Asymmetric Attention

Alexandre N. Kohlhas and Ansgar Walther*

This version: October 2018
First draft: January 2018

Abstract

We document simultaneous over- and under-responses to new information by households, firms, and professional forecasters in survey data. Such behavior is inconsistent with existing theories based on either behavioral bias or rational inattention. We develop a structural model of information choice in which people base expectations on observables that can reconcile the seemingly contradictory facts. We show that optimally-chosen, asymmetric attention to different observables can explain the co-existence of over- and under-responses. We then embed our model of information choice into a micro-founded macroeconomic model, which generates expectations consistent with the survey data. We demonstrate that our model creates over-optimistic consumption beliefs in booms and predictability in consumption changes.

JEL codes: D83, D84, E32 *Keywords:* Expectations, learning, inattention

1 Introduction

The past five decades have shown that expectations are central to macroeconomics. Despite their importance, a unified model of how households' and firms' expectations are formed remains elusive. A substantial body of work since Lichtenstein *et al.* (1977) has established at least two robust stylized facts, which reject the benchmark of full information and rational expectations. First, it appears that forecasters, even professionals, tend to extrapolate recent

*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, Nicola Gennaioli, Per Krusell, Kristoffer Nimark, Laura Veldkamp, Mathieu Taschereau-Dumouchel, Robert Ulbricht, and conference and seminar participants at the Barcelona Summer Forum, Cambridge University, Cornell University, CEPR Summer Symposium (Gerzensee) 2017, 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*.

trends and over-react to recent observations (see, for instance, Barberis *et al.*, 2016).[1] Second, forecast revisions are systematically too small, suggesting under-reactions to new information (see, for example, Coibion and Gorodnichenko, 2015).[2] In this paper, we argue that over- and under-responses can be understood in a consolidated framework, consistent with the survey data, which rests on the idea that agents pay imperfect but asymmetric attention to observable, structural variables in the the economy.

We document a series of stylized facts which indicate that output and inflation expectations simultaneously exhibit extrapolation and under-responses to new information. Recent research has advanced our understanding of what drives over- and under-responses, but it is not clear how their *simultaneous* occurrence can be rationalized. One can explain under-responses with models of imperfect information or imperfect attention (Sims, 2003a; Maćkowiak and Wiederholt, 2015; Coibion and Gorodnichenko, 2015), where rational agents down-weigh noisy information.[3] However, these agents tend to down-weigh all new information and do not over-respond or extrapolate. Similarly, one can explain extrapolation with behavioral models where agents over-weight the representativeness of recent conditions (Bordalo *et al.*, 2017), but such models do not generate under-responses.

Motivated by this friction between evidence and theory, we revisit the question of whether imperfect attention can reconcile the evidence. We first study a canonical forecasting problem to understand whether certain patterns of inattention can, in principle, explain these facts. We further ask whether these patterns can be the result of optimal information choice by boundedly rational agents. Moreover, we evaluate whether it is sensible to expect such behavior in the context of a micro-founded macroeconomic model, and examine the auxiliary implications of such a framework.

A simple example illustrates our results. Consider an economy where output is driven by two components. The first component is pro-cyclical, while the second one is counter-cyclical. For example, as we show below, the pro-cyclical component could represent firm productivity, while the counter-cyclical could capture strategic interactions amongst firms that offset the direct effects of the productivity shock. Agents receive noisy signals about each component, and rationally update their output expectations. We think of attention to each component as the precision of the associated signal. Attention is *asymmetric* if agents receive a relatively more precise signal about one component than the other.

[1] See also Cagan (1956), Kahneman and Tversky (1972), Cutler *et al.* (1990), De Long *et al.* (1990), Greenwood and Shleifer (2014), Bordalo *et al.* (2012), and Bordalo *et al.* (2017). Muth (1961) himself acknowledged the possibility that people in reality could have extrapolative forecasts in his formulation of the full information, rational expectations hypothesis.

[2] See also Coibion and Gorodnichenko (2012), Andrade and Le Bihan (2013), Dovern *et al.* (2015), Dovern (2015) and Fuhrer (2017) for closely related estimates.

[3] See also Sims (1998), Maćkowiak and Wiederholt (2009), and Wiederholt (2010).

We show that the combination of imperfect but asymmetric attention can reconcile the co-existence of over- and under-responsive expectations. If agents pay asymmetric attention to the first component, then their rational forecasts are over-responsive and exhibit extrapolation. Indeed, compared to the full information benchmark, agents who focus on pro-cyclical variables become over-optimistic in booms and over-pessimistic in busts. The measured over-response to recent output is, in effect, an under-response to counter-cyclical components. In addition, as long as attention remains imperfect, agents still exhibit under-responses to new information, due to their down-weighting of noisy information.

Our formal analysis starts by generalizing this argument to a class of forecasting problems, which in its broadest form nests most linearized DSGE models. We further establish a partial converse: Within a rational framework, imperfect attention can explain extrapolation *only if* it is asymmetric about structural components. Moreover, we show how simple behavioral models with naive extrapolation alone cannot explain systematic under-responses.

In the remainder of the paper, we address two further points. On one hand, our argument above takes attention as given, and we can go further by analyzing the reasons for asymmetric attention choices. On the other hand, we have also taken the dynamics of the economy as given, and ignored feedback effects from information choices to the macroeconomic variables themselves. To understand these effects, and the wider implications of asymmetric attention, we embed our argument in a workhorse macroeconomic model.

We micro-found asymmetric attention in a structural, component-based model of information choice. Agents choose signals of economic components optimally, so as to make good decisions. Paying attention is costly, in the sense that more precise signals reduce the agent's utility. There is a large and diverse literature on information choice, and solutions are often sensitive to the exact functional form of attention costs (see, for instance, Veldkamp, 2011). We illustrate that asymmetric attention can be optimal for a range of functional forms.

We demonstrate that asymmetric attention is commonly optimal, based on two insights. First, when attention is costly, it optimally gravitates towards variables that are the most decision-relevant. Second, when agents solve information choice problems, it is optimal for them to pay attention to components that strongly correlate with the unobserved fundamentals of the economy. Overall, our analysis suggests that is is not difficult to generate asymmetric attention as the outcome of optimal, boundedly rational choices.

Combining these results with our previous characterization of expectations, we show how models of optimal but costly information choice can rationalize the survey data. Indeed, if procyclical factors are more correlated with fundamentals, or if these factors are more important to agents' decisions then rational forecasters will acquire beliefs that are both extrapolative and under-responsive.

We revisit the data in order to study whether agents' expectations are more or less precise than those obtained from standard time series models of output. A central assumption in our model is that agents take into account the structural components of output, as opposed to inferring it only from its past time series. Therefore, a testable implication is that expectations should be more precise than pure time series forecasts. We update estimates from Stark (2010) to show that forecasters' expectations consistently outperform time series models, especially at short horizons. This supplementary evidence is consistent with our model. This result further suggests that asymmetric attention is likely to be optimally chosen, as opposed to resulting from strong behavioral biases.

Last, we embed our model of information choice into a benchmark, micro-founded macroeconomic model with flexible prices. In the model, firms choose output under imperfect information about labor productivity and a policy maker, the tax-authority, sets labor tax. This macroeconomic model closely resembles those proposed by Adam (2007), Maćkowiak and Wiederholt (2009), and Angeletos *et al.* (2016) to study the social value of public information. Firm output choices can in equilibrium be split into two components: (i) firm beliefs about *a local component*, their own labor productivity, and (ii) firm beliefs about *an economy-wide component*, here a linear combination of aggregate output and labor taxes. The latter matters for firm choices through its general equilibrium effects on prices and mark-ups. This connects the macroeconomic model to beauty-contest models studied in, for instance, Morris and Shin (2002) and Angeletos and Pavan (2007).

We show that, for benchmark parameter values, firms may optimally choose to pay asymmetric attention. Because local conditions are more variable and more important than economy-wide ones, firms may optimally choose to pay close attention to the local component. This, in turn, causes firm expectations of future output to appear both extrapolative and under-responsive. A simple calibration exercise shows that the model captures well the salient features of the survey data – in particular the size of over- and under-responses to new information and the fact that firms forecast better than simple ARMA models.

We last use the model to explore the quantitative implications of the simultaneous over- and under-responses of firm expectations. We show how they result in predictability of consumption and output changes, in line with the empirical evidence in Attanasio (1999).

Organization: We empirically document the over- and under-responses in economy-wide forecast data in Section 2. In Sections 3 and 4, we analyze a class of forecasting problems and establish that asymmetric attention choices can rationalize the evidence. We also discuss how our results extend across several different cost functions for attention. Section 5 maps the insights from the baseline model into a micro-founded macroeconomic model. We conclude in Section 6. Additional extensions and all proofs are in the Appendix.

2 Over- and Under-responses in the Data

We start by specifying two regressions which provide tests of specific deviations from full information and rational expectations, as used by Coibion and Gorodnichenko (2015) and Bordalo *et al.* (2017) among others. We first consider output expectations from the US Survey of Professional Forecasters, and demonstrate that forecasts simultaneously exhibit over-responses to current output levels and under-responses to new information as measured by forecast revisions. We then show that these patterns extend to inflation forecasts, across countries and to non-professional forecasters.

2.1 Tests of Deviations from Rational Expectations

Let y_t be a macroeconomic variable such as output at time t. We write $\hat{y}_{t+j|t}$ for the time t forecast of a future variable y_{t+j}. The j-period-ahead forecast error in this notation is $y_{t+j} - \hat{y}_{t+j|t}$. Positive and negative values of the forecast error, respectively, stand for under- and overestimates of y_{t+j}.

