Dynamic Inattention, the Phillips Curve and Forward Guidance

Hassan Afrouzi* Columbia U.

Choongryul Yang† UT Austin

Preliminary Draft

Abstract

We show that rationally inattentive firms are forward-looking in their information acquisition and study the implications of these incentives for inflation dynamics by deriving a new micro-founded Phillips curve. The Phillips curve is forward-looking and relates current inflation to the forecast errors of firms about future inflation and the growth rate of the output gap in the economy – a feature that is absent in sticky and noisy information models. Unlike the forward-looking Phillips curves derived under nominal rigidities, we show inflation is not necessarily increasing in expected inflation, and it can decrease with the forecast errors of firms about future inflation and output gap growth. We test this Phillips curve using the Survey of Professional Forecasters as a proxy for firms' expectations and show that forecast errors about future significantly affect current inflation in the direction that is predicted by the model. We apply our findings to examine the effectiveness of forward guidance policies in a general equilibrium model. News about future interest rates affects inflation more if firms are more rationally inattentive or if they discount future profits less. The model also survives the forward guidance puzzle as the initial response of inflation decreases with the horizon of forward guidance.

JEL Classification: E31, E32, E37, E52, E70

Keywords: Rational Inattention, Dynamic Information Acquisition, Phillips Curve, Forward Guidance, Survey of Professional Forecasters

*Columbia University, Department of Economics. 420 West 118th Street, New York, NY, 10027 U.S.A. Email: hassan.afrouzi@columbia.edu

†University of Texas at Austin, Department of Economics. 2225 Speedway C3100, Austin, TX, 78712 U.S.A. Email: c.yang@utexas.edu

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 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 micro-founded 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, we show that they induce a pricing behavior under which inflation does not depend on firms' expected future inflation. Therefore, each class of micro-founded 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 micro foundations, 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].

We derive a micro-found hybrid Philips curve without nominal rigidities. We call it the dynamic inattention Phillips curve. We show the inflation dynamics under this Phillips curve incorporates both of 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

[1]By reduced-form noisy information models, we 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 micro-founded 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, Reis, et al. (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 micro-foundation of these models in the first place.

[4]See, for instance, Christiano, Eichenbaum, and Evans (2005)

them. More importantly inflation dynamics depends 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.

The dynamic inattention Phillips curve has salient characteristics that can be tested using data. It is forward-looking and relates current inflation to the forecast errors of firms about future inflation and the growth rate of the output gap in the economy – a feature that is absent in sticky and noisy information models. Moreover, unlike the forward-looking Phillips curves derived under nominal rigidities, we show inflation is not necessarily increasing in expected inflation, and it can decrease with the forecast errors of firms about future inflation and output gap growth. We test this Phillips curve with nonlinear GMM estimation using the Survey of Professional Forecasters as a proxy for firms' expectations. We show that forecast errors about future significantly affect current inflation in the direction that is predicted by the model.

To manifest the importance of the forward-looking behavior in our model, we consider a news shock exercise, and compare it to sticky and reduced-form noisy information models. Since prices are flexible, firms within sticky and reduced-form noisy information models, which assume agents only see the current realizations of their fundamentals, do not respond at all to the news shocks. However, we show that when firms are allowed to choose their information structure endogenously, they optimally choose to pay attention to these 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.

We extend our model to a general equilibrium model to study the effects of the forward guidance policy. When the future expansionary monetary shock is announced, inflation and output increase immediately since firms' optimal signal incorporates their best forecast about the future marginal costs, and thus firms increase their prices to the news. When the forward guidance shock is actually realized, the nominal interest rate falls. In a standard Calvo sticky price model, output, inflation, and the interest rate all go back to steady-states immediately after the shock realization, since the model is completely forward-looking and the shock is transitory. In contrast, after the shock realization, inflation slowly converges to the steady-state in our model. This is because the rationally inattentive firms have noisy information about their fundamentals. In aggregate, the Phillips curve shows a backward-looking term, which is the past expectation about current inflation and output gap growth. This backward-looking nature of the Phillips curve enables our model to survive the forward guidance puzzle, established in Del Negro, Giannoni, and Patterson (2012). Unlike the standard sticky price model, which is completely forward-looking, the initial responses of inflation and output decrease with the horizons of forward guidance.

This paper contributes to the rational inattention literature in several dimensions. First, we 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, we 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 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 a Gaussian fundamental is one that allows the agent to form expectations over current and future fundamentals.

We also contribute to a literature on micro-foundations of Phillips curves. Especially, we derive the dynamic inattention Phillips curve, which is a micro-foundation for reduced-form hybrid models of the Phillips curve, such as sticky price with indexation(Christiano, Eichenbaum, and Evans (2005)) or rule-of-thumb firms of backward-looking pricing(Gali and Gertler (1999)). The hybrid Phillips curve has been an important way to explain both the inertial responses of inflation to shocks and the forward-looking behavior of inflation dynamics. Nimark (2008), Melosi (2017), and Šauer (2016) derive the Phillips curves in the models of imperfect-common knowledge with Calvo or Rotemberg sticky pricing. In their models, the Phillips curves are also forward-looking due to the nominal rigidities. Moreover, the inflation dynamic shows inertial responses to shocks because of higher-order beliefs from imperfect-common knowledge. Our model is different from theirs because we derive a micro-founded hybrid Phillips curve within a perfectly flexible pricing environment with rationally inattentive firms who face the dynamic incentives to process information.

Lastly, our application also contributes to a literature on forward guidance policy. Forward guidance policy is an optimal policy commitment at the zero lower bound (e.g. Krugman, Dominquez, and Rogoff (1998); Woodford (2003); Campbell, Evans, Fisher, and Justiniano (2012)). In our model, when firms are allowed to acquire information endogenously, the news about a future interest rate cut is expansionary: firms increase their prices to the news shocks because their optimal signal incorporates the best forecast of future marginal costs as they do not want to be "too" mistaken about future marginal costs when those days come. However, we show that this expansionary effect of the forward guidance policy is not a free lunch: after the transitory shock is actually realized, output contracts and slowly converges to the steady-state as firms still increase their prices due to the noisy signal that they optimally choose to observe. Our model weakens the power of forward guidance, suggesting a potential resolution to the forward guidance puzzle. Recent papers suggest some resolutions using the sticky information models, which weaken firms' forward-looking behavior.[5] For example, Kiley (2016) shows that in a zero-lower bound environment, forward guidance multipliers are small under the sticky information models. Carlstrom, Fuerst, and Paustian (2015) examine a general class of interest rate pegs in a variety of dynamic New Keynesian models and find that a sticky information model can ameliorate the forward guidance puzzle. Angeletos and Lian (2016) also show that imperfect common knowledge model lessen forward-guidance puzzle through the agents' higher-order beliefs about their future fundamentals. The effect of forward guidance in our model is also in line with these studies: the dynamic inattention Phillips curve depends not only on the expectations of future fundamentals, but also on the past average expectations about current fundamental. This backward-looking nature of the Phillips curve helps to resolve the forward guidance puzzle.

This paper builds on the rational inattention literature and the seminal work of Sims (2003). The rational inattention problem has been applied in various macroeconomic problem.[6] In particular, Maćkowiak and Wiederholt (2009, 2015), Paciello (2012) and Paciello and Wiederholt (2014) consider rationally inattentive firms or(and) households, and study the effects of standard monetary policy shocks on the dynamics of inflation and output. In contrast, our focus is to evaluate the effect of forward guidance policy, which is a type of unconventional monetary policy. Our forward guidance exercise clearly illustrates the dynamic incentives of firms' information acquisitions, and the importance of these incentives on the aggregate economy.

The dynamic model of this paper also relates to a recent literature on characterizing dynamic incentives in information acquisition. Steiner, Stewart, and Matejka (2017) solve a general dynamic rational inattention problems with discrete choice that an agent repeatedly acquires costly information about an evolving state. They consider general payoffs and distributions in discrete environments while we focus on Gaussian fundamentals in continuous choices. There are two recent works that consider the dynamic rational inattention problem under Gaussian fundamentals. Afrouzi (2016) develops a dynamic rational inattention problem, but focuses on the strategic incentives of firms. The model corresponds to our $\beta = 0$ case. One of the most related works is done by Maćkowiak, Matejka, and Wiederholt (2016). They also develop a dynamic rational inattention problem with Gaussian process. There are several differences in the formulation and the applications between their work and ours. For example, in their model, the time preferenceβ does not affect on agents' optimal actions while our model shows the time preference parameter has important implications on the form of optimal signals that agents choose to observe, and thus on

[5]Other resolutions suggest to lessen the forward-looking behavior of households. For example, McKay, Nakamura, and Steinsson (2016) introduce an incomplete market and a borrowing constraint to otherwise a standard NK model, and show that forward guidance is less powerful due to self-insurance motives of households. Farhi and Werning (2017) extend a standard NK model by assuming 1) households heterogeneity with incomplete markets and 2) bounded rationality in the form of level-k thinking. They show that interaction of these two frictions leads to a powerful resolution of forward guidance puzzle.

[6]For a comprehensive recent survey of this literature, see Sims (2010) and Luo and Young (2013).

their optimal actions. Moreover, we characterize the necessary and sufficient conditions for solving the dynamic rational inattention problem in a Linear-quadratic-Gaussian (LQG) framework. In terms of applications, they apply their model to a business cycle model with news shocks in technology while we apply the dynamic rational inattention model to a pricing theory to derive the dynamic inattention Phillips curve, and focus on the effect of forward guidance monetary policy in the dynamic rational inattention model.

