State-dependent attention and pricing decisions

Online Appendix

Javier Turen*

A. Profit Function Approximation

The derivation follows closely Alvarez and Lippi (2010). All firms share the same profit function $\Pi(P_t, Y_t, C_t) = Y_t P_t^{-\eta}(P_t - C_t)$. Where $\eta > 1$ represents the constant price elasticity, Y_t is the intercept of the demand (i.e. it's a demand shifter) and C_t is the marginal cost at time t. I assume that Y_t and C_t are perfectly correlated, i.e. when costs are high demand is also high. In order to approximate the objective function as (1), I compute a second order approximation of $\Pi(P_t, Y_t, C_t)$ around its frictionless price. In the RI context, the frictionless price is the optimal price under full information P_t^*.

The second order approximation of $\Pi(P_t, Y_t, C_t)$

$$\Pi(P_t, Y_t, C_t) \approx \Pi(P_t^*, Y_t, C_t) + \left.\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t}\right|_{P_t = P_t^*} (P_t - P_t^*) + \frac{1}{2} \left.\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\right|_{P_t = P_t^*} (P_t - P_t^*)^2$$

Which can be written:

$$\begin{aligned}\frac{\Pi(P_t, Y_t, C_t)}{\Pi(P_t^*, Y_t, C_t)} = 1 + &\frac{1}{\Pi(P_t^*, Y_t, C_t)} \left.\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t}\right|_{P_t = P_t^*} P_t^* \frac{(P_t - P_t^*)}{P_t^*} \\ &+ \frac{1}{2} \frac{1}{\Pi(P_t^*, Y_t, C_t)} \left.\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2}\right|_{P_t = P_t^*} (P_t^*)^2 \left(\frac{P_t - P_t^*}{P_t^*}\right)^2\end{aligned}$$

Taking the first and second order conditions:

*Instituto de Economía, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, 7820436, Santiago, Chile, (email: jturen@uc.cl)

$$
\begin{aligned}
\frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t} &= Y_t P_t^{-\eta} \left[-\eta \left(\frac{P_t - C_t}{P_t} \right) + 1 \right] \\
\frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2} &= -Y_t P_t^{-\eta-1} \eta \left[-\eta \left(\frac{P_t - C_t}{P_t} \right) + 1 \right] - Y_t \eta P_t^{-\eta-2} C_t
\end{aligned}
$$

From the first order conditions, the optimal price is simply a constant mark-up over marginal cost $P_t = \frac{\eta}{\eta-1} C_t$. Evaluating the first and second order conditions at the optimal price:

$$
\begin{aligned}
\left. \frac{\partial \Pi(P_t, Y_t, C_t)}{\partial P_t} \right|_{P_t^*} &= 0 \\
\left. \frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2} \right|_{P_t^*} &= -\eta Y_t C_t \left(\frac{1}{P_t^*} \right)^2 \left(\frac{\eta}{\eta - 1} C_t \right)^{-\eta}
\end{aligned}
$$

The maximized value of the profits:

$$
\Pi(P_t^*, Y_t, C_t) = Y_t \left(\frac{\eta}{\eta - 1} \right)^{-\eta} C_t^{1-\eta} \left(\frac{1}{\eta - 1} \right)
$$

Therefore, the term:

$$
\frac{1}{2} \frac{1}{\Pi(P_t^*, Y_t, C_t)} \left. \frac{\partial^2 \Pi(P_t, Y_t, C_t)}{\partial P_t^2} \right|_{P_t} (P_t^*)^2 = \frac{-\eta Y_t C_t \left(\frac{\eta}{\eta-1} C_t \right)^{-\eta}}{Y_t \left(\frac{\eta}{\eta-1} \right)^{-\eta} C_t^{1-\eta} \left(\frac{1}{\eta-1} \right)} = -\eta(\eta - 1)
$$

Finally, the second order approximation:

$$
\frac{\Pi(P_t, Y_t, C_t) - \Pi(P_t^*, Y_t, C_t)}{\Pi(P_t^*, Y_t, C_t)} = -\frac{1}{2} \eta(\eta - 1) \left(\frac{P_t - P_t^*}{P_t^*} \right)^2 + o \left(\frac{P_t - P_t^*}{P_t^*} \right)
$$

Where I can finally define $\gamma \equiv -\frac{1}{2}\eta(\eta-1)$, $\widehat{\Pi}(p_{it}, \widehat{p}_{it}) = log(\Pi(P_t, Y_t, C_t)) - log(\Pi(P_t^*, Y_t, C_t))$, $p_t = log(P_t)$ and $\widehat{p}_{it} = log(P_t^*)$ as stated in equation (1).

