1) and (2) using average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the dependent variable. $\square$ = GDP forecasts, $\diamond$ = CPI inflation forecasts, $\star$ = GDP deflator inflation forecasts, and $\circ$ = MSC inflation forecasts that have been instrumented. All estimates are for one-year ahead forecasts, and estimates of (2) use semi-annual revisions (LS Survey) or one-quarter revisions (all others). Inflation and deflator estimates use post-1992 forecasts to account for the potential of a structural break in the inflation series; GDP growth estimates by contrast employ the full sample. The Federal Reserve Bank of Philadelphia also took over ownership of the SPF in 1992.

D Auxiliary Test of Underreactions

Coibion and Gorodnichenko (2012) propose two regressions that can be used to provide an alternative test for the presence of underreactions to aggregate information (i.e. information frictions). Consistent with the notation in our paper, let η_t denote a structural shock and y_t output growth. Coibion and Gorodnichenko (2012) propose the following two regressions:

$$y_t = \alpha + \sum_{h=1}^{H} \beta_h y_{t-h} + \sum_{j=1}^{J} d_j \eta_{t-j} + e_t. \tag{OA1}$$

$$y_t - \bar{f}_{t-k}\left[y_t\right] = \alpha + \sum_{h=1}^{H} \beta_h \left(y_{t-h} - \bar{f}_{t-k-h}\left[y_{t-h}\right]\right) + \sum_{j=1}^{J} d_j \eta_{t-j} + e_t. \tag{OA2}$$

Under the null hypothesis of full information and rational expectations, there should be an immediate and complete adjustment of forecasts to shocks, and therefore zero systematic responses of forecast errors after any shock. By contrast, under the hypothesis of informational frictions, the conditional response of forecast errors to a shock should have the same sign as the response of the variable being forecasted to the shock.

We report the results from estimates of (OA1) and (OA2) in Figure D.1. To operationalize (OA1) and (OA2), we use identified productivity shocks, consistent with our quantitative model, as the structural shock η_t. As in Coibion and Gorodnichenko (2012), we use the identification approach from Gali (1999). Specifically, we estimate a trivariate VAR(4) on quarterly data for output, the change in labor productivity, and hours, using the same sample as Coibion and Gorodnichenko (2012). Technology shocks are identified from the restriction that only technology shocks have a long-run effect on productivity. In accordance with our baseline estimates, and as in Coibion and Gorodnichenko (2012), we consider one-year ahead forecasts ($k = 4$).

Consistent with models of information frictions, the correlation between the conditional response of forecast errors and the conditional response of output to identified productivity shocks is positive in Figure D.1. This lends credence to our estimates based on regression (2).

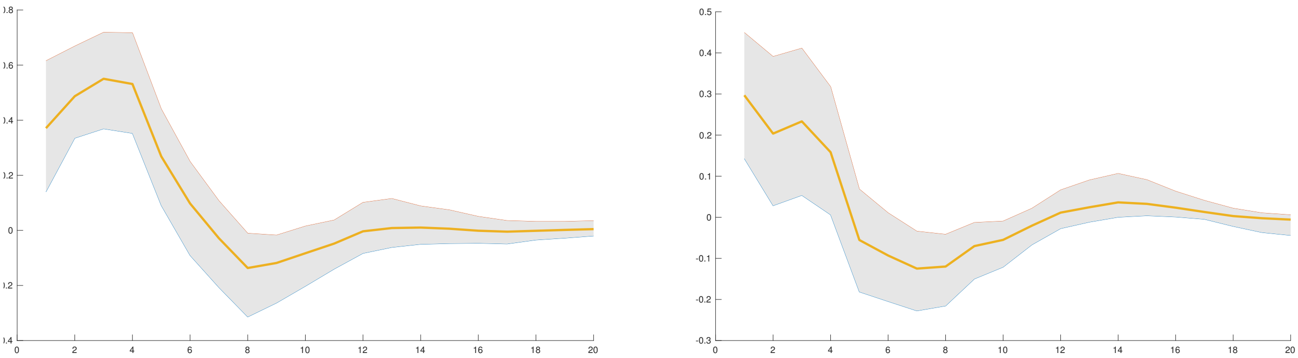
The estimates in Figure D.1 are in line with models of information frictions, and hence also our theory. We briefly document this result below for our baseline model.

Proposition D.1. *The average forecast error of future output $y_{t+k} - \bar{\mathbb{E}}_t y_{t+k}$ and output y_t itself are positively correlated in response to an innovation η_t to the latent factor θ_t.*

Proof of Proposition D.1: The proof is simple. Notice that we can write the average nowcast error of the latent factor θ_t (e.g. productivity) in our model as

$$\theta_t - \bar{\mathbb{E}}_t\theta_t = \rho\left(1 - \sum_j g_j a_j\right)\left(\theta_{t-1} - \bar{\mathbb{E}}_{t-1}\theta_t\right) + \left(1 - \sum_j g_j a_j\right)\eta_t - \sum_j g_j b_j u_{jt}, \tag{OA3}$$

Figure D.1: Coibion and Gorodnichenko (2012) test for information frictions



The left-hand panel depicts the ex-post output growth (measured as the year-over-year growth rate) response to a one unit identified productivity shock, based upon (OA1). The right-hand panel depicts the mean forecast error response to the same productivity shock, based upon (OA2), using the identification scheme from Gali (1999). The shaded area indicates one-standard deviation error bounds. Consistent with the baseline in Section 2.1, we set $k = 4$. Furthermore, as in Coibion and Gorodnichenko (2012), lag selection in (OA1) and (OA2) is done so as to ensure that there is no residual serial correlation, and standard errors are computed using a parametric bootstrap. We use the entire sample available from the SPF and the productivity shock series to estimate (OA1) and (OA2). Finally, as in Coibion and Gorodnichenko (2012), because forecasts of output growth are from time t to $t+k$, we drop the first k observations of the impulse response in (OA1) and (OA2).

where we have used that the average expectation of the latent factor equals

$$\bar{\mathbb{E}}_t\theta_t = \rho\bar{\mathbb{E}}_{t-1}\theta_{t-1} + \sum_j g_j\left(x_{jt} - \bar{\mathbb{E}}_{t-1}x_{ijt}\right),$$

with g_j characterized in Lemma 1 in the paper. The average forecast error of output is thus

$$y_{t+k} - \bar{\mathbb{E}}_t y_{t+k} = \alpha\left(\theta_t - \bar{\mathbb{E}}_t\theta_t\right) + \text{t.n.p},, \quad \alpha = \rho^k \sum_j a_j > 0 \tag{OA4}$$

where *t.n.p.* denotes *terms from next period* that are uncorrelated with information at time t.