To detect extrapolative expectations, we regress forecast errors on the current level of y_t:

$$y_{t+j} - \hat{y}_{t+j|t} = \text{constant} + \gamma y_t + \xi_t, \tag{2.1}$$

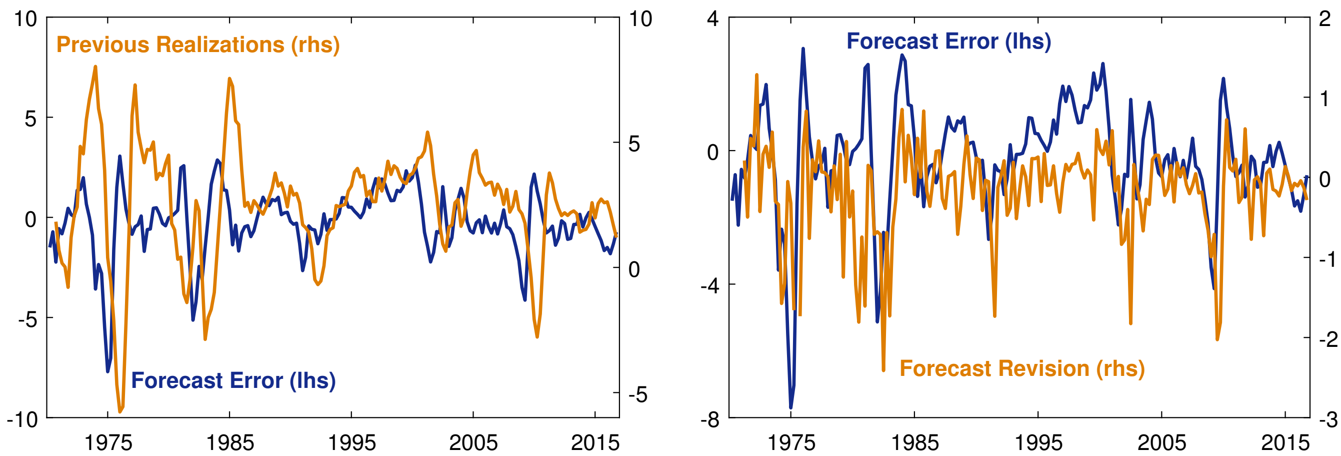
where ξ_t is an error term. Under full information and rational expectations, it should be impossible to predict next period's forecast error using information available today, and we have $\gamma = 0$. This is an instance of the *Projection Theorem.* Empirically, by contrast, $\gamma = 0$ is often rejected in favor of $\gamma < 0$ (see, for instance, Bordalo *et al.*, 2017 and references therein). This case suggests extrapolation, because agents systematically over-predicts y_{t+j} when y_t is currently high. An informed rational agent, by contrast, would adjust these estimates downwards.

To detect under-responsive expectations, Coibion and Gorodnichenko (2015) propose a regression of forecast errors on previous forecast revisions:

$$y_{t+j} - \hat{y}_{t+j|t} = \text{constant} + \delta \left(\hat{y}_{t+j|t} - \hat{y}_{t+j|t-1} \right) + \xi_t \tag{2.2}$$

The independent variable $\hat{y}_{t+j|t} - \hat{y}_{t+j|t-1}$ measures the ex-ante change in the forecast of y_{t+j} between times $t-1$ and t. Under full information and rationality, the projection theorem again implies $\delta = 0$: Today's revision should not predict next period's error. Empirically, we often find $\delta > 0$ (see (see Andrade and Le Bihan, 2013; Coibion and Gorodnichenko, 2015). This indicates under-responsive expectations since a positive revision today predicts

Figure 2.1: Over- and Under-responses in Output Growth Forecasts



Mean one-year ahead forecasts of output growth from the *SPF*, the associated mean forecast error, and the mean one-quarter forecast revision are shown. All scales are in percent year-over-year.

an underestimate next period, and the revision of a fully informed rational agent would have been larger to begin with.

2.2 Empirical Results

We first present results from the US *Survey of Professional Forecasters (SPF)*, collected by the Federal Reserve Bank of Philadelphia. The *SPF* is a common and conservative benchmark for testing rational expectations, since the null hypothesis of full information and rationality should be hardest to reject for professional forecasters.

We consider one-year-ahead forecasts of year-on-year output growth. Our data consists of GDP/GNP growth forecasts from the *SPF* starting from 1968:Q4 at a quarterly frequency (implying $j = 4$ for a one-year ahead forecast in the above regressions). We use real-time data to measure *ex-post* realizations to more precisely capture the precise definition of the output variable being forecast.

The raw data already hint at deviations from full-information rationality. Figure 2.2 plots one-year-ahead forecast errors against previous realizations of output and one-quarter forecast revisions. In the left panel, forecasts are frequently over-optimistic (with an associated negative forecast error, for instance at the onset of the 2008 financial crisis) when the previous level of output is high, consistent with extrapolation. Anticipating under-responses, the right panel suggests that forecast errors and past forecast revisions are positively correlated.

Table 2.1 confirms these impressions and reports estimates of γ and δ in regressions (2.1) and (2.2) from the *SPF* data on output growth. The estimated coefficients are significantly

Table 2.1: Estimated over- and under-responses in the *SPF*

	Forecast Error	*Forecast Error*	*Forecast Error*
Constant	-0.22*	-0.03	-0.07
	(0.12)	(0.02)	(0.10)
Realization (t-1)	-0.15**	–	-0.20***
	(0.06)		(0.05)
Forecast Revision	–	0.77***	0.85***
		(0.26)	(0.23)
Sample	01/70:10/17	01/70:10/17	01/70:10/17
F	6.50	15.2	13.8
R^2	0.04	0.08	0.13

(i) HAC standard errors used.

(ii) * p<0.10, ** p<0.05, *** p<0.01.

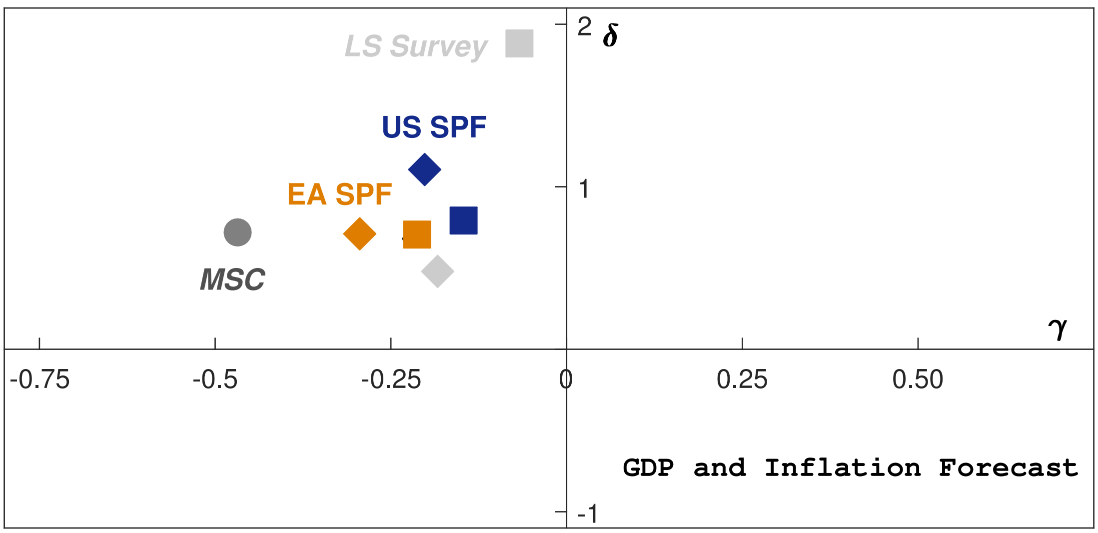
(iii) An HP-trend $\lambda = 1600$ has been deducted from y_{t-1} to account for potential structural changes.

different from zero and have opposite signs, $\gamma < 0$ while $\delta > 0$. These findings are consistent with both the large body of literature documenting extrapolation from estimates of (2.1), and the more recent work which documents the appearance of under-responses from (2.2). This, in turn, indicates that expectations are simultaneously over-responsive (or extrapolative) to past output growth and under-responsive to new information incorporated in forecast revisions.

Robustness of Evidence: This patterns we have found in the *SPF* are remarkably stable when we consider other macroeconomic variables, non-professional forecasts, or other countries. Figure 2.2 summarizes the results from estimating (2.1) and (2.2) across a range of survey data; we report the associated regression output in Appendix A. First, in addition to *SPF* output forecasts, we consider *SPF* inflation forecasts, which are a focal point in the literature on survey expectations because of the importance of inflation expectations for the New Keynesian Phillips Curve. The estimated coefficients have the same sign as in the output case and are and slightly larger in absolute value. Second, we extend beyond professional forecasts by using *(i)* output and inflation forecasts from the *Livingstone Survey*, which covers a broad range of agents, including academic institutions, investment banks, non-financial firms, and government agencies (Croushore, 1997); and *(ii)* inflation forecasts from the *Michigan Survey of Consumers*.[4] The estimated coefficients all have the same sign

[4]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

Figure 2.2: Estimated over- and under-responses across surveys



Estimates of coefficients γ and δ from (2.1) and (2.2). *US SPF* represents the estimates for the *US Survey of Professional Forecasters*, *EA SPF* the *ECB's Survey of Professional Forecasters*, *LS Survey* the *Livingstone Survey*, and last *MSC* the *Michigan Survey of Consumers.* $\square$ = GDP forecasts, $\diamond$ = Inflation forecasts, and $\circ$ = MSC inflation forecasts that have been instrumented. All estimates are for one-year ahead forecasts, and estimates of (2.2) use semiannual revisions (Livingstone Survey) or one-quarter revisions (all others).

as those for professional forecasts. Last, we use professional forecasts in the Eurozone as collected by the *ECB Survey of Professional Forecasts*, again finding coefficients of the same sign and similar magnitudes to the US *SPF*. Table A.1 shows that, with the exception of Eurozone and Livingstone inflation forecasts, all of the estimates shown in Figure 2.2 are statistically significant at the five percent level.

In sum, all estimated coefficients lie in the north-west quadrant in Figure (2.2). The data thus strongly suggest that extrapolation ($\gamma < 0$) and under-responsiveness ($\delta > 0$) occur simultaneously, both for output and inflation expectations. As we have argued in the introduction, and will show more formally below, models based purely on extrapolation or dispersed, noisy information cannot rationalize this coincidence of over- and under-responsive expectations. In the next Section, we present a simple model that takes a first step towards reconciling these disparate stylized facts.

3 Asymmetric Attention in Forecasting Problems

We start with a simple stylized model in which output is driven by a set of structural components, and where economic agents have imperfect information. We model agents' *attention* to a component as the quality of their information about its current level. Attention about output is then a composite of attention paid to individual components.

Our focus in this Section is to establish how the properties of attention map into the properties of agents' expectations. In particular, we show that imperfect but asymmetric attention to underlying components generates over-responses to current output and simultaneous under-responses to new information. Last, we generalize the results to a wide class of tracking problems,which nests most linear DSGE models (Fernández-Villaverde *et al.*, 2007).

We will take attention choices as given for now, and defer a study of their determination to Section 4. As such, the results of this Section could also be interpreted in terms of differentials in information quality that are not driven by attention, for example, those driven by the segmentation of markets in localized economies.

estimate (2.2), we follow Coibion and Gorodnichenko (2015) and replace the *ex-ante forecast revision* with the quarterly *ex-ante forecast change* and instrument this variable with the (log) oil price change. This approach provides an asymptotically consistent estimate of δ.