The estimation of our Phillips curve relates to a literature on New Keynesian Phillips curve estimation.[7] Using instrumental variables to compute a proxy for expectations for future inflation and growth of output gap, we estimate our dynamic inattention Phillips curve with GMM, which is widely used in the estimation of Phillips curves.[8] We use ex-ante forecast revisions as instruments for ex-post forecast errors in our Phillips curve. This choice of instruments is based on the theoretical predictions of noisy information models, established in Coibion and Gorodnichenko (2015a). The use of forecasts data for estimation also relates this paper to a literature on estimation of Phillips curves using survey data of inflation forecasts.[9]

The paper is organized as follows. Section 2 presents a dynamic model of attention allocation, and develops a tractable way to solve the dynamic rational inattention problem under the Linear-quadratic-Gaussian (LQG) framework. Applying the results to pricing theory, section 3 derives a dynamic inattention Phillips curve. Section 4 extends the simple model to a general equilibrium model to assess the effects of forward guidance shocks. Section 5 concludes. Moreover, all the technical derivations as well as the proofs of all the propositions and corollaries are included in Appendix A.

2 A Dynamic Model of Attention Allocation

2.1 Environment Given an Information Structure

Suppose the agent tracks a fundamental that is characterized by a covariance stationary[10] 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.

[7]See **?** for comprehensive reviews.

[8]See, for example, Gali and Gertler (1999), Roberts (2005), and Rudd and Whelan (2005) for the GMM estimation of the New Keynesian Phillips curve.

[9]For example, see Brissimis and Magginas (2008), Coibion (2010), and Coibion and Gorodnichenko (2015b) among others.

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

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 $s^t \equiv (s_0, s_1, \ldots, s_t) \in F^t$ be a possible realization of the signals until time t, and let $S^t \equiv \left\{ s^t \mid s^t \in F^t \right\}$ 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:

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

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

$$L_0 \equiv \min_{\tilde{a} \in \tilde{A}} \sum_{t=0}^{\infty} \beta^t \int_{s^t \in S^t} \int_{x_t \in \mathbb{R}} \left[\left(a_t(s^t) - x_t \right)^2 + \lambda \mathcal{I} \left(s_t, x_t \mid S^{t-1} \right) \right] f_t \left(x_t, s^t \right) dx_t ds^t$$

where for $s^t \in S^t$ and $x_t \in \mathbb{R}$, $f_t \left(x_t, s^t \right)$ is their joint density. $\mathcal{I} \left(s_t, x_t \mid S^{t-1} \right)$ is the mutual information between the signal and the state, and we will formally define it in the next section. Notice that given the information choice, the optimal action profile does not affect the amount of information flow. The first order condition with respect to $a_t(s^t)$ is then

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

where $\mathbb{E} \left[. \right]$ 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 variances of x_t minus the amount of information flow and:

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

Hence, the agent's objective in choosing her information structure is to 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:

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

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 we are going to use computational methods to solve for the solution, when a closed form does not exist.

For an arbitrarily large $T \in \mathbb{N}$, we 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 $\mathbf{u}_t = (u_t, u_{t-1}, u_{t-2}, \ldots, u_{t-T})'$. We 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 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.

Second, to specify the information constraint, following the rational inattention literature, we 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

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

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. We 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

$$\min_{\{f_t\}_{t=0}^{\infty}} \quad \mathcal{L}_0(f_{-1}) = \sum_{t=0}^{\infty} \beta^t \int_{s^t \in S^t} \int_{x_t \in \mathbb{R}} \left[\left(\mathbb{E}\left[x_t \mid s^t \right] - x_t \right)^2 + \lambda \mathcal{I}\left(s_t, \mathbf{u}_t \mid S^{t-1} \right) \right] f_t\left(x_t, s^t \right) dx_t ds^t$$

Maćkowiak 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. Since this result is valid in our framework, given a Gaussian initial prior at time zero, the agent will choose Gaussian signals over time. Thus, at any point in time, $t \geq 0$, the agent is born with a Gaussian prior over $\mathbf{u}_t$. Formally,

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

where $\Sigma_{t|t-1} \equiv \mathbb{E}_{t-1}\left[\left(\mathbf{u}_t - \mathbf{u}_{t|t-1} \right) \left(\mathbf{u}_t - \mathbf{u}_{t|t-1} \right)' \right]$ 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 $\mathbf{u}_t$:

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

Notice that we do not make any restrictions about how many signals firms can observe. The following Lemma establishes that observing one signal is optimal for each agent.

Lemma 1. *Every agent observes only one signal at any time.*

Proof. See Appendix. □

The intuition for the optimality of observing one signal is simple: since the agent's optimal action is a linear combination of signals, instead of seeing multiple signals separately and paying a high cost, the agent would like to see the combination of the signals. In the following Lemma, we rewrite the information capacity constraint.

Lemma 2. *At any $t \geq 0$, given that $\mathbf{u}_t | S^{t-1} \sim \mathcal{N}\left(\mathbf{u}_{t|t-1}, \Sigma_{t|t-1} \right)$, for any $s_t \in \mathcal{S}_t^F$, such that $s_t = \mathbf{y}'\mathbf{u}_t + e_t$, the mutual information of signal s_t and the fundamental $\mathbf{u}_t$ reduces to*

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

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:

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

Thus $\mathcal{I}\left(s_t, \mathbf{u}_t | S^{t-1} \right) \leq \kappa \Leftrightarrow \mathbf{y}'\Sigma_{t|t-1}\mathbf{y} \leq \left(1 - 2^{-2\kappa} \right) var\left\{ s_t | S^{t-1} \right\}$. 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 $var\left\{s_t|S^{t-1}\right\} = 1$, $\forall t \geq 0$. Let $\hat{\mathcal{S}}_t^F = \left\{\mathbf{y} \mid s_t = \mathbf{y}'\mathbf{u}_t + e_t \in \mathcal{S}_t^F,\ var\left\{s_t|S^{t-1}\right\} = 1\right\}$ 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, we 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, $\mathbf{u}_t|S^{t-1} \sim \mathcal{N}\left(\mathbf{u}_{t|t-1}, \Sigma_{t|t-1}\right)$, 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:

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

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[11], and $\mathbf{e}_1$ is the first column of the identity matrix. Since u_{t+1} is drawn at time $t+1$, it is orthogonal to all the agent's information until t. Hence,

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

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 $var\left\{x_t|S^t\right\} = \mathbf{w}'\Sigma_{t|t}\mathbf{w}$, we can rewrite the agent's attention

[11] $\mathbf{M}$ 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.

problem as[12]

$$
\begin{aligned}
\mathcal{L}_0\left(\Sigma_{0|-1}\right) &= \min_{\{\mathbf{y}_t\in\hat{\mathcal{S}}_t^F\}_{t=0}^{\infty}} \sum_{t=0}^{\infty}\beta^t\left(\mathbf{w}'\Sigma_{t|t}\mathbf{w}+\lambda\kappa_t\right) \\
s.t. \quad & \Sigma_{t|t}=\Sigma_{t|t-1}-\Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1} \\
& \Sigma_{t+1|t}=\mathbf{M}\Sigma_{t|t}\mathbf{M}'+\mathbf{e}_1\mathbf{e}_1' \\
& \frac{1}{2}\log_2\left(\frac{1}{1-\mathbf{y}_t'\Sigma_{t|t-1}\mathbf{y}_t}\right)\le\kappa_t \\
& \Sigma_{0|-1}\succeq 0 \text{ given.}
\end{aligned} \tag{4}
$$

The following Theorem characterize the necessary and sufficient condition for the optimal signal vector $\mathbf{y}_t$ and the covariance matrix $\Sigma_{t|t-1}$.

Theorem 1. *Given an initial prior, $\mathbf{u}_0 \sim \mathcal{N}\left(\mathbf{u}_{0|-1},\Sigma_{0|-1}\right)$, the 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*

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

where $\phi_t=\frac{\lambda}{2\ln 2}\left(\frac{1}{1-\mathbf{y}_t'\Sigma_{t|t-1}\mathbf{y}_t}\right)$ and $\mathbf{X}_t$ is the matrix of Lagrange multipliers on each constraints of the evolution of the prior. Let $\widehat{\mathbf{X}_t}=\mathbf{w}\mathbf{w}'+\mathbf{X}_t$. Then, this signals, $\{\mathbf{y}_t\}_{t=0}^{\infty}$, are optimal if and only if for every t, ϕ_t is the largest eigenvalue of $\widehat{\mathbf{X}_t}\Sigma_{t|t-1}$ and $\mathbf{y}_t$ is the corresponding eigenvector. 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:

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

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

Proof. See Appendix A. □

Now, we define a steady-sate prior of the agents' information choice problem.

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 $\Sigma_{t+1|t}=\Sigma,\forall t\ge 0$. This implies that $(\Sigma,\mathbf{y})$ should satisfy the

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

following conditions:

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

where $\phi = \frac{\lambda}{2\ln 2}\left(\frac{1}{1-\mathbf{y}'\Sigma\mathbf{y}}\right)$.

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, we 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 $\left\{(c_{i,t})_{i=0}^{p-1}, (d_{i,t})_{i=0}^{q-1}\right\}_{t=0}^{\infty}$. 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[13] 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:

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

[13] 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 $\mathbb{E}_t^f\{x_{t+i}\} = \mathbb{E}\{x_{t+i}|\mathbf{u}_t\}$.

where $\kappa = \frac{1}{2}\log_2\left(\frac{2\ln 2}{\lambda}\mathbf{w}'\Sigma\mathbf{w}\right)$ and e_t is the agent's rational inattention error.

This result follows directly from the Euler equation in Theorem 1[14]. 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, that is $\mathbb{E}_t^f\{x_{t+i}\} = \rho^i x_t$; 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

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

where $\mathbf{w} = \left(1, \rho, \rho^2, \dots\right)$, $\kappa = \frac{1}{2}\log_2\left(\frac{2\ln 2}{\lambda}\mathbf{w}'\Sigma\mathbf{w} + \beta\rho^2\right)$ and $\mathbf{w}'\Sigma\mathbf{w} = \frac{1}{1-\rho^2 2^{-2\kappa}}$. Also, agent's optimal action profile, $a_t^*\left(s^t\right)$, is given by

$$a_t^*\left(s^t\right) = 2^{-2\kappa}\rho a_{t-1}^*\left(s^{t-1}\right) + \left(\sqrt{\left(1-2^{-2\kappa}\right)\mathbf{w}'\Sigma\mathbf{w}}\right)s_t$$

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 $x_t = (0.95+\rho)\,x_{t-1} - \rho x_{t-2} + u_t$,[15] 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[16]. While having x_t in their signal helps

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

[15] 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.

