Online Appendix to:

"Asymmetric Attention"

Alexandre N. Kohlhas and Ansgar Walther

IIES, Stockholm University and Imperial College London

March, 2021

Contents of Online Appendix:

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

A Proofs and Derivations (cont'd)

A.3 Optimal Attention Choice

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

$$L_t^\star \equiv \mathbb{E}\left\{\min_{a_{it}} \mathbb{E}\left[(a_{it} - a_t^\star)^2 \middle| \Omega_{it}\right]\right\}. \tag{OA1}$$

The minimizer to this problem is

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

Substituting this expression into (OA1) shows that

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

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

$$L_t^\star = \mathbb{V}\mathrm{ar}\left[a_t^\star \mid \Omega_{it}, \theta_t\right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}\left[a_t^\star \mid \Omega_{it}, \theta_t\right] \mid \Omega_{it}\right]. \tag{OA2}$$

To complete the proof, we need to derive expressions for the two components of (OA2).

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 (OA2).

Computing the first term in (OA2):

$$\begin{aligned}
\mathbb{V}\text{ar}\left[a_t^\star \mid \Omega_{it},\theta_t\right] &= \mathbb{V}\text{ar}\left[w_\theta\theta_t+\sum_j w_{xj}x_{jt}\middle|\,\Omega_{it},\theta_t\right]=\mathbb{V}\text{ar}\left[\sum_j w_{xj}x_{jt}\middle|\,\Omega_{it},\theta_t\right] \\
&= \sum_j w_{xj}^2\mathbb{V}\text{ar}\left[x_{jt}\middle|\,\Omega_{it},\theta_t\right]+\sum_j\sum_{k\neq j}\underbrace{\mathbb{C}\text{ov}\left[x_{jt},x_{kt}\middle|\,\Omega_{it},\theta_t\right]}_{=0} \\
&= \sum_j w_{xj}^2 b_j^2(1-m_j). \qquad \text{(OA3)}
\end{aligned}$$

Computing the second term in (OA2):

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

so that

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\mathbb{E}\left[a_t^\star \mid \Omega_{it},\theta_t\right]\mid\Omega_{it}\right] &= \mathbb{V}\text{ar}\left[w_\theta\theta_t+\sum_j w_{xj}\left((1-m_j)a_j\theta_t+m_jz_{ijt}\right)\middle|\,\Omega_{it}\right] \\
&= \mathbb{V}\text{ar}\left[\left(w_\theta+\sum_j w_{xj}(1-m_j)a_j\right)\theta_t\middle|\,\Omega_{it}\right] \\
&= \left[w_\theta+\sum_j w_{xj}(1-m_j)a_j\right]^2\mathbb{V}\text{ar}\left[\theta_t\mid\Omega_{it}\right]. \qquad \text{(OA4)}
\end{aligned}$$

Substituting (OA3) and (OA4) into (OA2) then yields the desired expression. □

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

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

The Lagrangian for this problem is

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

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

A.4 Macroeconomic Example

Proof of Proposition 4: We start with a firm's output choice,[1]

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

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

$$v\left(y_i, y, a_i, \omega\right) \approx v_1 y_i + \frac{v_{11}}{2} y_i^2 + v_{12} y_i y + v_{13} y_i a_i + v_{14} y_i \omega + t.i.a, \tag{OA5}$$

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

As a result of (OA5), a firm's optimal, full-information choice of output is

$$y_i^\star = \frac{v_{12}}{\mid v_{11} \mid} y + \frac{v_{13}}{\mid v_{11} \mid} a_i + \frac{v_{14}}{\mid v_{11} \mid} \omega, \tag{OA6}$$

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

$$y_i = \mathbb{E}_i \left[y_i^\star \right]. \tag{OA7}$$

It remains to derive the optimal output choice under full information in (OA6). A few simple but tedious derivations combine to show that

$$y_i^\star = r a_i + \alpha r \left(\sigma^{-1} y - \omega \right) \equiv x_{i1} + x_2. \tag{OA8}$$

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 (OA6) and (OA7) to derive the difference between a firm's valuation of its profits $v_i = v\left(y_i, y, a_i, \omega\right)$ and those that would have arisen under full information $v_i^\star$:

$$\begin{aligned} v_i - v_i^\star &= \frac{v_{11}}{2} y_i^2 - \frac{v_{11}}{2} y_i^{\star 2} + \left(v_{12} y + v_{13} a_i + v_{14} \omega \right) \left(y_i - y_i^\star \right) \\ &= \frac{v_{11}}{2} y_i^2 - \frac{v_{11}}{2} y_i^{\star 2} - v_{11} y_i^\star \left(y_i - y_i^\star \right) = \frac{v_{11}}{2} \left(y_i - y_i^\star \right)^2, \end{aligned} \tag{OA9}$$

where we have used the first-order condition for optimal output in (OA5). $\square$

Proof of Proposition 5: Follows immediately from (OA7) and (OA9). $\square$

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

B Over- and Underreactions in a General Linear Model

We extend the results from Section 2 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

$$y_t = D\theta_t + Ex_t + Fu_t \tag{OA10}$$

$$x_t = A\theta_t + Bx_{t-1} + Cu_t, \tag{OA11}$$

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),

$$\theta_t = M\theta_{t-1} + Nu_t, \tag{OA12}$$

where M and N are conformable matrices.