3.1 A Simple Structural Model

Consider an economy where output y_t is determined as the sum of N intermediate components:

$$y_t = \sum_{j=1}^{N} x_{jt} \tag{3.1}$$

For concreteness, we refer to y_t as output in this Section, but it can also stand for inflation or other macroeconomic aggregates. We obtain (3.1), for example, from a generalized Cobb-Douglas production function, where y_t denotes the log of output.

Each output component x_{jt} is driven by a persistent state variable θ_t and a component-specific transitory shock u_{jt}:

$$x_{jt} = a_j \theta_t + b_j u_{jt} \tag{3.2}$$

The state θ_t follows the autoregressive process

$$\theta_t = \rho \theta_{t-1} + \eta_t \tag{3.3}$$

We assume without loss of generality that $\frac{dy_t}{d\theta_t} = \sum_j a_j > 0$, so that y_t is increasing in θ_t.

There is a continuum of forecasters $i \in [0, 1]$. These agents do not observe output or its factors directly. Each forecaster i instead observes a signal process

$$z_{ijt} = x_{jt} + q_j \epsilon_{ijt} \tag{3.4}$$

where q_j is a constant measuring the noise contained agents' signals about x_{jt}, and ϵ_{ijt} is an agent-specific shock. We assume that η_t, u_{jt} and ϵ_{ijt} are i.i.d. Standard Normal random variables, and that $\theta_1 \sim \mathcal{N}(0, \tau_\theta^{-1})$. We can choose τ_θ so that the forecasters' signal extraction problem is initialized in steady state.

To distill our central insights, we first treat forecasters as passive agents who make predictions about economic variables given their information z_{ijt}, so as to minimize their mean-square error. We link this analysis to the behavior of active, utility-maximizing agents in Section 4. In Section 5, we further account for the fact that agents' learning itself alters the economic laws of motion in a general equilibrium model.

We note that the information structure in (3.4) is not without loss of generality: Given complete freedom to choose signals as in Sims (2010), agents may not acquire *orthogonal* information ϵ_{ijt} about individual factors unless doing so helps them to make decisions. We start with the case of orthogonal signals because it makes for a particularly clean exposition. In Section 4, we return to the non-orthogonal case, and show that our main intuitions continue to apply.

Below, we extend this framework to allow for many state variables, correlated disturbances and lags in the equation driving components x_t. This allows us to encapsulate linearized macroeconomic models, in which the state θ_t can itself contain expectations about future conditions.

Attention: We formalize what we mean by attention. Suppose that agents exogenously learn the fundamental state θ_t, and we ask each agent to predict each factor x_{jt} based on her noisy information z_{ijt}. Conditional on θ_t, this prediction satisfies

$$\mathbb{E}[x_{jt}|z_{ijt}, \theta_t] = \mathbb{E}[x_{jt}|\theta_t] + m_j \left(z_{ijt} - \mathbb{E}[z_{ijt}|\theta_t]\right), \tag{3.5}$$

where $m_j = \frac{b_j^2}{q_j^2 + b_j^2}$ is the Bayesian weight on new information.

We interpret m_j as a measure of attention to component j. It will play a key role in the over- and under-responsiveness of expectations.

Over-response Expectations and Extrapolation: Consider the "extrapolation" regression equation (2.1). The estimated coefficient on the current level y_t of output is

$$\gamma \overset{\text{sign}}{=} \text{Cov}\left[y_{t+1} - \bar{\mathbb{E}}_t y_{t+1}, y_t\right]$$

where we define the operator $\bar{\mathbb{E}}_t X = \int_0^1 \mathbb{E}_{it} X di$ as the average expectation of any random variable X across agents in the economy.

For simplicity, we now derive this covariance only for period $t = 1$. The extension to later periods is straightforward, and we subsume it in Proposition 1 below. We first note that

$$\mathbb{C}\text{ov}\left[y_2 - \bar{\mathbb{E}}_1 y_2, y_1\right] = \left(\Sigma_j a_j\right) \rho \times \text{Cov}\left[\theta_1 - \bar{\mathbb{E}}_1 \theta_1, y_1\right]$$

and since we have imposed $\Sigma_j a_j > 0$, we need only to characterize the covariance between the current level y_1 and the average forecast error $\theta_1 - \bar{\mathbb{E}}_1 \theta_1$ about θ_1.

Using the standard Gaussian updating formula, and noting that the effective precision of signal z_{ijt} for θ_1 is $\tau_j = \frac{a_j^2}{b_j^2 + q_j^2}$, we find that

$$\mathbb{E}_1^i\left[\theta_1\right] = \frac{1}{\sum_{j=1}^N \tau_j + \tau_\theta} \sum_{j=1}^N \frac{\tau_j}{a_j} z_{ijt} \tag{3.6}$$

Averaging over $i \in [0, 1]$, and evaluating the covariance of interest, we find that

$$\left(\sum_{j=1}^{N}\tau_j+\tau_\theta\right)\times\mathbb{C}\mathrm{ov}[\theta_1-\bar{\mathbb{E}}_1\theta_1,y_1]=\mathbb{C}\mathrm{ov}\left[\tau_\theta\theta_1-\sum_{j=1}^{N}\tau_j\frac{b_j}{a_j}u_{jt},\sum_{j=1}^{N}(a_j\theta_1+b_ju_{jt})\right]$$
$$=\sum_{j=1}^{N}a_j(1-m_j)$$

Extending this logic beyond time $t=1$, it is easy to show that:

Lemma 1. *The regression (2.1) of ex-post mean forecast errors of y_{t+1} on current levels of y_t suggests extrapolation ($\gamma<0$) if and only if*

$$\sum_{j=1}^{N}a_j(1-m_j)<0 \tag{3.7}$$

.

Condition (3.7) links extrapolation to asymmetric attention. Indeed, extrapolation occurs if and only if agents fail to pay sufficient attention to countercyclical factors. To see how (3.7) supports this intuition, consider an example where there are $N=2$ factors and the agent completely ignores the first ($m_1=0$) while paying full attention to the second ($m_2=1$). By Lemma 1, we then obtain extrapolation if and only if $a_1<0$, that is, if the ignored factor is countercyclical.

This result is intuitive: If an agent pays attention chiefly to procyclical factors, and ignores countercyclical ones, she will tend to be more optimistic at the peak of the cycle than she would be if paying full attention. Effectively, over-responses are driven by under-responses to countercyclical factors. More generally, the left-hand side of (3.7) corresponds to a covariance, taken across factors j, of the agents' inattention $1-m_j$ with the factor's loading a_j on the aggregate state. This covariance is negative when attention centers on pro-cyclical components of output.

Under-responsive Expectations: We now turn to the "under-responsiveness" regression (2.1). The estimated coefficient on forecast revisions is

$$\delta\overset{\text{sign}}{=}\mathbb{C}\mathrm{ov}\left[y_{t+1}-\bar{\mathbb{E}}_ty_{t+1},\bar{\mathbb{E}}_ty_{t+1}-\bar{\mathbb{E}}_{t-1}y_{t+1}\right]$$

We once more start with the covariance for time $t=1$, which then readily extends to later periods. Since the prior expectation is $\mathbb{E}_0^iy_2=0$, the average forecast revision is simply

$$\bar{\mathbb{E}}_1y_2-\bar{\mathbb{E}}_0y_2=\bar{\mathbb{E}}_1y_2=(\Sigma_ja_j)\,\bar{\mathbb{E}}_1\theta_1.$$

It now follows that the covariance determining the sign of δ is

$$\mathbb{C}\text{ov}\left[y_2 - \bar{\mathbb{E}}_1 y_2, \bar{\mathbb{E}}_1 y_2 - \bar{\mathbb{E}}_0 y_2\right] = \left(\Sigma_j a_j\right)^2 \mathbb{C}\text{ov}\left[\theta_1 - \bar{\mathbb{E}}_1 \theta_1, \bar{\mathbb{E}}_1 \theta_1\right],$$

so that it is sufficient to analyze the covariance between the average forecast error $\theta_1 - \bar{\mathbb{E}}_1\theta_1$ about the fundamental state θ_1, and the forecast revision $\left(\Sigma_j a_j\right) \bar{\mathbb{E}}_1\theta_1$.

Averaging the updating equation (3.6) , we obtain

$$\begin{aligned}\left(\sum_{j=1}^{N} \tau_j + \tau_\theta\right)^2 \times \mathbb{C}\text{ov}\left[\theta_1 - \bar{\mathbb{E}}_1\theta_1, \bar{\mathbb{E}}_1\theta_1\right] &= \mathbb{C}\text{ov}\left[\sum_{j=1}^{N} \tau_j \left(\theta_1 + \frac{b_j}{a_j} u_{jt}\right), \tau_\theta \theta_1 - \sum_{j=1}^{N} \tau_j \frac{b_j}{a_j} u_{jt}\right] \\ &= \sum_{j=1}^{N} \tau_j (1 - m_j)\end{aligned}$$

which is positive as soon as $0 < m_j < 1$ for at least one factor j. Repeating these steps for later time periods yields:

Lemma 2. *The regression (2.2) of ex-post mean forecast errors of y_{t+1} on the mean ex-ante forecast revisions suggests under-responsiveness ($\delta > 0$) if and only if there is at least one factor x_{jt} where $0 < m_j < 1$.*

Lemma 2 states that the average data exhibit under-responsiveness as soon as agents pay imperfect but positive attention to economic factors. The reason is that each person down-weights her noisy information, but on average the noise terms cancel across agents, leading to an apparent under-response to new information.

Crucially, and in contrast to Coibion and Gorodnichenko (2015), our model allows under-responsiveness and extrapolation to coexist, as long as attention is imperfect and satisfies Condition (3.7). Before moving on, we make this point more explicit by contrasting our results with existing theories of expectation formation, which do not consider asymmetric attention.

Comparison to Representativeness Bias: A substantial debate since Goodwin (1947) has emphasized extrapolation and related it to biased beliefs, in particular, to systematic overconfidence in the extent to which future resembles the present. This has more recently been linked to Kahnemann and Tversky's (1972) more fundamental *representativeness heuristic* (Bordalo *et al.*, 2012 and Bordalo *et al.*, 2017). We now consider a simple version of representativeness bias in our framework. Of course, the literature on behavioral expectations is much richer than this example, and extrapolation can be generated from careful psychological foundations. However, our example illustrates a common thread, namely that it is difficult to reconcile simultaneous over- and under-responses.