[16] 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.

the agent to both predict the current realization of the fundamental, and prepare themselves with 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 we 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.[17]

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 the agent's incentive to infer about future realizations of the fundamental decreases.[18] 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, which corresponds to full-information rational expectations, 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.

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

[18] 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, $\kappa \to \infty$.

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.

In this section, we 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 difference 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 $u_{t-i} \sim \mathcal{N}(0,1), \forall i \geq 0$. Now let $\mathbf{M}$ be the infinite dimensional lower shift matrix[19]. Thus we can write

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

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 $\mathbf{dw}'\left(\sum_{i=0}^{\infty} \mathbf{M}'^i\right)$ 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

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

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

[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 $L.u_t = u_{t-1}$.

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 sub-diagonal. Second, observe that

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

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

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

Thus,

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

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

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

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

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

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 $\tilde{\mathbf{u}}_0\sim\mathcal{N}\left(\tilde{\mathbf{u}}_{0|-1},\Sigma_{0|-1}\right)$, $\Sigma_{0|-1}\succeq 0$.

[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 $\tilde{\mathbf{u}}_t$.

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

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

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

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

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

$$\begin{aligned}
\mathcal{L}_0\left(\hat{\Sigma}_{0|-1}\right) \quad = \quad & \min_{\{\mathbf{dy}_t \in \hat{\mathcal{S}}_t^F\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left(\mathbf{dw}'\Sigma_{t|t}\mathbf{dw} + \lambda\kappa_t \right) \\
s.t. \quad & \Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}\mathbf{dy}_t\mathbf{dy}_t'\Sigma_{t|t-1} \\
& \Sigma_{t+1|t} = \left(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1'\right)\Sigma_{t|t}\left(\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1'\right) + \mathbf{e}_1\mathbf{e}_1' \\
& \frac{1}{2}\log_2\left(\frac{1}{1 - \mathbf{dy}_t'\Sigma_{t|t-1}\mathbf{dy}_t}\right) \leq \kappa_t \\
& \Sigma_{0|-1} \succeq 0 \text{ given.}
\end{aligned} \tag{5}$$

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:

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

where $\phi_t = \frac{\lambda}{2\ln 2}\left(\frac{1}{1-\mathbf{dy}_t'\Sigma_{t|t-1}\mathbf{dy}_t}\right)$ and $\mathbf{X}_t$ is the matrix of Lagrange multipliers on the evolution of the priors.

Lemma 3. *Suppose that the agent's fundamental follows a first order integrated ARIMA process x_t. Then the optimal signals are of the form*

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

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 $\left\{(b_{j,t})_{j=0}^{\infty}\right\}_{t=0}^{\infty}$ 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 $\left\{(c_{k,t})_{k=0}^{p-1}, (d_{l,t})_{l=0}^{q-1}\right\}$ *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, $a_t^*\left(s^t\right) = \mathbb{E}\left\{x_t | s^t\right\}$, 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 A Rational Inattention Phillips Curve

In the previous section, we develop a tractable method to solve a dynamic rational inattention problem under Linear-quadratic-Gaussian (LQG) set-up. In this section, we apply the results to a simple pricing model to derive a dynamic inattention Phillips curve, which has novel characteristics of inflation dynamics.

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 aggregater. This implies that the demand function of firm i is given by

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

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

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

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 $P_t^* = P^*\left(P_t, Y_t\right) \equiv \arg\max_x \Pi\left(x; P_t, Y_t\right)$ be the maximizer of this flow profit function at time t. Thus,

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

[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.

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

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

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

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

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.

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

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], we assume that the aggregate nominal GDP, $Q_t \equiv P_tY_t$, 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, we assume that the growth rate of nominal GDP follows an ARIMA(1,1,0)[26]:

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

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-

[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.

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

[26]In the section 3.4, we assume that the growth rate of nominal GDP has a *news shock* component to show how the firms' dynamic incentives to process information affects inflation dynamics.

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

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

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. we 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, we 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 A Rational Inattention Phillips Curve

In this section, we derive the dynamic rational inattention Phillips curve by solving the above rational inattention equilibrium. We first define agents' average forecast errors, which are the novel components in our dynamic inattention Phillips curve.

Definition 2. *For any stochastic process $\{x_t | t \geq 0\}$, we define firms' average forecast error of x at time t for horizon τ as*

$$\mathbb{FE}_t[x_{t+\tau}] \equiv \tilde{\mathbb{E}}_t\left[x_{t+j}\right] - \mathbb{E}_t^f\left[x_{t+j}\right]$$

where $\tilde{\mathbb{E}}_t[\cdot]$ is the average expectations at time t and $\mathbb{E}_t^f[\cdot]$ is the full-information rational expectation at time t.

With this definition, we derive a semi-analytical form of the dynamic inattention Phillips curve in the following lemma.

Lemma 4. ***(Dynamic Inattention Phillips Curve where $\beta > 0$)*** *By Lemma 3 that the optimal signals under dynamic inattention has the form $s_t^* = \sum_{j=0}^{\infty} \beta^j b_j \mathbb{E}_t^f \left[p_{t+j}^*\right] + e_t$. Then, given the sequence $(b_j)_{j=0}^{\infty}$, the Phillips curve under dynamic inattention is given by*

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\left[\pi_t + \alpha \Delta y_t\right] + c_0 \alpha y_t - \sum_{j=1}^{\infty} c_j \mathbb{F}\mathbb{E}_t\left[\pi_{t+j} + \alpha \Delta y_{t+j}\right] \tag{8}$$

where $c_j = 2^{2\kappa}\delta_0 \left(\sum_{k=j}^{\infty} \beta^k b_k\right)$ for every $j \geq 0$ and $\delta_0 \equiv \mathbf{dw}_{p^*}' \Sigma \mathbf{dy}$.

Proof. See Appendix A. □

We call (8) the dynamic inattention Phillips curve (DIPC). While we have to solve for $(c_j)_{j=0}^{\infty}$ numerically, this representation illustrate how dynamic inattention introduces a forward-looking behavior among agents. Unlike the forward-looking Phillips curves derived under nominal rigidities, current inflation is not necessarily increasing in expected inflation, and it can decrease with the forecast errors of firms about future inflation and output gap growth. More importantly, current inflation depends not only on current output, but also on all the future forecast errors. Therefore, by altering these expectations, any forward guidance policy can have immediate effects on inflation, and consequently output. Before we go through the inflation dynamics to forward guidance shocks in this model, we consider some special cases that we can derive analytical solutions to get intuitions of our Phillips curve.

Example 4. ***(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:

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

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 2 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 6 we show that even with a random walk fundamental, the optimal signal depends on β when shocks are announced beforehand.

Example 5. ***(Myopic Inattention Phillips Curve where*** $\beta = 0$***)*** 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

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\left[\pi_t + \alpha \Delta y_t\right] + \alpha\left(2^{2\kappa} - 1\right) y_t \tag{9}$$

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

Proof. See Appendix A. □

In contrast to (8), we call (9) the myopic inattention Phillips curve since it eliminates firms' dynamic incentives of information acquisition. 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 3 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 4 shows the impulse responses of inflation and output for different levels of capacity of processing information[27]. Higher capacity corresponds to a higher response of inflation 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 The Effect of News Shocks

3.4.1 Attention Allocation to News Shocks

In the previous section, we have considered some special cases of the model in which either β was assumed to be zero (example 5) or it was irrelevant due to strong assumptions on the nature of the fundamental (example 4). In this section, we 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. We start with a simple example of monetary news shocks, for which a closed form Phillips curve can be derived.

Example 6. ***(News Shock 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 4, 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

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

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:

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

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

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 5 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 4.

Figure 6 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 4: 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. □

3.4.2 Comparison with Sticky/Noisy Information Phillips Curves

Now, we compare our dynamic rational inattention Phillips curve with the other Phillips curves derived from other information friction models. This comparison highlights the novel forward-looking behavior in the dynamic inattention Phillips curve (8). First, let's consider the myopic inattention Phillips curve we derived in (9), which is the case of $\beta = 0$. This corresponds to a setting when firms choose to only observe their current fundamental, as shown above in example 5. This setting 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 5 that the myopic inattention Phillips curve is given by

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\left[\pi_t + \alpha \Delta y_t\right] + \alpha\left(2^{2\kappa} - 1\right) y_t, \tag{10}$$

where $\tilde{\mathbb{E}}_{t-1}\left[.\right] \equiv \int_0^1 \mathbb{E}\left[.|S^{i,t}\right] di$ denotes the average expectation of firms conditional on their time $t-1$ information given by the signal vector $S^{i,t} = (p_0^* + e_{i,t}, \ldots, p_{t-1}^* + e_{i,t})$, 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, we use the Phillips curve derived in Mankiw, Reis, et al. (2002):

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

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

We now compare the impulse responses of the three models to a forward guidance shock. Suppose 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.

Figure 7 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 we have set $\kappa = 0.5$, $\rho = 0.5$, and $\alpha = 0.8$. Moreover, in the sticky information model, we have set $\lambda = 0.2$, so that the peak of output and inflation in this model and noisy information (myopic inattention) one would be the same[28]. In both sticky information and myopic 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 myopic inattention model. Therefore, both inflation and output demonstrate more inertial behavior under dynamic inattention.

