Dynamic Inattention and the Phillips Curve under Optimal Dynamic Information Acquisition

Hassan Afrouzi*

<Preliminary Draft>

March 17, 2016

Abstract

This paper characterizes the solution for optimal information acquisition under rational inattention for Gaussian processes, and shows that limited attention creates a forward looking behavior among agents, even when their decisions depend only on the current realizations of the shocks. Applied to the pricing theory, the results of the paper imply a forward looking pricing scheme. In an economy with perfectly flexible prices, where firms' optimal prices depend only on the current realization of their marginal cost, agents face the following trade-off in choosing their information structure: on one hand, they want to learn the current level of their marginal cost, to which their current decision depends. On the other hand, however, they also have an incentive to learn about the best possible forecasts of future marginal costs as they do not want to make large mistakes in pricing when those periods arrive. Thus, they choose signals that not only inform them about their current marginal cost, but also about the best forecasts of its future realizations, which leads to a forward looking Phillips curve. This implies that, under forward guidance, when future monetary policy shocks are announced before their realization, firms optimally choose to acquire information about those shocks, and react to them before they happen, a feature that other models of information rigidity, such as sticky information or reduced-form noisy information models, miss.

1 Introduction

This paper proposes a new tractable approach for characterizing the optimal solution for dynamic rational inattention models with Gaussian fundamentals, and shows that rationally inattentive agents have a forward looking behavior in their information acquisition, even when their decisions only depend on the current realization of their fundamentals. In particular, agents face the following trade-off in gathering information about their fundamental: on one hand they want to know

*UT Austin, Department of Economics. 2225 Speedway C3100, Austin TX 78712 U.S.A. Email: afrouzi@utexas.edu

the current realization of the fundamental as their contemporaneous payoff depends on it; however, on the other hand, they also want to learn about the future path of their fundamental to minimize the distance of their perception when those periods arrive. This leads agents to choose signals that not only includes the current value of the fundamental, but also the best possible estimates of its future values. Therefore, when agents choose their optimal actions under such signals, a forward looking pattern in actions emerge as at each period the agent's information set incorporates the future path of the fundamental.

Applied to the pricing theory, this introduces a forward looking Phillips curve, an important feature that has been missing from the sticky and reduced-form noisy information models[1]. The importance of expectations of future inflation on its current realization has been the cornerstone of the modern analysis of monetary policy. This forward looking behavior has been microfounded in the economic literature by introducing price rigidities such as sticky prices or menu cost models. These models, however, has been criticized for not being able to match the inertial response of inflation to monetary policy shocks, a feature that has been shown to be consistent with sticky or noisy information models[2].

While noisy and sticky information models are consistent with the inertial response of inflation, I show that they induce a pricing behavior under which inflation does not depend on firms' expected future inflation. Therefore, each class of microfounded models of pricing fail to capture an important feature of the pricing behavior of the firms. Perhaps it is because of these shortcomings that, despite the lack of strong microfoundations, reduced form hybrid models of the Phillips curve, such as sticky prices with indexation[3], are widely used to assess different policies, as they have proven to be much more consistent with inflation dynamics observed in the data[4].

The results of this paper, applied to the pricing theory, introduces a microfoundation for inflation dynamics that incorporates both these features, even within a perfectly flexible pricing environment: inflation has an inertial response to shocks due to the fact that rationally inattentive firms have noisy information about them. More importantly inflation dynamics depend on firms' expectations of future inflation, as they optimally choose to allocate some attention to those and form a prior about it before the period arrives. Intuitively, when faced with limited attention, firms' optimal response is to prepare themselves for future to avoid big losses in profits over their lifetime.

To manifest the importance of this forward looking behavior in analyzing the effects of monetary policy, I consider a forward guidance exercise under rational inattention and compare it to sticky and reduced-form noisy information models. Forward guidance is modeled by assuming that shocks

[1]By reduced-form noisy information models, I refer to models in which agents are assumed to observe noisy signals of the fundamentals where their signal structure is assumed to be exogenously determined. In this sense, rational inattention models with Gaussian signals are microfounded noisy information models that endogenize information structure of the agents by allowing them to choose their signal structure in an optimal manner.

[2]See, for instance, Mankiw and Reis (2002); Woodford (2003).

[3]These are models that assume within sticky price models, firms who do not get to re-optimize, change their prices with a rule of thumb. They have been widely criticized as the rule of thumb pricing neglects the assumption of sticky prices that is the microfoundation of these models in the first place.

[4]See, for instance, Christiano et al. (2005)

to monetary policy can be observed beforehand. Since prices are flexible, firms within sticky and reduced-form noisy information models[5] do not respond at all to this sort of policy. However, I show that when firms are allowed to choose their information structure endogenously they optimally choose to pay attention to these news shocks, and incorporate that information in their pricing scheme. This leads to dynamics in which inflation responds to these shocks, even before they affect the fundamentals of firms.

While the form of the solution is given by the formulation of the attention problem, similar to rational expectation models, the solution has be solved numerically. However, in special cases, a closed form solution can be derived for the Phillips curve that allows us to understand how rational inattention affects the dynamics of inflation and output. For instance, when the aggregate demand follows a random walk, the Phillips curve is simply a proportional relationship between output and inflation, where its slope is an increasing function of the firms' capacity of processing information. This implies that both the magnitude and the persistence of the real effects of monetary policy are directly related to this capacity.

Moreover, a semi-closed form solution for the Phillips curve can be derived for the case when shocks to aggregate demand are announced one period ahead, and the aggregate demand, again, follows a random walk. The solution shows that rationally inattentive firms, who do not fully discount future losses, optimally choose signals that incorporates these future shocks. This leads to a Phillips curve in which inflation responds contemporaneously to these announced shocks. The degree of this response depends on two key parameters: the discount factor of the firms and their capacity of processing information. More patient firms chooses signals that put higher weights on announced shocks and therefore react more strongly to them. The effect of higher capacity, however, is more interesting. When capacity of processing information increases, firms choose to be less informed of future shocks. This is due to the fact that firms with higher capacity are more confident that they will know about these shocks when the time comes, and since these shocks do not affect their current fundamental they choose to simply ignore them.

This paper also contributes to the rational inattention literature in several dimensions. First, I characterize and solve the attention problem of the agent as a sequential problem of choosing priors and posteriors over time, for a given initial prior over the state of the economy. The solution method relies on the fact that any stationary Gaussian process can be approximated by an MA(T) process for an arbitrarily large T, and thus the attention problem boils down to choosing a vector of weights over the last T innovations of the process. The Euler equation of the attention problem is derived based on this approximation, which is then can be used to solve for the set of optimal signals. Furthermore, I show that even when the fundamentals are not stationary, they can be transformed to choosing an stationary part for the optimal signals based on the stationary parts of the fundamental. Thus, the method introduced in the paper can be used for any ARIMA process.

Second, this formulation sheds some light on the economic trade-off of the agent in choosing their information structure. Rationally inattentive agents are aware that they will never perfectly

[5]that assume agents only see the current realizations of their fundamentals.

observe the realizations of their fundamentals. Therefore, any signal that they get at a given period will serve them in two dimensions: first, it will give them a posterior about the current level of their fundamental, according to which they choose their optimal action, and second, it will equip them with a prior over future realization of that fundamental, so that when those periods arrive they would be able to better estimate what that fundamentals are. This dynamic trade-off manifests itself in the optimal signal that agents choose at every period: the signal not only incorporates information about the current fundamental, but also includes information about the best possible estimates of future fundamentals that can be formed at that period. In fact, the optimal signal will be a linear combination of the current fundamental and the estimates of its future realizations. Thus, the optimal signal of an agent for Gaussian processes is one that allows the agent to form expectations over current and future fundamentals: for example, in the case of an AR(2) fundamental, the optimal signal can be written as a linear combination of the current value of the fundamental and its lag, as those are sufficient in terms of forming expectations over any horizon. 1 shows that this result can be extended for the case of any ARMA process.

Section 2 characterizes the rational inattention problem of an agent who follows a single fundamental over time, Section 3 applies the results to the pricing theory.

2 Model

This section characterizes the attention problem of an agent who only follows a single stochastic process over time. For example, this could be a firm that only follows its marginal cost over time to decide on its optimal pricing strategy. While this section characterizes this problem in an abstract manner, in later sections I come back to this example, and study the economic implications of dynamic inattention in pricing.

The problem of an agent with limited attention, who follows a single stochastic process over time, has two stages: at each point in time, the agent first decides what information they want to gather about the stochastic process, and second, based on the information that they get, they decide on an optimal action.

I solve this problem in a backward manner: in Section 2.1, I characterize the optimal action profile for any arbitrary information structure, and then in Section 2.2, I present and solve the attention problem of the agent to an stationary process, where they optimize over a set of feasible information structures. Section 2.3 extends the results to an environment where the agent follows a difference stationary process to prepare the model to tackle the pricing problem of firms whose fundamentals are integrated of order one processes.

2.1 Environment Given an Information Structure

Suppose the agent tracks a fundamental that is characterized by a covariance stationary[6] Gaussian process $\{x_t : t = 0, 1, 2, \dots\}$. At each time t, x_t realizes, and then the agent chooses an action $a_t \in \mathbb{R}$. For a possible realization of the fundamentals $\tilde{x} \in \tilde{X} \equiv \{(x_t)_{t=0}^{\infty} | x_t \in \mathbb{R}, \forall t \geq 0\}$, and for a given sequence of actions $\tilde{a} = (a_0, a_1, a_2, \dots)$, the agent's realized payoff is

$$L_0(\tilde{a}, \tilde{x}) \equiv -\sum_{t=0}^{\infty} \beta^t (a_t - x_t)^2.$$

The agent does not observe $\{x_t : t \geq 0\}$ directly, but sees another stochastic process $\{s_t \in F : t = 0, 1, 2, \dots\}$ that is jointly distributed with the process x_t, where F is the set on which the signals are realized. Note that s_t can be a vector of signals instead of a single signal that are realized at time t. For any $t \geq 0$, let $s^t \equiv (s_0, s_1, \dots, s_t) \in F^t$ be a possible realization of the signals until time t, and let $S^t \equiv \{s^t | s^t \in F^t\}$ be the set of all possible realizations of signals until time t.

At each time t, having observed $s^t \in S^t$ the agent chooses an action $a_t \in \mathbb{R}$. Therefore, an action profile is a sequence of functions that map the set of signals to an action in $\mathbb{R}$. Let $\tilde{A}$ be the set of all possible action profiles:

$$\tilde{A} \equiv \{\tilde{a} = (a_t)_{t=0}^{\infty} | a_t : S^t \to \mathbb{R}, \forall t \geq 0\},$$

then the agent's problem in choosing the optimal action profile is

$$L_0 \equiv \min_{\tilde{a} \in \tilde{A}} \sum_{t=0}^{\infty} \beta^t \int_{s^t \in S^t} \int_{x_t \in \mathbb{R}} (a_t(s^t) - x_t)^2 f_t(x_t, s^t) dx_t ds^t$$

where for $s^t \in S^t$ and $x_t \in \mathbb{R}$, $f_t(x_t, s^t)$ is their joint density. The first order condition with respect to $a_t(s^t)$ is then

$$\begin{aligned} & \int_{x_t \in \mathbb{R}} (a_t^*(s^t) - x_t) f_t(x_t, s^t) = 0 \\ \Rightarrow \quad & a_t^*(s^t) = \int_{x_t \in \mathbb{R}} x_t \frac{f_t(x_t, s^t)}{\int_{x_t \in \mathbb{R}} f_t(x_t, s^t) dx_t} dx_t \\ \Rightarrow \quad & a_t^*(s^t) = \mathbb{E}\{x_t | s^t\}, \end{aligned}$$

where $\mathbb{E}\{.\}$ is the mathematical expectation operator. Under this optimal action profile, the expected net present value of all future losses boils down to a weighted average of the conditional

[6] This assumption will be relaxed in later sections.

variances of x_t:

$$\begin{aligned}\mathcal{L}_0 &= \sum_{t=0}^{\infty} \beta^t \int_{s^t \in S^t} \int_{x_t \in \mathbb{R}} (\mathbb{E}\{x_t|s^t\} - x_t)^2 f_t(x_t, s^t) dx_t ds^t \\ &= \sum_{t=0}^{\infty} \beta^t var\{x_t|S^t\}. \end{aligned} \tag{1}$$

Hence, the agent's objective in choosing her information structure would be minimize this weighted average of conditional variances over time subject to the informational constraints that she faces.

2.2 The Information Choice Problem

To characterize the agent's attention problem we need to specify two things; (1) the set of the objects to which the agent can pay attention at each time, and (2) the constraint that she faces in allocating her attention among those objects.

To specify the first one, since x_t is a covariance stationary Gaussian process, by Wold's theorem it can be decomposed to its innovation process:

$$x_t = \sum_{j=0}^{\infty} w_j u_{t-j},$$

where u_{t-j}'s are uncorrelated and the unconditional distribution of each of them is the standard normal. Since $\{x_t : t \geq 0\}$ is stationary, $\sum_{j=0}^{\infty} w_j^2$ is finite. This implies that for any arbitrary $\epsilon > 0$, $\exists T \in \mathbb{N}$ such that $\sum_{j=T+1}^{\infty} w_j^2 < \epsilon$, meaning that x_t can be approximated in a probabilistic sense by an MA(T) process:

$$\forall \epsilon > 0, \exists T \in \mathbb{N}, Pr(|x_t - \sum_{j=0}^{T} w_j u_{t-j}|) < \epsilon.$$

This approximation will be helpful in later sections in avoiding infinite dimensional covariance matrices, which may not exist or may not inherit the properties of their finite counterparts. Also, it justifies using a truncation of the process as I am going to use computational methods to solve for the solution, when a closed form does not exist.

For an arbitrarily large $T \in \mathbb{N}$, I use this approximation for the rest of the paper. Now, In matrix notation

$$x_t \approx \mathbf{w}'\mathbf{u}_t,$$

where $\mathbf{w} = (w_0, w_1, w_2, \ldots, w_T)'$ is the vector of weights, and $\mathbf{u}_t = (u_t, u_{t-1}, u_{t-2}, \ldots, u_{t-T})'$. I assume that at time zero, in addition to u_0, the nature also draws a sequence of $(u_{-i})_{i=1}^{T}$ from the standard normal. This decomposition gives us the finest set of independently distributed set of random variables that the agent *might want* to know, depending on her optimal attention strategy. Intuitively, since u_{t-i}'s are independent, paying attention to each of them does not reveal any information about the rest. Moreover, since at any given time $\forall \tau > t$, u_τ is not drawn by the nature yet, the vector $\mathbf{u}_t$ contains all the elements that agent *can* pay attention to at time t.

Also, to specify the information constraint, following the rational inattention literature, I assume that at any given point in time the agent cannot process more than κ bits of information, as measured by the reduction in entropy. Formally, this constraint is given by

$$\begin{aligned}\mathcal{I}(s_t, \mathbf{u}_t | S^{t-1}) &= \int_{(s_t,\mathbf{u}_t)} \log_2\left(\frac{f_t(\mathbf{u}_t, s_t)}{f_t(s_t) f_{t-1}(\mathbf{u}_t)}\right) d(s_t, \mathbf{u}_t) \\ &\leq \kappa.\end{aligned}$$

where $f_t(.)$ and $f_{t-1}(.)$ denote densities generated by S^t and S^{t-1} respectively. The information choice of the agent can in fact be viewed as choosing these joint distributions over time: at any time t, the agent inherits her chosen distribution, f_{t-1}, which gives her a prior about $\mathbf{u}_t$, and then chooses a new f_t subject to the above information constraint. I assume that at the beginning of time, $t = 0$, as the nature draws $\mathbf{u}_0$, the agent is born with a prior $f_{-1}(.)$ over $\mathbf{u}_{-1}$.

Therefore, the information problem of the agent at time zero is

$$\begin{aligned}&\min_{\{f_t\}_{t=0}^\infty} \mathcal{L}_0(f_{-1}) = \textstyle\sum_{t=0}^\infty \beta^t \int_{s^t \in S^t} \int_{x_t \in \mathbb{R}} (\mathbb{E}\{x_t | s^t\} - x_t)^2 f_t(x_t, s^t) dx_t ds^t \\ s.t.\quad &\mathcal{I}(s_t, \mathbf{u}_t | S^{t-1}) \leq \kappa\end{aligned}$$

Mackowiak and Wiederholt (2009) show that when the period loss functions are quadratic and priors are Gaussian, then the optimal signals under rational inattention are also Gaussian. While this result remains to be proven in this setting, for now, I assume that given a Gaussian initial prior at time zero the agent will choose Gaussian signals over time.

Hence, at any point in time, $t \geq 0$, the agent is born with a Gaussian prior over $\mathbf{u}_t$. Formally,

$$\mathbf{u}_t | S^{t-1} \sim \mathcal{N}(\mathbf{u}_{t|t-1}, \Sigma_{t|t-1})$$

where $\Sigma_{t|t-1} \equiv \mathbb{E}_{t-1}\{(\mathbf{u}_t - \mathbf{u}_{t|t-1})(\mathbf{u}_t - \mathbf{u}_{t|t-1})'\}$ is the covariance matrix of the agent's prior over $\mathbf{u}_t$ at time t.

Moreover, the set of all signals at time t is given by all the stationary Gaussian signals over $\mathbf{u}_t$:

$$\mathcal{S}_t^F \equiv \{s_t = \mathbf{y}'\mathbf{u}_t + e_t | \mathbf{y} \in \mathbb{R}^T, e_t \sim \mathcal{N}(0, \sigma_e^2), e_t \perp \mathbf{u}_t\}.$$

While it remains to be proven rigorously, for now I assume that at any $t \geq 0$, the agent only chooses to see a single signal in $\mathcal{S}_t^F$[7].