Because the effective Kalman Gain weights $g_j a_j$ sum to less than one,[1] output y_t and average forecast error of the latent factor $\theta_t - \bar{\mathbb{E}}_t\theta_t$ in (OA3) react in the same direction in response to an innovation to η_t. However, because the average forecast error of future output $y_{t+k} - \bar{\mathbb{E}}_t y_{t+k}$ is simply proportional to that of the fundamental in (OA4), this also implies that the responses of the average forecast error of output and output itself are positively correlated. □

E Analysis of Alternative Models

E.1 Expectations of Output in Maćkowiak and Wiederholt (2009)

Maćkowiak and Wiederholt (2009) model nominal log-output (q_t in their notation) as an exoge-

[1] To see this result, first normalize the signals $\tilde{z}_{ijt} = \theta_t + b_{jt}/a_j u_{jt} + q_j/a_j\epsilon_{ijt}$, and then use the standard result that when signals are of the form "latent factor + noise", then the sum of Kalman Gain coefficients is less than one (see, for example, Anderson and Moore, 2012 or Lemma 1).

nous, stationary process. In their second case with an analytical solution, it is an AR(1) process. Firms rationally allocate attention to acquire information about an economy-wide component $\Delta_t = k_0 q_t$, for some coefficients k_0, and about idiosyncratic productivity shocks z_{it}, which also follow an independent AR(1) process. In their paper, Maćkowiak and Wiederholt conjecture and later verify (see the discussion after their Proposition 4) that it is optimal for firms to acquire two separate signals that convey "truth plus white noise" for each component:

$$s_{1it} = \Delta_t + \varepsilon_{it}, \quad s_{2it} = z_{it} + \psi_{it}. \tag{OA5}$$

Furthermore, Maćkowiak and Wiederholt (2009) show that the price level p_t is a linear function of q_t in equilibrium (see their equation (38)). Using $y_t = q_t - p_t$, it follows that output y_t is also proportional to q_t, and thus that the signal structure in (OA5) is equivalent to

$$\tilde{s}_{1it} = y_t + \tilde{\varepsilon}_{it}, \quad s_{2it} = z_{it} + \psi_{it} \tag{OA6}$$

for some shock $\tilde{\varepsilon}_{it}$ with a different variance to ε_{it}. We note that because output y_t is proportional to an AR(1) process it too follows an AR(1) in reduced form.

The only difference between the information structure in (OA6) and our equations in Section 2.2 is the second signal s_{2it}, which informs firms about idiosyncratic shocks. Notice that these shocks are uncorrelated with aggregate variables by design. If agents (firms) are asked to forecast output, these forecasts will be independent of s_{2it}. Thus, forecast errors behave as if they were determined by the noisy rational expectations case in Section 2.2:

Proposition E.1. *Expectations about output in the analytical version of Maćkowiak and Wiederholt (2009) underreact to output and average forecast revisions ($\gamma > 0$ in (1) and $\delta > 0$ in (2)).*

E.2 Expectations of Output in Lucas (1973)

Lucas (1973) considers a continuum of measure one of islands $i \in [0, 1]$. The supply of output on island i is assumed to follow the supply equation:

$$y_{it}^s = \alpha \left(p_{it} - \mathbb{E}\left[p_t \mid \Omega_{it}\right]\right) + \lambda y_{it-1}, \quad \alpha, \lambda > 0, \tag{OA7}$$

where $p_t = \int_0^1 p_{it} di$ denotes the economy-wide price level, and $\mathbb{E}\left[\cdot \mid \Omega_{it}\right]$ island inhabitants' expectations conditional on their information set Ω_{it} (described below).

The price level on island i is *exogenous* and equal to

$$p_{it} = p_t + \epsilon_{ipt}, \quad \epsilon_{ipt} \sim \mathcal{N}\left(0, \tau_p^{-1}\right),$$

while the central bank directly sets nominal demand m_t, so that

$$m_t = y_t^d + p_t = m_{t-1} + \epsilon_{mt}, \quad \epsilon_{mt} \sim \mathcal{N}\left(0, \tau_m^{-1}\right).$$

Finally, the information structure is as follows: On each island, all agents observe the (infinite) history of local prices, in addition to m_{t-1} and y_{t-1}, so that

$$\Omega_{it} = \{p_{i\tau}, p_{\tau-1}, m_{\tau-1}, y_{\tau-1}\}_{\tau=-\infty}^{\tau=t}.$$

As is well-known, the equilibrium price level for this economy follows[2]

$$p_t = \pi_1 m_t + \pi_2 m_{t-1} + \pi_3 y_{t-1},$$

where the coefficients π_k solve

$$\pi_1 = \frac{1}{1+\gamma w}, \quad \pi_2 = \frac{\gamma w}{1+\gamma w}(\pi_1 + \pi_2), \quad \pi_3 = \frac{\gamma w}{1+\gamma w}\pi_3 - \frac{\lambda}{1+\gamma w}.$$

and where w denotes the weight on island inhabitants' prior expectation about p_t at time t.

As a result, economy-wide output, our key variable of interest, equals

$$y_t = m_t - p_t = (1-\pi_1) m_t - \pi_2 m_{t-1} - \pi_3 y_{t-1} \equiv k_0 m_t + k_1 m_{t-1} + k_2 y_{t-1},$$

where the coefficients k_j satisfy $k_0 > 0$, $k_1 < 0$, $k_2 > 0$, and $k_0 + k_1 = 0$.

We conclude that output follows the AR(1) process

$$y_t = k_0 \epsilon_{mt} + k_2 y_{t-1}. \tag{OA8}$$

We now turn to agents' expectations about future output. To start, notice that the expectation of the nominal demand shock ϵ_{mt} in (OA8) is

$$\mathbb{E}_{it}[\epsilon_{mt}] = \mathbb{E}[\epsilon_{mt} \mid p_{it}] = \mathbb{E}[\epsilon_{mt} \mid s_{it}] = v\left(\epsilon_{mt} + \frac{1}{\pi_1}\epsilon_{ipt}\right),$$

where we have defined

$$s_{it} \equiv \frac{1}{\pi_1}(p_{it} - \pi_1 m_{t-1} - \pi_2 m_{t-2} - \pi_3 y_{t-1}) = \epsilon_{mt} + \frac{1}{\pi_1}\epsilon_{ipt},$$

and v denotes the associated signal extraction weight.

Thus, agent i's expectation of next period's output equals

$$\mathbb{E}_{it}[y_{t+1}] \quad = \quad k_2(k_0 \mathbb{E}_{it}[\epsilon_{mt}] + k_2 y_{t-1}) = k_2\left(k_0 v \epsilon_{mt} + k_2 y_{t-1} + k_0 v \frac{1}{\pi_1}\epsilon_{ipt}\right)$$

so that her forecast error becomes

$$y_{t+1} - \mathbb{E}_{it}[y_{t+1}] = k_2 k_0 (1-v)\epsilon_{mt} + k_0 \epsilon_{mt+1} - k_2 k_0 v \frac{1}{\pi_1}\epsilon_{ipt}. \tag{OA9}$$

[2]See, for example, Veldkamp (2011) Chapter 6.