3.5 Estimation of the Dynamic Inattention Phillips Curve

To assess the empirical validity of our dynamic inattention Phillips curve, we estimate it using historical US inflation data. We are interested in that whether US inflation dynamics indeed shows both forward- and backward-looking behavior. Recall that the DIPC is written as,

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\left[\pi_t + \alpha \Delta y_t\right] + c_0 \alpha y_t + \sum_{j=1}^{\infty} c_j \left(\mathbb{E}_t^f\left[\pi_{t+j} + \alpha \Delta y_{t+j}\right] - \tilde{\mathbb{E}}_t\left[\pi_{t+j} + \alpha \Delta y_{t+j}\right]\right)$$

where $\tilde{\mathbb{E}}_t[\cdot]$ is firms' average expectation at time t and $\mathbb{E}_t^f[\cdot]$ is full-information rational expectation at time t. Notice that $\mathbb{E}_t^f\left[\pi_{t+j} + \alpha \Delta y_{t+j}\right] = \pi_{t+j} + \alpha \Delta y_{t+j} - v_{t+j}$ where v_{t+j} is independent and

[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.

identically distributed (i.i.d) rational expectation error. Let $\tilde{\mathbb{FE}}_t(X_{t+j}) = \tilde{\mathbb{E}}_t(X_{t+j}) - X_{t+j}$ be an ex-post average forecast error of variable X_{t+j} at time t. Then, we can rewrite the DIPC as

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\left[\pi_t + \alpha\Delta y_t\right] + c_0\alpha y_t - \sum_{j=1}^{\infty} c_j\left(\tilde{\mathbb{FE}}_t(\pi_{t+j}) + \alpha\tilde{\mathbb{FE}}_t(\Delta y_{t+j})\right) + \tilde{v}_{t+1}^{\infty} \tag{11}$$

where $\tilde{v}_{t+1}^{\infty} = -\sum_{j=1}^{\infty} c_j v_{t+j}$ is the weighted sum of rational expectation errors which are dated $t+1$ and later.

There are two main challenges to consistently estimate the above DIPC. First, the infinite amount of regressors on the right-hand side should be truncated. Second, although we use some proxies for firms' average forecasts, there might be measurement errors since it is difficult to directly observe firms' forecasts of inflation and changes in output gap. We now address each of these difficulties in turn.

In practice, we need to truncate the infinite amount of regressors in RHS to estimate our DIPC. This truncation will tend to provide a source of error. Specifically, equation (11) should be written as

$$\pi_t = \tilde{\mathbb{E}}_{t-1}\left[\pi_t + \alpha\Delta y_t\right] + c_0\alpha y_t - \sum_{j=1}^{J} c_j\left(\tilde{\mathbb{FE}}_t(\pi_{t+j}) + \alpha\tilde{\mathbb{FE}}_t(\Delta y_{t+j})\right) + \tilde{v}_{t+1}^{\infty} + \xi_{t,t+J}$$

where $\xi_{t,t+J} = -\sum_{j=J+1}^{\infty} c_j\left[\tilde{\mathbb{FE}}_t(\pi_{t+j}) + \alpha\tilde{\mathbb{FE}}_t(\Delta y_{t+j})\right]$. Because this additional source of error is dated t, the orthogonality condition will generally fail if we estimate the equation using ordinary least squares or nonlinear least squares. However, consider the covariance of any variable z with $\xi_{t,t+J}$:

$$cov\left(z, \xi_{t,t+J}\right) = -\sum_{j=J+1}^{\infty} c_j\left[cov\left(z,\ \tilde{\mathbb{FE}}_t(\pi_{t+j} + \alpha\Delta y_{t+j})\right)\right].$$

This covariance will be nonzero unless z is uncorrelated with all ex-post forecast errors dated t of future inflation and changes in the output gap. However, because each covariance is weighted by c_j, it follows that the covariance of any regressors with $\xi_{t,t+j}$ will converge to 0 as J goes to infinity as long as c_j converges to zero as J increases and the covariance of z with ex-post forecast errors is not too explosive. Quantitatively, truncating ex-post forecast errors should thus have little effect on the estimation for a large enough J. In Figure 9, we numerically show that the coefficient $\{c_j\}$vanishes quickly to zero as the forecast horizons increase. For example, if the nominal demand follows a ARIMA(1,1,0) process, the coefficients $\{c_j\}$ of DIPC converge to zero in 3-4 periods. Based on this quantitative results, we set our estimation equation as the following:

$$\pi_t = \beta\tilde{\mathbb{E}}_{t-1}\left[\pi_t + \alpha\Delta y_t\right] + c_0\alpha y_t - \sum_{j=1}^{J} c_j\left(\tilde{\mathbb{FE}}_t(\pi_{t+j}) + \alpha\tilde{\mathbb{FE}}_t(\Delta y_{t+j})\right) + \tilde{v}_{t+1}^{\infty} \tag{12}$$

where $\tilde{v}_{t+1}^{\infty}$ is the error term which includes rational expectation errors dated $t+1$ and later.

Another source of difficulty for consistent estimation of the DIPC is a potential measurement error of expectation terms in RHS. Since firms' true expectations of inflation and changes in output gap are not observed in data, we need to proxy the average expectations of inflation and changes in output gap. We use inflation and GDP forecasts data from Survey of Professional Forecasters(SPF). However, one might have a concern about whether the forecasts data are relevant proxies for the firms' expectations. Especially, the ex-post forecast errors in RHS are likely to be vulnerable to measurement error problem since the dispersion of forecasts of variables in SPF is increasing as the forecast horizons increase. This measurement error issue can be addressed using the instrumental variable approach. A set of instruments should be correlated with the RHS variables, but not with the error term which is dated $t+1$ and later. Notice that lags of output gap and past expectations of current inflation and changes in output gap are valid instruments. However, it is not obvious to find a set of instruments for ex-post forecast errors of future inflation and changes in output gap. We argue that forecast revisions for future inflation and changes in output gap are valid instruments for the average ex-post forecast errors. This choice of instruments is based on a theoretical prediction of general noisy information models like ours. We exploit the results of Coibion and Gorodnichenko(AER 2015) that the average forecast errors can be predicted by ex-ante forecast revisions in models with information frictions. For example, if the variable of interest for forecast follows AR(1) process, then noisy information model implies that

$$\tilde{\mathbb{FE}}_t\left(x_{t+j}\right) = \frac{1-G}{G}\left(\mathbb{F}_t x_{t+j} - \mathbb{F}_{t-1} x_{t+j}\right) + \nu_{t,t+j}$$

where G is the implied Kalman gain, $\mathbb{F}_t(\cdot)$ is an average forecast at time t and $v_{t,t+j}$ is rational expectations error. This tells us that average ex-post forecast errors are systemically related with current average forecast revision if the forecasters are subject to a noisy or sticky information friction.[29] In general, under the information friction models like ours, average forecast errors can be systemically predicted by current and past average forecast revisions. Based on this prediction, we choose the set of instruments: 1) past forecasts of inflation and changes in output gap, 2) lags of output gap, 3) current and lags of forecast revisions for inflation and changes in output gap.

3.5.1 Data

We use mean expectations data from the Survey of Professional Forecasters (SPF).[30] The SPF data provide an ideal source of expectations because they are a direct measure of what economists were forecasting and are available on a quarterly basis. To generate expectations of changes in the output gap, we follow Coibion (2006) and assume that forecasters knew the actual changes in the

[29] Coibion and Gorodnichenko (2015) test the noisy/sticky information models predictions by assuming that true inflation dynamics follows AR(1) process and setting up the following estimation equation: $\tilde{\mathbb{FE}}_t\left(x_{t+j}\right) = \beta\left(\mathbb{F}_t x_{t+j} - \mathbb{F}_{t-1} x_{t+j}\right) + \nu_{t,t+j}$. The null hypothesis is $\beta = 0$ and they find that β is significantly different from zero and positive. This tells that the forecasters are subject to sticky or noisy information. This results also hold if one assumes that inflation follows AR(p) process or VAR(p) process or if one uses other variables such as GDP growth.

[30] SPF data are available at the Philadelphia Federal Reserve Board: https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters. Median forecasts were also used and yielded qualitatively similar results.

Congressional Budget Office (CBO) measure of potential output and derive expectations of future changes in the output gap as expected changes in output minus actual changes in the CBO measure of potential output. Although SPF forecasts data are provided for the next four quarters, since we instrument ex-post average forecast errors using ex-ante average forecast revisions, we lose one forecast horizon. Thus, we set $J = 3$.

We use four lags for the lags of instruments. The baseline sample period is 1972:Q1-2016:Q4, but we check the robustness using only Post-Volcker period(1979:Q4-2016:Q4). The baseline inflation measure we use is the annualized growth of GDP deflator. We use CPI data also for robustness check although CPI forecasts data in SPF are only available after 1982:Q4.

3.5.2 Results

We estimate the equation (12) with non-linear GMM using instrumental variables. The null hypothesis is that the coefficients $\{c_j\}$ on forward-looking terms are significantly different zero. It is a noble characteristic that our DIPC has compared to other Phillips curves from sticky/noisy information models. Table 1 shows the estimates of coefficients in our DIPC. First, the coefficients on forward-looking terms, $\{c_j\}$, are consistent to our theory: those are significantly different from zero. This result holds when we use the sample of post-Volcker period or CPI inflation. Second, the coefficient on backward-looking term in the DIPC is also significantly different from zero, and is not statistically different from one, which is consistent to the model. Thus, we verify that the US inflation dynamics is consistent to our DIPC and has the properties of the both forward- and and backward-looking behavior. Third, we can also estimate a structural parameter of the model. When we estimate using the entire sample, the degree of strategic complementarity is quite high, $\alpha = 0.03$. Post-Volcker period with CPI inflation gives the estimate of the low degree of strategic complementarity, $\alpha = 0.28$.

4 Forward Guidance under Rational Inattention

In this section, we study the effects of forward guidance policy in a general equilibrium model where each firm is rationally inattentive.

4.1 Households

We assume that households are fully informed about prices and wages and maximize their life-time utilities:

$$\max_{C_t, L_{it}, B_t} \mathbb{E}_0^f \left[\sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{\int_0^1 \left(L_{it}^S\right)^{1+\psi} di}{1+\psi} \right) \right] \tag{13}$$

$$\text{s.t.} \quad P_t C_t + B_t = \int_0^1 W_{it} L_{it}^S di + R_{t-1} B_{t-1} + T_t$$

where L_{it}^{S} is firm-specific labor supply, R_{t-1} is nominal interest rate, B_t is nominal risk-less one-period government bonds, and T_t is the aggregate profits of firms. The first-order optimal conditions are:

$$
\begin{aligned}
C_t^{-\sigma} &= \beta R_t \mathbb{E}_t^f \left[C_{t+1}^{-\sigma} \frac{P_t}{P_{t+1}} \right] \\
C_t^{\sigma} \left(L_{it}^{S} \right)^{\psi} &= \frac{W_{it}}{P_t}
\end{aligned}
\tag{14}
$$

4.2 Firms

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 aggregater.