[7] This should follow from the fact that the agent only chooses a single action, a_t, at any $t \geq 0$: Suppose there is a solution in which the agent gets more than one signal at any time. Since the agent chooses a single action at any time, ultimately they combine these signals through an optimal policy function to their action. From previous section we know that given any information structure this action is simply the conditional expectation of x_t. As conditional expectations of Gaussian variables, given Gaussian signals, are linear, the optimal policy function for the agent's action is also linear in the underlying signals. However, since $\mathcal{S}_t^F$ is closed under linear operations, there is a single signal in $\mathcal{S}_t^F$ that generates the same process for the conditional expectations, and hence the same

Lemma 1. *Assume that the agent gets Gaussian signals over time. At any $t \geq 0$, given that $\mathbf{u}_t|S^{t-1} \sim \mathcal{N}(\mathbf{u}_{t|t-1}, \Sigma_{t|t-1})$, for any $s_t \in \mathcal{S}_t^F$, such that $s_t = \mathbf{y}'\mathbf{u}_t + e_t$, the information capacity constraint reduces to*

$$\mathbf{y}'\Sigma_{t|t-1}\mathbf{y} \leq (1 - 2^{-2\kappa})var\{s_t|S^{t-1}\}.$$

Proof. We use the entropy definition of the mutual information, and the fact that entropy of a Gaussian is a constant plus the log of its variance:

$$\begin{aligned} \mathcal{I}(s_t, \mathbf{u}_t|S^{t-1}) &= h(s_t|S^{t-1}) - h(s_t|\mathbf{u}_t, S^{t-1}) \\ &= \frac{1}{2}\log_2\left(\frac{var\{s_t|S^{t-1}\}}{var\{s_t|S^{t-1}\} - \mathbf{y}'\Sigma_{t|t-1}\mathbf{y}}\right) \end{aligned}$$

Thus $\mathcal{I}(s_t, \mathbf{u}_t|S^{t-1}) \leq \kappa \Leftrightarrow \mathbf{y}'\Sigma_{t|t-1}\mathbf{y} \leq (1 - 2^{-2\kappa})var\{s_t|S^{t-1}\}$. Q.E.D. □

Moreover, since inference is independent of scale, meaning that for $\alpha \neq 0$, s_t and αs_t contain the same information about $\mathbf{u}_t$, we can normalize the signals such that $var\{s_t|S^{t-1}\} = 1, \forall t \geq 0$. Let $\hat{\mathcal{S}}_t^F = \{\mathbf{y}|s_t = \mathbf{y}'\mathbf{u}_t + e_t \in \mathcal{S}_t^F, \ var\{s_t|S^{t-1}\} = 1\}$ be the set of all feasible signals that satisfy this normalization. Notice that the objects of $\hat{\mathcal{S}}_t^F$ are vectors of $\mathbf{y}$'s and not signals, as there is a one to one mapping between the two: every signal is a weighted average of the elements of $\mathbf{u}_t$, plus a white noise. From now on, I will refer to signals through these weight vectors. Intuitively, choosing a signal for the agent is nothing more than choosing how much weight she wants to put on each of u_{t-i}'s, for $i \geq 0$.

To pin down the dynamics of the attention problem, we need to specify how priors evolve over time as a function of the signal choices of the agent. Given a prior at time t, $\mathbf{u}_t|S^{t-1} \sim \mathcal{N}(\mathbf{u}_{t|t-1}, \Sigma_{t|t-1})$, and a signal choice, $\mathbf{y}_t \in \hat{\mathcal{S}}_t^F$, the agent's posterior at time t is given by the Kalman filter:

$$\begin{aligned} &\mathbf{u}_t|S^t \sim \mathcal{N}(\mathbf{u}_{t|t}, \Sigma_{t|t}) \\ \text{such that} \quad &\mathbf{u}_{t|t} = \mathbf{u}_{t|t-1} + \Sigma_{t|t-1}\mathbf{y}_t(s_t - \mathbf{y}_t'\mathbf{u}_{t|t-1}) \\ , \quad &\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1} \end{aligned} \tag{2}$$

Also, to derive the law of motion for the prior, notice that $\mathbf{u}_t$ itself evolves according to

$$\begin{aligned} \mathbf{u}_{t+1} &= \begin{bmatrix} u_{t+1} \\ \mathbf{u}_t \end{bmatrix} \\ &= \mathbf{M}\mathbf{u}_t + u_{t+1}\mathbf{e}_1, \forall t \geq -1 \end{aligned}$$

where $\mathbf{M}$ is the lower shift matrix[8], and $\mathbf{e}_1$ is the first column of the identity matrix. Since u_{t+1} is

expected loss.

[8] $\mathbf{M}$ is a $T \times T$ matrix with ones on its sub-diagonal and zeros elsewhere. Operating from left, it shifts a vector down by 1 element and sets the first element of the new vector to zero.

drawn at time $t+1$, it is orthogonal to all the agent's information until t. Hence,

$$
\begin{aligned}
& \mathbf{u}_{t+1}|S^t \sim \mathcal{N}(\mathbf{u}_{t+1|t}, \Sigma_{t+1|t}) \\
\text{such that} \quad & \mathbf{u}_{t+1|t} = \mathbf{M}\mathbf{u}_{t|t} \\
, \quad & \Sigma_{t+1|t} = \mathbf{M}\Sigma_{t|t}\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1'
\end{aligned} \tag{3}
$$

Therefore, given $\Sigma_{t|t-1}$, and a signal $\mathbf{y}_t \in \hat{\mathcal{S}}_t^F$, the agent's prior at $t+1$ is given by (2) and (3).

Now, by (1), and the fact that $var\{x_t|S^t\} = \mathbf{w}'\Sigma_{t|t}\mathbf{w}$, we can rewrite the agent's attention problem as[9]

$$
\begin{aligned}
\mathcal{L}_0(\Sigma_{0|-1}) \quad = \quad & \min_{\{\mathbf{y}_t \in \hat{\mathcal{S}}_t^F\}_{t=0}^\infty} \sum_{t=0}^\infty \beta^t \mathbf{w}'\Sigma_{t|t}\mathbf{w} \\
s.t. \quad & \mathbf{y}_t'\Sigma_{t|t-1}\mathbf{y}_t \leq 1 - 2^{-2\kappa} \\
& \Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1} \\
& \Sigma_{t+1|t} = \mathbf{M}\Sigma_{t|t}\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1' \\
& \Sigma_{0|-1} \succeq 0 \text{ given.}
\end{aligned} \tag{4}
$$

Theorem 1. *Given an initial prior, $\mathbf{u}_0 \sim \mathcal{N}(\mathbf{u}_{0|-1}, \Sigma_{0|-1})$, the optimal signals, $\{\mathbf{y}_t\}_{t=0}^\infty$, that solve the agent's attention problem as specified in (4) are given by the following Euler equation*

$$
\begin{aligned}
\phi_t \mathbf{y}_t &= (\mathbf{w}'\Sigma_{t|t-1}\mathbf{y}_t)\mathbf{w} + \mathbf{X}_t\Sigma_{t|t-1}\mathbf{y}_t \\
\mathbf{X}_t &= \beta\mathbf{M}'(\mathbf{w}\mathbf{w}' - \phi_{t+1}\mathbf{y}_{t+1}\mathbf{y}_{t+1}' + \mathbf{X}_{t+1})\mathbf{M}.
\end{aligned}
$$

where $\beta^t\phi_t$ is the Lagrange multiplier on the information capacity constraint, and $\mathbf{X}_t$ is the matrix of Lagrange multipliers on each constraints of the evolution of the prior. Moreover, let $\mathbb{E}_t^f\{.\} \equiv \mathbb{E}\{.|\mathbf{u}_t\}$ be the mathematical expectation operator of an agent with full information about $\mathbf{u}_t$. Then, the optimal signal of the agent at time t is of the following form:

$$
s_t^* = \sum_{j=0}^\infty \beta^j b_{j,t} \mathbb{E}_t^f\{x_{t+j}\} + e_t.
$$

for a set of real coefficients $\{(b_{j,t})_{j=0}^\infty\}_{t=0}^\infty$, and where $e_t \perp \mathbf{u}_t$ is the rational inattention error of the agent.

Proof. See Appendix A. □

Definition 1. We call an initial prior, Σ, a steady state prior if it reproduces itself over time, meaning that for Σ, $\exists\mathbf{y}$ such that if $\Sigma_{0|-1} = \Sigma$, then the constant sequence $\{\mathbf{y}\}_{t=0}^\infty$ solves the agent's attention problem, and $\Sigma_{t+1|t} = \Sigma, \forall t \geq 0$. This implies that $(\Sigma, \mathbf{y})$ should satisfy the following

[9] $\Sigma_{0|-1} \succeq 0$ means that $\Sigma_{0|-1}$ is positive semi-definite.

conditions:

$$
\begin{aligned}
\phi\mathbf{y} &= (\mathbf{w}'\Sigma\mathbf{y})\mathbf{w} + \mathbf{X}\Sigma\mathbf{y} \quad &, \\
\mathbf{X} &= \beta\mathbf{M}'(\mathbf{w}\mathbf{w}' - \phi\mathbf{y}\mathbf{y} + \mathbf{X})\mathbf{M} \quad &, \\
\Sigma &= \mathbf{M}(\Sigma - \Sigma\mathbf{y}\mathbf{y}\Sigma)\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1' \quad &, \\
\mathbf{y}'\Sigma\mathbf{y} &= 1 - 2^{-2\kappa} \quad &.
\end{aligned}
$$

Since the agent's attention problem is deterministic, the steady state prior can be thought of as the prior that emerges when the agent sees a sufficiently large number of signals. In later sections, when using computational methods, I will use this steady state prior to avoid time varying signals.

Corollary 1. *Suppose $\{x_t : t \geq 0\}$ follows an ARMA(p,q) process. Then, the optimal signal depends only on $p-1$ lags of x_t and $q-1$ lags of u_t. Formally, if $x_t = \sum_{i=1}^{p} \rho_i x_{t-i} + \sum_{j=0}^{q} \theta_j u_{t-j}$. Then,*

$$
s_t^* = \sum_{i=0}^{p-1} c_{i,t} x_{t-i} + \sum_{i=0}^{q-1} d_{i,t} u_{t-i} + e_t.
$$

for a set of real coefficients $\{(c_{i,t})_{i=0}^{p-1}, (d_{i,t})_{i=0}^{q-1}\}_{t=0}^{\infty}$. Moreover, these coefficients are time invariant in the steady state of the attention problem.

Proof. See Appendix A. □

Theorem 1 shows that the optimal signals are chosen under a forward looking behavior: each signal not only gives the agent information about the current state of the fundamental, but also it will be useful by shaping the agent's future priors. Each period, while the agent wants to know the realized value of x_t as precisely as possible, they also do not want to be "too" mistaken about future x_{t+i}'s when those days come. As a result they choose a signal that incorporates an optimal amount of available information[10] about each of x_{t+i}'s at time t.

This trade-off is represented in the Euler equation: the vector $\mathbf{y}_t$, which includes the optimal weights that the agent puts on each innovation, is a combination of $\mathbf{w}$, which represents how each innovation will affects current periods fundamental, and matrix $\mathbf{X}_t$, which represents how today's information will affect the evolution of the agent's prior about each innovation in the next period.

While this solution does not have a closed form in general, the following examples illustrate some its properties.

Example 1. Suppose $\beta = 0$, meaning that the agent fully discounts the future losses; then, the agent's optimal signal at time t is to observe x_t with the highest possible precision allowed by their capacity:

$$
s_t^* = x_t + e_t,\ e_t \sim \mathcal{N}(0, \frac{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}{2^{2\kappa} - 1}),\ e_t \perp u_{t-i}, \forall i \geq 0.
$$

where e_t is the agent's rational inattention error.

[10] Since future innovations are not realized at time t, the best information that the agent can get about x_{t+i} at time t is $\mathbb{E}_t^f\{x_{t+i}\} = \mathbb{E}\{x_{t+i}|\mathbf{u}_t\}$.

This result follows directly from the Euler equation in Theorem 1[11]. Intuitively, when the agent fully discounts the future, the evolution of the prior becomes irrelevant for them. At each period, they only care about minimizing that period's loss, and accordingly, they weigh each innovation exactly according to how that innovation affects their fundamental.

Nevertheless, this is not the only case where the agent chooses to only see x_t. The following example shows that even when the agent does not fully discount the future, meaning that $\beta > 0$, if x_t follows an AR(1) process, the optimal signal is the same as the one above. The reason is based on the very specific nature of the AR(1) process: while $\beta > 0$ implies that the agent cares about the evolution of its prior and wants to infer the future realizations of the fundamental, at any given time, the best forecast of any future fundamental is simply proportional to today's fundamental[12]; therefore, seeing x_t as precisely as possible is sufficient for inferring how it will evolve over time.

Example 2. Suppose x_t follows an AR(1) process such that $x_t = \rho x_{t-1} + u_t$, then the optimal signal at time t is given by

$$s_t^* = x_t + e_t,\ e_t \sim \mathcal{N}(0, \frac{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}{2^{2\kappa}-1}),\ e_t \perp u_{t-i}, \forall i \geq 0.$$

where $\mathbf{w} = (1, \rho, \rho^2, \dots)$, and $\mathbf{w}'\Sigma_{t|t-1}\mathbf{w} = \frac{1-(\rho^2 2^{-2\kappa})^t}{1-\rho^2 2^{-2\kappa}} + (\rho^2 2^{-2\kappa})^t \mathbf{w}'\Sigma_{0|-1}\mathbf{w}, \forall t \geq 1$. Also, agent's optimal action profile, $a_t^*(s^t)$, is given by

$$a_t^*(s^t) = 2^{-2\kappa}\rho a_{t-1}^*(s^{t-1}) + (1 - 2^{-2\kappa})s_t$$

Proof. See Appendix A. □

This result immediately breaks down if we move on to other processes as, in general, seeing x_t alone is not sufficient for the best possible inference of its future realizations. This intuition sheds light on the result in Corollary 1: for any ARMA(p, q) process, all $\mathbb{E}_t^f\{x_{t+i}\}$'s break down to seeing $(x_t, x_{t-1}, \dots, x_{t-p+1}, u_t, u_{t-1}, \dots, u_{t-q+1})$.

Example 3. For instance, if x_t follows stationary AR(2) such that $x_t = (0.95+\rho)x_{t-1} - \rho x_{t-2} + u_t$,[13] in order to form expectations over x_{t+i}, the agent needs to see both x_t and x_{t-1}, and according to Corollary 1, their optimal signal in the steady state of the attention problem is simply a weighted average of the two:

$$s_t^* = x_t + \gamma x_{t-1} + e_t,$$

where e_t is the rational inattention error of the agent[14]. While having x_t in their signal helps the agent to both predict the current realization of the fundamental, and prepare themselves with

[11] Here, the optimal signal is normalized such that the coefficient on x_t is equal to 1.

[12] $\mathbb{E}_t^f\{x_{t+i}\} = \rho^i x_t$.

[13] Here ρ measures the degree of inertia in the AR(2) process. For instance $\rho = 0$ corresponds to an AR(1), and $\rho > 0$ corresponds to AR(2) with a humped shape response to u_t.

[14] Again, since inference is independent of scale, the signal is presented with a normalization such that the coefficient on x_t is equal to 1.

shaping a better prior for predicting its future values, the presence of the parameter γ is purely due to the agent's desire to infer about the future realizations of the fundamental. In fact, in absence of this desire, as we saw in Example 1, the agent will choose γ to be zero. Hence, the magnitude of γ is directly linked to the agent's intertemporal incentive to acquire information.

An interesting exercise is to analyze how this intertemporal incentive depends on the underlying parameters of the model, namely the discount factor, β; the inertia parameter, ρ; and the agent's capacity of processing information κ. To do so I solve the attention problem computationally and plot the magnitude of the coefficient on x_{t-1} in the optimal signal, γ, versus different values of these underlying parameters[15].

Figure 1a shows that the intertemporal incentive of the agent in acquiring information increases with β. A higher discount factor means that the agent values future losses more, and hence has a higher incentive to minimize those losses by being able to predict future fundamentals more precisely. This, in turns, leads to a higher coefficient on x_{t-1} in the optimal signal that the agent gets at time t.

Figure 1b shows the degree of the agent's forward looking behavior increases with the degree of inertia in the AR(2) process. To better understand this result, first let us consider the case of $\rho = 0$, which corresponds to the AR(1) case in Example 1. Recall that with an AR(1), knowing x_t is sufficient for predicting the future realizations of the fundamental conditional on time t information. Therefore, the agent chooses to only see x_t as precisely as possible. However, as ρ increases, x_t is no longer sufficient for predicting future realizations of the fundamental, and the agent needs to include x_{t-1} in their signal to be able to do so. Therefore, with higher ρ's the agent will choose a higher weight on x_{t-1}.

Finally, the most interesting case is to see how capacity of processing information affects the agent's intertemporal incentive in acquiring information. Figure 1c shows that as the capacity increases[16] the agent's incentive to infer about future realizations of the fundamental decreases. The higher the capacity of processing information, the less concerned the agent is about figuring out what is going to happen in the future, as they will have enough resources to acquire sufficient information when the time comes. The case of $\lambda = 1$ shows that an agent with infinite capacity[17] is completely ignorant of the evolution of x_t over time, and chooses to only see x_t at any given time t. Moreover, their infinite capacity, however, guarantees them a perfectly precise signal that minimizes their life time losses to zero. On the contrary, when the capacity of processing information is low, the agent's optimal strategy is to get a signal that reveals information not only about the current state of their fundamental, but also about what it will be in the future.

Example 4. It is also an interesting exercise to see how the optimal action of the agent responds to a shock to the fundamental. Figure 2a shows the impulse response of the agent's action[18] to

[15] The baseline values set for the parameters are as follows: $\beta = 0.95$, $\rho = 0.5$ and $\kappa = 1$.

[16] We consider a monotone transformation of the capacity defined as $\lambda \equiv 1 - 2^{-2\kappa}$. λ is strictly increasing in the capacity of processing information. $\lambda = 0$ corresponds to zero capacity, $\kappa = 0$, and $\lambda = 1$ corresponds to infinite capacity, $\kappa \to \infty$.

[17] which corresponds to full-information rational expectations.