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

$$z_{it} = x_t + Q\epsilon_{it}, \quad Q = diag\left(q\right), \tag{OA13}$$

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 (OA10) to (OA12), as

$$y_t = \alpha\bar{\theta}_t + \beta u_t, \tag{OA14}$$

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

$$\bar{\theta}_t = \bar{M}\bar{\theta}_{t-1} + \bar{N}u_t, \tag{OA15}$$

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 (OA13) as

$$z_{it} = L_0\bar{\theta}_t + L_1\bar{\theta}_{t-1} + Ru_t + Q\epsilon_{it}, \tag{OA16}$$

where L_0, L_1 and R are implicitly defined.

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

Proposition B.1. *If the economy evolves according to (OA10)-(OA13), then the population coefficients in the regression equations (1) and (2) satisfy:*

$$\gamma < 0 \iff \alpha \bar{M}^k \left(GQQ'E' + \Sigma_{\theta\bar{\theta}} D' + \Omega \right) < 0 \tag{OA17}$$

$$\delta > 0 \iff \exists q_j \in (0\,, \infty)\,, \tag{OA18}$$

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 B.1; $\delta > 0$ 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 ($D \neq \underline{0}$),[2] (ii) for any cross-correlation in errors between the signal vector and output ($\Omega \neq \underline{0}$); and lastly ($iii$) for any effects that lagged components may have on output (see the expression for $\bar{M}$). The business cycle model in Section 5 provides an example of a model in which the second extension is relevant.

Proof of Proposition B.1: 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 (OA16) 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

[2] 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 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 (OA17) 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 ($\rho_2 < 0$).

(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

$$G_k=\mathbb{C}ov\left(\bar{\theta}_{t+k}-\mathbb{E}_{it-1}\bar{\theta}_{t+k},\, z_{it}-\mathbb{E}_{it-1}z_t\right)\mathbb{V}\left[z_{it}-\mathbb{E}_{it-1}z_t\right]^{-1}. \tag{OA19}$$

We note that

$$\bar{\mathbb{E}}_t\left[y_{t+k}\right]=\bar{\mathbb{E}}_{t-1}\left[y_{t+k}\right]+\alpha G_k\left(x_t-\bar{\mathbb{E}}_{t-1}\left[x_t\right]\right). \tag{OA20}$$

We can now use (OA20) to show that

$$\bar{\mathbb{E}}_t\left[y_{t+k}\right]-\bar{\mathbb{E}}_{t-1}\left[y_{t+k}\right] = \alpha G_k\left(x_t-\bar{\mathbb{E}}_{t-1}\left[x_t\right]\right) \tag{OA21}$$

$$y_{t+k}-\bar{\mathbb{E}}_t\left[y_{t+k}\right] = \alpha\left(\bar{\theta}_{t+k}-\bar{\mathbb{E}}_t\left[\bar{\theta}_{t+k}\right]\right)+Fu_{t+k}. \tag{OA22}$$

This completes the first step.

We are now ready to derive the overreaction coefficient γ:

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

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

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

where the last two equalities follow from

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

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

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

This completes the second step of the proof.

Lastly, we compute the underreaction coefficient δ. Equation (OA21), (OA22) 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

$$\begin{aligned}
\delta &\propto \alpha\left(\bar{G}_k - G_k\right) \mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1} x_t\right] G_k' \alpha' \\
&= \left(\bar{d}_k - d_k\right) \mathbb{V}\left[x_t - \bar{\mathbb{E}}_{t-1} x_t\right] d_k',
\end{aligned} \tag{OA23}$$

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 (OA23) 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:

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

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

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

An application of Proposition B.1, with G evaluated according to the standard expression for Kalman gains (Anderson and Moore, 2012), then also establishes Proposition 2.

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			0.68	
		(0.24)			(0.19)	
Individual Revision			−0.19			−0.02
			(0.06)			(0.08)
Observations	7,343	7,303	5,469	7,104	7,065	5,281
R^2	0.02	0.05	0.02	0.02	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 period: 1970Q4-2019Q4.

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	-0.08	–	-0.05
	(0.05)		(0.04)	(0.10)		(0.06)
Average Revision	–	0.68	0.71	–	0.38	0.40
		(0.19)	(0.18)		(0.16)	(0.16)
Observations	7,104	7,065	7,008	7,622	7,546	7,546
F	169.2	449.6	363.8	33.66	303.9	162.0
R^2	0.02	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	0.02	-0.09	0.25	0.22	0.24	0.45
	(0.19)	(0.10)	(0.15)	(0.24)	(0.13)	(0.21)
Current Realization	-0.10	–	-0.13	-0.04	-	-0.08
	(0.05)		(0.05)	(0.07)		(0.06)
Average Revision	–	0.78	0.84		0.40	0.44
		(0.26)	(0.25)		(0.14)	(0.14)
Observations	196	195	194	196	196	196
F	3.29	16.6	11.9	0.48	14.6	8.66
R^2	0.02	0.08	0.11	0.00	0.07	0.08

Note: Estimates of regressions (1) and (2) using one-quarter ahead forecasts. 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: 1970Q4-2019Q4.

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	7,104	7,190	7,190
F	169.2	253.8	185.4
R^2	0.02	0.04	0.03
	Panel b: average forecast error		
	Benchmark	*Level detrend*	*Linear detrend*
Constant	0.02	0.10	0.02
	(0.19)	(0.18)	(0.19)
Current Realization	-0.10	-0.13	-0.10
	(0.05)	(0.05)	(0.05)
Observations	196	196	196
F	3.29	6.47	3.29
R^2	0.02	0.03	0.02

Note: Estimates of regressions (1) using different methods for detrending output growth. Column (1): No detrending. 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: 1970Q4-2019Q4.