B. Equivalence of mutual information

Information entropy is a measure about the uncertainty of a random a variable. Consider a random variable X with finite support Ω_x, which is distributed according to $f \in \Delta(\Omega_x)$. The entropy of X, is defined by:

$$\mathcal{H}(X) = -\sum_{x \in \Omega_s} f(x) log f(x)$$

With the convention that $0 log 0 = 0$. In RI, the acquired amount of information is measured by entropy reduction. Given the signal s_t, entropy reduction is measured by mutual information, which in the context of this dynamic model is:

$$\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1}) = \mathcal{H}(\widehat{p}_{it}|s_i^{t-1}) - E_{s_{it}}[\mathcal{H}(\widehat{p}_{it}|s_{it})|s_i^{t-1}]$$

Given the entropy, the target-price $\widehat{p}_{it} = \sigma_t \epsilon_{it} \in \Omega_{\widehat{p}}$, and the definition for mutual information we can prove:

$$\begin{aligned}
\mathcal{I}(\widehat{p}_{it}, s_{it}|s_i^{t-1}) &= \mathcal{H}(\widehat{p}_{it}|s_i^{t-1}) - E_{s_{it}}[\mathcal{H}(\widehat{p}_{it}|s_{it})|s_i^{t-1}] \\
&= \sum_{s_{it}} f(s_{it}|s_i^{t-1}) \left[\sum_{\sigma} \sum_{\epsilon} f(\widehat{p}_{it}|s_{it}, s_i^{t-1}) log(f(\widehat{p}_{it}|s_{it}, s_i^{t-1})) \right] \\
&- \sum_{\sigma_t} \sum_{\epsilon_{it}} g(\widehat{p}_t|s_i^{t-1}) log(g(\widehat{p}_t|s_i^{t-1})) \\
&= \sum_{s_{it}} \sum_{\sigma_t} \sum_{\epsilon_{it}} f(s_{it}, \widehat{p}_{it}|s_i^{t-1}) log(f(\widehat{p}_{it}|s_{it}, s_i^{t-1})) \\
&- \sum_{\sigma_t} \sum_{\epsilon_{it}} \left[\sum_{s_{it}} f(s_{it}, \widehat{p}_{it}|s_i^{t-1}) \right] log(g(\widehat{p}_{it}|s_i^{t-1})) \\
&= \sum_{s_{it}} \sum_{\sigma_t} \sum_{\epsilon_{it}} f(s_{it}, \widehat{p}_{it}|s_i^{t-1}) log \left(\frac{f(\widehat{p}_{it}|s_{it}, s_i^{t-1})}{g(\widehat{p}_{it}|s_i^{t-1})} \right) \\
&= \sum_{s_{it}} \sum_{\sigma_t} \sum_{\epsilon_{it}} f(s_{it}, \widehat{p}_{it}|s_i^{t-1}) log \left(\frac{f(s_{it}, \widehat{p}_{it}|s_i^{t-1})}{g(\widehat{p}_{it}|s_i^{t-1}) f(s_{it}|s_i^{t-1})} \right)
\end{aligned}$$

Using the notation $\sum_{x_t} = \sum_{x_t \in \Omega_x}$.

From the second to the third line of the equivalence we rely on the fact that the prior distribution (marginal) is characterized as the sum of the joint probability distribution $f(s_{it}, \widehat{p}_{it}|s_i^{t-1})$ across all potential signals. The final expression is then what is shown in equation (3).

C. Solution of the dynamic RI problem

In this section, I show how to derive the solution for the dynamic RI problem formally introduced in Section 2.4. Given prior beliefs $g_{it}(\widehat{p}_{it}|p_{it-1})$, firms choose the conditional probability distribution of prices $f_{it}(p_{it}|\widehat{p}_{it})$ (equivalent of choosing $f(p_{it},\widehat{p}_{it})$) in each point of the simplex $\Omega_p \times \Omega_\sigma \times \Omega_\epsilon$. To simplify notation, I will omit the lagged price conditioning and focus on a representative firm $\lambda_i = \lambda$.