In our model, output can be written in reduced form as:

$$y_t = \left(\sum_{j=1}^{N} a_j\right)\theta_t + \zeta_t,$$

where ζ_t cumulates the relevant *i.i.d.* components. In order to explain extrapolation, we could specify a behavioral model where agents observe y_t, but overstate the persistence of the true driving force θ_t. Suppose that forecasters act *as if* the autoregressive coefficient in θ_t was $\kappa \in (\rho, 1)$, instead of the true coefficient $\rho < \kappa$. It is easy to show that, since forecasters' bias is inherently extrapolative, the estimated coefficient on current output in (2.1) is negative ($\gamma < 0$). The coefficient on forecast revisions is in this case

$$\delta \overset{\text{sign}}{=} \text{Cov}\left[y_{t+1} - \hat{\mathbb{E}}_t y_{t+1}, \hat{\mathbb{E}}_t y_{t+1} - \hat{\mathbb{E}}_{t-1} y_{t+1}\right] = \left(\sum_{j=1}^{N} a_j\right)^2 (\rho - \kappa)\kappa(1 - \rho\kappa)\text{Var}\left[\theta_t\right] < 0, \quad (3.8)$$

where $\hat{\mathbb{E}}_t\left[y_{t+1}\right]$ denotes the extrapolative agents forecast of y_{t+1} based period t information. Equation (3.8) is inconsistent with the empirical evidence from Section 2. Intuitively, behavioral agents over-respond to all information, so that their forecast revisions are systematically too large as opposed to too small.

Comparison to Symmetric Inattention: Recently, an influential literature (e.g. Andrade and Le Bihan, 2013; Coibion and Gorodnichenko, 2015) has argued that under-responses to the *ex-ante* mean forecast revisions are consistent with information frictions, caused by either rational inattention (Sims, 1998, 2003a) or dispersed information (Lucas, 1972). Our model develops this approach by allowing attention to be asymmetric across the underlying components of output.

To highlight why a model of asymmetric attention is needed, consider an simpler model with imperfect information, where rational agents simply observe a noisy signal of output

$$z_{it} = y_t + \epsilon_{it}$$

In this case, attention is effectively symmetric across component, since only a signal of their sum (output) is observed. This model easily generates under-responses, that is $\delta > 0$ in regression (2.2), since agents down-weight their own noisy information. However, this the coefficient on current output in regression (2.1) becomes

$$\gamma \overset{\text{sign}}{=} \rho\hat{m}\text{Var}\left[\epsilon_{it}\right] > 0$$

where $\hat{m} = \frac{\text{Cov}[y_{t+1}, z_{it}]}{\text{Var}[z_{it}]}$ measures attention to output (i.e. the Kalman gain in the agent's signal extraction problem). Agents thus under-respond also to past output, because they internalize that their signals are noisy. Like the behavioral model, this model yields coefficients δ and γ that have the same sign, which is inconsistent with the evidence in Figure 2.2.

We further note that the inconsistency cannot be resolved by simply combining these two examples does not solve the problem. If we add imperfect observations of y_t to the case with representativeness bias, we still obtain $\delta < 0$, unless this imperfection is large enough to dominate the strictly negative coefficient.

This comparison to alternative models highlights the importance of asymmetric attention, and the additional insights that our model delivers. In particular, asymmetric attention to structural components can generate simultaneous under-responses ($\delta > 0$) and under-responses ($\gamma < 0$) that are consistent with the stylized facts and do not arise in a reduced form treatment.

3.2 A General Forecasting Problem

Consider a linear economy with the structural equations:

$$
\begin{aligned}
y_t &= D\theta_t + Ex_t + Fu_t \\
x_t &= A\theta_t + Bx_{t-1} + Cu_t \\
\theta_t &= P\theta_{t-1} + \Sigma u_t,
\end{aligned}
$$

where θ_t is an $n_\theta \times 1$ vector, x_t is an $n_x \times 1$ vector, u_t is a $n_u \times 1$ vector of *i.i.d.* standard normal random variables, and y_t is a scalar variable. Most linear DSGE models can be written in this form (see Fernández-Villaverde *et al.*, 2007). Each forecaster receives a signal

$$
z_{it} = x_t + Q\epsilon_{it}
$$

where ϵ_{it} is an $n_x \times 1$ vector of standard normal random variables which are uncorrelated with u_t. It is useful to re-write this system using following compact notation,

$$
\begin{aligned}
\bar{\theta}_t &= \begin{bmatrix} \theta_t' & x_t' \end{bmatrix}' = \bar{P}\bar{\theta}_{t-1} + \bar{\Sigma}u_t \\
y_t &= \alpha\bar{\theta}_t + \beta u_t,
\end{aligned}
$$

where $\bar{P} = \begin{bmatrix} P & \underline{0} \\ AP & B \end{bmatrix}$, $\bar{\Sigma} = \begin{bmatrix} \Sigma \\ A\Sigma + C \end{bmatrix}$, and $\alpha = \begin{bmatrix} D & E \end{bmatrix}$, and where each agent $i \in [0,\, 1]$ now receives the noisy signal,

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

with L_0, L_1 and R implicitly defined. We can then extend the previous results which characterize necessary and sufficient conditions for a simultaneous over- and under-response to new information. Proposition 1 summarizes the conditions.

Proposition 1. *The population coefficients in the regression equations (2.1) and (2.2) satisfy, in the general case,*

$$\gamma \overset{sign}{=} \alpha \bar{P} \left\{ K Q Q' E' + \Sigma_{\theta\bar{\theta}} D' + \left[\bar{\Sigma} - K \left(L_0 \bar{\Sigma} + M \right) \right] \right\}, \tag{3.9}$$

$$\delta \overset{sign}{=} \alpha \left(\bar{G} - G \right) \mathbb{V} \left[x_t - \bar{\mathbb{E}}_{t-1} x_t \right] G' \alpha' \tag{3.10}$$

where $\Sigma_{\theta\bar{\theta}} = \mathbb{C}ov\left(\theta_t,\, \bar{\theta}_t\right)$, $G = PK$ is the Kalman gain on z_t^i in agents' expectation about $\bar{\theta}_{t+1}$, and $\bar{G}$ is the hypothetical Kalman gain if agents instead observed the full history of x_t.

Similar to Lemma 2, expectations are generically under-responsive in Proposition 1, $\delta > 0$. The hypothetical weight that agents attach the observation of x_t exceeds that attached to the signal vector z_{it} because of the private noise component ϵ_{it}, $\left|\bar{G}\right| > |G|$. Likewise, informative, counter-cyclical components – that is, those that are assigned a large weight in K and negative elements in E – push expectations towards extrapolation. But unlike in Lemma 2, we now also have to adjust the condition for extrapolation for (i) the direct impact of the persistent fundamental on output ($D > 0$), and (ii) for the cross-correlation in errors between the signal vector and output ($R \neq \underline{0}$). The second and third component of (3.9) can be shown to illustrate these additional terms (see also Section 5).

In summary, we have established that asymmetric attention may be able to reconcile the stylized facts in a large class of models. We now evaluate whether such asymmetries can be micro-founded using models of optimal information choice.

4 Optimal Information Choice

In this Section, we consider rational agents' optimal choices when attention is costly. We show that attention costs paired with rational choices can easily generate asymmetric attention, and this effect becomes even stronger when agents' utility depends asymmetrically on the underlying components of the economy. In Section 4.2, we present some supplementary evidence which suggests that optimal information choice may be a reasonable framework for explaining the expectations data.

4.1 Optimal Costly Attention

Consider an agent who wishes to choose an action a_t each period. Her action yields von-Neumann-Morgenstern utility

$$U(a_t, x_t) = -\frac{1}{2}\left(a_t - \sum_{j=1}^{m} w_j x_{jt}\right)^2$$

so that the ideal action $a_t^\star = \sum_j w_j x_{jt}$ is a linear combination of the underlying factors driving the economy. This nests the case where the agent wants to take actions directly tracking the aggregate variable y_t (this corresponds to $w_j \equiv 1$).

The agent chooses the vector m of attention levels to each factor, and observes the associated vector of signals z_t, before choosing an action. Attention is chosen to maximize ex ante expected utility, and incurs a utility cost $\lambda K(m)$, for which we consider various cases below.

To economize on notation, we focus on the problem of an agent who lives for only one period $t = 1$. The properties of her choice map directly to the steady-state choices of an infinitely lived agent who faces static optimization problem in the presence of dynamic fundamentals. This case arises naturally in many macroeconomic applications of inattention (e.g. Maćkowiak and Wiederholt, 2009), as well as in our general equilibrium model below.

Case 1: Fixed Costs of Model Complexity: Popular model selection criteria, such as the Akaike and Schwarz information, penalize complex models according to the number of variables that feature in estimation. In our context, the corresponding attention cost is the counting norm $K(m) = \sum_{j=1}^{p} \mathbf{1}\{m_j > 0\}$. Here, the agent effectively pays a fixed cost for every variable that enters her attention.

Fixed costs induce optimally sparse representations, where some components are omitted altogether with $m_j = 0$. For our purposes, a model with $N = 2$ components suffices to convey the important ideas.[5] Depending on the marginal cost λ of attention, the agent will pay attention to either zero, one or all two factors. In the interesting case where she pays

[5]The general problem with fixed-cost penalties is that the statistician must estimate 2^N models, one with every combination of factors, to select the best subset of factors to include. One quickly runs into the *curse of dimensionality*. However, an iterative algorithm such as forward or backward selection could easily be obtained in general by using inequalities such as (4.1) below.

attention to only one, it is easy to show that she chooses x_{1t} if and only if (see Appendix B)

$$\frac{w_1}{w_2} > \frac{\text{St.dev.}[x_2]}{\text{St.dev.}[x_1]} \tag{4.1}$$

Asymmetric attention is paid to x_{1t} if it is relatively important in the utility function (high w_1) or relatively volatile (high standard deviation of x_{1t}). By Lemma 1, the agent's output expectations exhibit extrapolation whenever (4.1) holds and the omitted component x_{2t} is countercyclical. Moreover, as long as there is an upper bound on the amount of attention the agent can pay to her preferred factor, so that m_j is strictly below 1, this model also generates under-responses by Lemma 2. For brevity, we do not introduce such an upper bound explicitly, but it is easy to micro-found it by assuming that there is irreducible background noise in the agent's observations, as in the case with segmented markets discussed above.

Importantly, asymmetric utility weights are sufficient but not necessary for asymmetric attention in the fixed cost case. Indeed, even if the agent aims to track total output ($w_1 = w_2 = 1$), she chooses to pay attention to x_{1t} only as long as this component is more volatile. This is a general feature of sparsity-inducing cost functions, including fixed costs or the *Lasso Penalty* (see Gabaix, 2014).