There are intermediate good producers in the monopolistically competitive market. This implies that the demand function of intermediate goods firm i is given by

$$
Y_{it} = \left(\frac{P_{it}}{P_t} \right)^{-\sigma} Y_t
$$

where σ is the elasticity of substitution across goods, Y_{it} 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. Let each firm has a linear production technology: $Y_{it} = L_{it}^{D}$ where L_{it}^{D} is labor demand of firm i. Firm's (real) marginal cost is:

$$
mc_{it} = \frac{W_{it}}{P_t}
$$

Given the firm's optimal information choice, the firm i maximizes its flow profit:

$$
\max_{P_{it}} \ \mathbb{E}_t^i \left[\Pi \left(P_{it}; P_t, Y_t \right) \right] = \mathbb{E}_t^i \left[\left(\frac{P_{it}}{P_t} - mc_{i,t} \right) \left(\frac{P_{it}}{P_t} \right)^{-\varepsilon} Y_t \right]
$$

where $\mathbb{E}_t^i [\cdot] = \mathbb{E}_t \left[\cdot \mid S_t^i \right]$ given the history of signals S_t^i. The first-order optimal condition gives:

$$
P_{it} = \frac{\varepsilon}{(\varepsilon - 1)} \mathbb{E}_t^i \left[P_t mc_{it} \right]
$$

Using market clearing conditions, stated in Definition 3, we can rewrite firm i's optimal price given her information set as

$$
P_{it}^{1+\varepsilon\psi} = \frac{\varepsilon}{(\varepsilon - 1)} \mathbb{E}_t^i \left[P_t^{1+\varepsilon\psi} Y_t^{\sigma+\psi} \right]
$$

A Taylor expansion of this first order condition around an optimal non-stochastic steady-states gives

$$
p_t^* = p_t + \alpha y_t
$$

where small letters denote the log-deviation from the optimal non-stochastic point around which we have linearized the equation, and $\alpha = \frac{\sigma+\psi}{1+\varepsilon\psi}$ is the degree of strategic complementarity. Now, define the function $\mathcal{L}(P_{it}; P_t, Y_t)$ as the flow loss in the profit of firm i for any given price P_{it}:

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

This loss function, up to a second order approximation, is proportional to the quadratic difference between p_{it} and p_t^*

$$\mathcal{L}(P_{i,t}; P_t, Y_t) = \frac{1}{2}(p_{it} - p_t^*)^2.$$

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.

4.3 Monetary Policy and General Equilibrium

We assume that monetary authority is fully rational and monetary policy is given by the standard Taylor rule:

$$R_t = (R_{t-1})^\rho \left(\left(\frac{P_t}{P_{t-1}} \right)^{\phi_\pi} \left(\frac{Y_t}{\bar{Y}} \right)^{\phi_y} \right)^{(1-\rho)} \exp(u_{t-k}) \tag{15}$$

where u_{t-k} is the k-periods ahead monetary news shock.

Let $\mathbf{u}_t = (u_t, u_{t-1}, u_{t-2}, \cdots)'$ be the vector of monetary policy shocks until time t. From 2.3, we can define the random walk part of the shocks as $\tilde{\mathbf{u}}_t \equiv (\mathbf{I} - \mathbf{M}')^{-1} \mathbf{u}_t$. Let's denote the perceived covariance matrix as $\Sigma^i_{t|t-1} = \mathbb{E}^i_{t-1}\left\{ \left(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}^i_{t|t-1} \right) \left(\tilde{\mathbf{u}}_t - \tilde{\mathbf{u}}^i_{t|t-1} \right)' \right\}$.

Definition 3. A general equilibrium for the economy is an allocation for the household $\Omega^H = \left\{ C_t, B_t, (L_{it})_{i \in I} \right\}_{t=0}^\infty$, a optimal signal and allocation profile for firms given an initial set of signals, $\Omega^F = \left\{ \left(\mathbf{dy}_{it} \in \hat{\mathcal{S}}^F_t, p_{it}, L^D_{it} \right)_{t=0}^\infty \right\}_{i \in I} \times \left\{ S_i^{-1} \right\}_{i \in I}$, and a set of prices $\left\{ R_t, P_t, (W_{it})_{i\in I} \right\}_{t=0}^\infty$ such that

1. Households: given prices and Ω^F, the household's allocation solves their problem as specified in Equation (13).

2. Firms: given Ω^H, and the implied labor supply and output demand curves, $\{\mathbf{dy}_{it}\}_{t=0}^\infty$solves

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

where $p_t^* = \mathbf{dw}'_{p^*} \tilde{\mathbf{u}}_t$, $p_{it} = \mathbb{E}^i_t [p_t^*]$, and $L^D_{it} = Y_{it} = \left(\frac{P_{it}}{P_t} \right)^{-\sigma} Y_t$.

3. Monetary Policy: given Ω^H and Ω^F, the monetary authority follows the Taylor rule specified in Equation (15).

4. Market clears: $C_t = Y_t,\ B_t = 0,\ L_{it} = L_{it}^D\ \ \forall i$.

4.4 Solution Algorithm

The linearized version of IS equation (14) around non-stochastic steady-state, combined with the market clearing conditions, gives:

$$y_t = \mathbb{E}_t^f y_{t+1} - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t^f \pi_{t+1}\right). \tag{16}$$

Also, the linearized version of Taylor rule (15) is given by:

$$\hat{R}_t = \rho \hat{R}_{t-1} + (1-\rho)\,\phi_\pi \pi_t + \phi_y y_t + u_{t-k}. \tag{17}$$

Now, we guess $\pi_t = \mathbf{dw}'_\pi \tilde{\mathbf{u}}_t$, $y_t = \mathbf{dw}'_y \tilde{\mathbf{u}}_t$ and $\hat{R}_t = \mathbf{dw}'_R \tilde{\mathbf{u}}_t$. We guess all these three variables are stationary so that the tails of $(\mathbf{dw}_\pi, \mathbf{dw}_y, \mathbf{dw}_R)$ all converge to zero. We truncate these vectors with an arbitrary large length T. Then using state-space formulation, we can rewrite the equation (16) and (17) as the following:

$$\begin{aligned}
(\mathbf{I} - \rho\mathbf{M})\,\mathbf{dw}_i &= (1-\rho)\,\phi_\pi \mathbf{dw}_\pi + (1-\rho)\,\phi_y \mathbf{dw}_y + (\mathbf{I} - \mathbf{M})\left(\mathbf{M}'\right)^k \mathbf{e}_1 \\
\mathbf{dw}_i &= \sigma\left(\mathbf{M}'_p - \mathbf{I}\right)\mathbf{dw}_y + \mathbf{M}'_p \mathbf{dw}_\pi
\end{aligned}$$

where $\mathbf{M}_p = \mathbf{M} + \mathbf{e}_1\mathbf{e}'_1$ and $\mathbf{e}_1$ is the first column of the $T \times T$ identity matrix. Now, by combining the above two equations, we have

$$\begin{aligned}
&\left[\sigma\,(\mathbf{I} - \rho\mathbf{M})\left(\mathbf{M}'_p - \mathbf{I}\right) - (1-\rho)\,\phi_y \mathbf{I}\right]\mathbf{dw}_y \\
&\quad = \left[(1-\rho)\,\phi_\pi \mathbf{I} - (\mathbf{I} - \rho\mathbf{M})\,\mathbf{M}'_p\right]\mathbf{dw}_\pi + (\mathbf{I} - \mathbf{M})\left(\mathbf{M}'\right)^k \mathbf{e}_1.
\end{aligned}$$

We consider a symmetric steady-state rational inattention equilibrium to the model, where $\mathbf{dy}_{it} = \mathbf{dy}$ and $\Sigma^i_{t|t-1} = \Sigma$ for every i and t. To solve the equilibrium, first, we start from guessing the optimal pricing rule: $p_t^* = \mathbf{dw}'_{p^*}\tilde{\mathbf{u}}_t$. Second, given $p_t^* = \mathbf{dw}'_{p^*}\tilde{\mathbf{u}}_t$, each firm solves the rational inattention problem and finds the optimal signal $s^*_{i,t} = \mathbf{dw}'_s \tilde{\mathbf{u}}_t + e_{i,t}$. Third, given the set of $\{p_{it}\}_{i\in I}$, where each p_{it} is implied by the Kalman filtering of the sequence of the optimal signals, we find that $p_t = \mathbf{dw_p}'\tilde{\mathbf{u}}_t$. Note that $\pi_t = p_t - p_{t-1} = \mathbf{dw_p}'\,(\mathbf{I} - \mathbf{M}')\,\tilde{\mathbf{u}}_t$, thus $\mathbf{dw}_\pi = (\mathbf{I} - \mathbf{M})\,\mathbf{dw_p}$. Fourth, using $\mathbf{dw_p}$ and equation (8), find $\mathbf{dw_y}$ and then calculate $p_t^* = p_t + \alpha y_t \Longleftrightarrow \mathbf{dw}_{p^*} = \mathbf{dw}_p + \alpha\mathbf{dw}_y$. Finally, we update our guess $\mathbf{dw}_{p^*}$ and iterate until we have a convergence.

4.5 Results

4.5.1 Parameterization

Table 2 contains numerical values we used for the parameters of the model. The parameterization is standard. In the baseline, we use log preference ($\sigma = 1$) as well as a unit Frisch elasticity of labor supply $\left(\frac{1}{\psi} = 1\right)$. The elasticity of substitution between goods is 6, which implies about 20% of steady-state markup. The implied strategic complementarity is also standard in literature($\alpha = 0.29$). We set the capacity of processing information $\kappa = 0.5$, which is consistent with the estimate from Coibion and Gorodnichenko (2015a). For the monetary policy feedback rule, we assume that there is no interest rate smoothing ($\rho = 0$), and set $\phi_\pi = 1.2$ and $\phi_y = 0.2$.