Since the prior belief about the volatility distribution $m_t(\sigma_L)$ is the state variable of the problem, we can write the Bellman equation:

$$V(m_t(\sigma_L)) = \max_{f_t(p_t|\widehat{p}_t)} \sum_\sigma \sum_\epsilon \sum_p [\widehat{\Pi}(p_t,\widehat{p}_t) + \beta V(m_{t+1}(\sigma_L))] f_t(p_t|\widehat{p}_t) g_t(\widehat{p}_t) \; - \; \lambda \mathcal{I}(\widehat{p}_t, p_t)$$

Where:

$$\mathcal{I}(\widehat{p}_t, p_t) = f_t(p_t, \widehat{p}_{it}) log\left(\frac{f_t(p_t,\widehat{p}_t)}{g_t(\widehat{p}_t) f_t(p_t)}\right) = f_t(p_t|\widehat{p}_t) g_t(\widehat{p}_t)[log(f_t(p_t|\widehat{p}_t)) - log(f_t(p_t))]$$

The function is also maximized subject to the constraint on the prior (7). The first order condition of $V(m_t(\sigma_L))$ with respect to $f_t(p_t|\widehat{p}_{it})$:

$$\begin{aligned} g_t(\widehat{p}_t)\left[\widehat{\Pi}(p_t,\widehat{p}_{it}) + \beta V(m_{t+1}(\sigma_L)) + \beta\left[\frac{\partial V(m_{t+1}(\sigma_L))}{\partial m_{t+1}(\sigma_L)} \times \frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)}\right]\right] \\ -\lambda g_t(\widehat{p}_t)[log(f_t(p_t|\widehat{p}_t)) + 1 - log(f_t(p_t)) - 1] - g_t(\widehat{p}_t)\mu(\widehat{p}_t) = 0 \end{aligned} \tag{A.1}$$

The last term on the left hand side of equation (A.1), $\mu(\widehat{p}_{it})$, corresponds to the Lagrange multiplier of the constraint attached to the prior, equation (7) in the main text.

Embedded in equation (A.1) is the effect of the current information strategy on posterior beliefs, $\frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)}$. As discussed, posterior beliefs will later become the prior for $t+1$, $g_{t+1} = m_{t+1}(\sigma)h(\epsilon)$. The known i.i.d. structure of the idiosyncratic shocks ϵ_t implies that the chosen information strategy is not going to affect beliefs about this marginal distribution. Moreover, as stressed by SSM (2017), we can treat the effects of current information on future beliefs about the persistent state σ_t as fixed. The authors shows that a dynamic RI problem such as the one presented in this paper, is equivalent to a control problem without uncertainty about persistent states.[1] Therefore $\frac{\partial m_{t+1}(\sigma_L)}{\partial f_t(p_t|\widehat{p}_t)} = 0$ and given $g_t(\widehat{p}_t) \geq 0$ and $\lambda > 0$, equation (A.1) becomes:

$$\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L)) - \mu(\widehat{p}_t)}{\lambda} = log\left(\frac{f(p_t|\widehat{p}_t)}{f_t(p)}\right)$$
$$exp\left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) exp\left(\frac{-\mu(\widehat{p}_t)}{\lambda}\right) = \frac{f(p_t|\widehat{p}_t)}{f_t(p)}$$
$$\Rightarrow f(p_t|\widehat{p}_t) = \ exp\left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) f_t(p_t)\phi(\widehat{p}_t)$$

Where:

$$\phi(\widehat{p}_t) \equiv exp\left(\frac{-\mu(\widehat{p}_t)}{\lambda}\right) \tag{A.2}$$

Finally, due to the restriction on the prior:

$$\begin{aligned} g_t(\widehat{p}_t) &= \sum_{p_t'} f_t(p_t'|\widehat{p}_t) g(\widehat{p}_t) \\ &= \sum_{p_t'} exp\left(\frac{\Pi(p_t', \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) f_t(p_t')\phi(\widehat{p}_t) g(\widehat{p}_t) \\ \Rightarrow \phi(\widehat{p}_t) &= \frac{1}{\sum_{p_t'} \ exp\left(\frac{\Pi(p_t', \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda}\right) f_t(p_t')} \end{aligned}$$