Finally, using (OA8) and (OA9) it immediately follows that

$$\gamma \propto \mathbb{C}ov\left(y_{t+1} - \mathbb{E}_{it}\left[y_{t+1}\right], y_t\right) = k_2 k_0^2 \left(1 - v\right) \tau_m^{-1} > 0.$$

A standard argument based on the dispersion of information (e.g., Coibion and Gorodnichenko, 2015) further implies that $\delta > 0$. We conclude that:

Proposition E.2. *Expectations about future output in Lucas (1973) underreact to both current output and average forecast revisions (i.e. $\gamma > 0$ in (1) and $\delta > 0$ in (2)).*

Intuitively, s_{it} provides island inhabitants with a noisy signal of the money supply shock, and hence with a noisy signal of the innovation to output (see equation OA8). In this sense, the Lucas (1973) island model is closely related to our results from the noisy rational expectations case in Section 2. In fact, the only differences are that island inhabitants observe a private signal of the *innovation* to output today rather than the *level* of output itself, and that island inhabitants are assumed to observe the previous period's output without noise. Despite these distinctions, the intuitions from the noisy rational expectations case in Section 2 carry over, so that we find both $\gamma > 0$ and $\delta > 0$ for all admissible parameters.

E.3 Expectations of Output in Lorenzoni (2009)

Lorenzoni (2009) considers a continuum of measure one of islands $i \in [0, 1]$. The model can be log-linearized around a non-stochastic steady state, yielding the following equilibrium conditions (see e.g. Lorenzoni, 2009; Nimark, 2014; Kohlhas, 2019):

1. An Euler equation determining the intertemporal allocation of consumption:

$$c_{it} = \mathbb{E}\left[c_{it+1} \mid \Omega_{it}\right] - i_t + \mathbb{E}\left[\pi_{\mathcal{B},it+1} \mid \Omega_{it}\right], \tag{OA10}$$

   where $\pi_{\mathcal{B},it+1}$ is the inflation of the goods basket consumed on island i in period $t+1$ (defined below), and Ω_{it} denotes the information set on island i (also defined below).

2. A labor supply condition equating the marginal disutility of labor supply with the marginal utility of consumption multiplied by the real wage:

$$w_{it} - p_{\mathcal{B},it} = c_{it} + \psi n_{it}, \tag{OA11}$$

   where ψ denotes the inverse Frisch-elasticity of labor supply, and n_{it} labor supplied.

3. A demand schedule for the good produced on island i,

$$y_{it} = \int_{\mathcal{C},i,t} c_{mt} dm - \sigma\left(p_{it} - \int_{\mathcal{C},i,t} \bar{p}_{mt} dm\right), \tag{OA12}$$

   where $\int_{\mathcal{C},i,t} \bar{p}_{mt} dm$ is the logarithm of the relevant price subindex for consumers from other islands buying goods from island i.

4. An interest rate rule

$$i_t = \rho_m i_{t-1} + \phi \tilde{\pi}_t, \quad \tilde{\pi}_t = \pi_t + \epsilon_t^\pi, \quad \epsilon_t^\pi \sim \mathcal{N}\left(0, \sigma_\pi^2\right), \tag{OA13}$$

where $\tilde{\pi}$ denotes the publicly observable noisy signal of inflation.

5. Lastly, a Phillips curve relating inflation on each island i to the nominal marginal cost on island i and expected future inflation on island i,

$$p_{it} - p_{it-1} = \kappa\left(p_{\mathcal{B}.it} + c_{it} - p_{it} - a_{it}\right) + \kappa\psi\left(y_{it} - a_{it}\right) + \beta\mathbb{E}\left[p_{it+1} - p_{it} \mid \Omega_{it}\right], \tag{OA14}$$

where $\kappa = \frac{(1-f)(1-f\beta)}{\beta}$ denotes the slope of the Phillips curve and f the Calvo parameter.

Information Structure: As in Nimark (2014), we adopt the information structure from Lorenzoni (2009) but adjust the mean of the normally distributed shocks so that all signals are conditionally stationary. This does not change any of the economics of what follows, but simplifies the representation of agents' filtering problems as all variables (except for the price level) are stationary. Agents on island i observe the following signals:

1. Their own island-specific productivity

$$a_{it} = \theta_t + \epsilon_{it}^a, \quad \epsilon_{it}^a \sim \mathcal{N}\left(0, \sigma_a^2\right)$$

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

2. The demand for island goods ($\mathcal{C}$ is drawn such that the below is true)

$$y_{it} = y_t - \sigma\left(p_{it} - p_t\right) + \epsilon_{it}^y, \quad \epsilon_{it}^y \sim \mathcal{N}\left(\sigma\left(p_{it-1} - p_{t-1}\right), \sigma_y^2\right).$$

3. The price index for the goods basket consumed on island i ($\mathcal{B}$ is drawn such that)

$$p_{\mathcal{B},it} = p_t + \epsilon_{it}^p, \quad \epsilon_{it}^p \sim \mathcal{N}\left(p_{it-1} - p_{t-1}, \sigma_p^2\right).$$

4. The public signal of inflation

$$\tilde{\pi}_t = \pi_t + \epsilon_t^\pi, \quad \epsilon_t^\pi \sim \mathcal{N}\left(0, \sigma_\pi^2\right).$$

5. The public signal of the common, persistent component of productivity

$$s_t = \theta + \epsilon_t^s, \quad \epsilon_t^s \sim \mathcal{N}\left(0, \sigma_s^2\right).$$

6. The interest rate i_t.

Thus,

$$\Omega_{it} = \{a_{it}, y_{it}, p_{\mathcal{B}.it}, \tilde{\pi}_t, s_t, i_t, \Omega_{it-1}\}.$$

Model Solution: We solve the model using the *truncated state-space solution method* proposed in Nimark (2017). For the details of this method applied to the Lorenzoni (2009) model, see Nimark (2014) and Kohlhas (2019).

Simulation and Calibration: We simulate the model for one million periods, discarding the first 100,000 observations. We then estimate regression (1) and (2) from our paper, using one-year ahead forecasts of output growth.