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	–	-0.20	–
	(0.06)	(–)	(0.08)	(–)
Average Revision	–	0.76	–	0.54
	(–)	(0.24)	(–)	(0.32)
Observations	2,284	2,245	4,574	4,574
F	93.1	186.5	161.2	161.7
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	–	-0.11	–
	(0.07)		(0.08)	
Average Revision	–	0.94	–	0.56
		(0.37)		(0.34)
Observations	60	59	120	120
F	2.83	6.62	1.83	5.72
R^2	0.05	0.10	0.02	0.05

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: 1970Q4-2019Q4 (split into 1970Q4-1985Q1 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. Constant term is included in Panel b.

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

	(1) *Individual errors*	(2) *Average errors*
Constant	–	0.28
		(0.39)
Avr. Forecast from Time t (δ_0)	0.70	0.84
	(0.20)	(0.26)
Avr. Forecast from Time $t-1$ (δ_1)	-0.65	-0.96
	(0.28)	(0.31)
Observations	7,151	195
F Statistic	249.5	8.959
R^2	0.07	0.09

Model	Df.	χ^2	$Pr(>\chi^2)$
(1) Individual Forecast Errors	1	0.14	0.71
(2) Average Forecast Errors	1	0.92	0.34

Note: *Upper table*: Estimates of $y_{t+k} - f_{it}y_{t+k} = \alpha_i + \delta_0 \bar{f}_t y_{t+k} + \delta_1 \bar{f}_{t-1} y_{t+k} + \epsilon_{it}$. 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: 1970Q4-2019Q4. *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	–	–	0.73	–	–
			(0.17)		
Observations	7,104	7.247	7,151	3,276	7,247
R^2	0.02	0.01	0.09	0.07	0.02
F	169.2	98.2	326.5	220.5	146.4

Note: Estimates of (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 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: 1970Q4-2019Q4.

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	–	-0.13	–	-0.16	–	-0.10	–
	(0.05)		(0.07)		(0.06)		(0.05)		(0.12)		(0.04)		(0.03)		(0.10)	
Average Revision	–	0.67	–	0.20	–	1.04	–	0.73	–	0.38	–	0.22	–	0.68	–	0.66
		(0.19)		(0.23)		(0.22)		(0.22)		(0.36)		(0.10)		(0.15)		(0.33)
Observations	7,141	7,102	3,995	4,067	5,543	5,470	4,118	4,017	4,273	4,103	1,960	1,920	1,827	1,787	147	151
Sample	1970Q4:2020Q1		1982Q3:2020Q1		1970Q4:2020Q1		2000Q1:2020Q1		2000Q1:2020Q1		1992Q1:2020Q1		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	–	-0.01	–	-0.12	–	-0.16	–	-0.10	–
	(0.05)		(0.07)		(0.04)		(0.05)		(0.13)		(0.12)		(0.10)		(0.10)	
Average Revision	–	0.78	–	0.28	–	1.12	–	0.62	–	0.50	–	0.27	–	1.18	–	0.66
		(0.26)		(0.23)		(0.29)		(0.22)		(0.43)		(0.20)		(0.72)		(0.33)
Observations	197	196	147	151	198	196	83	79	82	78	56	55	56	55	147	151
Sample	1970Q4:2020Q1		1982Q3:2020Q1		1970Q4:2020Q1		1999Q4:2020Q1		1999Q4:2020Q1		1992Q1:2020Q1		1992Q1:2020Q1		1982Q1:2020Q1	

Note: Estimates of (1) and (2) across different surveys: US SPF, Euro Area SPF, Livingston Survey, and Michigan Survey of Consumers. *Inflation* is the percentage change in the CPI; *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. The top and bottom one percent of forecast errors and revisions have been removed in each survey in Panel a pre-estimation. For the Michigan Survey: Regression (2) is estimated using instrumental variables (see footnote 8 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. For the Livingston Survey: We use three-quarter (annualized) forecasts and semi-annual forecast revisions. We also cluster at the respondent level only because of few time series observations (Cameron and Miller, 2015). For the Euro Area SPF: We use annual forecast revisions, and because of limited variation we do not trim inflation observations during the GFC (2009:Q3 Panel a). For the US SPF: See footnote 7 about the current realization of output. Robust standard errors (double clustered in Panel a) in parentheses. Regressions in Panel b include a constant term.

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	−0.20	−0.21	−0.16
	(0.04)	(0.07)	(0.05)	(0.05)	(0.19)	(0.03)	(0.03)
Average Revision	0.70	0.28	1.11	0.97	0.90	0.38	0.70
	(0.18)	(0.22)	(0.23)	(0.23)	(0.47)	(0.09)	(0.15)
Observations	7,045	3,995	5,470	4,017	4,103	1,920	1,787
Sample	70Q4:20Q1	82Q3:20Q1	70Q4:20Q1	00Q1:20Q1	00Q1:20Q1	92Q1:20Q1	92Q1:20Q1
			Panel b: average forecast error				
		US SPF		EA SPF		LS Survey	
	Output	Inflation	Deflator	Output	Inflation	Output	Inflation
Current Realization	−0.13	−0.18	0.04	−0.29	−0.21	−0.05	−0.17
	(0.05)	(0.07)	(0.04)	(0.07)	(0.18)	(0.14)	(0.10)
Average Revision	0.84	0.16	1.06	0.88	1.07	0.25	1.16
	(0.24)	(0.22)	(0.28)	(0.22)	(0.52)	(0.21)	(0.69)
Observations	195	147	196	79	78	55	55
Sample	70Q4:20Q1	82Q3:20Q1	70Q4:20Q1	99Q4:20Q1	99Q4:20Q1	92Q1:20Q1	92Q1:20Q1