[18] As the agent's optimal action depends also on their rational inattention errors over time, I consider the mean

a 1% shock to an AR(1) fundamental, $x_t = 0.9x_{t-1} + u_t$. While the fundamental jumps to 1 on impact and decays exponentially, the agent responds with inertia. The reason is due to the noisy signal that the agent gets. On impact the agent sees a high signal, that, from their perspective, can be either due to an increase in their fundamental, or simply a rational inattention error. However, as time passes and they keep seeing high signals, they become more sure of the fact that the high signals are due to an increase in the fundamental, and their action catches up with the fundamental.

Moreover, Figure 2b shows the impulse response of the agent's average action to a 1% percent shock to an AR(2) fundamental, $x_t = 1.45x_{t-1} - 0.5x_{t-2} + u_t$. Recall from example 3 that unlike the case of the AR(1) fundamental, in this case the agent's optimal signal depends on how much they discount future losses. The dashed curve with circle markers shows the impulse response of the fundamental to a 1% shock; the dashed curve with square markers shows the path of the agent's action when they fully discount their future losses, and the dash-dotted curve with star markers shows the impulse response of the agent's action when they discount the future losses with rate $\beta = 0.99$. Again, the agent responds with inertia due to the same reasons as above.

An interesting observation is the difference between the response of the actions for the two different discount rates: first notice that both agents operate with the same capacity, so none of them can act systematically better than the other, meaning that it can not be the case the one of them has a path for their action that is always closer to the fundamental compared to the other one. This is represented in the graph by the fact that the patient agent reacts more confidently on the impact, but falling behind in later periods. In contrast, the impatient agent is relatively more uncertain of the increase on impact, as a result of which reacts with more hesitations, but they get to be closer to their ideal action in later periods. This difference is due to their information acquisition incentives: the patient agent is more forward looking in getting their signal than the impatient one. Hence, on impact when the fundamental is going to *increase*, the patient agent chooses a *higher* action than the impatient one. However, as soon as fundamental peaks and starts to *decrease*, the same forward looking behavior causes the patient agent to choose a *lower* action than the impatient one. Thus, the forward looking behavior created by the rational inattention motives, causes the agent to change the magnitude of their response due to the future path of the fundamental, even though that the agent does not face any kind of rigidities in choosing their action.

2.3 Attending to Difference Stationary Processes

So far we have only considered the case of stationary fundamentals, and characterized the solution of the attention problem under this assumption. However, in many economic problems agent's do not necessarily follow a stationary process. For instance, firms in the economy track their nominal marginal costs, whose levels are not stationary.

of their action over time.

In this section, I relax the stationarity assumption, and characterize the attention problem when the fundamental has a unit root. Suppose the environment is the same as the previous sections, but with the different that the agent follows a difference stationary Gaussian process $\{x_t : t = 0, 1, 2, \dots\}$, which implies that x_t is integrated of order 1. Therefore, since Δx_t is a stationary process, by Wold's theorem it can be decomposed to its innovations over time:

$$\Delta x_t = \mathbf{dw}'\mathbf{u}_t,$$

where $\mathbf{dw}' = (dw_0, dw_1, dw_2, \dots) \in \ell^2$ is a square-summable sequence and $\mathbf{u}_t = (u_t, u_{t-1}, u_{t-2}, \dots)'$ is the sequence of independently distributed innovations of Δx_t, with $u_{t-i} \sim \mathcal{N}(0, 1), \forall i$. Now let $\mathbf{M}$ be the infinite dimensional lower shift matrix[19]. Thus we can write

$$\begin{aligned} x_t &= \sum_{i=0}^{\infty} \mathbf{dw}'\mathbf{u}_{t-i} \\ &= \sum_{i=0}^{\infty} \mathbf{dw}'\mathbf{M}'^i\mathbf{u}_t \\ &= \mathbf{dw}'(\sum_{i=0}^{\infty} \mathbf{M}'^i)\mathbf{u}_t. \end{aligned}$$

where $\mathbf{M}'$ is the transpose of $\mathbf{M}$, and the second equality is derived from the fact that[20] $\mathbf{u}_{t-i} = (u_{t-i}, u_{t-i-1}, u_{t-i-2}, \dots) = \mathbf{M}'^i\mathbf{u}_t$. Notice that $\sum_{i=0}^{\infty} \mathbf{M}'^i$ is the upper triangular matrix whose i, j'th element is zero if $i > j$, and 1 if $i \leq j$, $\forall i, j$. Also, notice that $\mathbf{dw}'(\sum_{i=0}^{\infty} \mathbf{M}'^i)$ is a well-defined infinite dimensional vector whose i'th element is sum of the first i elements of $\mathbf{dw}$. Thus we can define the vector $\mathbf{w}$ such that

$$\mathbf{w} \equiv (\sum_{i=0}^{\infty} \mathbf{M}^i)\mathbf{dw}.$$

and

$$x_t = \mathbf{w}'\mathbf{u}_t.$$

Since the matrix $\sum_{i=0}^{\infty} \mathbf{M}^i$ is infinite dimensional, we have to be careful about inverting it, since, in general, infinite dimensional matrices do not necessarily inherit the properties of their finite dimensional counterparts. Let $\mathbf{I}$ be the infinite dimensional identity matrix: first, observe that $\mathbf{I} - \mathbf{M}$ is a well-defined matrix with ones on its diagonal and -1's on its subdiagonal. Second, observe that

$$[(\mathbf{I} - \mathbf{M})(\sum_{i=0}^{\infty} \mathbf{M}^i)]_{j,k} = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$$

[19] $\mathbf{M}$ is matrix with ones in its sub-diagonal and zero elsewhere. Operated from left, it shifts an infinite dimensional vector one element down, and replaces the first element of the new vector with zero.

[20] $\mathbf{M}'$ is simply the matrix representation of the lag-operator: $\mathbf{M}'\mathbf{u}_t = \mathbf{u}_{t-1}$ as $L.u_t = u_{t-1}$.

Thus, $(\mathbf{I}-\mathbf{M})\sum_{i=0}^{\infty}\mathbf{M}^i = \mathbf{I}$. Thus, we now define $(\sum_{i=0}^{\infty}\mathbf{M}^i)^{-1} \equiv \mathbf{I}-\mathbf{M}$, and we have[21]

$$\mathbf{dw} = (\mathbf{I}-\mathbf{M})\mathbf{w},\ \mathbf{w} = (\mathbf{I}-\mathbf{M})^{-1}\mathbf{dw}$$

Thus,

$$\begin{aligned} x_t &= \mathbf{w}'\mathbf{u}_t \\ &= \mathbf{dw}'(\mathbf{I}-\mathbf{M})^{-1}\mathbf{u}_t. \end{aligned}$$

Now, let $\tilde{x}_t$ be a random walk such that $\tilde{x}_t = \tilde{x}_{t-1} + u_t = \sum_{i=0}^{\infty} u_{t-i}$. Define $\tilde{\mathbf{u}}_t \equiv (\mathbf{I}-\mathbf{M})^{-1}\mathbf{u}_t$ and observe that

$$(\mathbf{I}-\mathbf{M})^{-1}\mathbf{u}_t = (\tilde{x}_t, \tilde{x}_{t-1}, \tilde{x}_{t-2}, \dots)',$$

Now, since $\mathbf{dw} \in \ell^2$, we can again truncate the process of x_t as follows[22]

$$x_t \approx \mathbf{dw}'\tilde{\mathbf{u}}_t$$

Where, with some abuse of notations, $\mathbf{dw} = (dw_0, dw_1, dw_2, \dots, dw_T)'$ and $\tilde{\mathbf{u}}_t = (\tilde{x}_t, \tilde{x}_{t-1}, \tilde{x}_{t-2}, \dots, \tilde{x}_{t-T})'$. Similarly, truncate the matrix $\mathbf{M}$ to a $T \times T$ lower shift matrix. Finally, observe that

$$\tilde{\mathbf{u}}_t = \begin{bmatrix} \tilde{x}_{t-1} + u_t \\ \tilde{x}_{t-1} \\ \tilde{x}_{t-2} \\ \vdots \\ \tilde{x}_{t-T} \end{bmatrix} = (\mathbf{M} + \mathbf{e}_1\mathbf{e}_1')\tilde{\mathbf{u}}_{t-1} + u_t\mathbf{e}_1$$

Where $\mathbf{e}_1$ is the first column of the $T \times T$ identity matrix and $u_t \sim \mathcal{N}(0,1)$, $u_t \perp \tilde{\mathbf{u}}_{t-1}$ is the time t innovation to the process. This brings us back to a problem similar to the previous section, but now the agent chooses a signal over $\tilde{\mathbf{u}}_t$. Similar to before, we assume that the agent starts with an initial prior over $\tilde{\mathbf{u}}_0 \sim \mathcal{N}(\tilde{\mathbf{u}}_{0|-1}, \Sigma_{0|-1})$, $\Sigma_{0|-1} \succeq 0$.

Now, to specify the agent's choice set of signals, we allow them to choose any signal over $\tilde{\mathbf{u}}_t$:

$$\mathcal{S}_t^F \equiv \{s_t = \mathbf{dy}'\tilde{\mathbf{u}}_t + e_t | \mathbf{dy} \in \mathbb{R}^T, e_t \sim \mathcal{N}(0, \sigma_e^2), e_t \perp \tilde{\mathbf{u}}_t\}.$$

Moreover, we again normalize the set of signals such $var_{t-1}(s_t) = 1$, as inference is independent of the scale of the signal. The set of signals at time t become

$$\hat{\mathcal{S}}_t^F = \{\mathbf{dy} | s_t = \mathbf{dy}'\tilde{\mathbf{u}}_t + e_t \in \mathcal{S}_t^F,\ var\{s_t | S^{t-1}\} = 1\}.$$

[21] In fact, the matrix $\mathbf{M}$ is the matrix representation of the lag-operator. The equation $(\mathbf{I}-\mathbf{M})\mathbf{w} = \mathbf{dw}$, simply corresponds to the fact that for any difference stationary process x_t, $(1-L)\phi(L)x_t = u_t$, where $\phi(L)$ is an invertible lag-polynomial.

[22] An argument similar to the case of stationary processes gives us this result, for any given prior over $\tilde{\mathbf{u}}_t$.

The agent's attention problem can now be re-written as

$$\begin{aligned}
\mathcal{L}_0(\hat{\Sigma}_{0|-1}) \quad &= \quad \min_{\{\mathbf{dy}_t \in \hat{\mathcal{S}}_t^F\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \mathbf{dw}' \Sigma_{t|t} \mathbf{dw} \\
s.t. \qquad & \mathbf{dy}_t' \Sigma_{t|t-1} \mathbf{dy}_t \leq 1 - 2^{-2\kappa} \\
& \Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} \mathbf{dy}_t \mathbf{dy}_t' \Sigma_{t|t-1} \\
& \Sigma_{t+1|t} = (\mathbf{M} + \mathbf{e}_1 \mathbf{e}_1') \Sigma_{t|t} (\mathbf{M}' + \mathbf{e}_1 \mathbf{e}_1') + \mathbf{e}_1 \mathbf{e}_1' \\
& \Sigma_{0|-1} \succeq 0 \text{ given.}
\end{aligned} \tag{5}$$

This is now a choice problem within stationary signals as before, meaning that we have re-written the agent's problem in terms of choosing the stationary part of their signal $\mathbf{dy}_t$, given the stationary part of their fundamental $\mathbf{dw}$. Therefore, we can use the method presented in Theorem 1 to derive the Euler equation of the agent's problem:

$$\begin{aligned}
\phi_t \mathbf{dy}_t &= (\mathbf{dw}' \Sigma_{t|t-1} \mathbf{dy}_t) \mathbf{dw} + \hat{\mathbf{X}}_t \Sigma_{t|t-1} \mathbf{dy}_t \\
\hat{\mathbf{X}}_t &= \beta (\mathbf{M}' + \mathbf{e}_1 \mathbf{e}_1') (\mathbf{dw}\mathbf{dw}' - \phi_{t+1} \mathbf{dy}_{t+1} \mathbf{dy}_{t+1}' + \hat{\mathbf{X}}_{t+1}) (\mathbf{M} + \mathbf{e}_1 \mathbf{e}_1').
\end{aligned} \tag{6}$$

Where ϕ_t is the Lagrange multiplier on the information processing constraint, and $\hat{\mathbf{X}}_t$ is the matrix of Lagrange multipliers on the evolution of the priors.

Lemma 2. *Suppose that the agent's fundamental follows a first order integrated ARIMA process x_t. Then the optimal signals are of the form*

$$s_t^* = \sum_{j=0}^{\infty} \beta^j b_{j,t} \mathbb{E}_t^f \{x_{t+j}\} + e_t$$

where $\mathbb{E}_t^f\{.\} \equiv \mathbb{E}\{.|\tilde{\mathbf{u}}_t\}$ is the expectation operator of an agent with full information at time t, and e_t is the agents rational inattention error, and $\{(b_{j,t})_{j=0}^{\infty}\}_{t=0}^{\infty}$ is a set of sequences of real coefficients that are given by the Euler equation above.

Proof. See Appendix A. □

Corollary 2. *Suppose the agent's fundamental follows an ARIMA$(p,1,q)$ process, then the optimal signal is a linear combination of $(x_t, x_{t-1}, \ldots, x_{t-p+1}, u_t, u_{t-1}, \ldots, u_{t-q+1})$:*

$$s_t^* = \sum_{k=0}^{p-1} c_{k,t} x_{t-k} + \sum_{l=0}^{q-1} d_{l,t} u_{t-k},$$

where $\{(c_{k,t})_{k=0}^{p-1}, (d_{l,t})_{l=0}^{q-1}\}$ are a set of real coefficients.

Proof. See Appendix A. □

After solving for the optimal signal through the Euler equation, the evolution of the optimal action, $a_t^*(s^t) = \mathbb{E}\{x_t|s^t\}$, will be then given by the Kalman filter. Also, similar to before we can define a steady state for the problem as an initial prior that reproduces itself over time.

3 Application: A Rational Inattention Phillips Curve

In this section, we apply the results to derive a Rational Inattention Phillips curve.

3.1 Environment

Assume that there is a measure 1 of firms indexed by $i \in [0,1]$. There is a price taking final good producer that assembles the products of these firms to a single consumption good through a CES aggregator. This implies that the demand function of firm i is given by

$$Y_{i,t} = Y_t(\frac{P_{i,t}}{P_t})^{-\sigma}$$

where $Y_{i,t}$ is i's output, $P_{i,t}$ is its chosen price, Y_t is the aggregate output and P_t is the aggregate level of prices. Firm i's flow profit function is given by

$$\Pi(P_{i,t}; P_t, Y_t) = P_{i,t}^{1-\sigma} P_t^{\sigma} Y_t - TC(P_{i,t}; P_t, Y_t)$$

where the first term is the firm's revenue and the second term is a function that maps the firms price, and the aggregate variables, to its total cost of production[23]. Let $P_t^* = P^*(P_t, Y_t) \equiv \arg\max_x \Pi(x; P_t, Y_t)$ be the argmax of this flow profit function at time t. Thus,

$$\Pi_1(P_t^*; P_t, Y_t) = 0.$$

A Taylor expansion of this first order condition around an optimal non-stochastic point $(P; P, Y)$[24] gives

$$p_t^* = p_t + |\frac{\Pi_{13}}{\Pi_{11}}\frac{Y}{P}| y_t$$

where small letter denote the log-deviation from the optimal non-stochastic point around which we have linearized the equation, and $\alpha \equiv |\frac{\Pi_{13} Y}{\Pi_{12} P}|$ is the degree of strategic complementarity. Now, define the function $L(P_{i,t}; P_t, Y_t)$ as the flow loss in the profit of firm i for any given price $P_{i,t}$:

$$L(P_{i,t}; P_t, Y_t) = \Pi(P_t^*; P_t, Y_t) - \Pi(P_{i,t}; P_t, Y_t).$$

[23]We assume this function is twice differentiable in all its arguments, and convex in $P_{i,t}$ so that the maximum exists. Moreover, assume that $TC(.;.,.)$ is homogeneous of degree zero in its first two arguments so that only relative price of the firm matters.

[24]The CES aggregation implies that, due to symmetry, in a non-stochastic optimal point all firms charge the same price which turns to be the aggregate price.

It is straight forward to show that this loss function, up to a second order approximation is proportional to the quadratic difference between $p_{i,t}$ and p^*_t.

$$L(P_{i,t}; P_t, Y_t) = \frac{1}{2}|\Pi_{11} P^2|(p_{i,t} - p^*_t)^2.$$

Thus $p^*_t = p_t + \alpha y_t$ is the firms' fundamental, and given its process, the firm's problem is the same as the one in section 2.1. Finally, to close the model, following the literature[25], I assume that the aggregate nominal GDP, $Q_t \equiv P_t Y_t$, is exogenous to the decision of firms, and is set by the monetary authority. This implies

$$p^*_t = (1-\alpha)p_t + \alpha q_t$$

Specifically, I assume that the growth rate of nominal GDP follows an ARIMA(1,1,0):

$$\Delta q_t = \rho \Delta q_{t-1} + u_t.$$

3.2 The Equilibrium

Let $\tilde{\mathbf{u}}_t$ be the vector of the random walk part of the nominal GDP until time t, as defined in section 2.3, for an arbitrarily large truncation of the process $T \in \mathbb{N}$. Thus,

$$q_t \approx \mathbf{dw}'_q \tilde{\mathbf{u}}_t \text{ s.t. } \mathbf{dw}_q \equiv (1, \rho, \rho^2, \ldots, \rho^T).$$

Each firm takes the process of p^*_t as given and given a prior over $\tilde{\mathbf{u}}_0$ solves a rational inattention problem as defined in previous sections. We assume that agents' rational inattention errors are orthogonal in the cross section so that the aggregate price only depends on $\tilde{\mathbf{u}}_t$. Since q_t follows a difference stationary process, the attention problem of the agents are similar to the one discussed in section 2.3.