Table E.1: Empirical Estimates Using Different Calibrations

	Lorenzoni 2009	*Nimark 2014*	*Kohlhas 2019*	*Calibrated*
Constant	0.00	0.00	0.00	0.00
Current Realization γ	0.03	0.05	0.02	0.10

The table below shows that we consistently find $\gamma > 0$ in regression (1) (including in several alternative, unreported calibrations). The first three columns consider the baseline parameterizations in (i) Lorenzoni (2009),[3] (ii) Nimark (2014), and (iii) Kohlhas (2019). While these columns show $\gamma > 0$, we note that the estimates of δ in (2) are an order of magnitude below our estimates in Table 1. This is because, across all the three calibrations, the public signals of productivity and inflation are substantially more precise than any of the private signals (see, for example, Lorenzoni, 2009 and Nimark, 2014). As a result, island inhabitants put very little weight on private information. The final column in the above table attempts to account for this feature. Specifically, we directly calibrate the noise in individual-specific productivity to target a $\delta-$coefficients of 0.70 (see Table I of our paper), and mute all public signals (that is, we let the standard deviation of the noise tend towards infinity). The rest of the parameters are set as in Kohlhas (2019). We once more find that $\gamma > 0$, which is inconsistent with our empirical results.

E.4 Expectations about Output in Angeletos *et al.* (2018)

Angeletos *et al.* (2018) study a simple deviation from rational expectations. In the version of their model that is solved analytically, output in equilibrium is

$$Y_t = A_t + \Lambda_z \bar{z}_t + \Lambda_\xi \xi_t, \quad \Lambda_z, \Lambda_\xi > 0,$$

where A_t denotes TFP, $\bar{z}_t$ the average signal of TFP, and ξ_t an exogenous process for agents' confidence. The true data generating process is that $\log A_t$ is a random walk, $\xi_t = \rho\xi_{t-1} + \zeta_t$, and the average signal is $\bar{z}_t = A_t$. Agents believe wrongly that $\bar{z}_t = A_t + \xi_t$.

[3]Because our solution method requires the model to be stationary, we set the persistence of θ_t to that in Kohlhas (2019). Indeed for $\rho = 1$ the above model is identical to that in Lorenzoni (2009). The only difference is the adjustment of the mean of the signals.

Thus, the common forecast errors of next-period output (for concreteness) is

$$\begin{aligned} Y_{t+1} - \mathbb{E}_t Y_{t+1} &= A_{t+1} - \mathbb{E}_t A_{t+1} + \Lambda_z \left(A_{t+1} - \mathbb{E}_t A_{t+1} - \mathbb{E}_t \xi_{t+1} \right) + \Lambda_\xi \left(\xi_{t+1} - \mathbb{E}_t \xi_{t+1} \right) \\ &= -\Lambda_z \rho \xi_t + \text{shocks at date } t+1. \end{aligned}$$

As a result, the equivalent of the coefficient in regression (1) in our paper is

$$\gamma \propto \mathbb{C}\text{ov}(Y_{t+1} - \mathbb{E}_t Y_{t+1}, Y_t) = -\rho \Lambda_z \mathbb{C}\text{ov}\left(\xi_t, Y_t\right) < 0.$$

The corresponding forecast revision is

$$\begin{aligned} \mathbb{E}_t Y_{t+1} - \mathbb{E}_{t-1} Y_{t+1} &= \left(1 + \Lambda_z\right)\left(A_t - A_{t-1}\right) + \Lambda_z \left(\xi_t - \mathbb{E}_t \xi_{t-1}\right) \\ &= \left(1 + \Lambda_z\right)\left(A_t - A_{t-1}\right) + \Lambda_z \zeta_t. \end{aligned}$$

Hence, the equivalent of the coefficient in regression (2) in our paper is

$$\delta \propto \mathbb{C}\text{ov}\left(Y_{t+1} - \mathbb{E}_t Y_{t+1}, \mathbb{E}_t Y_{t+1} - \mathbb{E}_{t-1} Y_{t+1}\right) = -\rho \Lambda_z^2 \mathbb{C}\text{ov}\left(\xi_t, \zeta_t\right) < 0$$

Angeletos *et al.* (2018) do not view ξ_t literally as a deviation from rationality, but rather as a reduced form of higher-order uncertainty akin to that in models of dispersed information. However, its implication for forecasts is that it generates overreactions across the board.

Proposition E.3. *Expectations about output in the analytical version of Angeletos et al. (2018) overreact to output and average forecast revisions ($\gamma < 0$ in (1) and $\delta < 0$ in (2)).*

F Extension of the Baseline Model with Overconfidence

We consider our baseline model in Section 3, but assume that instead of the Bayesian Kalman filter in Lemma 1, agents form their forecasts of the latent factor θ_t according to

$$f_{it}\theta_t = \mathbb{E}_{it-1}\left[\theta_t\right] + (1+\omega)\sum_j g_j \left(z_{ijt} - \mathbb{E}_{it-1}\left[z_{ijt}\right]\right). \tag{OA15}$$

We assume that the bias parameter $\omega > 0$, so that agents overreact to each signal z_{ijt} relative to the associated Bayesian update. This specification is similar to the model in Bordalo *et al.* (2018) and, more broadly, to the literature on overconfidence (e.g., Broer and Kohlhas, 2019). As long as the bias ω is not too large, the model replicates all of our findings, as well as the overreactions to individual information documented in Bordalo *et al.* (2018) and others:

Proposition F.1. *Suppose that attention to the components x_{jt} of output is asymmetric, with $\sum_j a_j(1-m_j) < 0$. There exists a $\bar{\omega}$ so that for all overconfidence parameters $\omega \in (0, \bar{\omega})$, the coefficients of regressions (1), (2), and (14) in the paper satisfy $\delta > 0$, $\delta^{ind} < 0$, and $\gamma < 0$.*

This proposition extends the argument in Bordalo *et al.* (2018) to the case with asymmetric attention, showing that agents with bias parameter $\omega > 0$ overreact to individual information, consistent with $\delta^{ind} < 0$ in regression (14). We show in the paper that asymmetric attention explains $\delta > 0$ and $\gamma < 0$ simultaneously in a rational model with $\omega = 0$. By continuity, we can explain all three sets of facts as long as the bias parameter ω is not too large.

Finally, we reiterate that, even in this extended model, asymmetric attention to different components of output is necessary to generate this result: Our analysis in Section 2 shows that if agents receive a signal directly of current output y_t, then, for for all values of $\omega > 0$, the coefficients δ and γ in regressions (1) and (2) have the same sign. This underlines the main insight of our paper: A model with asymmetric attention can be consistent with several properties of survey expectations, in particular the coexistence of extrapolation and underreactions.

Proof of Proposition F.1: The coefficient in regression (14) is

$$\delta^{ind} = \frac{\mathbb{C}\text{ov}\left[y_{t+k} - f_{it}y_{t+k}, f_{it}y_{t+k} - f_{it-1}y_{t+k}\right]}{\mathbb{V}\text{ar}\left[f_{it}y_{t+k} - f_{it-1}y_{t+k}\right]} = d_1\mathbb{C}\text{ov}\left[\theta_t - f_{it}\theta_t, f_{it}\theta_t - f_{it-1}\theta_t\right]$$

where $d_1 \equiv \left(\rho^k \sum_j a_j\right)^2 \mathbb{V}\text{ar}\left[f_{it}y_{t+k} - f_{it-1}y_{t+k}\right]^{-1} > 0$.