Case 2: Entropy Costs with Orthogonal Signals: A popular alternative since Sims (2003b) is to assume that the costs of attention are proportional to the reduction in entropy, a measure of uncertainty, that is brought about by observing signals z. In the context of our model, this amounts to defining the cost function

$$\begin{aligned} K(m) &= I(z; \{x, \theta\}) \\ &= \frac{1}{2}\log\left[\tau_\theta + \sum \tau_j\right] + \frac{1}{2}\sum_j \log\frac{1}{1-m_j} + \text{constant} \end{aligned} \tag{4.2}$$

where $I(z; \{x, \theta\})$ denotes the Shannon information, that is, the reduction in the entropy of state variables $\{x_t, \theta_t\}$ when signals z_t are observed. We refer the reader to Appendix B.2 for formal definitions and derivations, but the intuition behind (4.2) is clear. Indeed, there are two parts to the entropy reduction: On one hand, entropy is reduced by learning about the fundamental θ_t, whose posterior precision is $\tau_\theta + \sum \tau_j$. Second, uncertainty is reduced by learning about the idiosyncratic noise in specific components x_{jt} , and the rate of learning increases in the attention parameter m_j.

The optimal attention choice satisfies the first-order condition

$$w_j^2 b_j^2 + \mu_\tau \frac{a_j^2}{b_j^2} + \mu_\alpha w_j a_j = \lambda \frac{1}{1-m_j} \tag{4.3}$$

where μ_τ and μ_α are Lagrange multipliers that do not depend on j. Unlike in rational inattention models with uncorrelated components, there are learning spillovers. Learning about x_{jt} resolves uncertainty about that component, for example, but also resolves uncertainty about the fundamental θ_t, which in turn reduces the residual variance of x_{kt} for $k \neq j$. The multipliers μ_τ and μ_α measure the shadow value of increasing the posterior precision of θ_t, and of reducing the residual uncertainty about the ideal action $\sum w_j x_j$ that is attributable to θ_t.

Equation (4.3) shows that attention m_j tends to increase with the utility weight w_j, as well as with the rate of exchange a_j^2/b_j^2 at which attention to x_{jt} spills over into information about θ_t. Indeed, since the costs of inattention to each x_{jt} are convex, the agent optimally pays attention even to components with $w_j = 0$, which play no role in her decision, so as to better learn about fundamentals.

Furthermore, (4.3) reveals that attention will be imperfect, with $m_j < 1$ for all j. Combining this with Lemma 2 establishes that expectations are under-responsive. Moreover, combining these insights with Lemma 1, we learn that expectations are extrapolative if *(i)* utility weights are sufficiently skewed towards procyclical components, and *(ii)* procyclical components do not have too much of a disadvantage at generating spillovers. In practice, certain procyclical components such as investment or asset prices move more strongly with the overall business cycle, so that the latter requirement is likely to be satisfied.

Case 3: Entropy Costs with Unconstrained Signals: In Case 2, we have imposed on the agent that signals z take the shape in (3.4). As discussed above, this is not optimal when the agent has complete freedom to pick the conditional density $p(z|\{x,\theta\})$ of her information. In this case, with the entropy-based penalty $K(m) = I(z;\{x,\theta\})$, it is well known (see, for instance, Sims, 2010) that the optimal z consists of a single noisy signal of the optimal action:

$$z_{it} = \sum w_j x_{jt} + \epsilon_{it} \tag{4.4}$$

where the variance of ϵ^i is chosen to equalize the marginal (entropy) cost with the benefit of a more accurate action.

We omit a full analysis of expectations in this case. However, we note that this specification generates asymmetric attention even more naturally than the constrained case, because learning about different components is entirely determined by the agent's utility weights

w_j. Indeed, consider the limiting case where the agent's utility depends only on tracking one component x_{1t}, that is, $w_j = 0$ for all $j \geq 2$. Then the signal in (4.4) corresponds to (3.4) with $m_j = 0$ for $j \geq 2$, while $0 < m_1 < 1$. If components $j \geq 2$ are jointly countercyclical, our results imply that the agent's expectations will be extrapolative as well as under-responsive. By continuity, asymmetric attention can reconcile over- and under-responses in the unconstrained case as long as procyclical components have sufficiently large utility weights.

The first two cases in this Subsection highlight how attention gravitates to components of output that are most salient, i.e. most strongly correlated with the fundamental state θ_t. In these cases, moreover, asymmetric attention results even when utility weights are symmetric, as long as components are differentially salient. In addition, agents pay more attention to components with higher utility weights. This results in asymmetric attention even in the third case, where agents have complete freedom to choose signal distributions.

4.2 Are Attention Choices Optimal? Supplementary Evidence

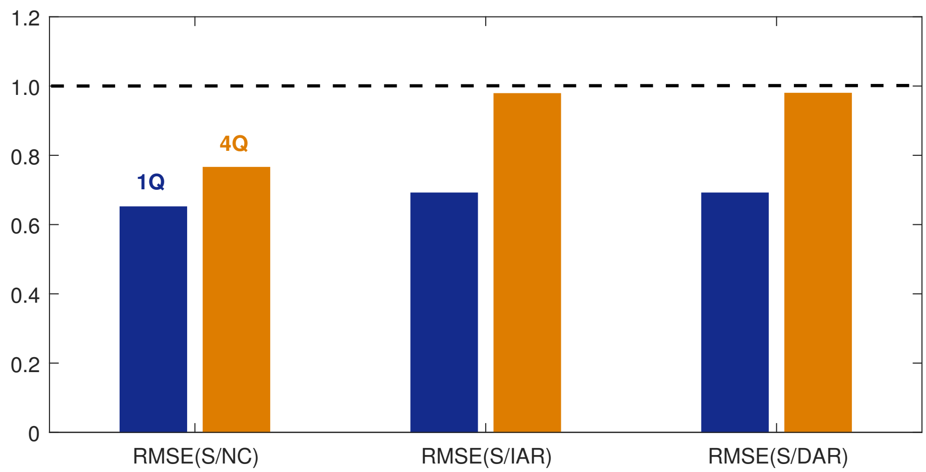
We now return to the data in order to compare the quality of forecasters' expectations to that of standard time series models. Figure 4.1 shows updated values from Stark (2010), available from the *Federal Reserve Bank of Philadelphia's* website.[6] The chart illustrates the *relative root mean-squared error* of one-quarter and four-quarter ahead forecasts of output growth from US *SPF* relative to three time-series models. A *RRMSE* ratio below unity indicates that the *SPF* consensus forecast is more accurate. *NC* denotes a Random Walk forecast, *IAR* forecasts from an ARMA model chosen to minimize one-quarter ahead forecast errors, and *DAR* forecasts from ARMA models chosen to minimize forecast errors at each forecast horizon. The sample period is 1985Q1:2015Q2.

The dashed line indicates parity between time series models and *SPF* forecasts. 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. Our unreported calculations indicate that the relative RMSE is statistically different from one at a one percent level for all one-quarter forecasts, while the four-quarter ahead forecasts for *NC* is statistically different at the five percent level.

This supplementary evidence suggests that forecasters do better than time series models at forecasting output. As discussed in the Introduction, this is consistent with our model, where agents pay attention to underlying components of the economy, and inconsistent with a model where agents consider only the past time series of output levels. Moreover, the

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

Figure 4.1: 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 which are described in the main text. A *RRMSE* ratio below unity indicates that the *SPF* consensus forecast is more accurate. The sample period is 1985Q1:2015Q2.

strong performance of *SPF* forecasts suggests that inattention is unlikely to be driven by strong behavioral bias, but is instead consistent with the models of optimal attention choice developed in this Section.

5 Asymmetric Attention in a Business Cycle Model

We have used an abstract forecasting problem to show that asymmetric attention choices can rationalize the simultaneous over- and under-responses to new information that we observe in economy-wide forecast data. We now demonstrate how these results extend to a workhorse business cycle model. We develop a simple explanation of asymmetric attention, in the spirit of Lucas (1972) and Maćkowiak and Wiederholt (2009): Firms may optimally choose to pay relatively more attention to local components, as opposed to economy-wide ones that arise from general equilibrium considerations. This model, combined with benchmark parameter values, naturally leads to extrapolative beliefs. We then show how this model is also consistent with *(i)* the precision of agents' forecasts consistently beating those from simple time-series models, and *(ii)* an inherent over-confidence in future consumption and income at the start of recessions, consistent with the empirical literature (see, for instance, Attanasio, 1999).

5.1 A Flexible Price Model

With the exception of the introduction of information choice, we base the analysis on a modified version of the model described in Angeletos and LaO (2010).

The economy consists of a representative household, a continuum of representative, monopolistically competitive firms $i \in [0, 1]$, which specialize in the production of differentiated goods, and a tax authority. Each period is comprised of two stages. In the first stage, firms pre-set their output choices. At this stage, firms receive imperfect information about aggregate productivity and subsequent taxes, and hence have imperfect information about the supply of goods from other firms. After output choices are sunk, the economy transitions to the second stage. The household now meets with firms to produce previous output choices. The wage adjust to clear the labor market. Goods markets open, goods prices adjust, and the household consumes.

Households: A representative household has preferences,

$$\mathcal{U} = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log\left(C_t\right) - \frac{1}{1+\eta} N_t^{1+\eta} \right], \tag{5.1}$$

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 η parametrizes the Frisch elasticity of labor supply. The consumption index and the associated welfare-based price index are

$$C_t = \left(\int_0^1 C_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}}, \quad P_t = \left(\int_0^1 P_{it}^{1-\sigma} di \right)^{\frac{1}{1-\sigma}}, \tag{5.2}$$

where C_{it} is the amount the household consumes of goods produced by firm $i \in [0, 1]$, with $\sigma > 1$, and P_{it} the price set by firm i.

Since the household receives all labor income and profits, its per-period budget constraint is

$$\int_0^1 P_{it} C_{it} di + B_{t+1} \leq \int_0^1 \Pi_{it} di + W_t N_t + (1+R_t) B_{t+1} + T_t^h, \tag{5.3}$$

where Π_{it} denotes the profits of firm $i \in [0, 1]$, W_t the nominal wage, R_t the net nominal rate of return on riskless bonds, B_t the amount of riskless bonds held, and T_t^h lump-sum nominal transfers. The representative household's seeks to maximize its utility (5.1) subject to (5.3).