Thus, a symmetric steady state rational inattention equilibrium to the model is a pair of steady state prior and signal $(\Sigma, \mathbf{dy})$, and a set of vectors $\{\mathbf{dw}_{p^*}, \mathbf{dw}_p\}$ such that

1. Given that $p^*_t = \mathbf{dw}'_{p^*} \tilde{\mathbf{u}}_t$, the constant sequence $\{(\Sigma_{t+1|t} = \Sigma, \mathbf{dy}_t = \mathbf{dy})\}_{t=0}^{\infty}$ is a solution to each firms' rational inattention problem

$$\begin{aligned}
\mathcal{L}_0(\hat{\Sigma}) \quad &= \quad \min_{\{\mathbf{dy}_t \in \hat{\mathcal{S}}^F_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \mathbf{dw}'_{p^*} \Sigma_{t|t} \mathbf{dw}_{p^*} \qquad (7) \\
s.t. \quad & \quad \mathbf{dy}'_t \Sigma_{t|t-1} \mathbf{dy}_t \leq 1 - 2^{-2\kappa} \\
& \quad \Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} \mathbf{dy}_t \mathbf{dy}'_t \Sigma_{t|t-1} \\
& \quad \Sigma_{t+1|t} = (\mathbf{M} + \mathbf{e}_1 \mathbf{e}'_1) \Sigma_{t|t} (\mathbf{M}' + \mathbf{e}_1 \mathbf{e}'_1) + \mathbf{e}_1 \mathbf{e}'_1 \\
& \quad \Sigma_{0|-1} = \Sigma \text{ given.}
\end{aligned}$$

[25] See, for instance, Mackowiak and Wiederholt (2009); Woodford (2003); Mankiw and Reis (2002).

2. Given the the set of $\{p_{i,t}\}_{i\in[0,1]}$, where each $p_{i,t}$ is implied by the Kalman filtering of the sequence of optimal signals $\{s^*_{i,t} = \mathbf{dy}'\tilde{\mathbf{u}}_t + e_{i,t}\}_{t=0,\forall i\in[0,1]}^{\infty}$, $\mathbf{dw}_p$ is such that

$$p_t = \int_0^1 p_{i,t} di = \mathbf{dw}_p'\tilde{\mathbf{u}}_t.$$

3. Given $\mathbf{dw}_q$ and $\mathbf{dw}_p$, $\mathbf{dw}_p^*$ is such that

$$p_t^* = (1-\alpha)p_t + \alpha q_t \Leftrightarrow \mathbf{dw}_{p^*} = (1-\alpha)\mathbf{dw}_p + \alpha\mathbf{dw}_q.$$

Such a solution can be derived by iteration. I start with guessing a process for $\mathbf{w}_{p^*}$, in particular, $\mathbf{w}_{p^*} = \mathbf{w}_q$, the solution to the model when $\kappa \to \infty$. Given the guess, I solve for $(\Sigma, \mathbf{y})$ using the steady state Euler equation of the attention problem derived in equation 6. Then, given the sequence of optimal signals implied by $(\Sigma, \mathbf{y})$, we find $\mathbf{dw}_p$ such that $\int_0^1 p_{i,t} di = \mathbf{dw}_p'\tilde{\mathbf{u}}_t$ using the Kalman filter. Finally, given $\mathbf{dw}_p$, we update our guess of $\mathbf{dw}_{p^*} = (1-\alpha)\mathbf{dw}_p + \alpha\mathbf{dw}_q$, and iterate until convergence.

3.3 Some Special Cases

While the solution to the model does not have a nice representation in general, we can investigate its properties through some special cases of the model, for which a representation of the solution can be derived.

Example 5. ***(The case of a random walk with no strategic complementarity)*** Suppose that the aggregate demand is a random walk, meaning that $\Delta q_t = u_t$. Also assume that there is no strategic complementarity in pricing, $\alpha = 1$, then the Phillips curve under rational inattention is given by

$$\pi_t = (2^{2\kappa} - 1)y_t$$

which together with the evolution of the aggregate demand implies that output and inflation both follow an AR(1) process:

$$\begin{aligned} y_t &= 2^{-2\kappa} y_{t-1} + 2^{-2\kappa} u_t \qquad , \\ \pi_t &= 2^{-2\kappa}\pi_{t-1} + (1 - 2^{-2\kappa})u_t \end{aligned}$$

Proof. See Appendix A. □

Woodford (2003) made the argument that noisy information models are well-equipped for matching the persistence of the real effects of monetary policy, as observed in the data; a feature that early models of information rigidity, such as Lucas (1972), failed to generate. In spite of its very restrictive parameterization, the closed form solution of this example sheds light on how rational inattention can create an endogenous real and persistent effect for monetary policy, where both directly depend on firms' capacity of processing information.

Figure 3 shows the impulse responses of inflation and output to a 1% shock to the aggregate demand, for different values of κ. Lower capacity corresponds to a smaller response of output on impact, which is accompanied by a larger response for inflation. The persistence of the effect is lower for both inflation and output, when capacity is higher. For instance, for a very large capacity, the shock has no effect on output at all. Moreover, inflation responds one to one to the shock and is zero after the first period, meaning that there is zero persistence in its response.

A shortcoming of this example, however, is that it fails to present the dynamic effects of rational inattention, as the solution is independent of β, the discount factor of the firms. The reason for this independence relates to the very specific nature of a random walk. Each innovation has a symmetrically permanent effect on the optimal price of the firms, which translates into a signal that is independent of how patient the firms are. Nevertheless, this is not true in general. For instance, later in Example 7 I show that even with a random walk fundamental, the optimal signal depends on β when shocks are announced beforehand.

Example 6. ***(The effect of strategic complementarity when*** $\beta = 0$***)*** Suppose $\beta = 0$, then the optimal set of signals is given by $s_t^i = p_t^* + e_t^i, \forall t, \forall i$. Also the Phillips curve under rational inattention is

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\{\pi_t + \alpha \Delta y_t\} + \alpha(2^{2\kappa} - 1) y_t$$

where $\tilde{\mathbb{E}}_t\{.\} \equiv \int_0^1 \mathbb{E}\{.|S^{i,t}\} di$ is the average expectation of firms given the optimal signal structure.

Proof. See Appendix A. □

The full discounting of future profit losses, $\beta = 0$, leads firms to choose signals of their current fundamental, p_t^*, and gives rise to a Phillips curve without any forward looking behavior. The semi-closed form of the Phillips curve, however, allows us to visualize the effect of firms' capacity and strategic complementarity on dynamics of inflation, as the slope of curve depends only on the two. This slope increases with higher capacity or strategic complementarity, which leads to the intuition that inflation should be more inertial when either of these parameters are lower.

Moreover, this is an example with endogenous feedback in formation of firms' expectations: firms get a signal of their fundamental $p_t^* = (1 - \alpha) p_t + \alpha q_t$, and choose their price at each period given the sequence of their signals over time. This, in turn, shapes the path of the fundamental as it depends on the aggregate price through α. Thus, intuitively, p_t^* should follow a more inertial path as the strategic complementarity increases, which would lead to a more inertial path for aggregate prices.

Figure 4 shows the impulse responses of inflation and output to a 1% shock to aggregate demand for two different values of strategic complementarity, when $\kappa = 0.2$. As expected, inflation follows a more inertial path in presence of higher strategic complementarity, which in turn translates to a more amplified response for output.

Moreover, Figure 5 shows the impulse responses of inflation and output for different levels of capacity of processing information[26]. Higher capacity corresponds to a higher response of inflation

[26]The value of strategic complementarity in these responses is set to $\alpha = 0.5$.

on impact, and less persistence as well as less humped-shape behavior for it over time. In the extreme case of a very high capacity the response of inflation corresponds one to one to the response of the growth of the aggregate demand, which is an AR(1) by assumption. Output, on the other hand, responds more strongly to monetary policy when capacity is lower. In fact, monetary policy is neutral when capacity is very high.

3.4 Attention Allocation under Forward Guidance

So far we have considered some special cases of the model in which either β was assumed to be zero (Example 6) or it was irrelevant due to strong assumptions on the nature of the fundamental (Example 5). In this section, I present examples and results that illustrate the forward looking behavior that a positive β induces through rational inattention, and compare it to other models of information rigidity such as noisy information and sticky information models. I start with a simple example of forward guidance, for which a closed form Phillips curve can be derived.

Example 7. ***(Forward Guidance under Rational Inattention)*** Suppose that there is no strategic complementarity in pricing and $\Delta q_t = u_{t-1}$. This corresponds to a monetary policy in which the shocks are announced a period before they take effect. While the fundamental of the firms has the same process as in example 5, the difference is that here firms have the option to pay attention to the shock that is going to take effect in the following period. In fact, the optimal signal incorporates information about Δq_{t+1}, and is given by

$$s_t = q_t + \gamma \Delta q_{t+1} + e_t$$

where γ is implicitly characterized by the following two equations as a function of the discount factor, β, and the capacity of processing information, κ:

$$\begin{gathered} \tfrac{(1-\beta)\gamma+\beta\delta}{1-\gamma} = \tfrac{\beta}{1-\gamma\delta} \\ (1-\delta(1-\gamma))(1-\delta\gamma) = 2^{-2\kappa} \end{gathered}$$

Here, γ is the optimal weight that firms put on $u_t = \Delta q_{t+1}$, the news shock about the next period monetary policy, relative to q_t, their current fundamental. The purpose of this example is therefore to see how this optimal weight depends on the two parameters of the model, and whether inflation or output respond to the news shock. Even though, δ, which is shown below to be related to persistence of the response of output, cannot be eliminated in deriving a closed form solution for γ, we can derive a closed form solution of the Phillips curve:

$$\pi_t = \delta \frac{\gamma\delta}{1-\delta(1-\gamma\delta)}(y_{t-1} + \gamma u_{t-1}) + \delta \frac{1-\gamma\delta}{1-\delta(1-\gamma\delta)}(y_t + \gamma u_t)$$

which implies that inflation not only responds to current output y_t, and the current shock to aggregate demand, u_{t-1}, but also to the news shock u_t. Notice that the response of inflation to

u_t is proportional to γ, the optimal weight on u_t in the optimal signal. Moreover, we can also characterize the joint equilibrium path of inflation and output over time:

$$\pi_t = \delta y_{t-1} + (1 - 2^{-2\kappa})u_{t-1} + \gamma\delta(1 - \gamma\delta)u_t$$

$$y_t = (1 - \delta)y_{t-1} + 2^{-2\kappa}u_{t-1} - \gamma\delta(1 - \gamma\delta)u_t$$

The response of output and inflation to the news shock is proportional to γ, and zero in net as the shock is set to affect the aggregate demand in the next period. Moreover, we now have an interpretation for $1 - \delta$: it is the persistence of the response of output to the shocks. Figure 6 shows the equilibrium values of γ and $1-\delta$ for different levels of capacity and patience: the dashed blue curves depict iso-capacity curves in the $(\gamma, 1-\delta)$ space, and the red solid lines are iso-patience curves. Each intersection is an equilibrium that corresponds to that particular level of capacity and patience. Notice that higher β always corresponds to higher value of γ: more patient firms have a higher incentive to know about the future path of their fundamental. The more interesting observation is that higher capacity always corresponds to a lower γ: firms with a larger capacity are more confident that when the time comes they will be able to recognize their fundamental and therefore choose to ignore the news shock, and pay a higher portion of their capacity to their current fundamental. Moreover, higher capacity also translates to a lower persistence in response of output to shocks, an observation similar to example 5.

Figure 7 shows the impulse responses of output and inflation in this setting under full discounting of future losses, $\beta = 0$, and $\beta = 0.99$. When $\beta = 0$, the model behaves the same as in example 5: firms completely ignore the news shock and wait until the time that the shock hits to get information and react to it. However, when $\beta > 0$, firms optimally choose to pay attention to the news shock and increases with an announced positive monetary policy shock: at the time of the announcement firms get a high signal, but as they are not able to perfectly differentiate between the current shock and the future shock they start increasing their prices immediately. Since the aggregate demand has not increased yet at time zero, output falls to compensate for the increase in prices. Intuitively, this result can be interpreted as follows: a rationally inattentive firm that cares about its future losses will optimally choose to be informed about news of monetary policy; however, this does not imply that they will have sufficient information to perfectly differentiate the news about future monetary policy from current policy. Accordingly, news about future shocks will have a real affect on the current state of the economy.

Proof. See Appendix A. □

To depart from this example, suppose now that $\Delta q_t = \rho \Delta q_{t-1} + u_{t-\tau}$, where τ is the degree of forward guidance: shocks to aggregate demand are announced τ periods before taking effect. The goal here is to compare dynamic inattention with reduced-form noisy information and sticky information models.

For reduced-form noisy information models, I simply consider the case of $\beta = 0$, which corresponds to a setting when firms choose to only observe their current fundamental, as shown above in example 6. This corresponds to a setting that is similar to the noisy information models, which exogenously assume an information structure for the agents in which the agent sees their current fundamental with an observation error. Recall from Example 6 that the Phillips curve in this setting is given by

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\{\pi_t + \alpha \Delta y_t\} + \alpha(2^{2\kappa} - 1)y_t,$$

Where $\tilde{\mathbb{E}}_{t-1}\{.\} \equiv \int_0^1 \mathbb{E}\{.|S^{i,t}\}di$ denotes the average expectation of firms conditional on their time $t-1$ information given by the signal vector $S^{i,t} = (p_0^* + e_{i,t}, \dots, p_{t-1}^* + e_{i,t})$, and α is the strategic complementarity in pricing.

Sticky information models assume that at each period only a fraction of firms update their information, but those who do acquire perfect information about the state of the economy and their expectations correspond to those of fully informed agent. For these models, I use the Phillips curve derived in Mankiw and Reis (2002):

$$\pi_t = \hat{\mathbb{E}}_{t-1}\{\pi_t + \alpha \Delta y_t\} + \alpha \frac{\lambda}{1-\lambda} y_t,$$

where λ is the fraction of the firms that update their information at each period, and $\hat{\mathbb{E}}_{t-1}\{.\} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E}_{t-j-1}^f\{.\}$ is the average expectation of firms at time $t-1$.

The similarity of the two Phillips curves is not a coincidence. In both models the response of inflation, and the real effects of monetary policy, depends on two things: the apriori expected changes in marginal cost, represented by time $t-1$ expectation term, and a surprise element represented by the coefficient on y_t: in both models as the degree of friction reduces, either by a higher capacity of processing information or a higher fraction of firms updating their information, the slope becomes steeper, and in the limit converges to a vertical Phillips curve, in which there can be no surprises in monetary policy and therefore no real effects. The fact that only y_t appears in the Phillips curve corresponds to the fact that there is no forward looking behavior in neither of these models: in the noisy information models it is by the assumption that firms have no incentive to do so by construction. In the sticky information model, it is because of the fact that firms who update their information can perfectly differentiate between current shocks, and future ones. Accordingly, in choosing their prices, they only incorporate information that is relevant for their current prices, and keep the information about future shocks out of their decision.

The Phillips curve under dynamic inattention is harder to derive analytically, but possible to a degree:

Lemma 3. ***(The Phillips Curve under Dynamic Inattention)*** *By Lemma 2 that the optimal signals under dynamic inattention has the form $s_t^* = \sum_{j=0}^{\infty} \beta^j b_j \mathbb{E}_t^f\{p_{t+j}^*\} + e_t$. Then, given the sequence $(b_j)_{j=0}^{\infty}$, the Phillips curve under dynamic inattention is given by*

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\{\pi_t + \Delta y_t\} + 2^{2\kappa}\delta_0 \sum_{j=0}^{\infty} \beta^j b_j (\mathbb{E}_t^f\{p_{t+j}^*\} - \tilde{\mathbb{E}}_t\{p_{t+j}^*\}),$$

where $\mathbb{E}_t^f\{.\}$ *is the time* t *expectation operator of a fully informed agent, and* $\tilde{\mathbb{E}}_t\{.\}$ *is the average expectation of firms at time* t*.*

Proof. See Appendix A. □

While we have to solve for $(b_j)_{j=0}^{\infty}$ numerically, this representation illustrate how dynamic inattention introduces a forward looking behavior among agents. Each element of the infinite sum represents the gap between a fully informed agent's expectation and that of the firms. The larger these gaps, the larger is the effect of monetary policy. More importantly, current inflation not only to current output, which is embedded in the first element of the infinite sum[27], but also on all the future expectation gaps. Therefore, by altering these gaps, any forward guidance policy can have immediate effects on inflation, and consequently output.

Figure 8 shows the impulse responses of the three models to an announced 1% shock to aggregate demand that is going to take effect in three periods ($\tau = 3$). For this exercise I have set $\kappa = 0.5$, $\rho = 0.5$, and $\alpha = 0.8$. Moreover, in the sticky information model, I have set $\lambda = 0.2$, so that the peak of output and inflation in this model and noisy information (naive inattention) one would be the same[28]. In both sticky information and naive inattention models the announced shock has no effect, as in the first firms completely ignore it due the fact that it does not affect their current fundamental, and in the latter firms who have updated their information can perfectly differentiate it from the current shocks. In both these models, it is only after the shock takes effect that firms start to respond to it. Output and inflation are more persistent in the sticky information model because this model needs a relatively large amount of friction, $\lambda = 0.2$, to have the same peak effect on output and inflation.

Unlike the former models, the dynamic inattention model exhibits immediate effects for the announced shock: rationally inattentive firms who care about their future losses optimally choose to be informed about future policy, but are not able to perfectly differentiate future shocks from current ones due to their limited capacity in processing information. Therefore, upon getting a high signal at time zero, they immediately respond by increasing which translates to an immediate increase in response of inflation. Since the shock has not taken effect yet, output starts to fall to compensate for the increase in prices. The peak of output is larger, however, when the shock takes effect. This is due to the fact that firms are actively paying attention to future shocks, which comes at the cost of being less informed about past ones compared to the naive inattention model. Therefore, both inflation and output demonstrate more inertial behavior under dynamic inattention.

---

[27]Observe that the first element of the infinite sum is $p_t^* - \tilde{\mathbb{E}}_t\{p_t^*\} = \alpha y_t$.

[28]There is no clear way that how these models should be compared. However, since the goal is to eventually match these models to the observed behavior of output and inflation in the data, it seems reasonable to compare them in such a manner.