Combining this expression with (A.2), and adding the conditioning on lagged prices, we get the expression for the optimal posterior distribution of prices given the unobserved target, (11) in the main text:

$$f_t(p_t|\widehat{p}_t, p_{t-1}) = \frac{exp\left[\left(\Pi(p_t, \widehat{p}_t) + \beta V(m_{t+1}(\sigma_L|p_t))\right)/\lambda\right] f_t(p_t|p_{t-1})}{\sum_{p_t'} \ exp\left[\left[\Pi(p_t', \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L|p_t))\right)/\lambda\right] f_t(p_t'|p_{t-1})}$$

[1] The intuition behind the result is the following. In the control problem, while firms have full information about the current and past history of shocks, they face a trade-off of optimizing their flow utility $\widehat{\Pi}(p_t, \widehat{p}_t)$ against a control cost given by: $E_{f(p_t|\widehat{p}_{it})}[log(f(p_t|\widehat{p}_t)) - log(q(p_t|\widehat{p}_t)|z^t]$. The variable z^t stands for the entire history of past shocks and prices. The cost is determined by the deviation of the final action with respect to some default action $q(p_t|\widehat{p}_{it})$. By relying on properties of the entropy, the paper shows an equivalence between a control and a dynamic Rational Inattention problem. Thus the inattention problem is solved by initially solving the control problem with observable states, characterizing the optimal conditional probability for each default rule $f(p_t|\widehat{p}_t)$, and then choosing q. As states are observable in the control problem, the solution ignores the effects of information acquisition on future beliefs (i.e., treat them as fixed) when solving the dynamic RI problem.

The expression for the value function is then simply given by plugging this expression into equation (5) in the main text:

$$\begin{aligned} V(m_t(\sigma_L)) &= \lambda \sum_{\sigma_t} \sum_{\epsilon_t} \sum_{p_t} f(p_t, \widehat{p}_t) \log \left(\sum_{p} exp \left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda} \right) f(p_t|p_{t-1}) \right) \\ &= \lambda E \left[\log \left(\sum_{p_t} \ exp \left(\frac{\Pi(p_t, \widehat{p}_t) \ + \beta V(m_{t+1}(\sigma_L))}{\lambda} \right) f(p_t|p_{t-1}) \right) \right] \end{aligned}$$

D. Dynamic RI Algorithm

The algorithm to solve the dynamic RI problem is as follows:

1. Fix a value for the idiosyncratic information acquisition cost, e.g. λ_1.

2. Given λ_1 and the belief simplex, compute prior beliefs $g(\widehat{p}_{it}) = m(\sigma_t)h(\epsilon_{it})$.

3. With $g(\widehat{p}_{it})$, the model is solved by Value Function Iteration.

   3.1. Starting with a guess for the vector $V(m_{t+1}(\sigma_L))$, we first solve the static RI problem. The algorithm computes $f(p_{it}, \widehat{p}_{it}|p_{it-1}) \in \Delta(\Omega_p \times \Omega_\sigma \times \Omega_\epsilon)$ which is the solution for the system of nonlinear equations (7), (11) and $f(p_{it}|p_{it-1}) = \sum_\sigma \sum_\epsilon f(p_{it}, \widehat{p}_{it}|p_{it-1})$.

   3.2. Given $f(p_{it}, \widehat{p}_{it}|p_{it-1})$, the prior $g_{it}(\widehat{p}_{it})$ and using Bayes Law, we can compute the conditional probability $f(\sigma|p_{it}, p_{it-1}) = \sum_\epsilon f(\sigma, \epsilon|p_{it}, p_{it-1})$ for each $p_{it} \in \Omega_p$. Through equation (10), posterior beliefs become the prior beliefs for the next period. With this we update $V(m_{t+1}(\sigma_L))$.

   3.3. Relying on the definition for $V(m_{it}(\sigma_L|p_t))$ in (12), the algorithm iterates the value function until convergence when, within each iteration, it re-estimates $f(p_{it}, \widehat{p}_{it}|p_{it-1})$.

4. Repeat point 3 for all possible values in $\Delta(\Omega_\sigma)$, i.e. setting different priors $g(\widehat{p}_{it})$.

5. Repeat 2, 3, and 4 for all possible values for λ_i.

The setting of the model and the decision of the shape of the joint probability distribution resembles a filtering problem. The numerical discrepancies between filtering with discrete variables relative to continuous outcomes are not significant and depend on the nature of the approximation, Farmer (2016) and Farmer and Toda (2017).