Firms: A representative firm $i \in [0, 1]$ produces output according to

$$Y_{it} = X_t N_{it},$$

where N_{it} denotes the amount of labor input used and X_t the (common) level of total factor productivity. Firm productivity is comprised of two separate components,

$$x_t = \theta_t + u_t^x, \quad \theta_t = \rho\theta_{t-1} + u_t^\theta,$$

where lower-case letters denote logs of their upper-case counterparts, θ_t the persistent component of total factor productivity, and u_t^x a purely transitory productivity shock. This is consistent with the decomposition of total factor productivity used in Blanchard *et al.* (2013), for example. Both $u_t^x \sim \mathcal{N}(0, 1/\tau_x)$ and $u_t^\theta \sim \mathcal{N}(0, 1/\tau_\theta)$ are assumed independent of each other and all other stochastic disturbances. The representative firm seeks to set its labor input choice to maximize its own expectation of the household's valuation of its profits, using the stochastic discount factor $1/(P_t C_t)$. Profits at time t are given by

$$\Pi_{it} = P_{it} Y_{it} - (1 + T_t^w) W_t N_{it}, \tag{5.4}$$

where the demand for a firm's product is consistent with the definition of the consumption basket in (5.2) and where

$$\mathcal{T}_t^w = \log T_t^s = \phi\theta_t + u_t^{\mathcal{T}}, \quad \phi > 0 \tag{5.5}$$

denotes an *economy-wide pay-roll tax* that is rebated lump-sum to households. The disturbance $u_t^{\mathcal{T}} \sim \mathcal{N}(0, 1/\tau_\theta)$ here denotes an economy-wide tax shock.

5.2 Equilibrium, Information, and Output Choices

We define an equilibrium in familiar manner as a sequence of prices, output choices, household labor supply, firm labor demand, and wage rates, such that at each point in time: (i) firms maximize profits and the representative household maximizes utility subject to constraints, including informational, and (ii) all goods market clear $Y_{it} = C_{it}$ for all $i \in [0, 1]$ and t, and *(3)* so too does the labor and bond markets. We focus below on the determination of firm output choices, since these are directly affected by imperfect attention. The remaining quantities and prices are straightforward to compute (see Appendix C).

Firm Output Choices: We follow the same steps as in Angeletos and LaO (2010) to show that a firm's output choice is pinned down by the fixed-point relation

$$\mathbb{E}_{it}\left[\left(\frac{Y_{it}}{X_t}\right)^\eta\right] = \mathcal{M}^{-1}\mathbb{E}_{it}\left[X_t (1 - T_t^w) Y_{it}^{-\frac{1}{\sigma}} Y_t^{\frac{1}{\sigma}-1}\right], \tag{5.6}$$

where $\mathbb{E}_{it}[\cdot] = \mathbb{E}[\cdot \mid \Omega_{it}]$ denotes firm i's expectation based upon its information set Ω_{it} (defined below), $\mathcal{M} = \frac{\sigma}{\sigma-1}$ the standard wedge caused by the presence of monopolistic competition, and $Y_t = C_t$ aggregate output.

Lemma 3. *An individual firm's output choice satisfies*

$$y_{it} = \mathbb{E}_{it}\left[rx_t + (1-r)y_t - \mathcal{T}_t^w\right], \tag{5.7}$$

where $\bar{\mathbb{E}}_t[\cdot] = \int_0^1 \mathbb{E}_{it}[\cdot]\, di$, $r = \frac{(1+\eta)\sigma}{1+\eta\sigma} > 1$ *and* $\mathcal{T}_t^w = \log T_t^w$.

Lemma 3 connects individual firm choices to the optimal actions studied in the large class of beauty-contest models analyzed by, for instance, Morris and Shin (2002) and Angeletos and Pavan (2007). Similar to those models, optimal output choices are a weighted average of firms' expectations about unobserved fundamentals and firms' expectations about the average choice in the economy. Unlike in those models, however, the unobserved fundamentals are dynamic. Equation (5.7) thus closely mirrors the price-setting models proposed by Woodford (2002) and Maćkowiak and Wiederholt (2009), in which firm choices are static but underlying fundamentals dynamic.

Endogenous Information and Attention: Combining (5.6) and Lemma 3 provides a useful characterization of a representative firm $i \in [0, 1]$'s problem, which will also be useful later to solve a firm's optimal choice of attention. Specifically, after a few simple derivations, we can show that (see Appendix C)

$$\mathbb{E}_{it}\left[\frac{1}{P_t C_t}\Pi_{it}\right] \propto \mathbb{E}_{it}\left[y_{it} - y_t^\star\right]^2, \tag{5.8}$$

where $y_t^\star$ denotes a firm's output choice under full information,

$$y_t^\star = rx_t + (1-r)y_t - \mathcal{T}_t^s = rx_t + \Delta_t. \tag{5.9}$$

We can therefore assess how well a firm maximizes profits by how well-aligned its output choice is to its full information counterpart.

We use the following convenient decomposition, similar to that proposed in Maćkowiak and Wiederholt (2009): We define *local conditions* as those that *directly* affect a firm's output choice. In (5.7), these simply equal a firm's own productivity, x_t. By contrast, we define *economy-wide conditions* as those that affect a firm's output choice *indirectly*, for instance through the demand for the firm's products in general equilibrium. We denote the economy-wide components in (5.7) as $\Delta_t = (1-r)y_t - \mathcal{T}_t^s$. We can then use (5.7) to write a firm's output choice as

$$y_{it} = \mathbb{E}_{it}\left[rx_t + \Delta_t\right]. \tag{5.10}$$

We assume that each firm observes signals of both local and economy-wide components of the form "truth plus white noise", either because of firm-specific productivity and market-segmentation or inattention,

$$z^x_{it} = x_t + q_x \epsilon^x_{it}, \quad z^\Delta_{it} = \Delta_t + q_\Delta \epsilon^\Delta_{it}, \tag{5.11}$$

where $\{q_x,\, q_\Delta\} \in \mathbb{R}_+$ and both ϵ^x_{it} and ϵ^x_{it} are standard white noise normal and independent of all other stochastic disturbances. A firm's information set is therefore, in sum equal to

$$\Omega_{it} = \left\{z^x_{it-k},\ z^\Delta_{it-k}\right\}_{k=-\infty}^{k=0}. \tag{5.12}$$

5.3 Analytical Solutions with Perfect State Verification

We now characterize equilibrium output in the economy. We will then use our results from Section 3 to detail the properties of firms' expectation about current and future output.

To start, we use Lemma 3 to derive a simple expression for aggregate output

$$y_t = \bar{\mathbb{E}}_t\left[rx_t + (1-r)y_t - \mathcal{T}^w_t\right] = \bar{\mathbb{E}}_t \sum_{i=0}^{\infty} (1-r)^i \bar{\mathbb{E}}^i_t\left[rx_t - \mathcal{T}^w_t\right], \tag{5.13}$$

where $\bar{\mathbb{E}}^j\left[\cdot\right] = \bar{\mathbb{E}}\left[\bar{\mathbb{E}}^{j-1}\left[\cdot\right]\right] \neq \bar{\mathbb{E}}\left[\cdot\right]$ for $j > 1$ since the *Law of Iterated Expectations* does not for hold the average expectation operator (see, for instance Morris and Shin, 2002). Combining this expression with that from Lemma 3 then shows that,

$$y_{it} = \mathbb{E}_{it}\left[rx_t - \mathcal{T}^w_t\right] + (1-r)\mathbb{E}_{it}\left[\bar{\mathbb{E}}_t \sum_{i=0}^{\infty} (1-r)^i \bar{\mathbb{E}}^i_t\left[rx_t - \mathcal{T}^w_t\right]\right]. \tag{5.14}$$

An individuals firm's output choice depends on its own expectation about the entire infinite hierarchy of expectations about $rx_t - \mathcal{T}^w_t$. Without further assumptions, this problem does not admit a known finite state-space representation (see, for example, Townsend, 1983 and Nimark, 2017). Below, we circumvent this problem by first augmenting firms' information sets with the public signal $s_t = \theta_{t-1}$ that *perfectly* reveals last period's realization of the persistent productivity shock. The assumption of *one-period perfect state verification* here collapses higher-order expectations to first-order expectations, and hence permits us to derive an analytical solution to (5.14). The next subsection, by contrast, finds an arbitrarily precise approximate solution without need for perfect state verification.

The analytical solutions below allows us to map our micro-founded model onto a two-component version of our prediction problem in Section 3. Because of the strategic substitutability between individual firm output choices in (5.13), however, other firms' information choice will now matter for an individual firm's information acquisition. Hellwig and Veldkamp (2009) and Colombo *et al.* (2014) study how the incentives to acquire private information depend upon the extent of strategic complementarities between individual players.

We restrict ourself to symmetric linear Bayesian equilibria, in accordance with the literature on noisy rational expectations. The standard approach to find linear equilibria in models with endogenous signals is the *method of undetermined coefficients* (see, for instance, Amador and Weill, 2010). Applying this approach results in Lemma 4.

Lemma 4. *Aggregate output in the economy is uniquely determined by*

$$y_t = d_0 x_t + d_1 \Delta_t, \tag{5.15}$$

where a_0, $a_1 > 0$*, and in which* $b_\theta < 0$ *in*

$$x_t = \theta_t + u_t^x, \quad \Delta_t = a_\theta \theta_t + c_x u_t^x + c_{\mathcal{T}} u_t^{\mathcal{T}}. \tag{5.16}$$

Lemma 4 provides a micro-founded, two-component example of the setup used in Sections 3 and 2: It shows how firm productivity contributes positively to output in equilibrium, while the economy-wide component through its general equilibrium effect on prices and wages dampens the increase.

We can now use the results from Section 3 to show that:

Proposition 2. *Firms' mean one-period ahead aggregate output forecasts exhibit extrapolation (*$\gamma < 0$*) in equilibrium when* q_x/q_Δ *is sufficiently small. Moreover, firms' forecasts exhibit under-responsiveness (*$\delta > 0$ *) for all* $\{q_x, q_\Delta\} \in \mathbb{R}_+^2$*.*

The conditions in Proposition 2 are sharp. First, because of imperfect information firms under-respond to the *average* new information received between period $t-1$ and t. Second, firms output forecasts can also exhibit extrapolation, in the sense that past realizations of output are negatively correlated with future forecast errors. This occurs whenever firms predominantly attend to their own local conditions. Coibion *et al.* (2018) provide convincing evidence that firms focus on local conditions rather than economy-wide variables.

As in Section 4, we can micro-found asymmetric attention choices by considering how firms optimally allocate their costly attention to local *versus* economy-wide components. This is similar to the rationale provided by Maćkowiak and Wiederholt (2009) for why firms mainly attend to their own productivity level when setting prices rather than the aggregate price level. Proposition 3 provides a simple example with an entropy based cost function for attention, using the restatement of the firm's problem in (5.8) and (5.9).

Proposition 3. *The optimal q_x/q_Δ is small with an entropy cost of attention whenever a_θ and c_x are small in absolute magnitude.*

Combined Proposition 2 and 2 rationalize forecasts that appear extrapolative. They show why a firm that optimally chooses its own information to best determine its own output can have forecast errors of economy-wide output that correlated with previous realizations. This, in turn, provides a clear example of how the combination of limited attention and asymmetric correlations between components of output and fundamentals can help rationalize the stylized facts from Section 2.