Using a parallel argument to Bordalo *et al.* (2018, Proposition 2), shows that

$$\theta_t - f_{it}\theta_t = \theta_t - \mathbb{E}_{it}\theta_t - \omega\left(\mathbb{E}_{it}\theta_t - \mathbb{E}_{it-1}\theta_t\right)$$

and

$$f_{it}\theta_t - f_{it-1}\theta_t = (1+\omega)\left(\mathbb{E}_{it}\theta_t - \mathbb{E}_{it-1}\theta_t\right) - \rho\omega\left(\mathbb{E}_{it-1}\left[\theta_{t-1}\right] - \mathbb{E}_{it-2}\left[\theta_{t-2}\right]\right).$$

Thus,

$$\begin{aligned}\delta^{ind} &\propto -\omega\mathbb{C}\text{ov}\left[\mathbb{E}_{it}\theta_t - \mathbb{E}_{it-1}\theta_t, f_{it}\theta_t - f_{it-1}\theta_t\right] \\ &= -\omega(1+\omega)\mathbb{V}\text{ar}\left[\mathbb{E}_{it}\theta_t - \mathbb{E}_{it-1}\theta_t\right].\end{aligned}$$

We conclude $\delta^{ind} < 0$ for all $\omega > 0$. Proposition 2 in the paper shows that $\gamma \propto \sum_j a_j(1-m_j)$ and $\delta > 0$ for $\omega = 0$, so the claim follows because γ and δ are continuous functions of ω. □

G Optimal Attention Choice with Entropy Costs

Suppose that the costs of attention are equal to the reduction in relative entropy:[4]

$$\mathcal{I} = \mu \lim_{T\to\infty} \frac{1}{T}\left\{H\left(\theta^T, x^T\right) - H\left(\theta^T, x^T \mid z_i^T\right)\right\}. \tag{OA16}$$

where $H(x|y)$ denotes the conditional entropy of x given y, and x^T denotes the history of the process $\{x_t\}_{t=-\infty}^{T}$. In this appendix, we first show that $\mathcal{I} = K(m)$ for a well-defined cost function

[4]See, for example, Maćkowiak *et al.* (2018).

$K(\cdot)$, so that the reduction in entropy is merely a special case of our analysis in Proposition 3. We then derive the comparable first-order condition to that in Proposition 3.

We use the following properties of conditional entropy:

Lemma G.1. *Let X, Y, and Z be random vectors. Then:*

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

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

Proof of Lemma G.1: See Cover and Thomas (2012). □

To start, let $s = \{\theta, x\}$. Symmetry and the chain rule for entropy, then allows us to write

$$\begin{aligned} H\left(s^T\right) - H\left(s^T \mid z_i^T\right) &= H(z_i^T) - H(z_i^T|s^T) \\ &= \sum_{t=1}^{T} H(z_{it}|z_i^{t-1}) - H(z_{it}|z_i^{t-1}, s^T). \end{aligned} \tag{OA17}$$

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

$$\begin{aligned} H(z_{it}|z_i^{t-1}) - H(z_{it}|z_i^{t-1}, s^T) &= H(z_{it}|z_i^{t-1}) - H(z_{it}|z_i^{t-1}, s_t) \\ &= H(s_t|z_i^{t-1}) - H(s_t|z_i^t) \\ &= H(\theta_t|z_i^{t-1}) - H(\theta_t|z_i^t) + H(x_t|z_i^{t-1}, \theta_t) - H(x_t|z_i^t, \theta_t), \end{aligned} \tag{OA18}$$

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

For *the first term* in (OA18), since all variables are jointly Gaussian, we have that

$$H(\theta_t|z_i^{t-1}) - H(\theta_t|z_i^t) = \frac{1}{2}\log\left[\frac{\mathbb{V}ar_{t-1}\left[\theta_t\right]}{\mathbb{V}ar_t\left[\theta_t\right]}\right].$$

Now focus on the steady state where $\mathbb{V}ar_t\left[\theta_t\right] = \mathbb{V}ar_{t-1}\left[\theta_{t-1}\right] = V(\tau)$, with τ defined in (18). Using the AR(1) dynamics of θ_t, we have that

$$\mathbb{V}ar_{t-1}\left[\theta_t\right] = \rho^2 V(\tau) + \sigma_\theta^2,$$

which after substituting gives us

$$H(\theta_t|z_i^{t-1}) - H(\theta_t|z_i^t) = \frac{1}{2}\log\left[\rho^2 + \frac{\sigma_\theta^2}{V(\tau)}\right] \equiv \mathcal{K}(\tau), \tag{OA19}$$

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

For *the second term* in (OA18), note that x_t is independent of z^{t-1} conditional on θ_t, so that

$$\begin{aligned}
H(x_t|z_i^{t-1},\theta_t) - H(x_t|z_i^t,\theta_t) =& H(x_t|\theta_t) - H(x_t|z_{it},\theta_t) \\
=& \frac{1}{2}\log\left[\frac{\det\left(\mathbb{V}ar[x_t|\theta_t]\right)}{\det\left(\mathbb{V}ar[x_t|\theta_t,z_{it}]\right)}\right] = \frac{1}{2}\log\left[\frac{\prod_{i=1}^m b_i^2}{\prod_{i=1}^m b_i^2(1-m_i)}\right] \\
=& \frac{1}{2}\log\left[\frac{1}{\prod_j(1-m_j)}\right] = -\frac{1}{2}\sum_{j=1}^m \log(1-m_j).
\end{aligned} \tag{OA20}$$

Substituting (OA20) and (OA19) into (OA18) then shows that

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

which is well-defined since τ is a function of m. Finally, combining (OA17) with (OA16) and using stationarity, we find that our cost function satisfies $K(m) = \mathcal{I}$.

We can now use Proposition 3 to see that the first-order condition for m_j at an interior optimum satisfies:

$$w_j^2 b_j^2 + \hat{\mu}_\tau \frac{a_j^2}{b_j^2} + \mu_\alpha w_j a_j = \frac{1}{2}\frac{1}{1-m_j}, \tag{OA21}$$

where the adjusted multiplier measuring learning spillovers is

$$\hat{\mu}_\tau = \mu_\tau - \mathcal{K}'(\tau),$$

with μ_τ defined as in Proposition 3. The second term in (OA21) is specific to the entropy cost formulation, because entropy reductions also depend on the sufficient statistic τ. The comparative statics remain the same as in our version with a generic cost function: It is optimal to pay attention to important components (high w_j), and to volatile components (high b_j) as long as spillovers are not too strong. In addition, we see that an entropy cost function naturally yields $m_j < 1$ for all j: Attention is always imperfect because the entropy costs of full attention $m_j \to 1$ are infinite. We summarize these results in Proposition G.1.