E. Sensitivity analysis

In this section, we provide a further description of the identification strategy of our baseline model. As discussed we targeted four moments, $E(|\Delta p|)$, $E(Dispersion|\sigma_L)$, $E(Dispersion|\sigma_H)$ and β_{IR}, using for parameters $\theta = \{\sigma_L, \phi, \overline{\lambda}, \sigma_\lambda\}$.

To provide a sensitivity analysis of which data moments are more informative of which parameters we perform the following exercise: starting from the first parameters in θ, i.e., σ_L, we reduce its magnitude by 0.1% while keeping *all* the remaining parameters fixed at their calibrated values. With this new parametrization we solve the model again, and report the four targeted moments. We repeat this exercise for different values of σ_L, where its original magnitude is reduced or increased by $0.x\%$ where $x = 1, \ldots, 5$. We repeat the procedure for each parameter in θ one at a time, keeping all of the remaining parameters fixed in their original calibrated values. The results are shown in Figure A.1.

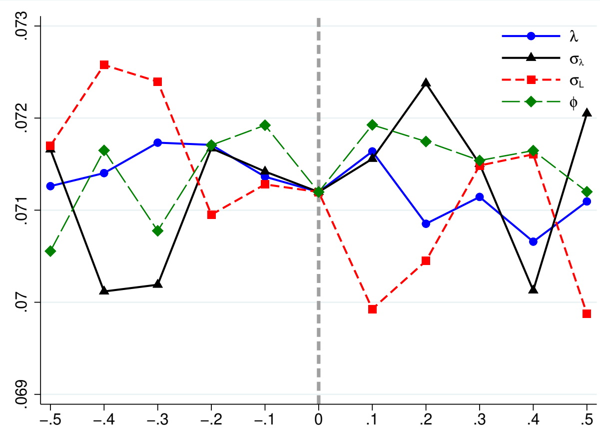
Although we allow for some marginal perturbations of the parameters, in some cases the response of the targeted moments is sizable. We conjecture that this is because of the interplay between the effects of the new parameters on the learning and pricing strategies combined with the discretization assumed for the simplex of each variable.

The average magnitude of price revisions $E(|\Delta p|)$ (top left panel of Figure A.1) seems very sensible to volatility changes in the more persistent (and therefore the most likely) aggregate state. Changes in σ_L, represented by the connected line with squares, affects the intensive margin of prices significantly. This moment is also very sensible to the dispersion of information costs, shown by the black line with triangles. In line with was described in Section 3.3, heterogeneous costs leads to heterogeneous pricing strategies creating a direct mapping between the array of values for λ_i and the magnitude of price adjustments. The degree by which $E(|\Delta p|)$ is affected by σ_λ is simply a direct implication of this result.

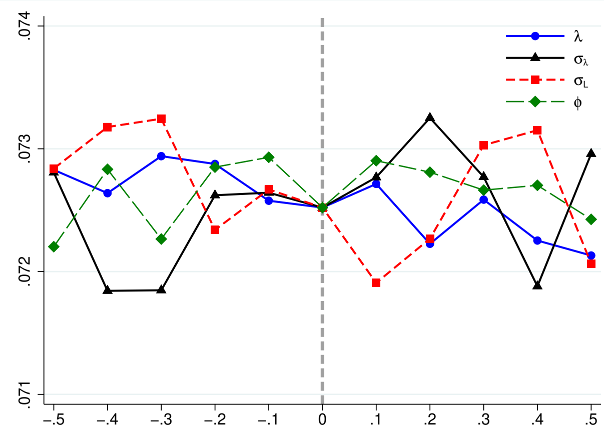
As the dispersion of information costs σ_λ disciplines the relative magnitude of price adjustments between firms, the two dispersion moments (top right and bottom left figures) are also highly responsive to this parameter. As expected, the price dispersion in the low volatility state is also affected by σ_L. Intuitively, the relevance of this last parameter is more muted when we look at the reaction of price dispersion in the high volatility state. In this latter case, and by construction, the role of ϕ in identifying this moment is more relevant relative to σ_L. However, for these two moments the implications are not so clear, suggesting that they are jointly identified by the set of parameters.

Finally, the overall degree of information rigidity β_{IR} in the model seems very sensible to average value of information costs $\overline{\lambda}$. As discussed in the main text, this parameter is identified by regressing the forecast error on the forecast revision over the cross-section of firms. As shown by CG (2015) there is a direct mapping between this parameter and the degree of information rigidity faced by agents. As $\overline{\lambda}$ reflects the average magnitude of the information rigidity across firms it is not surprising that this moment is very responsive to this parameter.