5.4 A Quantitative Exploration

We now return to the model without perfect state verification. Unlike in Subsection 3.3, the equilibrium dynamics for output can no longer be derived analytically. The entire infinite hierarchy of beliefs will now matter for equilibrium firm choices. We therefore solve the model numerically. Specifically, we employ the sequential method of undetermined coefficients technique developed in Nimark (2017) in Appendix C, to find an arbitrarily precise approximate equilibrium solution for the dynamics of output.

Lemma 5. *The equilibrium dynamics of output approximately follow*

$$y_t \simeq d_0 x_t + d_1 \Delta_t = \alpha \bar{\theta}_t^{(0:\bar{k})}, \quad \bar{\theta}_t = \begin{bmatrix} \theta_t & u_t^x & u_t^{\mathcal{T}} \end{bmatrix}', \tag{5.17}$$

where $\bar{\theta}_t^{(j)} = \bar{\mathbb{E}}_t \left[\bar{\theta}_t^{(j-1)} \right]$ for $j \geq 1$ and

$$\bar{\theta}_t^{(0:\bar{k})} = P \bar{\theta}_{t-1}^{(0:\bar{k})} + \Sigma \begin{bmatrix} u_t^\theta & u_t^x & u_t^{\mathcal{T}} \end{bmatrix}'. \tag{5.18}$$

We set the number of higher-order expectations in the approximation to $\bar{k} = 11$. Numerical simulations indicate that the equilibrium dynamics of output are stable already at $\bar{k} = 6$, consistent with the results in Nimark (2014).

Numerical Simulations: We now explore the quantitative implications of the model. The model is too stylized for a full quantitative investigation, so we only take a first pass at some basic quantitative questions: Can the model create similar amounts of extrapolation and under-responsiveness to new information to those that we saw for output growth and

Table 5.1: Over- and Under-responses in the Business Cycle Model

	Forecast Error	*Forecast Error*	*Forecast Error*
Constant	0.00	0.00	0.00
	(–)	(–)	(–)
Realization (t-1)	-0.05	–	-0.29
	(–)		(–)
Forecast Revision	–	0.46	1.85
		(–)	(–)
Sample	10^6	10^6	10^6
Relative RMSE (S/AR(1))		0.95	

(i) Results are for simulated 1Q ahead estimates.

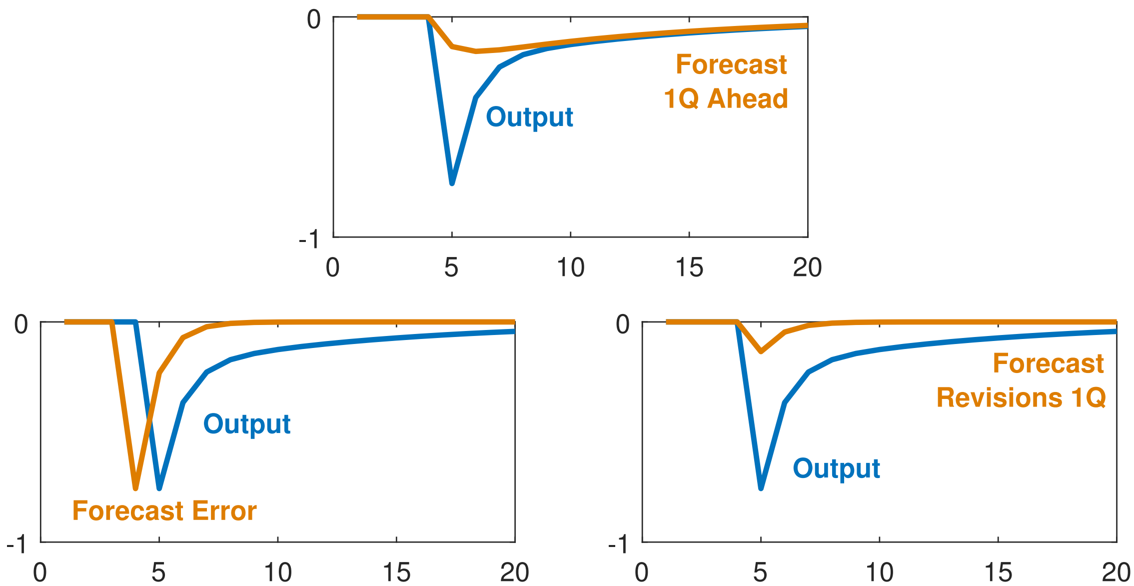
inflation in Section 2? And if so, are firms still able to forecast economy-wide output better than a simple ARMA model?

To tackle these questions, we will first parameterize the model and compare estimates of (2.1) and (2.2) to those from the data, using a simulation of length $T = 1,000,000$. We set the relative degree of risk-aversion $\sigma = 1$, and the inverse of Frisch elasticity $\eta = 2$. The persistence of the unobserved fundamental θ_t is set to 0.90 and the standard deviation of shocks to it $\frac{1}{\sqrt{\tau_\theta}} = 1$. The standard deviation of the transitory component of productivity is set to $\frac{1}{\sqrt{\tau_x}} = 1.5$. Both standard deviations are similar to the estimates used in Blanchard *et al.* (2013). We set $\phi = 1.5$ and $\frac{1}{\sqrt{\tau_\mathcal{T}}} = 0.75$, such that taxes are pro-cyclical and the standard deviation of transitory tax shocks is small, consistent with evidence from the NBER.[7] We still need to determine q_x and q_Δ. We do this by matching the implied γ and δ estimates from the simulation as closely as possible to -0.15 and 0.50, respectively, consistent with the benchmark results from Section 2. Table 5.1 summarizes the results.

The model matches the under-responsiveness estimates from the data well, and also matches how those estimates increase once we condition on last period's realization. Although the model also closely matches the conditional extrapolation estimates from Section 2, it can only account for about 1/3 of the unconditional extrapolation. The reason is as follows: Since firms predominantly pay attention to productivity x_t in the benchmark calibration, aggregate output chiefly reflects this component. The counter-cyclical offset provided by the economy-wide component Δ_t is therefore small. That is, b_θ is close to zero. This drives the γ estimate closer to zero. On balance though, the stylized model does remarkably well at capturing the salient features of the data – in particular, that firms' forecasts still

[7] http://users.nber.org/~taxsim/allyup/ally.html

Figure 5.1: Impulse Responses to a Persistent Productivity Shock



Impulse responses to a two standard deviation shock to the persistent component of productivity. The chart shows the responses of output, mean one-quarter ahead forecasts, forecast errors, and forecast revisions.

outperform those from a simple time-series model.

Last, we perturb the model with a two standard deviation persistent productivity shock and trace out the behavior of output, mean one-quarter ahead forecasts, and the associated forecast revisions. Figure 5.1 shows the results.

The impulse responses show that expectations are slow to adjust at the start of the downturn, and appear clearly extrapolative and over-optimistic. This, in turn, is consistent with the sizable empirical literature, following Attanasio (1999), which documents over-confidence in consumption and income expectations at the start of recessions. Besides this over-confidence in future output and hence consumption, Figure 5.1 visualizes how firms' asymmetric attention choices cause forecast errors to be clearly *negatively* associated with lagged realizations, but *positively* related to forecast revisions, consistent with the estimates in Table 5.1 and the survey data reported in Section 2.

6 Conclusion

We have documented a simultaneous over- and under-response of households', firms', and professional forecasters' expectations in economy-wide survey data. We have shown how such responses are inconsistent with benchmark models of extrapolation and rational inattention.

To resolve this friction, we have proposed a simple, components-based model of structural information choice that can reconcile the disparate facts.

We have characterized the properties of the resultant expectations, and shown that asymmetric attention choices can explain the co-existence of over- and under-responses. We have then embedded these expectations into a micro-founded macroeconomic model, in which firms choose to optimally attend more closely to local relative to economy-wide conditions. Firm expectations are found to line-up well with those from the survey data, and we have discussed how they help create momentum, persistence, and predictability in consumption changes. Finally, we have also confirmed several auxiliary implications from our model of information choice about the precision of agents' forecasts of future macroeconomic variables.

A Additional Empirical Results

Table A.1: Estimated over- and under-responses across surveys

	US SPF				EA SPF				LS Survey				MSC	
	Output		Inflation		Output		Inflation		Output		Inflation		Inflation	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Constant	-0.22*	-0.03	-0.09	-0.05	0.09	0.35**	0.11	0.18	0.09	0.20**	0.18	0.56***	-0.43***	-0.26**
	(0.12)	(0.02)	(0.08)	(0.08)	(0.16)	(0.15)	(0.10)	(0.13)	(0.09)	(0.09)	(0.12)	(0.11)	(0.11)	(0.12)
Realization (t-1)	-0.15**	–	-0.21**	–	-0.23***	–	-0.31*	–	-0.19***	–	-0.07	–	-0.47***	–
	(0.06)		(0.10)		(0.08)		(0.16)		(0.09)		(0.16)		(0.16)	
Forecast Revision	–	0.77***	–	1.14***	–	0.68***	–	0.70*	–	0.45***	–	1.91***	–	0.71***
		(0.26)		(0.28)		(0.25)		(0.40)		(0.16)		(0.24)		(0.26)
Sample	01/70:10/17				01/70:10/17				01/70:10/17				01/70:10/17	
F	6.50	15.24	7.20	42.6	5.17	13.5	4.56	3.64	4.67	11.5	0.31	67.13	19.3	13.6
R^2	0.04	0.08	0.04	0.19	0.07	0.17	0.06	0.05	0.05	0.12	<0.01	0.43	0.12	0.08

(i) Heteroskedasticity and autocorrelation consistent (HAC) standard errors used.

(ii) * p<0.10, ** p<0.05, *** p<0.01.

(iii) An HP-trend $\lambda = 1600/100$ for quarterly/semi-annual has been deducted from y_{t-1} and π_{t-1} to account for structural changes.

B Derivations for Section 4

B.1 Optimal Information Choice: Fixed Costs

If paying full attention to x_{1t} , action choice is

$$a = w_1 x_1 + w_2 \mathbb{E}[x_2|x_1]$$

so the utility loss is

$$\mathbb{E}(a - a^\star)^2 = w_2^2 \text{Var}[x_2|x_1]$$

By symmetry, choosing x_{1t} instead of x_{2t} minimizes loss if

$$\frac{w_1^2}{w_2^2} > \frac{\text{Var}[x_2|x_1]}{\text{Var}[x_1|x_2]}$$

To get Condition (4.1), note that the following is true of all Gaussian variables X, Y:

$$\frac{\text{Var}[X|Y]}{\text{Var}[Y|X]} = \frac{\text{Var}[X]}{\text{Var}[Y]}$$

since $\text{Var}[Y|X] = (1-\rho^2)\sigma_Y^2$ where $\rho = \text{Corr}[X, Y]$, and by symmetry, $\text{Var}[X|Y] = (1-\rho^2)\sigma_X^2$.