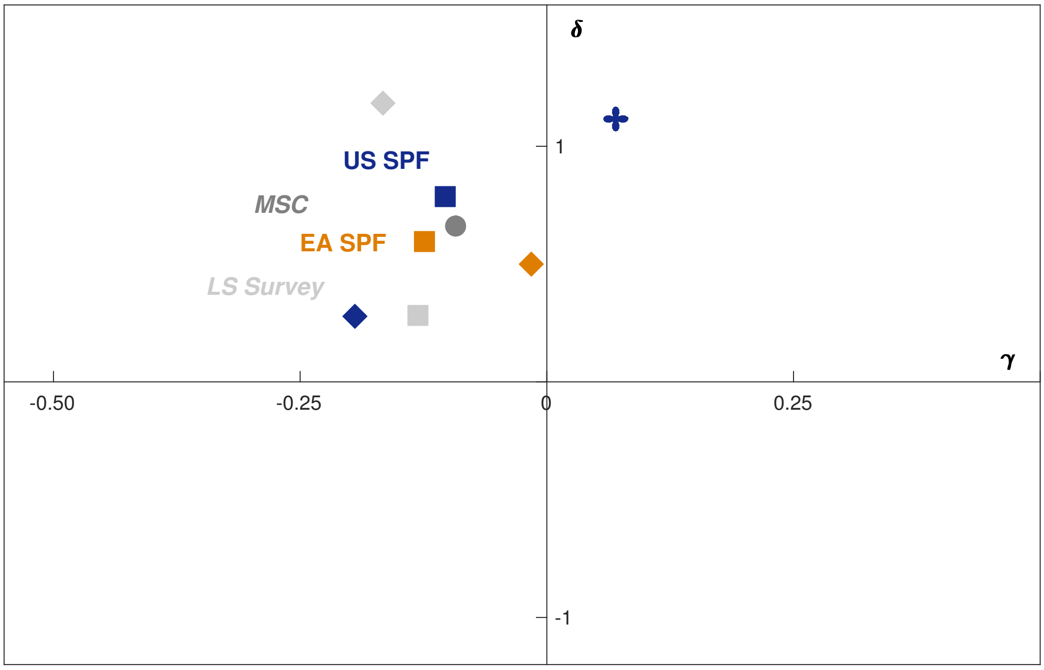
Note: Multivariate estimates of regressions (1) and (2) across surveys: the US SPF, the EA SPF, and the Livingston Survey. *Inflation* refers to the percentage change in the CPI index; *Deflator* to the percentage change in the GDP deflator. Panel a: Estimates with individual (respondent) fixed effects. 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. See the table note for Table C.7 for further comments. Regressions in Panel b include a constant.

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)
Current Realization	-0.10	–	-0.21	–	-0.17	–	-0.12	–	-0.01	–	-0.12	–	-0.16	–	-0.25	–
	(0.05)		(0.09)		(0.06)		(0.05)		(0.13)		(0.12)		(0.10)		(0.16)	
Average Revision	–	0.78	–	0.20	–	0.38	–	0.62	–	0.50	–	0.27	–	1.18	–	0.73
		(0.26)		(0.43)		(0.25)		(0.22)		(0.43)		(0.20)		(0.72)		(0.42)
Observations	197	196	113	113	113	113	83	79	82	78	56	55	56	55	113	113
Sample	1970Q4:2020Q1		1992Q1:2020Q1		1992Q1:2020Q1		1999Q4:2020Q1		1999Q4:2020Q1		1992Q1:2020Q1		1992Q1:2020Q1		1992Q1:2020Q1	

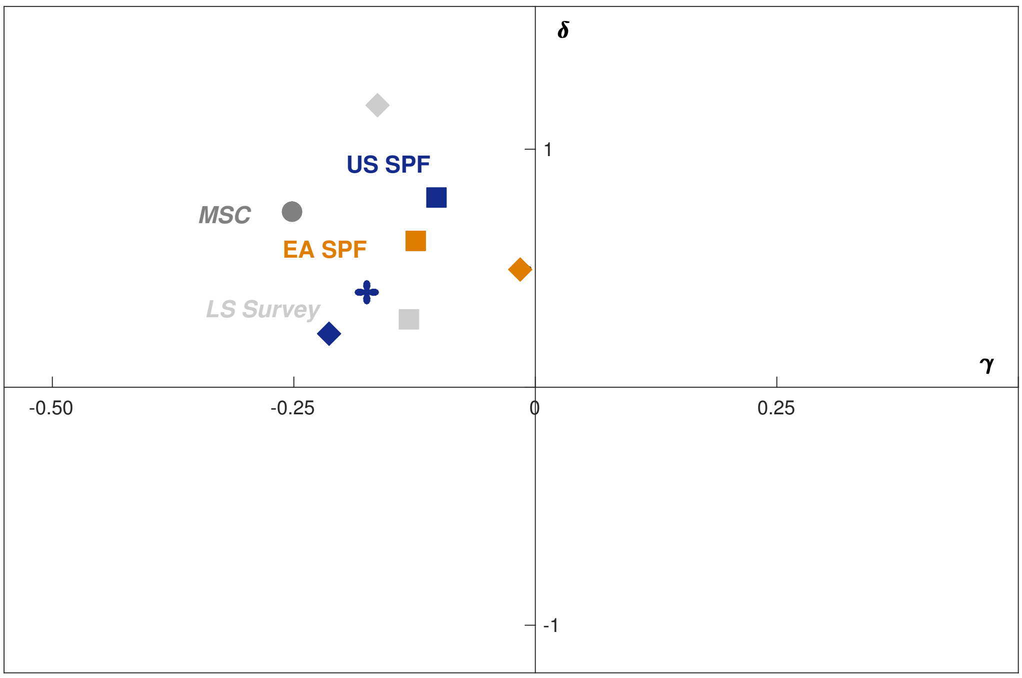
Note: Estimates of (1) and (2) across surveys using inflation data post 1992 and average forecast errors $y_{t+k} - \bar{f}_t y_{t+k}$ as the left-hand side variable: US SPF, Euro Area SPF, Livingston Survey, and Michigan Survey of Consumers. *Inflation* is the percentage change in the CPI index; *Deflator* is the percentage change in the GDP deflator. For the Michigan Survey: Regression (2) is estimated using instrumental variables (see footnote 8 in the paper). For the Livingston Survey: We use three-quarter (annualized) forecasts, and regression (2) uses semi-annual forecast revisions. For the Euro Area SPF: We use annual forecast revisions and one-year ahead forecasts. For the US SPF: See footnote 7 in the paper about the current realization of output. Robust standard errors in parentheses. All regressions in Panel b include a constant term.