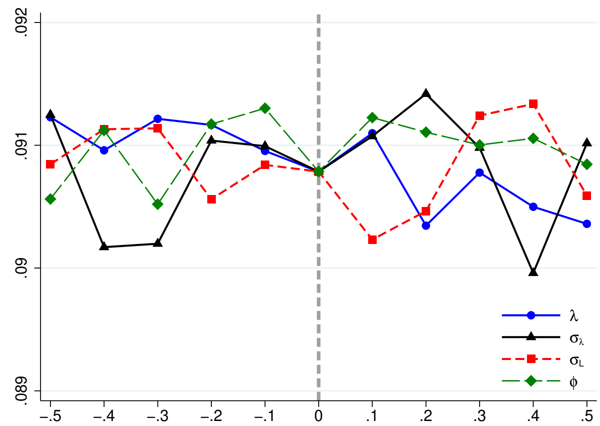
Figure A.1: Sensitivity of moments to parameters



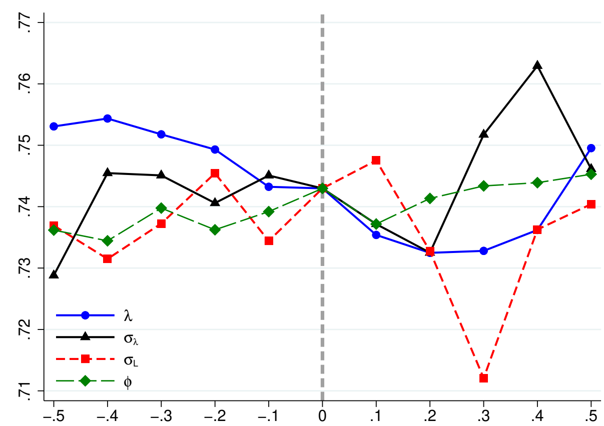
(a) $E(|\Delta p|)$



(b) $E(Dispersion|\sigma_L)$



(c) $E(Dispersion|\sigma_H)$



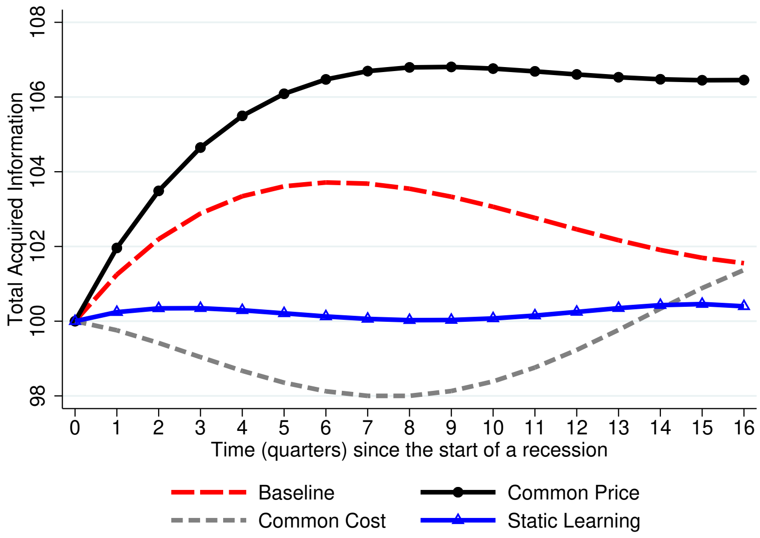
(d) β_{IR}

Notes: The figure presents the sensitivity of the four targeted moments to changes in one of the parameters in θ, while keeping the rest constant. The procedure is repeated for all four parameters where within each iteration the dynamic learning model is solved and the relevant moments are saved.

F. Information acquisition - Alternative Models

In this section we study the different implications for dynamic learning over the business cycle for all alternative specifications presented in Section 5. Figure A.2 shows the simulated evolution for the baseline model along with the static, homogeneous price, and common cost models. As in Figure 1 in the main text, we show the response of total acquired information after the economy enters into the high volatility state at quarter $t = 0$. Since the focus in on the underlying dynamics, we normalize the initial response to 100.