4 Conclusion

This paper proposes a new tractable method for solving dynamic rational inattention problems with Gaussian fundamentals and shows that rationally inattentive agents manifest a forward looking behavior in choosing their information. This forward looking behavior emerges due to a dynamic trade-off for the agents: at each period not only the information structure of the agent serves them by providing a posterior about their current fundamental, and hence their optimal decision, but also by forming a prior about future states of the fundamental by shaping their future priors. Faced by this trade-off, agents optimally choose to acquire information about both current and best possible estimates of future fundamentals. Acting on such an information structure, agents' actions exhibit a forward looking pattern: these actions respond to future expectations of fundamentals, even though agents do not face any rigidity in choosing them.

I apply this result to the pricing theory, and show that a Phillips curve that emerges under dynamic rational inattention relates current inflation to the expected future inflation, a feature that been missing from other models of information rigidity such as reduced form noisy information and sticky information models. Also, since agents choose their actions under imperfect information, this Phillips curve also replicates the inertial response of inflation and output to monetary policy shocks.

These two characteristics, the dependence of current inflation to expected future inflation, and the inertial behavior of it through depending on its past realizations, have been shown to be necessary to match the observed pattern of it in the data. However, current models of microfounded pricing, such as sticky prices, menu costs, and sticky and noisy inflation models fail to capture both of these features, with the former two missing the inertial pattern of inflation, and the latter two by missing its forward looking behavior. This has lead to the use of hybrid Phillips curves, such as sticky prices with indexation, that has been criticized for ignoring the underlying microfoundation through the assumption that firms with sticky prices change their prices with a rule of thumb.

In order to demonstrate the forward looking behavior that is microfounded under dynamic rational inattention, I implement a simple forward guidance model, in which shocks to aggregate demand is announced before taking effect. I show that while sticky information and reduced-form noisy information models fail to generate any response to these news shocks before they affect the aggregate demand, rationally inattentive firms optimally choose to attend to these news shocks, and respond to them before they take effect.

The huge interest in, and appeal to, forward guidance policies during the years after the Great Recession, lead by the belief that economies respond to news about future policies has been dampened by lack of adequate models to analyze the effects of such policies. Consequently, while other models of information rigidity fail to incorporate the dynamic effects of forward guidance policies, and therefore are incapable for any analysis of forward guidance policies, the dynamic rational inattention model poses as the sole rigorously microfounded information rigidity model that can fill this void.

References

Christiano, L. J., Eichenbaum, M., and Evans, C. L. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of political Economy*, 113(1):1–45.

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

Mackowiak, B. and Wiederholt, M. 2009. Optimal sticky prices under rational inattention. *The American Economic Review*, pages 769–803.

Maćkowiak, B. and Wiederholt, M. 2015. Business cycle dynamics under rational inattention*. *The Review of Economic Studies*, page rdv027.

Mankiw, N. G. and Reis, R. 2002. Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *Quarterly Journal of Economics*, pages 1295–1328.

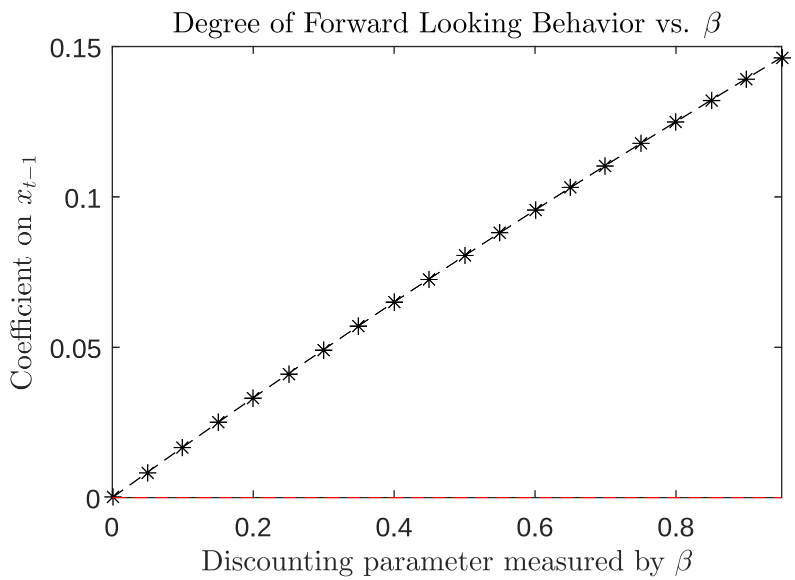
Matejka, F., Steiner, J., and Stewart, C. 2015. Rational inattention dynamics: Inertia and delay in decision-making. Technical report, CEPR Discussion Papers.

Petersen, K. B. and Pedersen, M. S. The matrix cookbook.

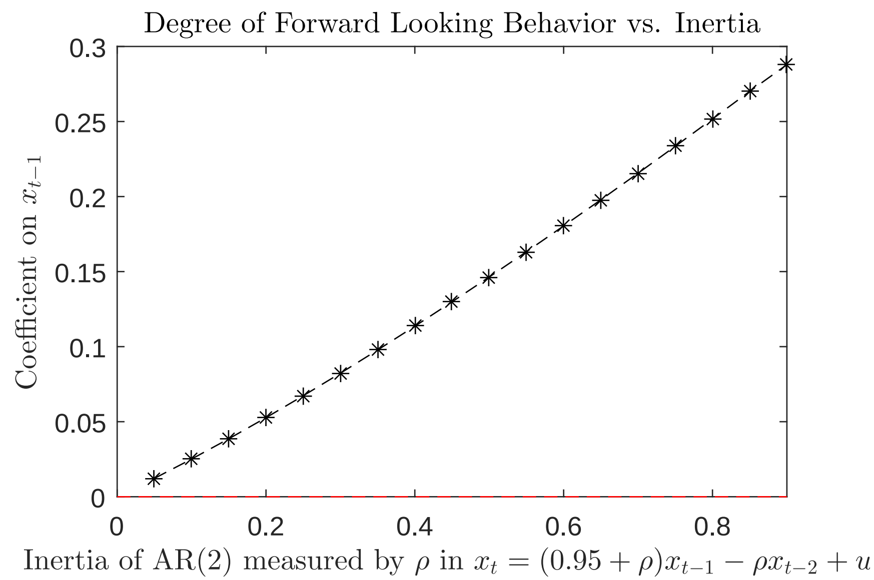
Sims, C. A. 2003. Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Woodford, M. 2003. Imperfect common knowledge and the effects of monetary policy. *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, page 25.

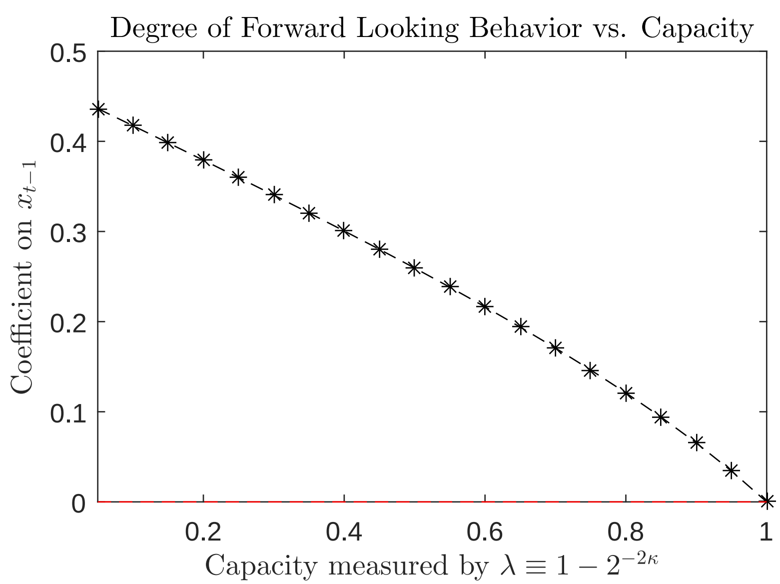
Woodford, M. 2009. Information-constrained state-dependent pricing. *Journal of Monetary Economics*, 56:S100–S124.



(a) This figure shows how the forward looking behavior of the agent in Example 3 depends on their discount parameter β. The higher the β, the higher is the agent's incentive to forecast future realizations of the fundamental. See Example 3 for detailed explanation.

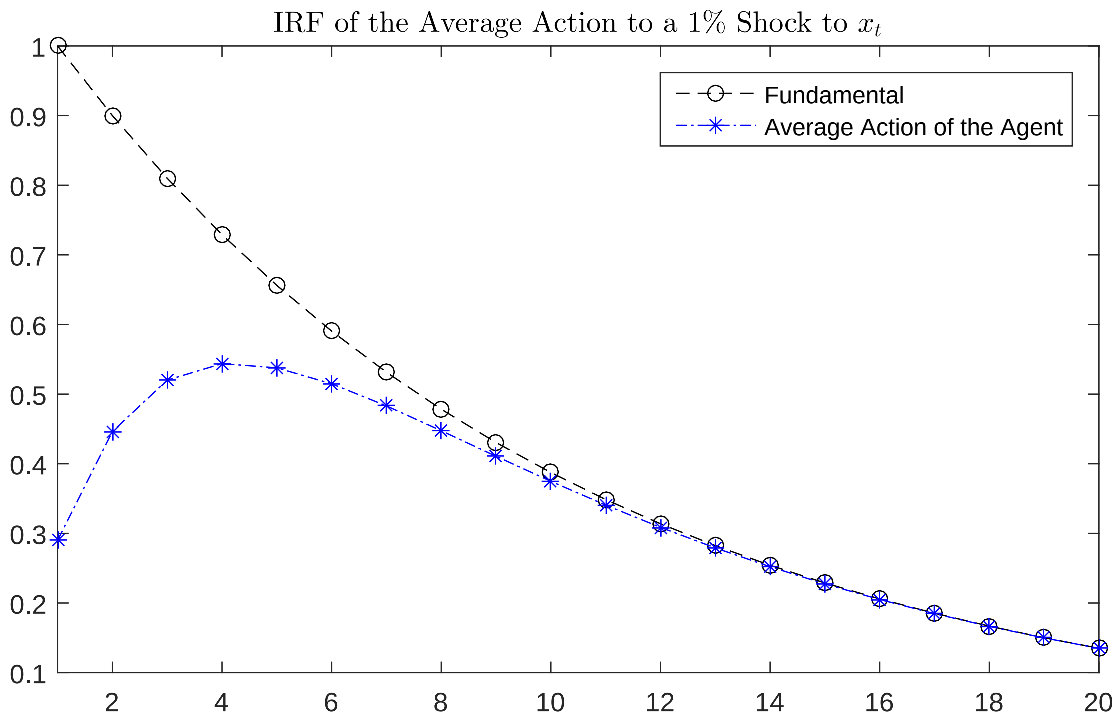


(b) This figure depicts how the forward looking behavior of the agent in Example 3 depends on the degree of inertia in the AR(2) process. The higher the inertia, the higher is the agent's incentive to forecast future realizations of the fundamental. See Example 3 for detailed explanation.

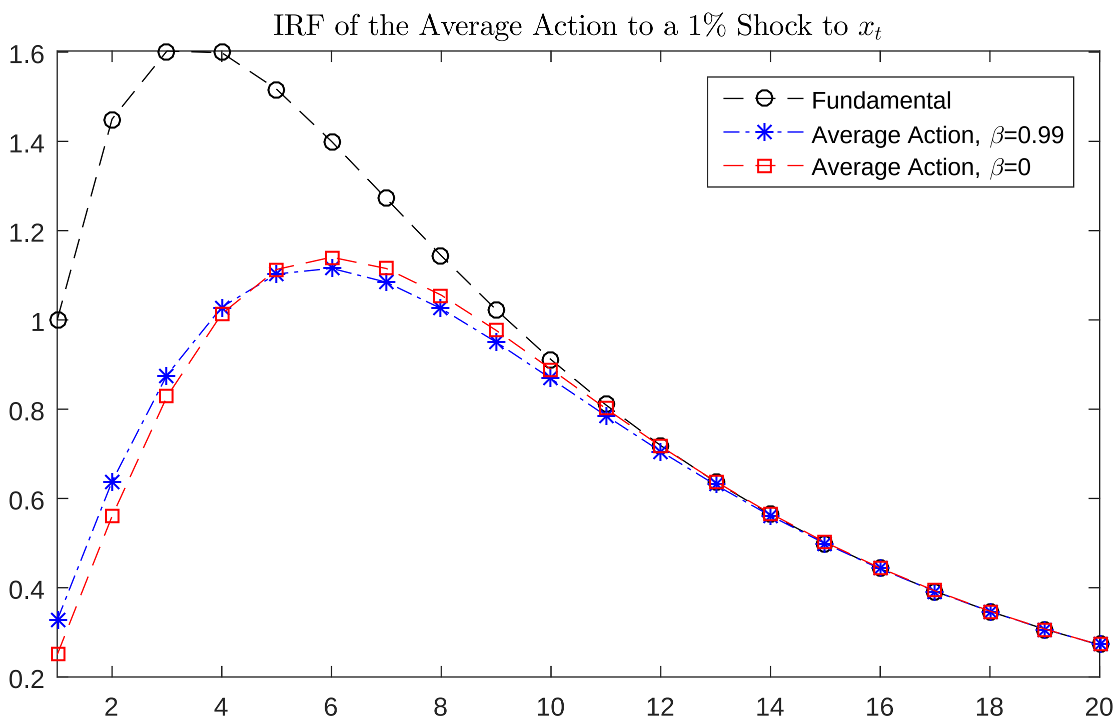


(c) This figure shows how the forward looking behavior of the agent in Example 3 depends on their capacity of processing information. The higher the capacity, the lower is the agent's incentive to forecast future realizations of the fundamental. See Example 3 for detailed explanation.

Figure 1: Degree of Forward Looking Behavior in Acquiring Information vs. Underlying Parameters in Example 3.



(a) The figure shows the IRF of the agent's average action to a 1% shock to the fundamental $x_t = 0.9x_{t-1} + u_t$, with $\kappa = 0.25$. See Example 4 for details.



(b) The figure shows the IRF of the agent's average action to a 1% shock to the fundamental $x_t = 1.45x_{t-1} - 0.5x_{t-2} + u_t$. See Example 4 for details.

Figure 2: Impulse Responses of the Agent's Action to a Shock to the Fundamental in Example 4.

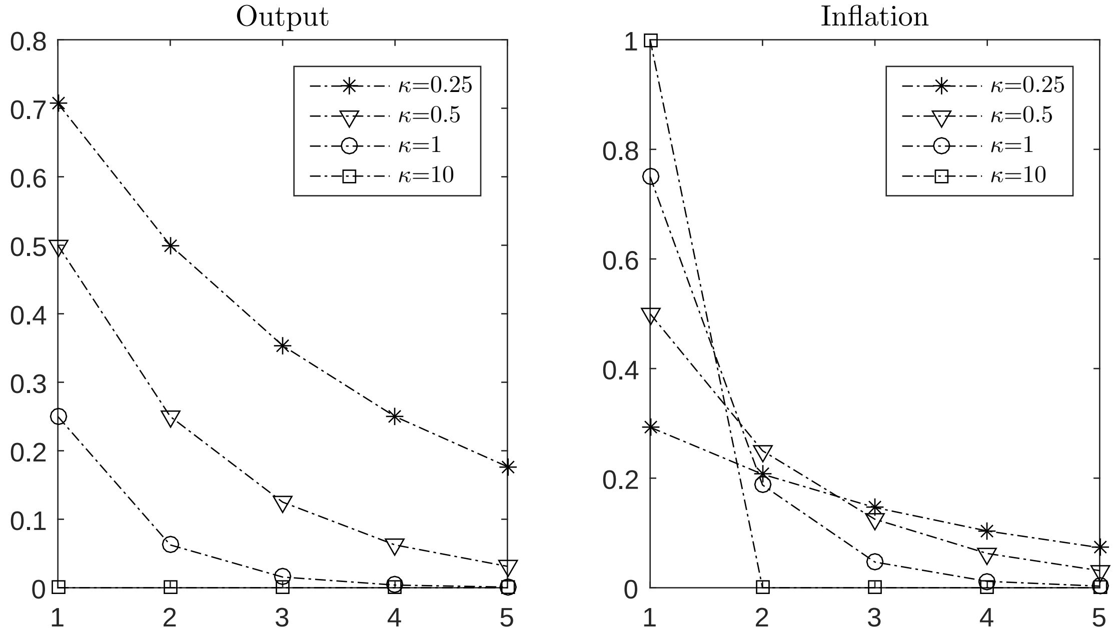


Figure 3: IRFs for Example 5: The figure shows the impulse responses of output and inflation to a 1% shock to the aggregate demand, for different levels of capacity of processing information. Rational inattention creates endogenous real and persistent effects for monetary policy. See Example 5 for details.

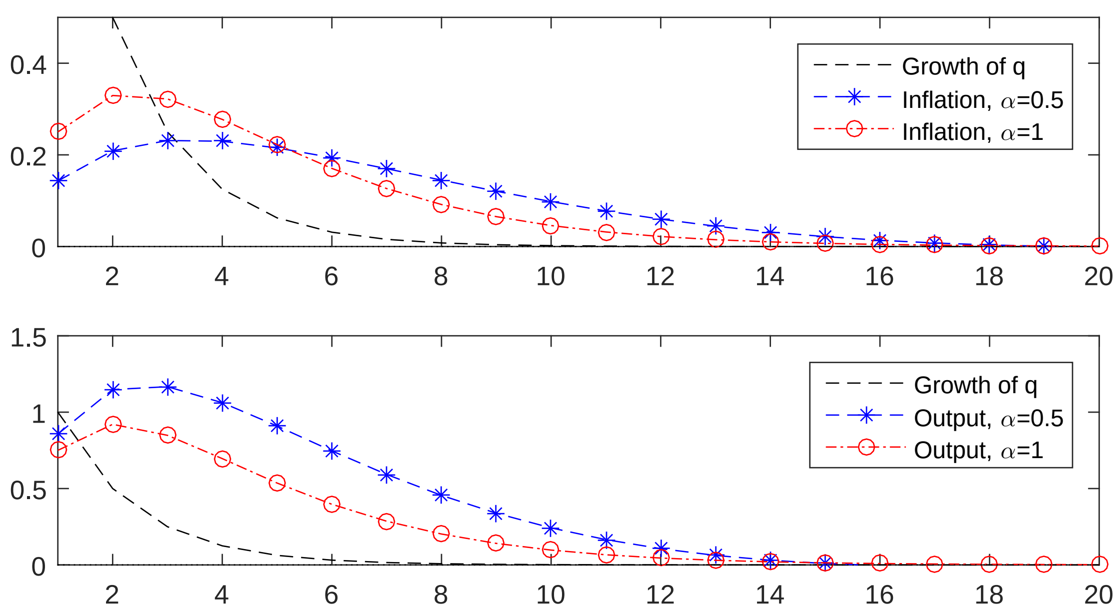


Figure 4: IRFs for Example 6: The figure shows impulse responses of output and inflation for a capacity parameter of $\kappa = 0.2$. The red curves with circle markers are the IRFs of the model with no strategic complementarity ($\alpha = 1$), and the blue curves with star markers are the IRFs when $\alpha = 0.5$. Higher strategic complementarity introduces higher inertia in response of inflation, and amplifies the response of the output. See example 6 for details.