4.5.2 Effects of the Forward Guidance Shock and Forward Guidance Puzzle

Impulse responses of variables to an expansionary monetary policy shock (of 1%) that will be realized 4 quarters later are shown in Figure 10. Let's first look at the impulse responses of the myopic inattention model ($\beta = 0$). Unlike the previous exogenous nominal demand exercise in Figure 7, even though firms are not forward-looking in their information acquisition, inflation responds immediately to the forward guidance shock. This is because households are fully rational and forward-looking, and thus they increase the demand for goods when the shock is announced. Since the increased demand pushes marginal costs to increase, firms increase their prices, and thus aggregate inflation increases. The nominal interest rate increases as well given the interest rate feedback rule. When the news shock is actually realized at 4 quarters, the nominal interest rate falls. Notice that in a standard Calvo sticky price model, output, inflation, and the short-term rate all go back to steady-state immediately after the shock realization since the model is completely forward-looking and the shock is transitory. In contrast, after the shock realization, inflation slowly converges to the steady-state in the myopic inattention model. The intuition is that the myopic inattention firms are backward-looking in their optimal price decisions because of the noisy signals that they choose. In aggregate, this backward-looking nature is clearly shown in the myopic Phillips curve in (10). Moreover, after the shock realization, output contracts for a while, and converges to the steady-state. Since firms are rationally inattentive, they optimally choose to observe noisy signal about the fundamentals, and increase their prices even after the shock realization. Real interest rate goes up through the the interest rate feedback rule, and thus output contract. Thus, unlike the standard New Keynesian sticky price model, the myopic inattention model implies that the expansionary effect of forward guidance policy is not a free lunch: it comes with output contractions which slowly converges to the steady-state as firms still increase their prices due to the noisy signal that they optimally choose to observe.

For the dynamic inattention model where $\beta = 0.99$, the overall dynamics are similar to the myopic inattention model. One notable difference is the large responses of inflation to the forward guidance shock until the shock is actually realized. The intuition is clear: here, firms are forward-looking in their optimal attention choices. In addition to the increased demand from the forward-looking households, firms expect that future fundamental (their marginal costs) will be large because of the future expansionary shock. This expectation passes through their optimal price decisions, and thus they increase their prices much more than the myopic inattention firms. The large response of inflation generates output booms by less compared to the myopic inattention model.

Our model survives the forward guidance puzzle, established in Del Negro, Giannoni, and Patterson (2012), in the standard New Keynesian models. The puzzle says the effectiveness of forward guidance shock is increasing as the horizons of the forward guidance increase because of the completely forward-looking behavior of households and firms. The literature proposes various resolutions to this puzzle, by reducing this forward-looking behavior of households or firms.[31] Our model also weakens the forward-looking behavior of firms because firms are rationally inattentive and choose to observe a noisy signal about their marginal costs. Figure 11 compares the initial response of inflation to the forward guidance shocks of different horizons in three different models: sticky price, myopic inattention, and dynamic inattention. The initial response of inflation is decreasing in the horizon of the forward guidance in the myopic and dynamic inattention models, while it is increasing in the Calvo sticky price model. In Figure 12, we find that initial output response is also small in the inattention models compared to the standard sticky price model.

5 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.

We apply this result to the pricing theory, and show that a Phillips curve that emerges under dynamic rational inattention relates current inflation to the future forecast errors about inflation and output gap growth, 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

[31] For example, McKay, Nakamura, and Steinsson (2016) introduce the incomplete market assumption for households, which make households are less forward-looking due to the self-insurance motive. Other papers introduce some information frictions in the model, which generates backward-looking behavior of firms' optimal decisions. (e.g., Angeletos and Lian (2016) for imperfect common knowledge, Gabaix (2016) for a behavioral model; and Carlstrom, Fuerst, and Paustian (2015); Kiley (2016) for sticky information.)

past realizations, have been shown to be necessary to match the observed pattern of it in the data. However, current models of micro-founded 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 micro foundation. In this paper, we develop a micro foundation of the hybrid Phillips curve from the rational inattention model with dynamic incentives of processing information. Our empirical estimation of the dynamic inattention Phillips curve using the Survey of Professional Forecasters as a proxy for firms' expectations confirms that the US inflation dynamics has both backward- and forward-looking behavior.

In order to demonstrate the forward-looking behavior that is micro-founded under our dynamic inattention model, we implement a simple forward guidance exercise, in which shocks to aggregate demand is announced before taking effect. We 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. In the general equilibrium model, we show that this forward-looking behavior of firms' information acquisition makes the forward guidance policy effective by increasing inflation and booming output. However, the initial responses of inflation and output decrease with the horizon of forward guidance because of the backward-looking term in the Phillips curve, which is coming from the noisy information that firms optimally choose to acquire. Thus, our model survives the forward guidance puzzle.

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 micro-founded information rigidity model that can fill this void.

References

Afrouzi, H. (2016): "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money," Job market paper, University of Texas at Austin.

Angeletos, G.-M., and C. Lian (2016): "Forward Guidance without Common Knowledge," Working Paper 22785, National Bureau of Economic Research.

Brissimis, S. N., and N. S. Magginas (2008): "Inflation Forecasts and the New Keynesian Phillips Curve," *International Journal of Central Banking*, 4(2), 1–22.

Campbell, J. R., C. L. Evans, J. D. Fisher, and A. Justiniano (2012): "Macroeconomic effects of Federal Reserve forward guidance," *Brookings Papers on Economic Activity*, 2012(1), 1–80.

Carlstrom, C. T., T. S. Fuerst, and M. Paustian (2015): "Inflation and output in New Keynesian models with a transient interest rate peg," *Journal of Monetary Economics*, 76, 230–243.

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

Coibion, O. (2010): "Testing the Sticky Information Phillips Curve," *The Review of Economics and Statistics*, 92(1), 87–101.

Coibion, O., and Y. Gorodnichenko (2015a): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105(8), 2644–78.

——— (2015b): "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, 7(1), 197–232.

Del Negro, M., M. Giannoni, and C. Patterson (2012): "The forward guidance puzzle," Staff Reports 574, Federal Reserve Bank of New York.

Farhi, E., and I. Werning (2017): "Monetary Policy, Bounded Rationality, and Incomplete Markets," Working Paper 23281, National Bureau of Economic Research.

Gabaix, X. (2016): "A Behavioral New Keynesian Model," Working Paper 22954, National Bureau of Economic Research.

Gali, J., and M. Gertler (1999): "Inflation dynamics: A structural econometric analysis," *Journal of monetary Economics*, 44(2), 195–222.

Kiley, M. T. (2016): "Policy paradoxes in the New Keynesian model," *Review of Economic Dynamics*, 21, 1–15.

Krugman, P. R., K. M. Dominquez, and K. Rogoff (1998): "It's baaack: Japan's slump and the return of the liquidity trap," *Brookings Papers on Economic Activity*, 1998(2), 137–205.

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

Luo, Y., and E. Young (2013): "Rational Inattention in Macroeconomics: A Survey," MPRA Paper 54267, University Library of Munich, Germany.

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

——— (2015): "Business Cycle Dynamics under Rational Inattention," *The Review of Economic Studies*, 82(4), 1502–1532.

Maćkowiak, B. A., F. Matejka, and M. Wiederholt (2016): "The rational inattention filter," CEPR Discussion Papers 11237, C.E.P.R. Discussion Papers.

Mankiw, N. G., R. Reis, et al. (2002): "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *The Quarterly Journal of Economics*, 117(4), 1295–1328.

McKay, A., E. Nakamura, and J. Steinsson (2016): "The Power of Forward Guidance Revisited," *American Economic Review*, 106(10), 3133–58.

Melosi, L. (2017): "Signalling Effects of Monetary Policy," *The Review of Economic Studies*, 84(2), 853–884.

Nimark, K. (2008): "Dynamic pricing and imperfect common knowledge," *Journal of Monetary Economics*, 55(2), 365 – 382.

Paciello, L. (2012): "Monetary Policy and Price Responsiveness to Aggregate Shocks under Rational Inattention," *Journal of Money, Credit and Banking*, 44(7), 1375–1399.

Paciello, L., and M. Wiederholt (2014): "Exogenous Information, Endogenous Information, and Optimal Monetary Policy," *The Review of Economic Studies*, 81(1), 356–388.

Petersen, K. B., and M. S. Pedersen (2012): "The Matrix Cookbook," .

Roberts, J. M. (2005): "How Well Does the New Keynesian Sticky-Price Model Fit the Data?," *Contributions in Macroeconomics*, 5(1), –.

Rudd, J., and K. Whelan (2005): "New tests of the new-Keynesian Phillips curve," *Journal of Monetary Economics*, 52(6), 1167 – 1181, The econometrics of the New Keynesian price equation.

Šauer, R. (2016): "The imperfect-common-knowledge Phillips curve: Calvo vs Rotemberg," *Economics Letters*, 148, 45 – 47.

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

Sims, C. A. (2010): "Rational Inattention and Monetary Economics," vol. 3 of *Handbook of Monetary Economics*, pp. 155 – 181. Elsevier.

Steiner, J., C. Stewart, and F. Matejka (2017): "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 85(2), 521–553.

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.*

——— (2009): "Information-constrained state-dependent pricing," *Journal of Monetary Economics*, 56, S100–S124.