Proposition G.1. *With the entropy attention costs in (OA16), the first-order condition for agents' optimal attention choice m_j at an interior optimum satisfies:*

$$w_j^2 b_j^2 + \hat{\mu}_\tau \frac{a_j^2}{b_j^2} + \mu_\alpha w_j a_j = \frac{1}{2}\frac{1}{1-m_j}, \tag{OA22}$$

where $\hat{\mu}_\tau = \mu_\tau - \mathcal{K}'(\tau)$ and μ_τ is defined in Proposition 3. We note that attention is always imperfect because the entropy costs of full attention $m_j \to 1$ are infinite.

H Flexible Information Design

An agent's *flexible information design problem* is

$$\min_{K,A,B,\Sigma_\psi} \mathbb{E}\left[\left(a_t^\star - \mathbb{E}\left[a_t^\star \mid \Omega_{it}\right]\right)^2\right] \tag{OA23}$$

subject to

$$\lim_{T\to\infty} \frac{1}{T}\left\{H\left(a^{\star,T} \mid \bar{a}_0^\star\right) - H\left(a^{\star,T} \mid \bar{a}_0^\star, s^{K,T}\right)\right\} \leq \kappa, \tag{OA24}$$

where $\bar{a}_0^\star$ denotes the vector of initial conditions, and

$$s_{it}^K = A a_t^{\star,M} + B v_t^N + \psi_{it}^K, \tag{OA25}$$

with $a_t^{\star,M} = \begin{bmatrix} a_t^\star & a_{t-1}^\star & \dots & a_{t-M+1}^\star \end{bmatrix}'$, $v_t^N = \begin{bmatrix} v_t' & v_{t-1}' & \dots & v_{t-N+1}' \end{bmatrix}'$, and $\psi_{it}^K \sim i.i.d.\mathcal{N}(0, \Sigma_\psi)$.

Lemma H.1. The information flow constraint (OA24) is equivalent to

$$\lim_{T\to\infty} \frac{1}{T}\left[H\left(\varphi_T \mid s_i^{K,T-1}\right) - H\left(\varphi_T \mid s_i^{K,T}\right)\right] \leq \kappa,$$

where the vector φ_T can be any vector with the following two properties: (i) $a_t^{\star,M}$ and v_t^N in (OA25) can be computed from φ_t; and (ii) φ_t contains no redundant elements.

Proof of Lemma H.1: The proof is identical to the proof of Lemma 1 in Maćkowiak *et al.* (2018). The proof relies on the symmetry of mutual information, the law of total mutual information, and the fact that $a_t^\star$ has a time series representation. The steps do not change when the innovations to $a_t^\star$ are a linear combination of white noise shocks (as in our model), or a single white noise innovation (as in Maćkowiak *et al.*, 2018). □

Lemma H.2. Any optimal signal vector s_{it}^K is a linear combination of $a_t^\star$ and v_t only.

Proof of Lemma H.2: This is a special case of Proposition 1 in Maćkowiak *et al.* (2018). To apply the steps in the proof of Proposition 1 in Maćkowiak *et al.* (2018), we first note that the signal vector in (OA25) has the following state-space representation:

$$\begin{aligned} s_{it}^K &= G'\varphi_t + \psi_{it}^K \\ \varphi_t &= F\varphi_{t-1} + w_t, \end{aligned} \tag{OA26}$$

where

$$\varphi_t = \begin{bmatrix} a_t^\star & \dots & a_{t-M+1}^\star & v_t' & \dots & v_{t-N+1}' \end{bmatrix}'$$

and, for example,

$$F_{M=N=2} = \begin{bmatrix} \rho & 0 & c_1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{I} & 0 \end{bmatrix}, \quad w_{t,M=N=2} = \begin{bmatrix} c_0 \\ \mathbf{0} \\ \mathbf{I} \\ 0 \end{bmatrix} v_t.$$

We note that the matrix $G_{M=N=2}$ is the matrix for which (OA26) equals (OA25). Such a matrix exists because $a_t^{\star,M}$ and v_t^N can be computed from φ_t.

Now, let

$$\xi_t = \begin{bmatrix} a_t^\star \\ v_t \end{bmatrix} = \begin{bmatrix} \rho & c_1 \\ 0 & \mathbf{0} \end{bmatrix} \begin{bmatrix} a_{t-1}^\star \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} c_0 \\ I \end{bmatrix} v_t \equiv F_\xi \xi_{t-1} + v_{\xi,t}$$

The same steps as in Proposition 1 in Maćkowiak *et al.* (2018) then show, using Lemma H.1, that both (i) the signal vector in (OA26), and (ii) a signal vector about ξ_t leave the objective function in (OA23) unchanged (both contain $a_t^\star$, and expectations about $a_t^\star$ can be computed from both). The only difference is the former is associated with more information flow. □

Lemma H.3. $K = 1$ optimally.

Proof of Lemma H.3: In this case, the proof is quite simple (and follows the same steps as those in Maćkowiak *et al.*, 2018 Proposition 3). Suppose the signal vector s_{it}^K consists of multiple signals of the *same* linear combination of $a_t^\star$ and v_t. Then, the signal vector is clearly equivalent to a one-dimensional signal with a higher precision. □

Combined, these three result imply that:

Proposition H.1. The optimal signal is

$$z_{it}^\star = \alpha_0 a_t^\star + \alpha_1' v_t + \psi_{it} \propto a_t^\star + h' v_t + q^\star \epsilon_{it}, \tag{OA27}$$

where h and $q^\star$ are defined in Section 4.3.

Finally, we discuss the comparative statics of $z_{it}^\star$ in (OA27) with respect to the weights w_θ and w_{xj} and the capacity constraint κ. To do so, we once more exploit and adapt a result from Cover and Thomas (2012) and Maćkowiak *et al.* (2018):

Corollary H.1. Suppose $w_\theta > 0$ and $w_{xj} = 0$ for all $j = \{1, 2, ..., N\}$.

Then, the optimal signal collapses to $z_{it}^\star = \alpha_0 \theta_t + q^\star \epsilon_{it}$.

Proof of Corollary H.1: The optimal action $a_t^\star$ is a standard AR(1) in this case. The result then follows from the well-known steps in, for example, Cover and Thomas (2012). □

Corollary H.2. As $\kappa \to \infty$, the optimal signal converges to only be about $a_t^\star$.