Figure C.1: Alternative version of Figure 3 based on Table C.7b (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.

Figure C.2: Alternative version of Figure 3 based on Table C.9 (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. 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 and LIV 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:

$$y_t = \alpha + \sum_{h=1}^{H} \beta_h y_{t-h} + \sum_{j=1}^{J} d_j \eta_{t-j} + e_t. \tag{OA24}$$

$$y_t - \bar{f}_{t-k}\left[y_t\right] = \alpha + \sum_{h=1}^{H} \beta_h \left(y_{t-h} - \bar{f}_{t-k-h}\left[y_{t-h}\right]\right) + \sum_{j=1}^{J} d_j \eta_{t-j} + e_t. \tag{OA25}$$

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 (OA24) and (OA25) in Figure D.1. To operationalize (OA24) and (OA25), 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).

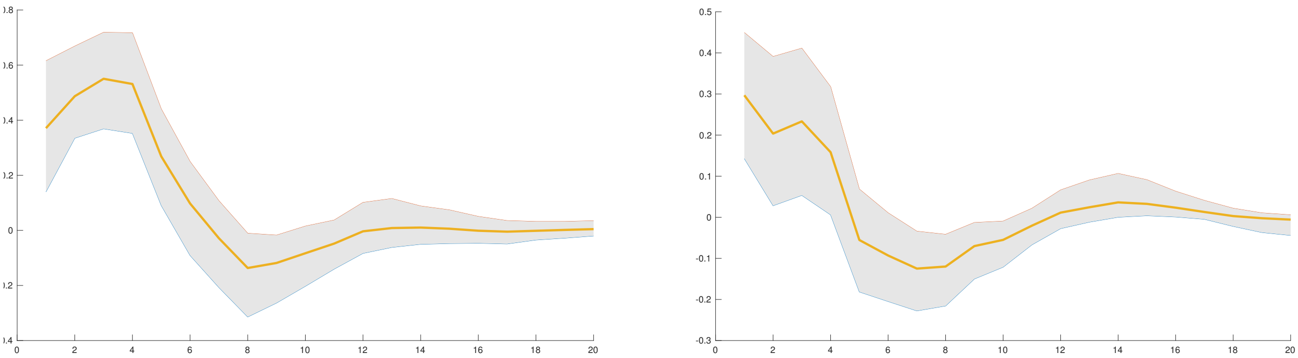
The estimates in Figure D.1 are in line with models of information frictions, and hence also our theory. We briefly document this result below for our baseline model.

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

$$\theta_t - \bar{\mathbb{E}}_t\theta_t = \rho\left(1 - \sum_j g_j a_j\right)\left(\theta_{t-1} - \bar{\mathbb{E}}_{t-1}\theta_t\right) + \left(1 - \sum_j g_j a_j\right)\eta_t - \sum_j g_j b_j u_{jt}, \tag{OA26}$$

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 (OA24). The right-hand panel depicts the mean forecast error response to the same productivity shock, based upon (OA25), 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 (OA24) and (OA25) 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 (OA24) and (OA25). 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 (OA24) and (OA25).

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

$$y_{t+k} - \bar{\mathbb{E}}_t y_{t+k} = \alpha\left(\theta_t - \bar{\mathbb{E}}_t\theta_t\right) + \text{t.n.p},, \quad \alpha = \rho^k \sum_j a_j > 0 \tag{OA27}$$

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,[3] output y_t and average forecast error of the latent factor $\theta_t - \bar{\mathbb{E}}_t\theta_t$ in (OA26) 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 (OA27), 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-

[3] 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:

$$s_{1it} = \Delta_t + \varepsilon_{it}, \quad s_{2it} = z_{it} + \psi_{it}. \tag{OA28}$$

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 (OA28) is equivalent to

$$\tilde{s}_{1it} = y_t + \tilde{\varepsilon}_{it}, \quad s_{2it} = z_{it} + \psi_{it} \tag{OA29}$$