6 Tables and Figures

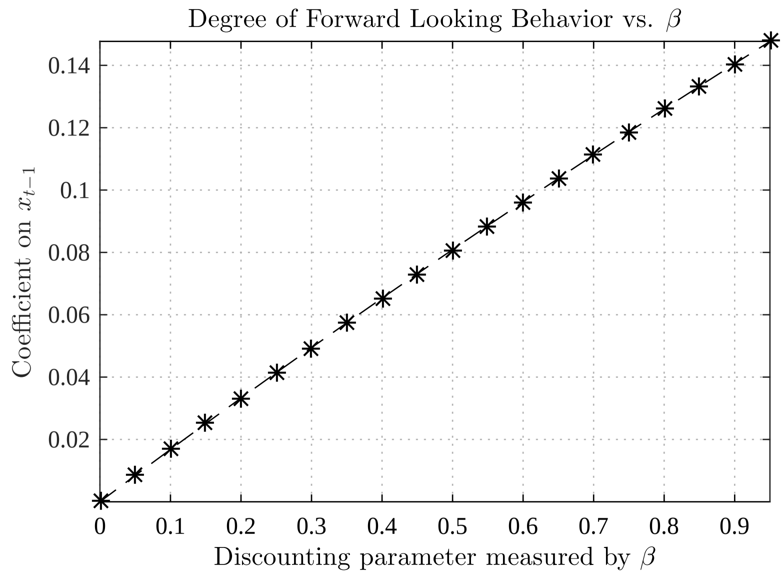
Table 1: Estimation of Dynamic Inattention Phillips Curve

	GDP deflator (72Q1:16Q4)		GDP deflator Post-Volcker (79Q4:16Q4)		CPI (82Q4:16Q4)	
	(1)	(2)	(3)	(4)	(5)	(6)
constant	-0.00		-0.03		0.81***	
	(0.04)		(0.05)		(0.15)	
β	1.01***	1.00***	1.03***	1.02***	0.78***	1.01***
	(0.01)	(0.00)	(0.01)	(0.01)	(0.04)	(0.02)
α	0.03***	0.03***	0.09***	0.09***	0.27***	0.28***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)
c_0	1.06*	1.10**	0.90***	0.94***	0.86***	0.69***
	(0.61)	(0.50)	(0.14)	(0.12)	(0.14)	(0.11)
c_1	0.45***	0.45***	0.10***	0.11***	0.09***	0.10***
	(0.07)	(0.07)	(0.04)	(0.04)	(0.03)	(0.03)
c_2	0.17***	0.17***	0.18***	0.18***	-0.33***	-0.34***
	(0.06)	(0.06)	(0.03)	(0.03)	(0.04)	(0.04)
c_3	0.13***	0.13***	0.18***	0.17***	-0.13***	-0.12***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
Over-identification Test $J-stat$	8.06 (p = 1.0)	8.08 (p = 1.0)	7.30 (p = 1.0)	7.30 (p = 1.0)	6.82 (p = 1.0)	6.82 (p = 1.0)
N	172	172	146	146	134	134

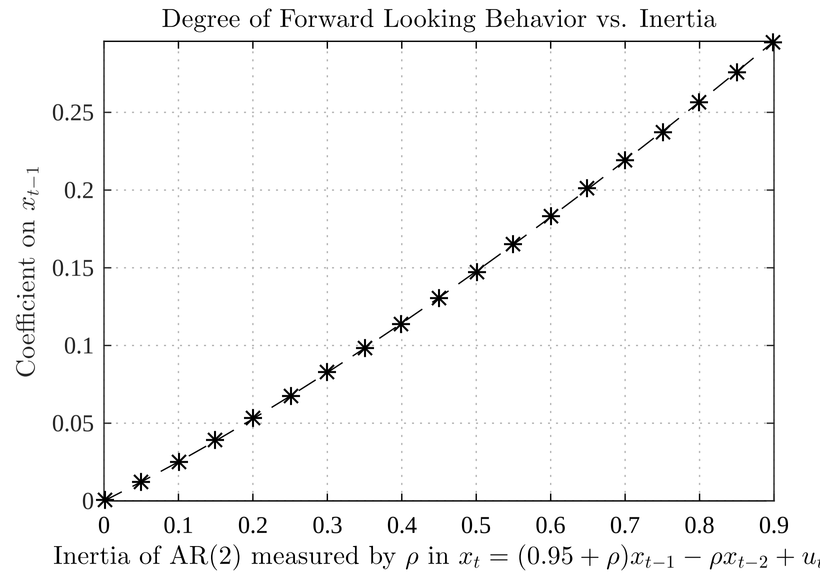
Newey-West robust standard errors are in parentheses.
***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table 2: Parameterization of the Model

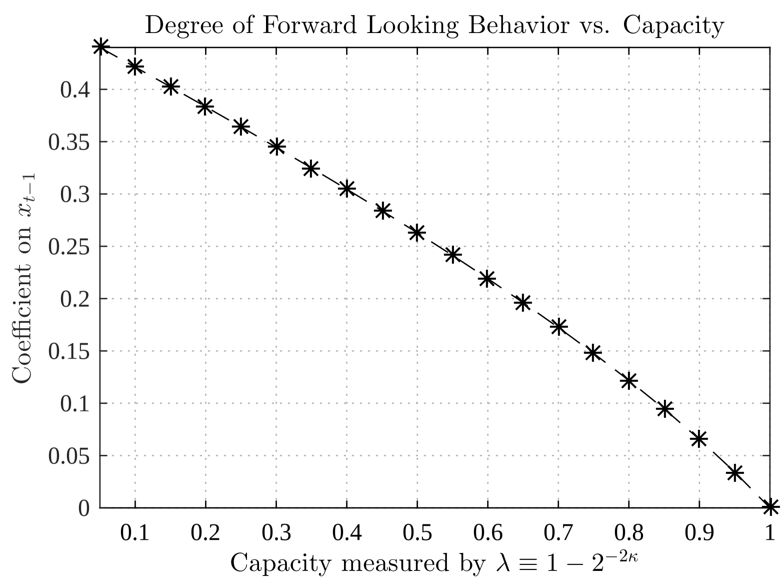
Parameter	Calibrated Value	Description
β	0.0 or 0.99	time preference
σ	1	log preference
ψ	1	inverse of elasticity of substitution for labor supply
ε	6	elasticity of substitution among goods
α	$\frac{\sigma+\psi}{1+\varepsilon\psi}$	strategic complementarity
λ	$1-2^{-2\kappa}=0.5$	information processing capacity parameter
ρ	0.0	interest rate smoothing parameter
ϕ_π	1.2	elasticity of interest rate to inflation
ϕ_y	0.2	elasticity of interest rate to output



(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.

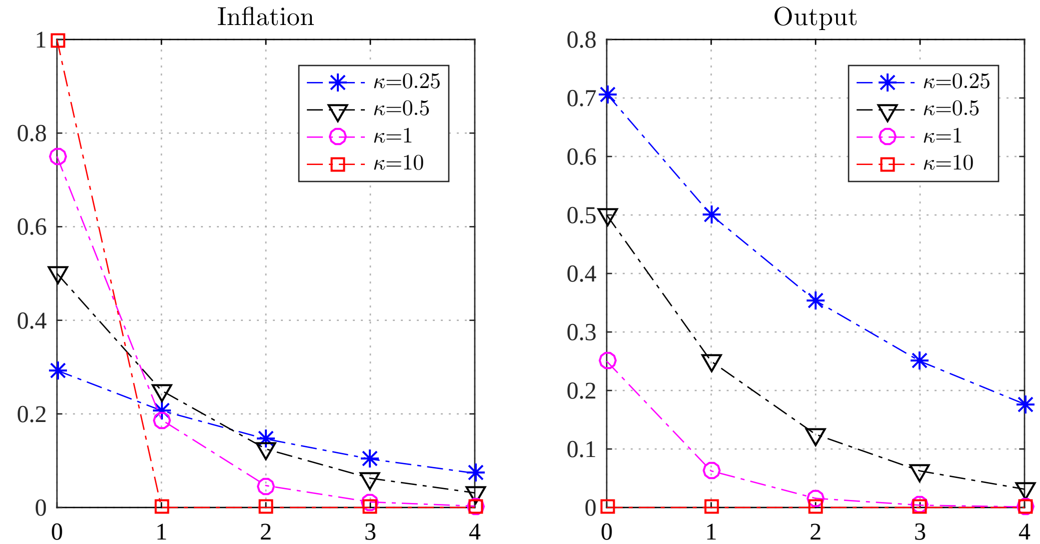


Figure 2: IRFs for example 4: The figure shows the impulse responses of output and inflation to a 1% shock to the aggregate demand, for different levels of capacity of processing information. Rational inattention creates endogenous real and persistent effects for monetary policy. See example 4 for details.

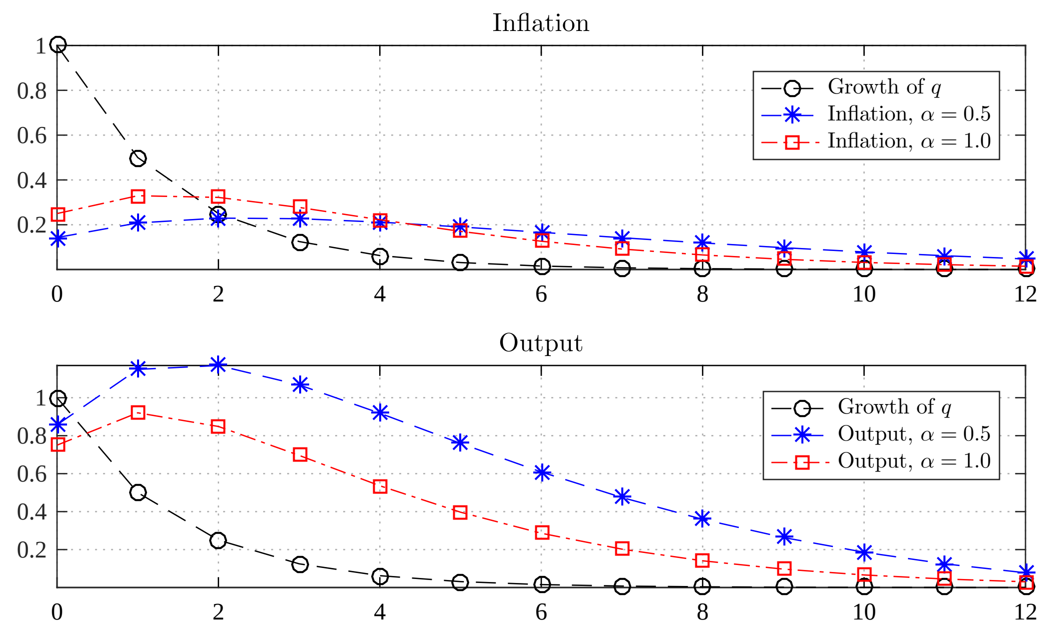


Figure 3: IRFs for example 5: The figure shows impulse responses of output and inflation for a capacity parameter of $\kappa = 0.2$. The red curves with circle markers are the IRFs of the model with no strategic complementarity ($\alpha = 1$), and the blue curves with star markers are the IRFs when $\alpha = 0.5$. Higher strategic complementarity introduces higher inertia in response of inflation, and amplifies the response of the output. See example 5 for details.