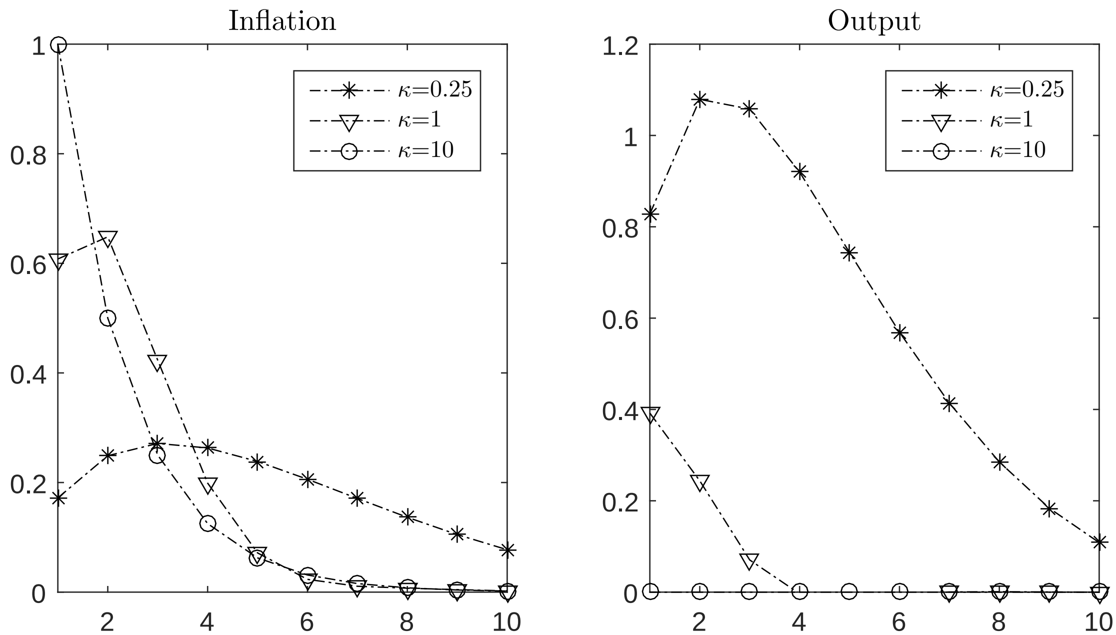


Figure 5: IRFs for Example 6: the figure shows impulse responses of output and inflation for different values of capacity. Higher capacity leads to less inertial response of inflation and a smaller and less persistent response of output. When capacity is very large, inflation exactly follows the AR(1) path of the growth of the aggregates demand, and output does not respond to monetary policy at all. See Example 6 for details.

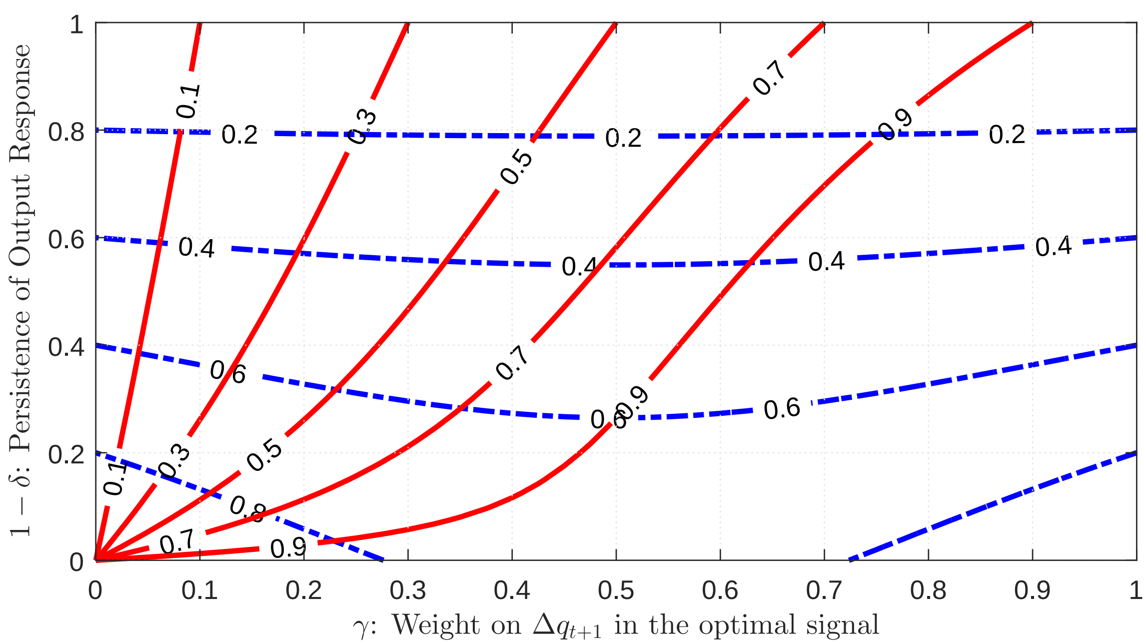


Figure 6: The figure depicts the iso-capacity (measured by $\lambda \equiv 1 - 2^{-2\kappa} \in [0, 1)$) curves in blue dashed lines, and iso-patience (measured by $\beta \in [0, 1)$) curves in red solid lines. Each intersection gives an equilibrium pair of (γ, δ). Higher capacity or lower patience correspond to a less forward looking behavior in the information acquisition of the firms when there is forward guidance. See Example 7 for details.

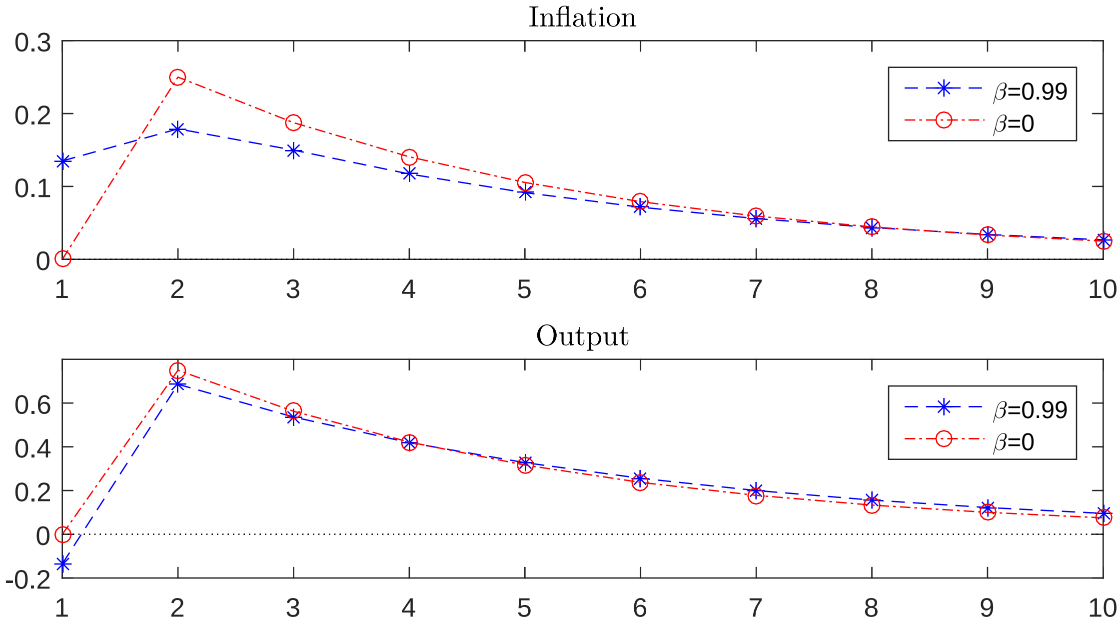


Figure 7: IRFs for Example 7: the figure shows the impulse response functions of output and inflation to 1% announced shock to the aggregate demand that will take effect in period one. When $\beta = 0$, firms choose to ignore the news about future policy, and the news has no effects on the economy at the time of announcement. However, when β is positive, firms include the news in their optimal signal, and react to it immediately before the shock affects the aggregate demand.

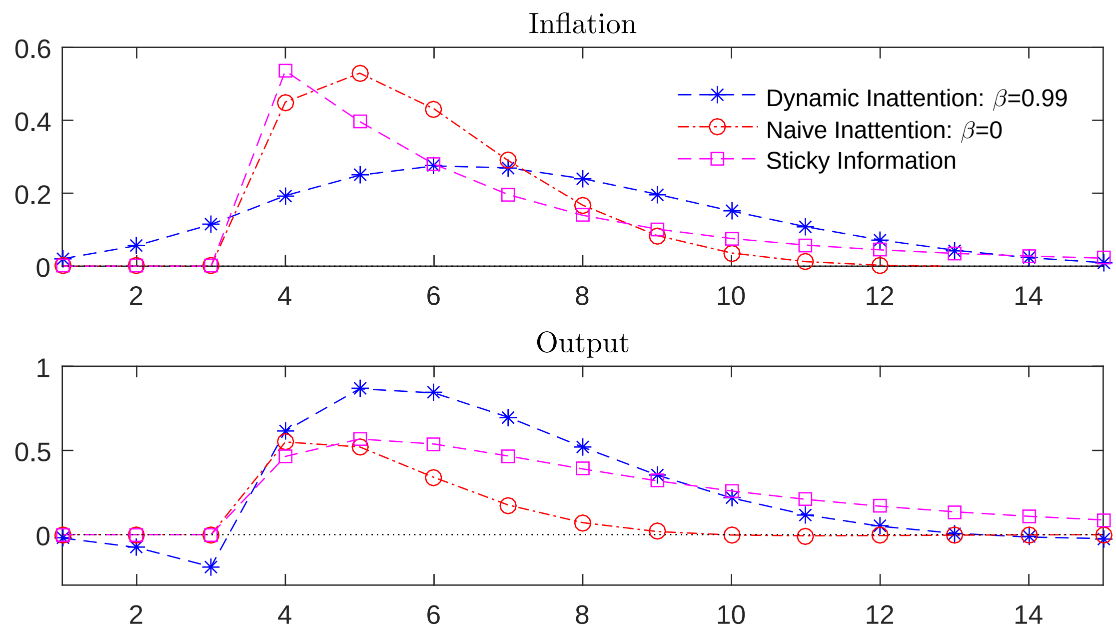


Figure 8: The figure shows impulse responses of three different models to a 1% news shock about aggregate demand that is going to take effect after three periods. Firms do not respond to this news shock neither in the sticky information model nor in the reduced-form noisy information model (naive intention), in which firms only observe their current fundamental. However, under dynamic inattention firms optimally choose to pay attention to the news shock and respond to it immediately.

A Proofs

Proof of Theorem 1.

Recall that the agent's problem is

$$\begin{aligned}
\mathcal{L}_0(\Sigma_{0|-1}) \quad &= \quad \min_{\{\mathbf{y}_t \in \hat{\mathcal{S}}_t^F\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \mathbf{w}'\Sigma_{t|t}\mathbf{w} \\
s.t. \quad & \quad \mathbf{y}_t'\Sigma_{t|t-1}\mathbf{y}_t \leq 1 - 2^{-2\kappa} \\
& \quad \Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1} \\
& \quad \Sigma_{t+1|t} = \mathbf{M}\Sigma_{t|t}\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1' \\
& \quad \Sigma_{0|-1} \text{ given.}
\end{aligned}$$

To simplify the problem, combine the law of motions for $\Sigma_{t|t}$ and $\Sigma_{t+1|t}$ to get a single law of motion for the priors:

$$\Sigma_{t+1|t} = \mathbf{M}(\Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1})\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1' \ , \forall t \geq 0.$$

Finally, as $\Sigma_{t|t-1}$ is a state variable at time t, we can consider the following change of variables: $\mathbf{z}_t = \Sigma_{t|t-1}\mathbf{y}_t$, and let agent choose $\mathbf{z}_t$. Notice that $\mathbf{z}_t$ is the covariance vector of agent's signal with $\mathbf{u}_t$, and if $\Sigma_{t|t-1}$ is invertible, choosing the covariance vector is equivalent to choosing a vector $\mathbf{y}_t$.[29]

This is a standard constrained optimization problem, with a countable number of constraints, that can be solved by maximizing the following Lagrangian: (for simplicity of notation, let $\Sigma_t \equiv \Sigma_{t|t-1}$ denote the agent's prior at time t.)

$$\begin{aligned}
L \quad &= \quad \sum_{t=0}^{\infty} \beta^t(-\mathbf{w}'\Sigma_t\mathbf{w} + \mathbf{w}'\mathbf{z}_t\mathbf{z}_t'\mathbf{w}) \\
&+ \quad \sum_{t=0}^{\infty} \beta^t\phi_t((1 - 2^{-2\kappa}) - \mathbf{z}_t'\Sigma_t^{-1}\mathbf{z}_t) \\
&+ \quad \sum_{t=0}^{\infty} \beta^t(\sum_{j=1}^{\infty} \eta_{j,t}'[\Sigma_{t+1} - \mathbf{M}(\Sigma_t - \mathbf{z}_t\mathbf{z}_t')\mathbf{M}' - \mathbf{e}_1\mathbf{e}_1']\mathbf{e}_j)
\end{aligned}$$

where $\eta_{j,t}$ is the vector of multipliers on the j'th column of the matrix constraint, and $\mathbf{e}_j$ is a vector with 1 as its j'th element and zero elsewhere. I start with the first order condition with respect to Σ_{t+1}:[30]

[29] Assuming that the initial prior, $\Sigma_{0|-1}$, is invertible, meaning that there is strictly positive entropy in agent's initial prior over the history of innovations, one can show that under finite capacity all future Σ_t's are also invertible for any set of signals. This is a direct implication of the fact that resolving all uncertainty about Gaussian variables requires infinite capacity.

[30] For a guide to taking the derivative of symmetric matrices, see for example Petersen and Pedersen.

$$
\begin{aligned}
0 &= -\beta^{t+1}[2\mathbf{w}\mathbf{w}' - diag(\mathbf{w}\mathbf{w}')] \\
&+ \beta^{t+1}\phi_{t+1}[2\Sigma_{t+1}^{-1}\mathbf{z}_{t+1}\mathbf{z}_{t+1}'\Sigma_{t+1}^{-1} - diag(\Sigma_{t+1}^{-1}\mathbf{z}_{t+1}\mathbf{z}_{t+1}'\Sigma_{t+1}^{-1})] \\
&+ \beta^{t}\sum_{j=1}^{T}[\eta_{j,t}\mathbf{e}_j' + \mathbf{e}_j'\eta_{j,t} - diag(\eta_{j,t}\mathbf{e}_j')] \\
&- \beta^{t+1}\sum_{j=1}^{T}\mathbf{M}'[\eta_{j,t+1}\mathbf{e}_j' + \mathbf{e}_j'\eta_{j,t+1} - diag(\eta_{j,t+1}\mathbf{e}_j')]\mathbf{M}
\end{aligned}
$$

take the diagonal of this identity and see that the $diag(.)$ terms sum up to zero, so after replacing $\mathbf{y}_t = \Sigma_t^{-1}\mathbf{z}_t$, we are left with

$$
\mathbf{X}_t = \beta\mathbf{M}'(\mathbf{w}\mathbf{w}' - \phi_{t+1}\mathbf{y}_{t+1}\mathbf{y}_{t+1}' + \mathbf{X}_{t+1})\mathbf{M}
$$

where $\mathbf{X}_t \equiv \frac{1}{2}\sum_{j=1}^{T}\mathbf{M}'\{\mathbf{e}_j\eta_{j,t}' + \eta_{j,t}\mathbf{e}_j'\}\mathbf{M}$.