B.2 Optimal Information Choice: Entropy Costs with Orthogonal Signals

The agent's action choice is

$$a = w'\mathbb{E}[x|z]$$

so the utility loss is

$$\mathbb{E}\left(a - a^{\star}\right)^2 = w'\mathbb{E}\left[\left(x - \mathbb{E}[x|z]\right)\left(x - \mathbb{E}[x|z]\right)'\right] w$$

Apply iterated expectations to the matrix in the middle to get

$$\begin{aligned}
\mathbb{E}\left[\left(x - \mathbb{E}[x|z]\right)\left(x - \mathbb{E}[x|z]\right)'\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(x - \mathbb{E}[x|z]\right)\left(x - \mathbb{E}[x|z]\right)' |z\right]\right] \\
&= \mathbb{E}\left[\mathrm{Var}\left[x|z\right]\right] \\
&= \mathrm{Var}[x|z]
\end{aligned}$$

where the last line follows because x and z are jointly Gaussian, making $\mathrm{Var}[x|z]$ is deterministic.

By the law of total variance, applied conditional on z, we have

$$\begin{aligned}
\mathrm{Var}[x|z] &= \mathbb{E}\left[\mathrm{Var}\left[x|z,\theta\right]|z\right] + \mathrm{Var}\left[\mathbb{E}\left[x|z,\theta\right]|z\right] \\
&= \mathrm{Var}\left[x|z,\theta\right] + \mathrm{Var}\left[\mathbb{E}\left[x|z,\theta\right]|z\right]
\end{aligned}$$

using Gaussianity again to note that $\mathrm{Var}\left[x|z,\theta\right]$ is deterministic.

For $\mathrm{Var}\left[x|z,\theta\right]$, consider the hypothetical Bayesian update if θ is already known and z becomes known in addition. The standard posterior variance formula, conditional on θ, is

$$\mathrm{Var}\left[x|z,\theta\right] = \mathrm{Var}[x|\theta] - \mathrm{Cov}\left[x,z|\theta\right]\mathrm{Var}\left[z|\theta\right]^{-1}\mathrm{Cov}\left[z,x|\theta\right]$$

Writing our system in stacked form,

$$\begin{aligned}
x &= a\theta + Bu \\
z &= a\theta + Bu + Q\epsilon
\end{aligned}$$

where $a = (a_1, ..., a_m)'$, $B = \mathrm{diag}(b_j)$ and $Q = \mathrm{diag}(q_j)$, we find that

$$\begin{aligned}
\mathrm{Var}[x|\theta] &= BB' = \mathrm{Cov}\left[x,z|\theta\right] \\
\mathrm{Var}[z|\theta] &= BB' + QQ'
\end{aligned}$$

and we get

$$\begin{aligned}\mathrm{Var}[x|z,\theta] &= BB' - BB'(BB' + QQ')^{-1}BB' \\ &= QQ'(BB' + QQ')^{-1}BB' \\ &= \mathrm{diag}\left(b_j^2(1-m_j)\right)\end{aligned}$$

For $\mathbb{E}\left[x|z,\theta\right]$, write the updating equation 3.5 in stacked form

$$\begin{aligned}\mathbb{E}\left[x|z,\theta\right] &= (I-M)\mathbb{E}\left[x|\theta\right] + Mz \\ &= (I-M)a\theta + Mz\end{aligned}$$

where $M = \mathrm{diag}(m_j)$. Now take the variance of this conditional on z:

$$\mathrm{Var}\left[\mathbb{E}\left[x|z,\theta\right]|z\right] = \mathrm{Var}\left[\theta|z\right] \times (I-M)aa'(I-M)$$

so that the (i,j) element of this matrix is

$$\frac{1}{\tau_\theta + \sum \tau_j} a_i a_j (1-m_i)(1-m_j)$$

Putting all this together, the utility loss is

$$\begin{aligned}w'\left[\mathrm{Var}[x|z,\theta] + \mathrm{Var}\left[\mathbb{E}\left[x|z,\theta\right]|z\right]\right]w &= \sum_j w_j^2 b_j^2(1-m_j) + \frac{1}{\tau_\theta + \sum \tau_j}\sum_{i,j} w_i w_j a_i a_j (1-m_i)(1-m_j) \\ &= \sum_j w_j^2 b_j^2(1-m_j) + \frac{1}{\tau_\theta + \sum \tau_j}\left(\sum_j w_j a_j (1-m_j)\right)^2\end{aligned}$$

We need to take into account the dependence

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

There is a fixed exchange rate for each component: One unit of attention to x_j brings $\frac{a_j^2}{b_j^2}$ units of precision about θ.

The entropy the state variables (θ, x) in the economy satisfies the chain rule

$$H(\theta, x) = H(\theta) + H(x|\theta)$$

before receiving signals z, and

$$H(\theta, x|z) = H(\theta|z) + H(x|\theta, z)$$

afterwards. Hence the mutual (Shannon) information of states and signals is

$$\begin{aligned}
I(\theta, x; z) &= H(\theta, x) - H(\theta, x|z) \\
&= H(\theta) - H(\theta|z) + H(x|\theta) - H(x|\theta, z) \\
&= I\left(\theta; z\right) + I(x; z|\theta)
\end{aligned}$$

The first term is

$$\begin{aligned}
I(\theta; z) &= \frac{1}{2}\log\left[\frac{\mathrm{Var}\left[\theta\right]}{\mathrm{Var}\left[\theta|z\right]}\right] \\
&= \frac{1}{2}\log\left[\frac{\tau_\theta + \sum \tau_j}{\tau_\theta}\right] \\
&= -\frac{1}{2}\log\tau_\theta - \frac{1}{2}\log\left[\frac{1}{\tau_\theta + \sum \tau_j}\right]
\end{aligned}$$

The second term is

$$\begin{aligned}
I(x; z|\theta) &= \frac{1}{2}\log\left[\frac{\det\left(\mathrm{Var}[x|\theta]\right)}{\det\left(\mathrm{Var}[x|\theta, z]\right)}\right] \\
&= \frac{1}{2}\log\left[\frac{\prod_{i=1}^{m} b_i^2}{\prod_{i=1}^{m} b_i^2(1-m_i)}\right] \\
&= \frac{1}{2}\log\left[\frac{1}{\prod_j (1-m_j)}\right] \\
&= -\frac{1}{2}\sum_{j=1}^{m}\log(1-m_j)
\end{aligned}$$

using the facts derived above, namely $\mathrm{Var}[x|\theta] = \mathrm{diag}(b_j^2)$ and $\mathrm{Var}[x|\theta, z] = \mathrm{diag}\left(b_j^2(1-m_j)\right)$.

Combining, the cost of attention is

$$\begin{aligned}
K(m) &= I(\theta, x; z) \\
&= \text{constant} - \frac{1}{2}\log\left[\frac{1}{\tau_\theta + \sum \tau_j}\right] - \frac{1}{2}\sum_j \log(1-m_j)
\end{aligned}$$

The agent's full maximization problem can now be written as

$$\max_{m,\tau,\alpha} - \sum_j w_j^2 b_j^2 (1-m_j) - \frac{\alpha^2}{\tau} + \lambda \left(\log\left[\frac{1}{\tau}\right] + \sum_j \log(1-m_j) \right)$$

subject to

$$\tau \leq \tau_\theta + \sum \frac{a_j^2}{b_j^2} m_j$$
$$\alpha \geq \sum_j w_j a_j (1-m_j)$$

The Lagrangian is

$$\mathcal{L} = - \sum_j w_j^2 b_j^2 (1-m_j) - \frac{\alpha^2}{\tau} + \lambda \left(\log\left[\frac{1}{\tau}\right] + \sum_j \log(1-m_j) \right)$$
$$+ \mu_\tau \left(\tau_\theta + \sum \frac{a_j^2}{b_j^2} m_j - \tau \right) + \mu_\alpha \left(\alpha - \sum_j w_j a_j (1-m_j) \right)$$

which yields the desired FOC for m_j:

$$w_j^2 b_j^2 + \mu_\tau \frac{a_j^2}{b_j^2} + \mu_\alpha w_j a_j = \lambda \frac{1}{1-m_j}$$

C Derivations for Section 5

To come

References

Adam, K. (2007). Optimal monetary policy with imperfect common knowledge. *Journal of Monetary Economics*, **54** (2), 267–301.

Amador, M. and Weill, P.-O. (2010). Learning from prices: Public communication and welfare. *Journal of Political Economy*, **118** (5), 866 – 907.

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

Angeletos, G.-M. and LaO, J. (2010). Noisy business cycles. *NBER Macroeconomics Annual*, **24** (1), 319–378.

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

— and Pavan, A. (2007). Efficient Use of Information and Social Value of Information. *Econometrica*, **75** (4), 1103–1142.

Attanasio, O. P. (1999). Consumption. *Handbook of macroeconomics*, **1**, 741–812.

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

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. and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly journal of economics*, **127** (3), 1243–1285.

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

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

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*.

Colombo, L., Femminis, G. and Pavan, A. (2014). Information acquisition and welfare. *The Review of Economic Studies*, **81** (4), 1438–1483.

Croushore, D. D. (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.

De Long, 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.

Dovern, J. (2015). A multivariate analysis of forecast disagreement: Confronting models of disagreement with survey data. *European Economic Review*, **80**, 16–35.

—, Fritsche, U., Loungani, P. and Tamirisa, N. (2015). Information rigidities: Comparing average and individual forecasts for a large international panel. *International Journal of Forecasting*, **31** (1), 144–154.

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.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, **129** (4), 1661–1710.

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

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

Hellwig, C. and Veldkamp, L. (2009). Knowing what others know: Coordination motives in information acquisition. *The Review of Economic Studies*, **76** (1), 223–251.

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

Lichtenstein, S., Fischhoff, B. and Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In *Decision making and change in human affairs*, pp. 275–324.

Lucas, R. J. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, **4** (2), 103–124.

Maćkowiak, B. 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.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, **92** (5), 1521–1534.

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

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.

Sims, C. (2010). *Rational Inattention and Monetary Economics*. mimeo., Faculty of Economics, Princeton University.

Sims, C. A. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, vol. 49, pp. 317–356.

— (2003a). Implications of rational inattention. *Journal of monetary Economics*, **50** (3), 665–690.

— (2003b). Implications of rational inattention. *Journal of monetary Economics*, **50** (3), 665–690.

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

Townsend, R. M. (1983). Forecasting the forecasts of others. *Journal of Political Economy*, **91** (4), 546–88.

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.