Figure A.2: Information acquisition over the business cycle - Alternative Models



Notes: The figure represents the impulse response function of the different simulated models after the economy enters into the less predictable state at time $t = 0$. The red dashed line represents the baseline model, the blue with triangles is the static learning model, the dotted black lines is the common price version and finally the dashed gray line is the homogeneous cost specification. The initial responses are normalized to 100.

Static Learning

Despite the muted predictions for β_{IR}, we can still assess if the static learning model is consistent with state-dependent attention and its dynamic evolution. Based on equation (15) it is straightforward to notice that total attention will increase immediately after the economy enters into the less predictable state at $t = 0$. As expected, the figure shows that κ_t will stays almost constant throughout the 16 simulated quarters. This result is completely in line with the average duration of the high volatility state given by the calibrated transition probabilities.[2] While the initial response is normalize to 100, the magnitude of $\overline{\kappa}$ in this case is significantly higher than in the other cases. In particular, $\overline{\kappa}$ is around 8 times higher compared to the baseline model. In the static version, price-setters does not waste any of their attention in noticing the actual state of the economy. Hence, the learning problem is simpler as firms only collect information to track the outcome of the target-price, leading them to rationally choose to collect more information.

[2] Actually, the two transitions probabilities $\tau_{LH} = 0.00882$ and $\tau_{HL} = 0.0196$ imply that the average duration of the high volatility state is 17 quarters.

The implied dynamics for κ in this case are completely at odds with the dynamic patters of both the data and the baseline model. This, along with the discussion in Section 5.5.1, reinforces the impossibility of a more stylized version of the model to consistently replicate the features of the data.

Common Cost

Given the same cost of information λ, the learning and pricing strategies will be common across firms. Hence, the rate by which firms will notice any new aggregate state will also be similar across them. Consistent with the results discussed in Section 3.3, a firm with an average cost will focus almost all of its attention on a subset of prices which are closer to the mean. This would prevent them from noticing any extreme realization of the target-price that could suggest that the economy is actually in the more volatile state. This intuition is reinforced by the results in figure A.2. The fact that $\overline{\kappa}$ slightly decreases until the 7-8th quarter after the recession starts suggest that firms confound the new state with the more predictable state, leading them to marginally collect less information. It is only after several quarters of being exposed to more extreme realization of $\widehat{p}_{it}$ that they start revising their beliefs in favor of the high volatility state. The rise in $\overline{\kappa}$ is therefore delayed due to their misperceived beliefs.

The simulated response of total information is not only inconsistent with the data, but also reinforces the relevance of allowing for idiosyncratic differences at the firm level as a relevant mechanism to capture the time-varying implication of attention.

Common Target Price

Among the alternative models, the version with a common optimal price successfully resembles the dynamic features of the data (black dotted line in figure A.2). Although this version of the model is not fully consistent with some key targeted price facts, this result highlights the relevance of having both dynamic information and firm heterogeneity to replicate state-dependent attention.

G. Robustness to first and second-moment shocks

This section provides further detailed results for the extended version of the model with both first and second-moments shocks. As discussed in Section 5.6, the consequences of adding a first-moment shock to our model are not obvious, particularly within a setup where information is endogenous and fully flexible. Since the entire learning strategy will change with the two shocks, it is hard to anticipate the changes in acquired information, which complicates the comparison with our baseline scenario. Hence, we solve the two shocks version using the original parameters in Table 2 (main text) and setting $\mu_H = 0.995\mu_L$. Table A.1 shows the pricing and the information moments.

Table A.1: Matched Moments and Alternative Specifications

Targeted moments	Data	Baseline	First & Second Moment
$E(\|\Delta p\|)$	0.077	0.071	0.061
$E(Dispersion\|\sigma_L)$	0.073	0.072	0.068
$E(Dispersion\|\sigma_H)$	0.090	0.090	0.086
β_{IR}	0.674	0.724	0.773
Non-Targeted moments			
Frequency	0.150	0.564	0.636
$Kurtosis(\|\Delta p\|)$	6.403	4.049	4.127
Fraction small	0.330	0.175	0.209
$Corr(Dis, Freq)$	0.506	0.630	0.516