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 (OA29) 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:

$$y^s_{it} = \alpha \left(p_{it} - \mathbb{E}\left[p_t \mid \Omega_{it}\right]\right) + \lambda y_{it-1}, \quad \alpha, \lambda > 0, \tag{OA30}$$

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

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

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[4]

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

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

$$y_t = k_0 \epsilon_{mt} + k_2 y_{t-1}. \tag{OA31}$$

We now turn to agents' expectations about future output. To start, notice that the expectation of the nominal demand shock ϵ_{mt} in (OA31) 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

$$y_{t+1} - \mathbb{E}_{it}[y_{t+1}] = k_2 k_0 (1-v)\epsilon_{mt} + k_0 \epsilon_{mt+1} - k_2 k_0 v \frac{1}{\pi_1}\epsilon_{ipt}. \tag{OA32}$$

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

Finally, using (OA31) and (OA32) 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 OA31). 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:

$$c_{it} = \mathbb{E}\left[c_{it+1} \mid \Omega_{it}\right] - i_t + \mathbb{E}\left[\pi_{\mathcal{B},it+1} \mid \Omega_{it}\right], \tag{OA33}$$

   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:

$$w_{it} - p_{\mathcal{B},it} = c_{it} + \psi n_{it}, \tag{OA34}$$

   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,

$$y_{it} = \int_{\mathcal{C},i,t} c_{mt} dm - \sigma\left(p_{it} - \int_{\mathcal{C},i,t} \bar{p}_{mt} dm\right), \tag{OA35}$$

   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

$$i_t = \rho_m i_{t-1} + \phi \tilde{\pi}_t, \quad \tilde{\pi}_t = \pi_t + \epsilon_t^\pi, \quad \epsilon_t^\pi \sim \mathcal{N}\left(0, \sigma_\pi^2\right), \tag{OA36}$$

   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,

$$p_{it} - p_{it-1} = \kappa \left(p_{\mathcal{B}.it} + c_{it} - p_{it} - a_{it}\right) + \kappa\psi \left(y_{it} - a_{it}\right) + \beta \mathbb{E}\left[p_{it+1} - p_{it} \mid \Omega_{it}\right], \tag{OA37}$$

   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	0.04	0.07	0.02	0.13

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),[5] (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 $\bar{z}_t = A_t + \xi_t$.

[5] 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

$$f_{it} \theta_t = \mathbb{E}_{it-1} \left[\theta_t \right] + (1 + \omega) \sum_j g_j \left(z_{ijt} - \mathbb{E}_{it-1} \left[z_{ijt} \right] \right). \tag{OA38}$$

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:[6]

$$\mathcal{I} = \mu \lim_{T\to\infty} \frac{1}{T}\left\{H\left(\theta^T, x^T\right) - H\left(\theta^T, x^T \mid z_i^T\right)\right\}. \tag{OA39}$$

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

[6] 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

$$\begin{aligned} H\left(s^T\right) - H\left(s^T \mid z_i^T\right) &= H(z_i^T) - H(z_i^T|s^T) \\ &= \sum_{t=1}^{T} H(z_{it}|z_i^{t-1}) - H(z_{it}|z_i^{t-1}, s^T). \end{aligned} \tag{OA40}$$

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

$$\begin{aligned} H(z_{it}|z_i^{t-1}) - H(z_{it}|z_i^{t-1}, s^T) &= H(z_{it}|z_i^{t-1}) - H(z_{it}|z_i^{t-1}, s_t) \\ &= H(s_t|z_i^{t-1}) - H(s_t|z_i^t) \\ &= H(\theta_t|z_i^{t-1}) - H(\theta_t|z_i^t) + H(x_t|z_i^{t-1}, \theta_t) - H(x_t|z_i^t, \theta_t), \end{aligned} \tag{OA41}$$

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

For *the first term* in (OA41), 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

$$H(\theta_t|z_i^{t-1}) - H(\theta_t|z_i^t) = \frac{1}{2}\log\left[\rho^2 + \frac{\sigma_\theta^2}{V(\tau)}\right] \equiv \mathcal{K}(\tau), \tag{OA42}$$

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

For *the second term* in (OA41), note that x_t is independent of z^{t-1} conditional on θ_t, so that

$$\begin{aligned} H(x_t|z_i^{t-1},\theta_t) - H(x_t|z_i^t,\theta_t) =& H(x_t|\theta_t) - H(x_t|z_{it},\theta_t) \\ =& \frac{1}{2}\log\left[\frac{\det\left(\mathbb{V}ar[x_t|\theta_t]\right)}{\det\left(\mathbb{V}ar[x_t|\theta_t,z_{it}]\right)}\right] = \frac{1}{2}\log\left[\frac{\prod_{i=1}^m b_i^2}{\prod_{i=1}^m b_i^2(1-m_i)}\right] \\ =& \frac{1}{2}\log\left[\frac{1}{\prod_j(1-m_j)}\right] = -\frac{1}{2}\sum_{j=1}^m \log(1-m_j). \end{aligned} \tag{OA43}$$

Substituting (OA43) and (OA42) into (OA41) 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 (OA40) with (OA39) 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:

$$w_j^2 b_j^2 + \hat{\mu}_\tau \frac{a_j^2}{b_j^2} + \mu_\alpha w_j a_j = \frac{1}{2}\frac{1}{1-m_j}, \tag{OA44}$$

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 (OA44) 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 (OA39), the first-order condition for agents' optimal attention choice m_j at an interior optimum satisfies:*

$$w_j^2 b_j^2 + \hat{\mu}_\tau \frac{a_j^2}{b_j^2} + \mu_\alpha w_j a_j = \frac{1}{2}\frac{1}{1-m_j}, \tag{OA45}$$

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

We show how to apply the dynamic rational inattention results in Maćkowiak *et al.* (2018) to our environment with flexible information choice (Section 4.3).