Moreover, the first order condition with respect to $\mathbf{z}_t$ is

$$
\begin{aligned}
& (\mathbf{w}'\mathbf{z}_t)\mathbf{w} - \phi_t\Sigma_t^{-1}\mathbf{z}_t + \mathbf{X}_t\mathbf{z}_t = 0 \\
\Rightarrow \quad & (\mathbf{w}'\Sigma_t\mathbf{y}_t)\mathbf{w} - \phi_t\mathbf{y}_t + \mathbf{X}_t\Sigma_t\mathbf{y}_t = 0
\end{aligned}
$$

Hence the FOCs reduce to

$$
\begin{aligned}
\phi_t\mathbf{y}_t &= (\mathbf{w}'\Sigma_t\mathbf{y}_t)\mathbf{w} + \mathbf{X}_t\Sigma_t\mathbf{y}_t \\
\mathbf{X}_t &= \beta\mathbf{M}'(\mathbf{w}\mathbf{w}' - \phi_{t+1}\mathbf{y}_{t+1}\mathbf{y}_{t+1}' + \mathbf{X}_{t+1})\mathbf{M}
\end{aligned}
$$

Now, substituting $\mathbf{X}_t$ recursively in the second equation gives us

$$
\begin{aligned}
\mathbf{X}_t &= \textstyle\sum_{j=1}^{\infty}\beta^j\mathbf{M}'^j(\mathbf{w}\mathbf{w}' - \phi_{t+j}\mathbf{y}_{t+j}\mathbf{y}_{t+j}')\mathbf{M}^j \\
\Rightarrow \quad \mathbf{X}_t\Sigma_t\mathbf{y}_t &= \textstyle\sum_{j=1}^{\infty}\beta^j(\mathbf{w}'\mathbf{M}^j\Sigma_{t+j}\mathbf{y}_{t+j})\mathbf{M}'^j\mathbf{w} - \sum_{j=1}^{\infty}(\mathbf{w}'\mathbf{M}^j\Sigma_{t+j}\mathbf{y}_{t+j})\mathbf{M}'^j(\phi_{t+j}\mathbf{y}_{t+j})
\end{aligned}
$$

Combining this with the first order condition for $\mathbf{y}_t$:

$$
\phi_t\mathbf{y}_t = \sum_{j=0}^{\infty}\beta^j(\mathbf{w}'\mathbf{M}^j\Sigma_{t+j}\mathbf{y}_{t+j})\mathbf{M}'^j\mathbf{w} - \sum_{j=1}^{\infty}\beta^j(\mathbf{y}_t'\mathbf{M}^j\Sigma_{t+j}\mathbf{y}_{t+j})\mathbf{M}'^j(\phi_{t+j}\mathbf{y}_{t+j}).
$$

Now guess that $\phi_{t+j}\mathbf{y}_{t+j} = \sum_{k=0}^{\infty} \beta^k a_{t+j,k}\mathbf{M}'^k\mathbf{w}$. Plugging in this guess in the above equation

$$\begin{aligned} 2^{-2\kappa}\phi_t\mathbf{y}_t &= \sum_{j=0}^{\infty} \beta^j(\mathbf{w}'\mathbf{M}^j\Sigma_{t+j}\mathbf{y}_{t+j})\mathbf{M}'^j\mathbf{w} - \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \beta^{j+k}a_{t+j,k}(\mathbf{y}_t'\mathbf{M}^j\Sigma_{t+j}\mathbf{y}_{t+j})\mathbf{M}'^{j+k}\mathbf{w} \\ &= \sum_{j=0}^{\infty} \beta^j[\mathbf{w}'\mathbf{M}^j\Sigma_{t+j}\mathbf{y}_{t+j} - \sum_{k=0}^{j} a_{t+k,j-k}(\mathbf{y}_t'\mathbf{M}^j\Sigma_{t+j-k}\mathbf{y}_{t+j-k})]\mathbf{M}'\mathbf{w} \end{aligned}$$

which verifies our guess and gives us a series of difference equations in terms of $\{(a_{t,j})_{j=0}^{\infty}\}_{t=0}^{\infty}$ where

$$a_{t,j} = 2^{2\kappa}\left[\mathbf{w}'\mathbf{M}^j\Sigma_{t+j}\mathbf{y}_{t+j} - \sum_{k=0}^{j} a_{t+k,j-k}(\mathbf{y}_t'\mathbf{M}^j\Sigma_{t+j-k}\mathbf{y}_{t+j-k})\right].$$

Finally, assuming that $\phi_t > 0$, meaning that the information constraint is binding, let $b_{t,j} \equiv \phi_t^{-1}a_{t,j}$, so that $\mathbf{y}_t = \sum_{j=0}^{\infty}\beta^j b_{t,j}\mathbf{M}'^j\mathbf{w}$. Now, the optimal signal is

$$\begin{aligned} s_t^* &= \mathbf{y}_t'\mathbf{u}_t + e_t \\ &= \sum_{j=0}^{\infty} \beta^j b_{t,j}\mathbf{w}'\mathbf{M}^j\mathbf{u}_t + e_t. \end{aligned}$$

but notice that $\mathbf{M}^j\mathbf{u}_t = \mathbb{E}\{\mathbf{u}_{t+j}|\mathbf{u}_t\}$, and $\mathbf{w}'\mathbf{M}^j\mathbf{u}_t = \mathbb{E}\{\mathbf{w}'\mathbf{u}_{t+j}|\mathbf{u}_t\} = \mathbb{E}\{x_{t+j}|\mathbf{u}_t\} = \mathbb{E}_t^f\{x_{t+j}\}$. Hence,

$$s_t^* = \sum_{j=0}^{\infty}\beta^j b_{t,j}\mathbb{E}_t^f\{x_{t+j}\} + e_t.$$

Q.E.D.

Proof of Corollary 1.

Recall

$$s_t^* = \sum_{j=0}^{\infty}\beta^j b_{t,j}\mathbb{E}_t^f\{x_{t+j}\} + e_t.$$

If x_t follows an $ARMA(p,q)$, then $\exists\{(\alpha_i^j)_{i=0}^{p-1}, (\gamma_i^j)_{i=0}^{q-1}\}_{j=0}^{\infty}$ such that

$$\mathbb{E}_t^f\{x_{t+j}\} = \sum_{i=0}^{p-1}\alpha_i^j x_{t-i} + \sum_{i=0}^{q-1}\gamma_i^j u_{t-i}$$

so

$$\begin{aligned} \sum_{j=0}^{\infty}\beta^j b_{t,j}\mathbb{E}_t^f\{x_{t+j}\} &= \sum_{j=0}^{\infty}\beta^j b_{t,j}\sum_{i=0}^{p-1}\alpha_i^j x_{t-i} + \sum_{j=0}^{\infty}\beta^j b_{t,j}\sum_{i=0}^{q-1}\gamma_i^j u_{t-i} \\ &= \sum_{i=0}^{p-1}(\sum_{j=0}^{\infty}\beta^j b_{t,j}\alpha_i^j)x_{t-i} + \sum_{i=0}^{q-1}(\sum_{j=0}^{\infty}\beta^j b_{t,j}\gamma_i^j)u_{t-i} \end{aligned}$$

Let $c_{i,t} = \sum_{j=0}^{\infty} \beta^j b_{t,j} \alpha_i^j, \forall i \in \{0, 1, \dots, p-1\}, \forall t \geq 0$ and $d_{i,t} = \sum_{j=0}^{\infty} \beta^j b_{t,j} \gamma_i^j, \forall i \in \{0, 1, \dots, q-1\}, \forall t \geq 0$. Then

$$s_t^* = \sum_{i=0}^{p-1} c_{i,t} x_{t-i} + \sum_{i=0}^{q-1} d_{i,t} u_{t-i} + e_t.$$

Q.E.D.

Proof of Example 2.

Proof. From Corollary 1 we know that the optimal signal is of the form $s_t = \alpha_t x_t + e_t$, for $\alpha_t \in \mathbb{R}$, meaning that $\mathbf{y}_t = \alpha_t \mathbf{w}$. α_t is directly implied by the information capacity constraint

$$\mathbf{y}_t' \Sigma_{t|t-1} \mathbf{y}_t = 1 - 2^{-2\kappa} \Rightarrow \alpha_t^2 = \frac{1 - 2^{-2\kappa}}{\mathbf{w}' \Sigma_{t|t-1} \mathbf{w}}.$$

Also, variance of e_t is given by the normalization that $var_{t-1}(s_t) = 1 \Rightarrow \frac{1-2^{-2\kappa}}{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}} var_{t-1}(x_t) + var(e_t) = 1$. Since $var_{t-1}(x_t) = \mathbf{w}'\Sigma_{t|t-1}\mathbf{w}$, $var(e_t) = 2^{-2\kappa}$.

Moreover, by $\Sigma_{t+1|t} = \mathbf{M}(\Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1})\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1'$, and by the fact that $x_t = \rho x_{t-1} + u_t$ implies $\mathbf{w} = (1, \rho, \rho^2, \dots)' \Rightarrow \mathbf{M}'\mathbf{w} = \rho\mathbf{w}$, $\forall t \geq 1$, we have

$$\begin{aligned}
\mathbf{w}'\Sigma_{t|t-1}\mathbf{w} &= \mathbf{w}'\mathbf{M}(\Sigma_{t-1|t-2} - (1 - 2^{-2\kappa})\frac{\Sigma_{t-1|t-2}\mathbf{w}\mathbf{w}'\Sigma_{t-1|t-2}}{\mathbf{w}'\Sigma_{t-1|t-2}\mathbf{w}})\mathbf{M}'\mathbf{w} + 1 \\
&= 1 + \rho^2 2^{-2\kappa}\mathbf{w}'\Sigma_{t-1|t-2}\mathbf{w} \\
&= \frac{1 - (\rho^2 2^{-2\kappa})^t}{1 - \rho^2 2^{-2\kappa}} + (\rho^2 2^{-2\kappa})^t \mathbf{w}'\Sigma_{0|-1}\mathbf{w}.
\end{aligned}$$

Finally, to get the law of motion for the optimal action, by Kalman filter

$$\begin{aligned}
a_t^*(s^t) &= \mathbb{E}\{x_t|s^t\} \\
&= \mathbb{E}\{x_t|s^{t-1}\} + \frac{cov(x_t, s_t|s^{t-1})}{var(s_t|s^{t-1})}(s_t - \mathbb{E}\{s_t|s^{t-1}\}) \\
&= \rho\mathbb{E}\{x_{t-1}|s^{t-1}\} + \sqrt{(1 - 2^{-2\kappa})\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}(s_t - \rho\sqrt{\frac{1 - 2^{-2\kappa}}{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}}\mathbb{E}\{x_{t-1}|s^{t-1}\}) \\
&= 2^{-2\kappa}\rho a_{t-1}^*(s^{t-1}) + \sqrt{(1 - 2^{-2\kappa})\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}} s_t \\
&= 2^{-2\kappa}\rho a_{t-1}^*(s^{t-1}) + (1 - 2^{-2\kappa})x_t + \sqrt{(1 - 2^{-2\kappa})\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}} e_t.
\end{aligned}$$

where $var(x_t|s^{t-1}) \equiv \mathbf{w}'\Sigma_{t|t-1}\mathbf{w}$. is the variance of x_t conditional on time t information of the agent. Q.E.D. □

Proof of Lemma 2.

To be included. Basically the same as theorem 1.

Proof of Corollary 2.

To be included. Basically the same as corollary 1.

Proof of Example 5.

The fact that $\Delta q_t = u_t$, implies that $q_t = \mathbf{e}_1' \tilde{\mathbf{u}}_t$, where $\tilde{\mathbf{u}}_t$ is a random walk vector as defined in section 2.3. The fact that there is no strategic complementarity implies that firms' optimal price is the nominal GDP itself: $\mathbf{dw}_{p^*} = \mathbf{e}_1$. Plugging this into the firms' steady state first order condition for the attention problem, we have

$$\phi \mathbf{dy} = (\mathbf{e}_1' \Sigma \mathbf{dy}) \mathbf{e}_1 + \mathbf{X} \Sigma \mathbf{dy}$$
$$\mathbf{X} = \beta(\mathbf{M}' + \mathbf{e}_1 \mathbf{e}_1')(\mathbf{e}_1 \mathbf{e}_1' - \phi \mathbf{dy} \mathbf{dy}' + \mathbf{X})(\mathbf{M} + \mathbf{e}_1 \mathbf{e}_1')$$

We guess that $\mathbf{dy} = \theta \mathbf{e}_1$, for some $\theta \in \mathbb{R}$. Intuitively, since the firms only care about the first element of $\tilde{\mathbf{u}}_t$, they choose to only see that element with the highest possible precision. To verify this guess, guess also that $\mathbf{X} = \zeta \mathbf{e}_1 \mathbf{e}_1'$ for some $\zeta \in \mathbb{R}$. Plugging these guesses in the second equation we have

$$\begin{aligned} \mathbf{X} &= \beta(1 - \phi\theta^2 + \zeta)(\mathbf{M}' + \mathbf{e}_1 \mathbf{e}_1') \mathbf{e}_1 \mathbf{e}_1' (\mathbf{M} + \mathbf{e}_1 \mathbf{e}_1') \\ &= \beta(1 - \phi\theta^2 + \zeta) \mathbf{e}_1 \mathbf{e}_1' \end{aligned}$$

Thus, $\zeta = \frac{\beta}{1-\beta}(1 - \phi\theta^2)$. Now, from the first equation

$$\phi \mathbf{dy} = (\mathbf{e}_1' \Sigma \mathbf{dy}) \mathbf{e}_1 + \frac{\beta}{1-\beta}(1 - \phi\theta^2)(\mathbf{e}_1' \Sigma \mathbf{dy}) \mathbf{e}_1$$

This verifies our guess that $\mathbf{dy}$ is proportional to $\mathbf{e}_1$. With that in mind, we can get θ directly from the capacity constraint, and the law of motion for the steady state prior:

$$\begin{aligned} & \theta^2 \mathbf{e}_1' \Sigma \mathbf{e}_1 = 1 - 2^{-2\kappa} \qquad , \\ & \mathbf{e}_1' \Sigma \mathbf{e}_1 = \mathbf{e}_1' (\Sigma - \theta^2 \Sigma \mathbf{e}_1 \mathbf{e}_1') \mathbf{e}_1' + 1 \\ \Rightarrow \quad & \theta^2 (\mathbf{e}_1' \Sigma \mathbf{e}_1)^2 = 1 \end{aligned}$$

Thus, $\theta = 1 - 2^{-2\kappa}$, and $\mathbf{e}_1' \Sigma \mathbf{e}_1 = \frac{1}{1-2^{-2\kappa}}$. Thus, every firm i gets a signal

$$\begin{aligned} s_{i,t} &= \sqrt{1 - 2^{-2\kappa}} \mathbf{e}_1' \tilde{\mathbf{u}}_t + e_t^i \\ &= \sqrt{1 - 2^{-2\kappa}} q_t + e_t^i \end{aligned}$$

meaning that they choose to see q_t with the highest possible precision, and where e_t^i is their rational inattention error.

Now, to get the evolution of prices and inflation, notice that

$$\begin{aligned} p_t &= \int_0^1 \mathbb{E}_t^i\{q_t\}di \\ &= \mathbf{e}_1' \int_0^1 \mathbb{E}_t^i\{\tilde{\mathbf{u}}_t\}di \end{aligned}$$

Let $\tilde{\mathbf{u}}_{t|t} = \int_0^1 \mathbb{E}_t^i\{\tilde{\mathbf{u}}_t\}di$, and $\tilde{\mathbf{u}}_{t|t-1} = \int_0^1 \mathbb{E}_{t-1}^i\{\tilde{\mathbf{u}}_t\}di$. By Kalman filtering,

$$\tilde{\mathbf{u}}_{t|t} = \tilde{\mathbf{u}}_{t|t-1} + \Sigma\mathbf{dy}\mathbf{dy}'(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t-1}).$$

Plugging in the solution for $\mathbf{dy}$, we have

$$\begin{aligned} p_t &= \mathbf{e}_1'\tilde{\mathbf{u}}_{t|t} \\ &= 2^{-2\kappa}\mathbf{e}_1'\tilde{\mathbf{u}}_{t|t-1} + (1 - 2^{-2\kappa})q_t \end{aligned}$$

Moreover, notice that since $\mathbb{E}_{t-1}^i\{u_t\} = 0$,

$$\begin{aligned} \mathbf{e}_1'\tilde{\mathbf{u}}_{t|t-1} &= \mathbf{e}_1' \int_0^1 \mathbb{E}_{t-1}^i\{\tilde{\mathbf{u}}_t\}di \\ &= \mathbf{e}_1'(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1') \int_0^1 \mathbb{E}_{t-1}^i\{\tilde{\mathbf{u}}_{t-1}\}di \\ &= \mathbf{e}_1'\tilde{\mathbf{u}}_{t-1|t-1} = p_{t-1} \end{aligned}$$

Thus,

$$\begin{aligned} & p_t = 2^{-2\kappa}p_{t-1} + (1 - 2^{-2\kappa})q_t \\ \Rightarrow \quad & \pi_t = (2^{2\kappa} - 1)y_t \end{aligned}$$

where $\pi_t \equiv p_t - p_{t-1}$ and $y_t \equiv q_t - p_t$. The law of motion for output is given by

$$\begin{aligned} \Delta y_t &= \Delta q_t - \pi_t \\ &= u_t - (2^{2\kappa} - 1)y_t \end{aligned}$$

which implies

$$y_t = 2^{-2\kappa}(y_{t-1} + u_t).$$

Also,

$$\begin{aligned}
\pi_t &= (2^{2\kappa}-1)y_t \\
&= (1-2^{-2\kappa})(y_{t-1}+u_t) \\
&= (1-2^{-2\kappa})\frac{\pi_{t-1}}{2^{2\kappa}-1}+(1-2^{-2\kappa})u_t \\
&= 2^{-2\kappa}\pi_t+(1-2^{-2\kappa})u_t.
\end{aligned}$$

Q.E.D.

Proof of Example 6.

Let $\mathbf{dw}_{p^*}$ be the equilibrium Wold decomposition of the firms' marginal cost, and consider the first order conditions of the attention problem in the steady state:

$$\phi\mathbf{dy} = (\mathbf{dw}'_{p^*}\Sigma\mathbf{dy})\mathbf{dw}'_{p^*} + \mathbf{X}\Sigma\mathbf{dy}$$
$$\mathbf{X} = \beta(\mathbf{M}'+\mathbf{e}_1\mathbf{e}'_1)(\mathbf{dw}_{p^*}\mathbf{dw}'_{p^*} - \phi\mathbf{dy}\mathbf{dy}' + \mathbf{X})(\mathbf{M}+\mathbf{e}_1\mathbf{e}'_1)$$

Notice that when $\beta = 0$, $\mathbf{X}$ is simply the zero matrix; thus,

$$\mathbf{dy} = \delta\mathbf{dw}_{p^*}$$

where $\delta \equiv \frac{\mathbf{dw}'_{p^*}\Sigma\mathbf{dy}}{\phi}$; meaning that firm's optimal signal is to see their marginal cost at every period with the highest possible precision allowed by their capacity:

$$\begin{aligned}
s^*_t &= \delta\mathbf{dw}'_{p^*}\tilde{\mathbf{u}}_t + e_t \\
&= \delta p^*_t + e_t
\end{aligned}$$

where e_t is the firm's rational inattention error and δ is such that

$$\mathbf{dy}'\Sigma\mathbf{dy} = 1-2^{-2\kappa} \Rightarrow \delta = \sqrt{\frac{1-2^{-2\kappa}}{\mathbf{dw}'_{p^*}\Sigma\mathbf{dw}_{p^*}}}.$$

Now, similar to the previous example, by the Kalman filter:

$$\begin{aligned}
& \tilde{\mathbf{u}}_{t|t} = \tilde{\mathbf{u}}_{t|t-1} + \delta^2\Sigma\mathbf{dw}_{p^*}\mathbf{dw}'_{p^*}(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t-1}) \\
\Rightarrow\quad & \mathbf{dw}_{p^*}\tilde{\mathbf{u}}_{t|t} = \mathbf{dw}_{p^*}\tilde{\mathbf{u}}_{t|t-1} + \delta^2\mathbf{dw}'_{p^*}\Sigma\mathbf{dw}_{p^*}\mathbf{dw}'_{p^*}(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t-1}) \\
\Rightarrow\quad & p_t = 2^{-2\kappa}\tilde{\mathbb{E}}_{t-1}\{p^*_t\} + (1-2^{-2\kappa})p^*_t \\
\Rightarrow\quad & \pi_t = 2^{-2\kappa}\tilde{\mathbb{E}}_{t-1}\{p^*_t - p^*_{t-1}\} + (1-2^{-2\kappa})(p^*_t - p_{t-1}) \qquad (\text{as } p_{t-1} = \tilde{\mathbb{E}}_{t-1}\{p^*_{t-1}\}) \\
\Rightarrow\quad & \pi_t = \tilde{\mathbb{E}}_{t-1}\{\pi_t + \alpha\Delta y_t\} + \alpha(2^{2\kappa}-1)y_t
\end{aligned}$$

Where the last line is derived from $p_t^* = p_t + \alpha y_t$. Q.E.D.