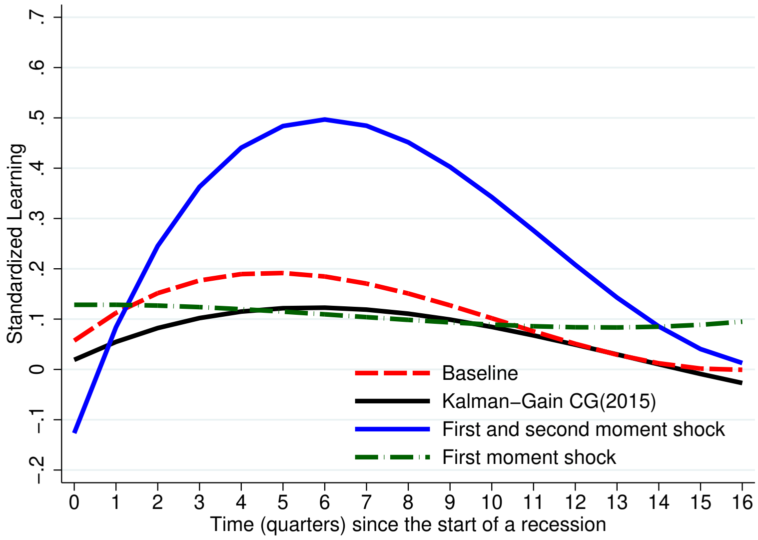
Although the moments are not directly targeted, the distance between the extended model and the data is not far. This is reassuring as it implies that the results are robust to more complex versions of the dynamic learning model, accounting for additional business cycle features. The model can still replicate (and even be closer to match) one of the key non-targeted moments, such as the correlation between dispersion and frequency of price adjustments. Turning to the learning responses, Figure A.3 shows the standardized response of acquired information over the cycle. From the Figure, we notice that, the presence of the two shocks brings an amplification in the learning reaction of firms relative to the scenario with just a second-moment shock. Intuitively, the different mean makes the learning problem easier for firms as now the perceived average of signals changes over time. This boosts learning right after the economy enters into a recession. Besides the different learning rates, the dynamic correlation between the extended model and the data is 0.83, a marginal improvement relative to the 0.79 correlation of the baseline data.[3]

First-moment shock only

Finally, we propose a different version that allows for a drop in the mean while keeping the volatility constant. In particular, we assume that $\widehat{p}_{it} = \mu_t + \sigma_t \epsilon_{it}$ where $\mu_H < \mu_L$ and $\sigma_t = \sigma_L$. The response of this specification to a recession is shown by the green dotted line in Figure A.3. Allowing for just a first-moment shock is not enough to generate any meaningful reaction in the learning rate. Although the states of the economy are still persistent, the only

[3]We conjecture that by calibrating the two-shock version of the model to match the targeted moments, both the empirical and the simulated responses would be closer to each other. This is because one of the targeted moments is the average information rigidity parameter β_{IR}. However, this section intended to address whether a more sophisticated model robust to business cycle features can be consistent with the data. We leave the challenge of calibrating such a model for future work.

Figure A.3: IRF Response



evidence suggesting a change of state is given by a shift in the average of acquired signals. Since the volatility of the target price remains constant, firms' learning problem is possibly more straightforward than the time-varying volatility setting. The level of total learning stays relatively constant over time. In particular, it randomly revolves around the initial attention level of 0.1 approximately.

As shown in Section 5.1 and consistent with Wiederholt (2010), in RI models, more attention is endogenously devoted to more volatile processes. Within a price-setting model, Maćkowiak and Wiederholt (2009) shows that the different attitudes towards volatility can explain why prices react more quickly to idiosyncratic relative to aggregate shocks. As idiosyncratic conditions are more volatile relative to aggregate ones, price-setters respond by focusing more attention on the former than the latter shock. The results with only a mean shock point exactly in this direction. While the results are robust to having both a mean and a variance shock, it is the presence of a rise in volatility that endogenously boosts learning over the business cycle, resembling the data.

References

Alvarez, Fernando E and Francesco Lippi, "A note on Price Adjustment with Menu Cost for Multi-product Firms," *Manuscript*, 2010.

Farmer, Leland E, "The Discretization Filter: A Simple Way to Estimate Nonlinear State Space Models," *SSRN Working Papers*, 2016.

_ **and Alexis Akira Toda**, "Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments," *Quantitative Economics*, 2017, *8* (2), 651–683.

Maćkowiak, Bartosz and Mirko Wiederholt, "Optimal sticky prices under rational inattention," *The American Economic Review*, 2009, *99* (3), 769–803.

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