To do so, first notice that an agent's optimal action can be written as follows:

$$\begin{aligned} a_t^\star = \underbrace{\left(w_\theta + \sum w_{xj} a_j\right)}_{\equiv \bar{w}_\theta} \theta_t + \sum \underbrace{w_{xj} b_j}_{\equiv \bar{w}_{xj}} u_{jt} &= \rho a_{t-1}^\star + \bar{w}_\theta \eta_t + \bar{w}_x' u_t - \rho \bar{w}_x' u_{t-1}. \\ &\equiv \rho a_{t-1}^\star + c_0' v_t + c_1' v_{t-1}. \end{aligned} \tag{OA46}$$

We conclude that $a_t^\star$ is an ARMA(1,1) process with a vector of white noise innovations $v_t \equiv [\eta_t \;\; u_t]'$. Define the weighted sum of innovations in this expression as the scalar process

$$\omega_t \equiv c_0' v_t + c_1' v_{t-1}.$$

Since the innovation vector v_t is independently and identically distributed across time, ω_t is a stationary process. The auto-covariance structure of this process is

$$\mathbb{V}\text{ar}\left[\omega_t\right] = c_0' \Sigma_v c_0 + c_1' \Sigma_v c_1, \quad \mathbb{C}\text{ov}\left(\omega_t, \omega_{t-1}\right) = c_0' \Sigma_v c_1, \quad \mathbb{C}\text{ov}\left(\omega_t, \omega_{t-j}\right) = 0,\ j \geq 2,$$

where $\Sigma_v \equiv \mathbb{V}\text{ar}\left[v_t\right]$. By Wold's Representation Theorem, ω_t has an MA(1) representation:

$$\omega_t = d_0 \xi_t + d_1 \xi_{t-1},$$

where ξ_t is a Gaussian white noise sequence, and $d_j \in \mathbb{R}, j = \{1, 2\}$.

We conclude that we can write $a_t^\star$ as the ARMA(1,1) process:

$$a_t^\star = \rho a_{t-1}^\star + d_0 \xi_t + d_1 \xi_{t-1}. \tag{OA47}$$

We are now ready to state an agent's *flexible information design problem.* Following Maćkowiak *et al.* (2018), we can specify this problem as follows:

$$\min_{K, A, B, \Sigma_\psi} \mathbb{E}\left[\left(a_t^\star - \mathbb{E}\left[a_t^\star \mid \Omega_{it}\right]\right)^2\right] \tag{OA48}$$

subject to

$$\lim_{T \to \infty} \frac{1}{T} \left\{ H\left(a^{\star,T} \mid \bar{a}_0^\star\right) - H\left(a^{\star,T} \mid \bar{a}_0^\star, s^{K,T}\right) \right\} \leq \kappa, \tag{OA49}$$

where $\bar{a}_0^\star$ denotes the vector of initial conditions, and the signal vector observed by agent i follows

$$s_{it}^K = A a_t^{\star,M} + B \xi_t^N + \epsilon_{it}^K, \tag{OA50}$$

with $a_t^{\star,M} \equiv \left[\begin{array}{cccc} a_t^\star & a_{t-1}^\star & \ldots & a_{t-M+1}^\star \end{array}\right]'$, $\xi_t^N \equiv \left[\begin{array}{cccc} \xi_t & \xi_{t-1} & \ldots & \xi_{t-N+1} \end{array}\right]'$, and $\epsilon_{it}^K \sim \mathcal{N}\left(0, \Sigma_\epsilon\right)$.

Proposition H.1 now follows from the characterizations of optimal signals derived in Maćkowiak *et al.* (2018), who consider the same problem as (OA47) - (OA50), but in a model in which the optimal action follows a general ARMA(p,q) process.

Proposition H.1 (*Maćkowiak et al., 2018*)**.** The optimal signal vector s_{it}^K has the properties:
(i) Any optimal signal vector s_{it}^K is a noisy signal of a linear combination of $a_t^\star$ and ξ_t only.
(ii) An agent can attain the optimum with a one-dimensional signal ($K = 1$), which satisfies

$$s_{it}^\star = a_t^\star + h\xi_t + q^\star \epsilon_{it}, \quad h \neq 0, \quad \epsilon_{it} \sim \mathcal{N}(0, 1). \tag{OA51}$$

(iii) Suppose $\kappa \to \infty$. Then, $h \to 0$, so that $s_{it}^\star$ is a signal only of $a_t^\star$.
(iv) Suppose $w_\theta > 0$ and $w_{xj} = 0$. Then, $s_{it}^\star$ satisfies $s_{it}^\star = \theta_t + q^\star \epsilon_{it}$.

Proof of Proposition H.1: We refer to the corresponding proofs in Maćkowiak *et al.* (2018).
(i) See the proof of Proposition 1 in Maćkowiak *et al.* (2018).
(ii) See the proof of Proposition 2 and 5 in Maćkowiak *et al.* (2018).
(iii) See the proof of Proposition 6 in Maćkowiak *et al.* (2018).
(iv) See the proof of Proposition 2 and 3 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).[7] 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 ($\eta = 0$).[8] 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:

[7] In addition to the features mentioned, Angeletos *et al.* (2016) include one additional layer of CES aggregation.
[8] 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.