Proof of Corollary H.2: Identical to the proof of Proposition 6 in Maćkowiak *et al.* (2018). □

I Macroeconomic Example and Angeletos *et al.* (2016)

Our macroeconomic example in Section 5 considers a model similar to those considered in Angeletos and La'O (2010, 2012) and Angeletos *et al.* (2016). To demonstrate why we view strategic substitutability as a natural assumption, we generalize our baseline model to encompass both our model from Section 5 as well as the features that determine the strategic considerations of firms in Angeletos *et al.* (2016).[5] Consider our model in Section 5. Assume that firm productivity follows a common process with $\epsilon_{it} = 0$ (as in our baseline calibration). Replace households' utility with $u(C, N) = \frac{C^{1-\psi}-1}{1-\psi} - \frac{1}{1+\eta} N_t^{1+\eta}$. Relative to this overarching model, our analysis in Section 5 restricts attention to log consumption utility ($\psi \to 1$) and linear disutility of labor ($\eta = 0$).[6] Angeletos *et al.* (2016) allow for general values for ψ and η, but set $\alpha = 1$ in firms' production function, so that it has constant returns to scale in labor. We below abstract from any labor supply shocks, which do not affect firms' strategic behavior, without loss of generality. We solve for the full-information equilibrium of this model:

Proposition I.1. *Let* $u(C, N) = \frac{C^{1-\psi}-1}{1-\psi} - \frac{1}{1+\eta} N_t^{1+\eta}$. *Under full information, firm i's optimal output choice satisfies the best response function*

$$y_{it} = k_0 a_t + k_1 y_t, \tag{OA28}$$

where $k_0 > 0$ *and the coefficient of strategic complementarity* k_1 *is*

$$k_1 = \frac{\alpha(1 - \sigma\psi)}{\alpha(1 - \sigma) + \sigma(1 + \eta)}. \tag{OA29}$$

We note that, because of certainty equivalence, we can use the full-information solution of the generalized model in (OA28) and (OA29) to determine whether output choices are strategic substitutes or complements even under imperfect information.

Equation (OA29) implies that firms' output choices are strategic substitutes ($k_1 < 0$) if and only if $\sigma\psi > 1$. Standard parameter choices in macroeconomics (see, for example, Gali, 2008, Chapter 3 p. 56) have $\sigma \in [4, 10]$ and $\psi \in [1, 4]$, so that $\sigma\psi \geq 4$ and $k_1 < 0$. Thus, we conclude that strategic substitutes are pervasive for most popular parameterizations.

[5] In addition to the features mentioned, Angeletos *et al.* (2016) include one additional layer of CES aggregation.

[6] We choose this parametrization for standard reasons. First, the calibration of $\psi \to 1$ is the only value within the iso-elastic utility class that is consistent with balanced growth (i.e. is within the well-known KPR-class). Second, the calibration of $\eta \to 0$ allows flex-price models to generate sufficient volatility in hours worked (e.g. Prescott and Wallenius, 2012). As shown by Hansen (1985) and Rogerson (1988), linear disutility of labor can arise from the iso-elastic framework (considered in Angeletos *et al.*, 2016) when one accounts for the fact that most of the variation in hours worked are due to changes in the extensive (rather than the intensive) margin. It thus allows our model to have a higher Frisch elasticity, without simultaneously being subject to the criticism that the labor supply elasticity is inconsistent with micro-evidence.

J Numerical Solution of Model with Imperfect Attention

We solve the model by repeated iteration of the two steps described in the main text. Below, we detail these steps in reverse order. First, we solve for the imperfect information equilibrium given a set of attention choices. Then, we solve for the optimal attention choices.

Step 2: Equilibrium Given Attention Choices:[7] Consider the equation for aggregate output that arises under imperfect attention:

$$y_t = \int_0^1 y_{it} di = \bar{\mathbb{E}}_t \left[x_{1t} + x_{2t} \right], \tag{OA30}$$

where $x_{1t} = \int_0^1 x_{i1t} di$ and

$$x_{1t} \quad = \quad r\theta_t + r u_t^x, \quad x_{2t} = \alpha r \sigma^{-1} y - \alpha r \left(\mathbb{E}_t^h \left[y_t \right] + u_t^n \right).$$

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

$$\boldsymbol{x}_t = A \boldsymbol{x}_{t-1} + B u_t, \quad u_t = \begin{bmatrix} u_t^\theta & u_t^x & u_t^n \end{bmatrix}', \tag{OA31}$$

where

$$A = \begin{bmatrix} 0 & 0 & r\rho_\theta & \mathbf{0} \\ & A_p & & \\ 0 & 0 & \rho_\theta & \mathbf{0} \\ & \mathbf{I} & & \end{bmatrix}, \quad B = \begin{bmatrix} r & r & 0 \\ & B_p & \\ 1 & 0 & 0 \\ & \mathbf{0} & \end{bmatrix}. \tag{OA32}$$

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

$$y_t = \psi \boldsymbol{x}_t, \quad x_{2t} = c_0 \boldsymbol{x}_t + c_1 u_t, \tag{OA33}$$

where ψ, c_0, and c_1 are vectors of coefficients.

Coefficients and Conjectures: It follows from (OA30) that

$$y_t = \begin{bmatrix} 1 & 1 & \mathbf{0} \end{bmatrix} \bar{\mathbb{E}}_t \left[\boldsymbol{x}_t \right] \mathrel{\mathring{=}} \psi \boldsymbol{x}_t, \tag{OA34}$$

where $\mathring{=}$ denotes "should equal". We conclude from (OA34) that to verify our conjecture we need to find a matrix Ξ such that

$$\bar{\mathbb{E}}_t \left[\boldsymbol{x}_t \right] = \Xi \boldsymbol{x}_t. \tag{OA35}$$

[7] The steps used to find this equilibrium are analogous to those described in Lorenzoni (2009).

Now since

$$\mathbb{E}_t^h\left[y_t\right]=\psi\left\{A\boldsymbol{x}_{t-1}+B\begin{bmatrix}1&0&0\\0&0&0\\0&0&1\end{bmatrix}u_t\right\}=\psi\boldsymbol{x}_t-\psi B\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}u_t,$$

it also follows that

$$x_{2t}=\alpha r\sigma^{-1}\psi\boldsymbol{x}_t-\alpha r\left\{\psi\boldsymbol{x}_t-\psi B\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}u_t+e_3u_t\right\}\mathrel{\mathring{=}}c_0\boldsymbol{x}_t+c_1u_t, \tag{OA36}$$

where e_l denotes a row vector with a one in the l's position but zeros elsewhere.