Proof of Example 7.

We start with the guess that the optimal signal has the following form

$$s_t = q_t + \gamma \Delta q_{t+1} + e_t$$

where e_t is the firm's rational inattention error and γ will be determined after the verification of the guess, from the optimal behavior of the firm. The fact that the firm can gather information about Δq_{t+1} is due to the forward-guidance policy that $\Delta q_{t+1} = u_t$ is announced at time t.

To translate this environment to our framework, notice that $\Delta q_t = \mathbf{e}_2' \mathbf{u}_t$, where $\mathbf{e}_2$ is the second column of the identity matrix and $\mathbf{u}_t$ is the vector of innovations at time t, with its first element being the innovation that is going to take effect one period ahead. Our guess of the optimal signal translate to

$$\mathbf{y} = \delta[(\mathbf{I} - \mathbf{M})^{-1}\mathbf{e}_2 + \gamma \mathbf{e}_1]$$
$$\mathbf{dy} = \delta[(1-\gamma)\mathbf{e}_2 + \gamma \mathbf{e}_1]$$

so that $s_t = \delta(\mathbf{y}'\mathbf{u}_t + e_t) = \delta(q_t + \gamma \Delta q_{t+1} + e_t)$, with δ being a normalization such that $var_{t-1}\{s_t\} = 1$. To verify the guess, we have to show that this signal solves the firms' first order conditions in the steady state:

$$\phi \mathbf{dy} = (\mathbf{e}_2' \hat{\Sigma} \mathbf{dy})\mathbf{e}_2 + \hat{\mathbf{X}} \hat{\Sigma} \mathbf{dy}$$
$$\hat{\mathbf{X}} = \beta(\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1')(\mathbf{e}_2\mathbf{e}_2' - \phi \mathbf{dy}\mathbf{dy}' + \hat{\mathbf{X}})(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1').$$

where $\hat{\Sigma}$ is such that

$$\hat{\Sigma} = (\mathbf{M} + \mathbf{e}_1\mathbf{e}_1')(\hat{\Sigma} - \hat{\Sigma}\mathbf{dy}\mathbf{dy}'\hat{\Sigma})(\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1') + \mathbf{e}_1\mathbf{e}_1'$$

and ϕ is such that $\mathbf{dy}'\hat{\Sigma}\mathbf{dy} = 1 - 2^{-2\kappa}$. To verify the guess for $\mathbf{dy}$, guess also that $\hat{\mathbf{X}} = \theta \mathbf{e}_1\mathbf{e}_1'$ for some θ. Now, plug in both these guesses in the law of motion for $\hat{\mathbf{X}}$, and observe that

$$\begin{aligned}
\hat{\mathbf{X}} &= \beta(\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1')((1 - \phi\delta^2(1-\gamma)^2)\mathbf{e}_2\mathbf{e}_2' - \phi\delta^2\gamma^2\mathbf{e}_1\mathbf{e}_1' \\
&- \phi\delta^2\gamma(1-\gamma)(\mathbf{e}_1\mathbf{e}_2' + \mathbf{e}_2\mathbf{e}_1'))(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1') \\
&+ (\mathbf{M} + \mathbf{e}_1\mathbf{e}_1')(\theta \mathbf{e}_1\mathbf{e}_1')(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1') \\
&= \beta(1 - \phi\delta^2(1-\gamma)^2)\mathbf{e}_1\mathbf{e}_1' - 2\beta\phi\delta^2\gamma(1-\gamma)\mathbf{e}_1\mathbf{e}_1' - \phi\delta^2\gamma^2\mathbf{e}_1\mathbf{e}_1' + \beta\theta\mathbf{e}_1\mathbf{e}_1' \\
&= \frac{\beta}{1-\beta}(1 - \phi\delta^2)\mathbf{e}_1\mathbf{e}_1'
\end{aligned}$$

Thus, $\theta = \beta(1-\phi\delta^2)$. Now, plug this into the first order condition for $\mathbf{dy}$ to get

$$\phi\mathbf{dy} = (\mathbf{e}_2'\hat{\Sigma}\mathbf{dy})\mathbf{e}_2 + \frac{\beta}{1-\beta}(1-\phi\delta^2)(\mathbf{e}_1'\hat{\Sigma}\mathbf{dy})\mathbf{e}_1.$$

which verifies our guess that $\mathbf{dy}$ is a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$. Before finding γ and δ, however, we need to find $\mathbf{e}_1'\hat{\Sigma}\mathbf{dy}$ and $\mathbf{e}_2'\hat{\Sigma}\mathbf{dy}$. To do so we need to use the steady state law of motion for $\hat{\Sigma}$:

$$\begin{aligned} \hat{\Sigma} &= (\mathbf{M}+\mathbf{e}_1\mathbf{e}_1')(\hat{\Sigma}-\hat{\Sigma}\mathbf{dy}\mathbf{dy}'\hat{\Sigma})(\mathbf{M}'+\mathbf{e}_1\mathbf{e}_1')+\mathbf{e}_1\mathbf{e}_1' \\ \Rightarrow \qquad \mathbf{e}_1'\hat{\Sigma}\mathbf{e}_1 &= \mathbf{e}_1'(\hat{\Sigma}-\hat{\Sigma}\mathbf{dy}\mathbf{dy}'\hat{\Sigma})\mathbf{e}_1 + 1 \\ \Rightarrow \qquad \mathbf{e}_1'\hat{\Sigma}\mathbf{dy} &= 1 \end{aligned}$$

Also, using the guess for $\mathbf{dy}$,

$$\begin{aligned} 1 &= \mathbf{e}_1'\hat{\Sigma}\mathbf{dy} \\ &= \delta\mathbf{e}_1'(\hat{\Sigma}-\hat{\Sigma}\mathbf{dy}\mathbf{dy}'\hat{\Sigma})\mathbf{e}_1 + \delta\gamma, \end{aligned}$$

which implies that $\delta(\mathbf{e}_1'\hat{\Sigma}\mathbf{e}_1 - \hat{\Sigma}\mathbf{dy}\mathbf{dy}'\hat{\Sigma}) = 1-\gamma\delta$. Finally, notice that

$$\begin{aligned} \mathbf{e}_2'\hat{\Sigma}\mathbf{dy} &= \delta\mathbf{e}_1'(\hat{\Sigma}-\hat{\Sigma}\mathbf{dy}\mathbf{dy}'\hat{\Sigma})\mathbf{e}_1 \\ &= 1-\gamma\delta \end{aligned}$$

Thus,

$$\begin{aligned} \phi\mathbf{dy} &= \frac{\beta}{1-\beta}(1-\phi\delta^2)\mathbf{e}_1 + (1-\gamma\delta)\mathbf{e}_2 \\ &= \phi\delta[\gamma\mathbf{e}_1 + (1-\gamma)\mathbf{e}_2] \end{aligned}$$

where the second line was our guess. This implies

$$\begin{aligned} \phi\delta\gamma = \tfrac{\beta}{1-\beta}(1-\phi\delta^2) \quad &, \\ \phi\delta(1-\gamma) = 1-\gamma\delta \quad &. \\ \Rightarrow \qquad \tfrac{(1-\beta)\gamma+\beta\delta}{1-\gamma} = \tfrac{\beta}{1-\gamma\delta} & \end{aligned}$$

The final equation for characterizing the solution comes from the capacity constraint:

$$\begin{aligned} 1-2^{-2\kappa} &= \mathbf{dy}'\hat{\Sigma}\mathbf{dy} \\ &= \delta\gamma\mathbf{e}_1'\hat{\Sigma}\mathbf{dy} + \delta(1-\gamma)\mathbf{e}_2'\hat{\Sigma}\mathbf{dy} \\ &= \delta\gamma + \delta(1-\gamma)(1-\gamma\delta). \end{aligned}$$

These two equations pin down γ and δ and hence characterize the optimal signal.

Finally, to derive the Phillips curve, let $\mathbf{u}_{t|t} = \int_0^1 \mathbb{E}_t^i\{\mathbf{u}_t\}di$, observe that

$$\begin{aligned} \mathbf{u}_{t|t} &= \mathbf{u}_{t|t-1} + \Sigma\mathbf{y}\mathbf{y}'(\mathbf{u}_t - \mathbf{u}_{t|t-1}) \\ \Rightarrow \quad (\mathbf{I}-\mathbf{M}')^{-1}\mathbf{u}_{t|t} &= (\mathbf{I}-\mathbf{M}')^{-1}\mathbf{u}_{t|t-1} + \hat{\Sigma}\mathbf{dy}\mathbf{dy}'(\mathbf{I}-\mathbf{M}')^{-1}(\mathbf{u}_t - \mathbf{u}_{t|t-1}) \end{aligned}$$

Multiply this once by $\mathbf{e}_1'$ and once by $\mathbf{e}_2'$ from left to get[31]

$$\begin{aligned} (\mathbf{e}_1' \times .): &\quad \tilde{\mathbb{E}}_t\{q_{t+1}\} = \tilde{\mathbb{E}}_{t-1}\{q_{t+1}\} + \mathbf{dy}'(\mathbf{I}-\mathbf{M}')^{-1}(\mathbf{u}_t - \mathbf{u}_{t|t-1}) \\ (\mathbf{e}_2' \times .): &\quad p_t = \tilde{\mathbb{E}}_{t-1}\{q_t\} + (1-\gamma\delta)\mathbf{dy}'(\mathbf{I}-\mathbf{M}')^{-1}(\mathbf{u}_t - \mathbf{u}_{t|t-1}) \end{aligned}$$

Where $\tilde{\mathbb{E}}_t\{.\} = \int_0^1 \mathbb{E}_t^i\{.\}di$. Now, notice that $\tilde{\mathbb{E}}_{t-1}\{q_{t+1}\} = \tilde{\mathbb{E}}_{t-1}\{q_t\}$, as u_t is not realized at $t-1$. Moreover, observe that

$$\mathbf{dy}'(\mathbf{I}-\mathbf{M}')^{-1}(\mathbf{u}_t - \mathbf{u}_{t|t-1}) = \delta(q_t + \gamma u_t - \tilde{\mathbb{E}}_{t-1}\{q_t\}).$$

Thus,

$$\begin{aligned} \tilde{\mathbb{E}}_t\{\Delta q_{t+1}\} &= (1-\delta)\tilde{\mathbb{E}}_{t-1}\{\Delta q_t\} + \delta(y_t + \gamma u_t) - (1-\delta)\pi_t \\ \pi_t &= \tilde{\mathbb{E}}_{t-1}\{\Delta q_t\} + \tfrac{(1-\gamma\delta)\delta}{1-\delta(1-\gamma\delta)}(y_t + \gamma u_t) \end{aligned} \tag{8}$$

Finally, substituting for $\tilde{\mathbb{E}}_{t-1}\{\Delta q_t\}$ in the the first equation using the second one we have

$$\begin{aligned} \tilde{\mathbb{E}}_t\{\Delta q_{t+1}\} &= (1-\delta)(\pi_t - \frac{(1-\gamma\delta)\delta}{1-\delta(1-\gamma\delta)}(y_t+\gamma u_t)) + \delta(y_t + \gamma u_t) - (1-\delta)\pi_t \\ &= \gamma\frac{\delta^2}{1-\delta(1-\gamma\delta)}(y_t + \gamma u_t) \end{aligned}$$

which implies that $\tilde{\mathbb{E}}_{t-1}\{\Delta q_t\} = \gamma\frac{\delta^2}{1-\delta(1-\gamma\delta)}(y_{t-1} + \gamma u_{t-1})$. Plugging this into 8 we get the following Phillip's curve:

$$\pi_t = \delta\frac{\gamma\delta}{1-\delta(1-\gamma\delta)}(y_{t-1} + \gamma\Delta q_t) + \delta\frac{1-\gamma\delta}{1-\delta(1-\gamma\delta)}(y_t + \gamma\Delta q_{t+1})$$

which implies

$$y_t = (1-\delta)(y_{t-1} + u_{t-1}) + \delta^2\gamma(1-\gamma)u_{t-1} - \gamma\delta(1-\gamma\delta)u_t$$

Using the fact that $u_{t-1} = \Delta q_t$ and $u_t = \Delta q_{t+1}$,we get the following laws of motion for inflation and output:

$$y_t = (1-\delta)y_{t-1} + 2^{-2\kappa}\Delta q_t - \gamma\delta(1-\gamma\delta)\Delta q_{t+1}$$

$$\pi_t = \delta y_{t-1} + (1-2^{-2\kappa})\Delta q_t + \gamma\delta(1-\gamma\delta)\Delta q_{t+1}$$

[31] We use the results form before that $\mathbf{e}_1'\hat{\Sigma}\mathbf{dy} = 1$ and $\mathbf{e}_2'\hat{\Sigma}\mathbf{dy} = 1-\gamma\delta$.

Proof of Lemma 3.

Let $\tilde{\mathbf{u}}_t$ be the random walk vector of shocks announced until time t, defined in section 2.3. Moreover, let $\mathbf{dw}_{p^*}$ be the Wold decomposition of the stationary part of the p_t^* in the equilibrium, and $\mathbf{dy} = \sum_{j=0}^{\infty} \beta^j b_j \mathbf{dw}'_{p^*}(\mathbf{M}+\mathbf{e}_1\mathbf{e}_1')$ be the representation of the optimal signal derived in that section. Notice that by Kalman filter

$$\begin{aligned} \tilde{\mathbf{u}}_{t|t} &= \tilde{\mathbf{u}}_{t|t-1} + \Sigma\mathbf{dy}\mathbf{dy}'(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t-1}) \\ \Rightarrow \quad \tilde{\mathbf{u}}_{t|t} &= \tilde{\mathbf{u}}_{t|t-1} + (\mathbf{I} - \Sigma\mathbf{dy}\mathbf{dy}')^{-1}\Sigma\mathbf{dy}\mathbf{dy}'(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t}) \end{aligned}$$

Moreover,

$$\begin{aligned} (\mathbf{I} - \Sigma\mathbf{dy}\mathbf{dy}')^{-1}\Sigma\mathbf{dy}\mathbf{dy}' &= \Sigma\mathbf{dy}\mathbf{dy}' \sum_{i=0}^{\infty} (\Sigma\mathbf{dy}\mathbf{dy}')^i \\ &= \Sigma\mathbf{dy}\mathbf{dy}' \sum_{i=0}^{\infty} (1 - 2^{-2\kappa})^i \\ &= 2^{2\kappa}\Sigma\mathbf{dy}\mathbf{dy}' \end{aligned}$$

where the second line is derived from the capacity constraint, $\mathbf{dy}'\Sigma\mathbf{dy} = 1 - 2^{-2\kappa}$. Thus,

$$\tilde{\mathbf{u}}_{t|t} = \tilde{\mathbf{u}}_{t|t-1} + 2^{2\kappa}\Sigma\mathbf{dy}\mathbf{dy}'(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t}).$$

Also, by the fact that $\mathbb{E}_t^f\{\tilde{\mathbf{u}}_{t+j}\} = (\mathbf{M} + \mathbf{e}_1\mathbf{e}_1')^j \tilde{\mathbf{u}}_{t+j}$, observe that

$$\begin{aligned} \mathbf{dy}'(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t}) &= \sum_{j=0}^{\infty} \beta^j b_j \mathbf{dw}'_{p^*}(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1')(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t}) \\ &= \sum_{j=0}^{\infty} \beta^j b_j (\mathbb{E}_t^f\{p_{t+j}^*\} - \tilde{\mathbb{E}}_t\{p_{t+j}^*\}). \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{dw}'_{p^*}\tilde{\mathbf{u}}_{t|t} &= \mathbf{dw}'_{p^*}\tilde{\mathbf{u}}_{t|t-1} + 2^{2\kappa}(\mathbf{dw}'_{p^*}\Sigma\mathbf{dy})\mathbf{dy}'(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}_{t|t}) \\ \Rightarrow \quad p_t &= \tilde{\mathbb{E}}_{t-1}\{p_t^*\} + 2^{2\kappa}\delta_0 \textstyle\sum_{j=0}^{\infty} \beta^j b_j (\mathbb{E}_t^f\{p_{t+j}^*\} - \tilde{\mathbb{E}}_t\{p_{t+j}^*\}) \end{aligned}$$

Where $\delta_0 \equiv \mathbf{dw}'_{p^*}\Sigma\mathbf{dy}$. Now, subtract $p_{t-1} = \tilde{\mathbb{E}}_{t-1}\{p_{t-1}^*\}$ from both sides of this equation to get

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\{\pi_t + \Delta y_t\} + 2^{2\kappa}\delta_0 \sum_{j=0}^{\infty} \beta^j b_j (\mathbb{E}_t^f\{p_{t+j}^*\} - \tilde{\mathbb{E}}_t\{p_{t+j}^*\}).$$

Q.E.D.