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*

$$y_{it} = k_0 a_t + k_1 y_t, \tag{OA52}$$

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

$$k_1 = \frac{\alpha(1-\sigma\psi)}{\alpha(1-\sigma)+\sigma(1+\eta)}. \tag{OA53}$$

We note that, because of certainty equivalence, we can use the full-information solution of the generalized model in (OA52) and (OA53) to determine whether output choices are strategic substitutes or complements even under imperfect information.

Equation (OA53) 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.

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:[9] Consider the equation for aggregate output that arises under imperfect attention:

$$y_t = \int_0^1 y_{it} di = \bar{\mathbb{E}}_t \left[x_{1t} + x_{2t}\right], \tag{OA54}$$

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

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

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

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + Bu_t, \quad u_t = \begin{bmatrix} u_t^\theta & u_t^x & u_t^n \end{bmatrix}', \tag{OA55}$$

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

where

$$A = \begin{bmatrix} 0 & 0 & r\rho_\theta & \mathbf{0} \\ & A_p & & \\ 0 & 0 & \rho_\theta & \mathbf{0} \\ & \mathbf{I} & & \end{bmatrix}, \quad B = \begin{bmatrix} r & r & 0 \\ & B_p & \\ 1 & 0 & 0 \\ & \mathbf{0} & \end{bmatrix}. \tag{OA56}$$

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

$$y_t = \psi \boldsymbol{x}_t, \quad x_{2t} = c_0 \boldsymbol{x}_t + c_1 u_t, \tag{OA57}$$

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

Coefficients and Conjectures: It follows from (OA54) that

$$y_t = \begin{bmatrix} 1 & 1 & \mathbf{0} \end{bmatrix} \bar{\mathbb{E}}_t \left[\boldsymbol{x}_t\right] \mathrel{\mathring{=}} \psi \boldsymbol{x}_t, \tag{OA58}$$

where $\mathring{=}$ denotes "should equal". We conclude from (OA58) that to verify our conjecture we need to find a matrix Ξ such that

$$\bar{\mathbb{E}}_t \left[\boldsymbol{x}_t\right] = \Xi \boldsymbol{x}_t. \tag{OA59}$$

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

$$x_{2t} = \alpha r \sigma^{-1} \psi \boldsymbol{x}_t - \alpha r \left\{ \psi \boldsymbol{x}_t - \psi B \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} u_t + e_3 u_t \right\} \mathrel{\mathring{=}} c_0 \boldsymbol{x}_t + c_1 u_t, \tag{OA60}$$

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

$$\begin{aligned} s_{it} &= \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \boldsymbol{x}_t + Q\epsilon_{it}, \quad Q = diag \begin{bmatrix} q_1 & q_2 \end{bmatrix} \\ &\equiv L\boldsymbol{x}_t + Q\epsilon_{it}. \end{aligned} \tag{OA61}$$

Thus,

$$\mathbb{E}_{it} \left[\boldsymbol{x}_t\right] = A\mathbb{E}_{it-1} \left[\boldsymbol{x}_t\right] + K \left(s_{it} - LA\mathbb{E}_{it-1} \left[\mathbf{z_{t-1}}\right]\right), \tag{OA62}$$

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

Then, from (OA59) and (OA62) it has to hold for all t that

$$\Xi \boldsymbol{x}_t = (I - KL)\, A \Xi \boldsymbol{x}_{t-1} + KL \boldsymbol{x}_t. \tag{OA63}$$

Fixed Point: We have from (OA58) and (OA60) that

$$\psi = \begin{bmatrix} 1 & 1 & \mathbf{0} \end{bmatrix} \Xi, \quad c_0 = \alpha r \left(\sigma^{-1} - 1\right) \psi, \quad c_1 = \alpha r \left(\psi B \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + e_3 \right). \tag{OA64}$$

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 $\boldsymbol{x}_t^{[T]} = \begin{bmatrix} \bar{x}'_{t-1} & \bar{x}'_{t-2} & \ldots & \bar{x}'_{t-T} \end{bmatrix}'$.

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 (OA61) and (OA62). This, in turn, allows us to find an expression for Ξ from (OA63) 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:

$$\Xi = (I - KL)\, A \Xi M + KL. \tag{OA65}$$

We can now use (OA64) 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 (OA56), 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

$$\min_{m_1, m_2} \mathbb{E}_{it} \left[y_{it} - y_{it}^{\star} \right]^2 + K(m), \quad K(m) = \mu \left(q_1^{-2} + q_2^{-2} \right), \tag{OA66}$$

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 (OA66), 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

$$\mathbb{V}\left[x_{it} \mid z_i^t\right] = \mathbb{V}\left[x_{it} \mid z_i^t, \boldsymbol{x}_t^{[T]}\right] + \mathbb{V}\left[\mathbb{E}\left[x_{it} \mid z_i^t, \boldsymbol{x}_t^{[T]}\right] \mid z_i^t\right] \tag{OA67}$$

by the *Law of Total Variance.*

It now follows that *the first component* in (OA67) 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 $\tilde{m} = (bb' + \bar{r}\bar{r}')\left[bb' + QQ' + \bar{r}\bar{r}'\right]^{-1}$.

To derive *the second component* in (OA67), 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 (OA62).

In sum, we have that the quadratic term (OA66) 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 (OA66).

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

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.

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*.

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

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.

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

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.