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

$$\begin{aligned}s_{it} &= \begin{bmatrix}e_1\\e_2\end{bmatrix}\boldsymbol{x}_t+Q\epsilon_{it}, \quad Q=diag\begin{bmatrix}q_1&q_2\end{bmatrix} \\ &\equiv L\boldsymbol{x}_t+Q\epsilon_{it}.\end{aligned} \tag{OA37}$$

Thus,

$$\mathbb{E}_{it}\left[\boldsymbol{x}_t\right]=A\mathbb{E}_{it-1}\left[\boldsymbol{x}_t\right]+K\left(s_{it}-LA\mathbb{E}_{it-1}\left[\mathbf{z_{t-1}}\right]\right), \tag{OA38}$$

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

Then, from (OA35) and (OA38) it has to hold for all t that

$$\Xi\boldsymbol{x}_t=\left(I-KL\right)A\Xi\boldsymbol{x}_{t-1}+KL\boldsymbol{x}_t. \tag{OA39}$$

Fixed Point: We have from (OA34) and (OA36) that

$$\psi=\begin{bmatrix}1&1&\mathbf{0}\end{bmatrix}\Xi, \quad c_0=\alpha r\left(\sigma^{-1}-1\right)\psi, \quad c_1=\alpha r\left(\psi B\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}+e_3\right). \tag{OA40}$$

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

To find the equilibrium, we apply the following algorithm: We start with some initial values for A_p and B_p (for simplicity, we use those from the corresponding full-information solution). We then use these values to compute K from (OA37) and (OA38). This, in turn, allows us to find an expression for Ξ from (OA39) since

$$\Xi \boldsymbol{x}_t^{[T]} = (I - KL) A\Xi M \boldsymbol{x}_t^{[T]} + KL\boldsymbol{x}_t^{[T]},$$

where

$$M = \begin{bmatrix} 0 & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

which gives us the following relationship that we solve for Ξ:

$$\Xi = (I - KL) A\Xi M + KL. \tag{OA41}$$

We can now use (OA40) to find an expression for ψ, c_0, and c_1.

Finally, we use these expressions to compute new values of A_p and B_p from (OA32), and then repeat these steps until convergence is achieved. The criterion used is the maximum absolute difference between the new and old elements of A_p and B_p.

Step 1: Attention Choices Given Equilibrium: Given the above aggregate equilibrium, we solve a firm's *ex-ante* attention choice problem. That is, we solve

$$\min_{m_1,m_2} \mathbb{E}_{it}\left[y_{it} - y_{it}^{\star}\right]^2 + K(m), \quad K(m) = \mu\left(q_1^{-2} + q_2^{-2}\right), \tag{OA42}$$

where $q_j = \mathbb{V}(x_{jt} \mid \theta_t)\left[m_j - \mathbb{V}(x_{jt} \mid \theta_t)\right]^{-1}$ for $j = \{1, 2\}$ and we have that

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

in which

$$\begin{aligned} x_{i1t} &= r\theta_t + ru_t^x + r\epsilon_{it}^a = re_3'\boldsymbol{x}_t^{[T]} + re_2'\sigma_x u_t \equiv a_1\boldsymbol{x}_t^{[T]} + b_1 u_t + \epsilon_{it}^a \\ x_{2t} &\equiv a_2\boldsymbol{x}_t^{[T]} + b_2 u_t, \end{aligned}$$

and where a_1 and b_1 are implicitly defined, while $a_2 = c_0$ and $b_2 = c_1$.

To minimize (OA42), we first derive an expression for the quadratic component

$$\mathbb{E}\left[y_{it}^{\star} - \mathbb{E}_{it}\left[y_{it}^{\star}\right]\right]^2 = 1'\mathbb{V}\left[x_{it} \mid z_i^t\right]1, \quad x_{it} = \begin{bmatrix} x_{i1t} & x_{2t} \end{bmatrix}'$$

where

$$\mathbb{V}\left[x_{it} \mid z_i^t\right] = \mathbb{V}\left[x_{it} \mid z_i^t, \boldsymbol{x}_t^{[T]}\right] + \mathbb{V}\left[\mathbb{E}\left[x_{it} \mid z_i^t, \boldsymbol{x}_t^{[T]}\right] \mid z_i^t\right] \tag{OA43}$$

by the *Law of Total Variance.*

It now follows that *the first component* in (OA43) is

$$\begin{aligned} \mathbb{V}\left[x_{it} \mid z_i^t, \boldsymbol{x}_t^{[T]}\right] = \mathbb{V}\left[x_{it} \mid z_{it}, \boldsymbol{x}_t^{[T]}\right] &= bb' + \bar{r}\bar{r}' - \left(bb' + \bar{r}\bar{r}'\right)\left[bb' + QQ' + \bar{r}\bar{r}'\right]^{-1}\left(bb' + \bar{r}\bar{r}'\right)' \\ &= bb' + \bar{r}\bar{r}' - \tilde{m}\left(bb' + \bar{r}\bar{r}'\right)', \end{aligned}$$

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

To derive *the second component* in (OA43), notice that

$$\begin{aligned} \mathbb{E}\left[x_{it} \mid z_i^t, \boldsymbol{x}_t^{[T]}\right] = \mathbb{E}\left[x_{it} \mid z_{it}, \boldsymbol{x}_t^{[T]}\right] &= \mathbb{E}\left[x_{it} \mid \boldsymbol{x}_t^{[T]}\right] + \tilde{m}\left(z_{it} - \mathbb{E}\left[z_{it} \mid \boldsymbol{x}_t^{[T]}\right]\right) \\ &= (I - \tilde{m})\, a\boldsymbol{x}_t^{[T]} + \tilde{m} z_{it}, \end{aligned}$$

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

$$\mathbb{V}\left[\mathbb{E}\left[x_{it} \mid z_i^t, \boldsymbol{x}_t^{[T]}\right] \mid z_i^t\right] = (I - \tilde{m})\, a\mathbb{V}\left[\boldsymbol{x}_t^{[T]} \mid z_i^t\right] a' (I - \tilde{m})',$$

in which $\mathbb{V}\left[\boldsymbol{x}_t^{[T]} \mid z_i^t\right]$ can be found from the Kalman Filter run in (OA38).

In sum, we have that the quadratic term (OA42) becomes

$$\begin{aligned} \mathbb{E}\left[y_{it}^{\star} - \mathbb{E}_{it}\left[y_{it}^{\star}\right]\right]^2 &= 1'\left[bb' + \bar{r}\bar{r}' - \tilde{m}\left(bb' + \bar{r}\bar{r}'\right)'\right] 1 \\ &+ 1' (I - \tilde{m})\, a\mathbb{V}\left[\boldsymbol{x}_t^{[T]} \mid z_i^t\right] a' (I - \tilde{m})' 1, \end{aligned}$$

which allows us to solve the problem in (OA42).

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

References

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

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

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

— and — (2012). Optimal monetary policy with informational frictions.

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

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

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

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.

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

Gali, J. (1999). Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? *American economic review*, **89** (1), 249–271.

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

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

Jacobson, M. and Occhino, F. (2012). Estimating real gdp growth trends.

Kohlhas, A. (2019). Learning-by-sharing: Monetart policy and common knowledge.

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

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

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.

Nimark, K. (2017). *Dynamic Higher Order Expectations*. Economics Working Papers 1118, Department of Economics, Cornell University.

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

Prescott, E. C. and Wallenius, J. (2012). Aggregate labor supply. *Quarterly Review*, **35** (2).

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

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