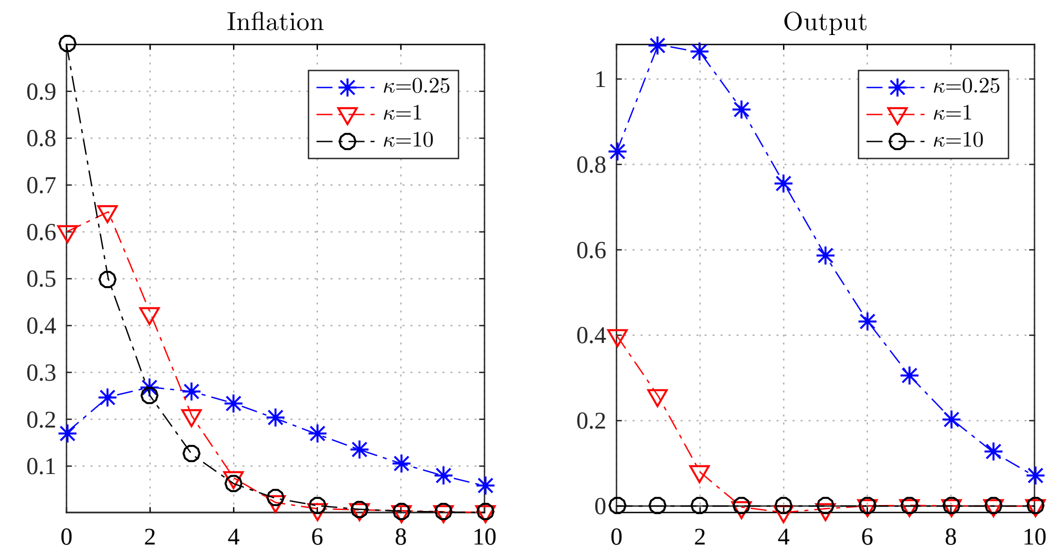


Figure 4: IRFs for example 5: the figure shows impulse responses of output and inflation for different values of capacity. Higher capacity leads to less inertial response of inflation and a smaller and less persistent response of output. When capacity is very large, inflation exactly follows the AR(1) path of the growth of the aggregates demand, and output does not respond to monetary policy at all. See example 5 for details.

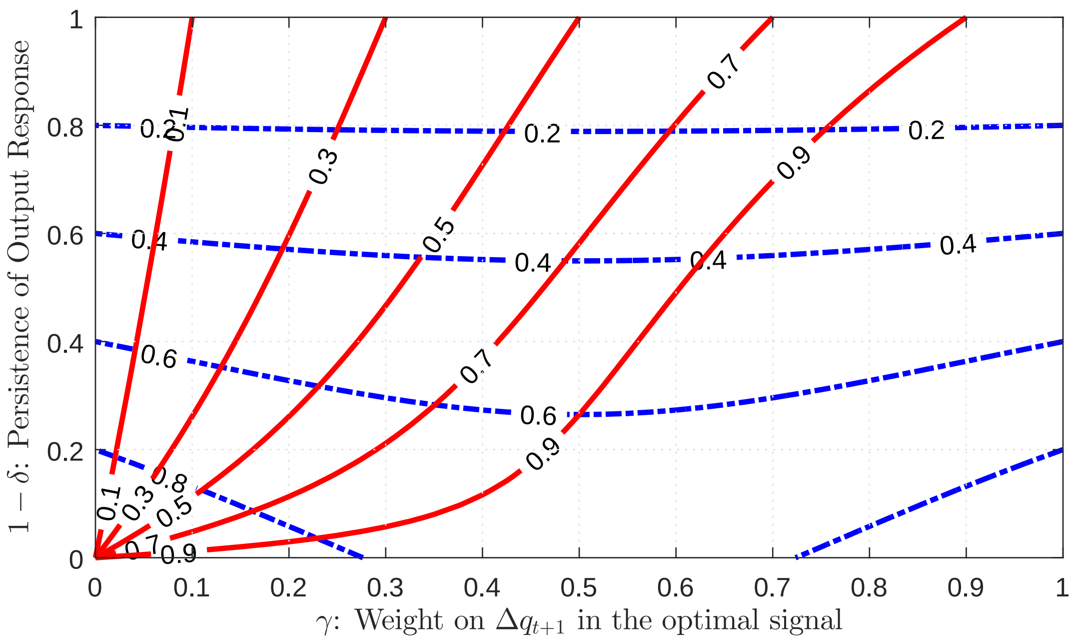


Figure 5: The figure depicts the iso-capacity (measured by $\lambda \equiv 1 - 2^{-2\kappa} \in [0, 1)$) curves in blue dashed lines, and iso-patience (measured by $\beta \in [0, 1)$) curves in red solid lines. Each intersection gives an equilibrium pair of (γ, δ). Higher capacity or lower patience correspond to a less forward-looking behavior in the information acquisition of the firms when there is forward guidance. See example 6 for details.

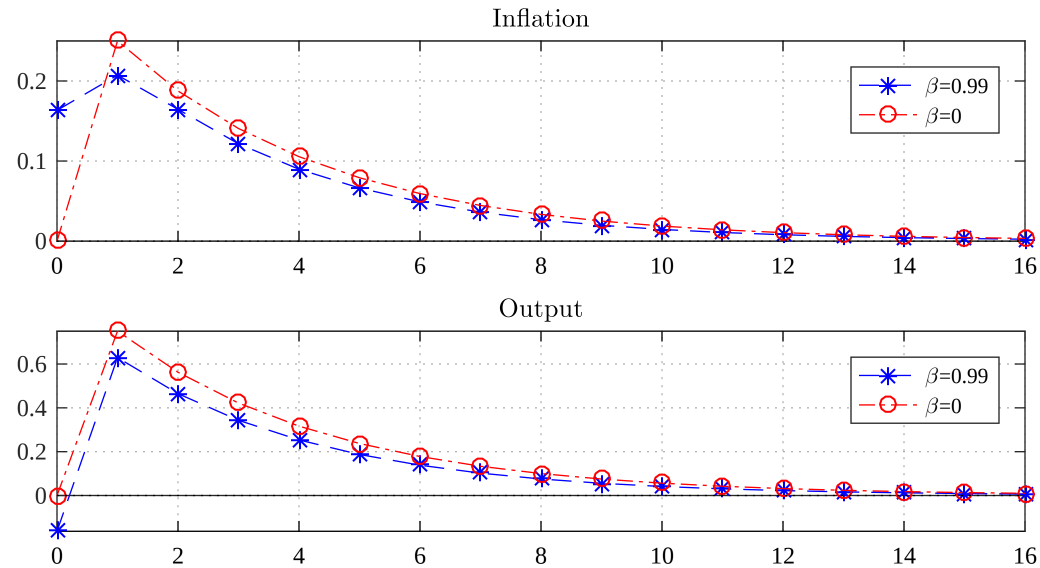


Figure 6: IRFs for example 6: the figure shows the impulse responses of output and inflation to a 1% announced shock to the aggregate demand that will take effect in period one. When $\beta = 0$, firms choose to ignore the news about future policy, and the news has no effects at the time of announcement. 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.

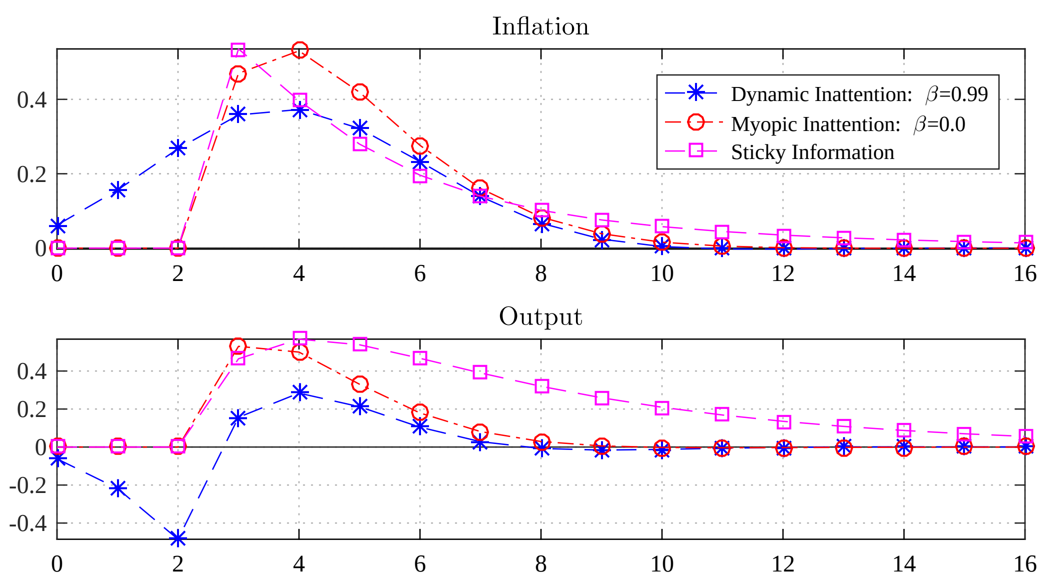


Figure 7: The figure shows impulse responses of three different models to a 1% news shock about aggregate demand that will take effect after three periods. Firms do not respond to this news shock neither in the sticky information model nor in the reduced-form noisy information model (myopic inattention), where 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.

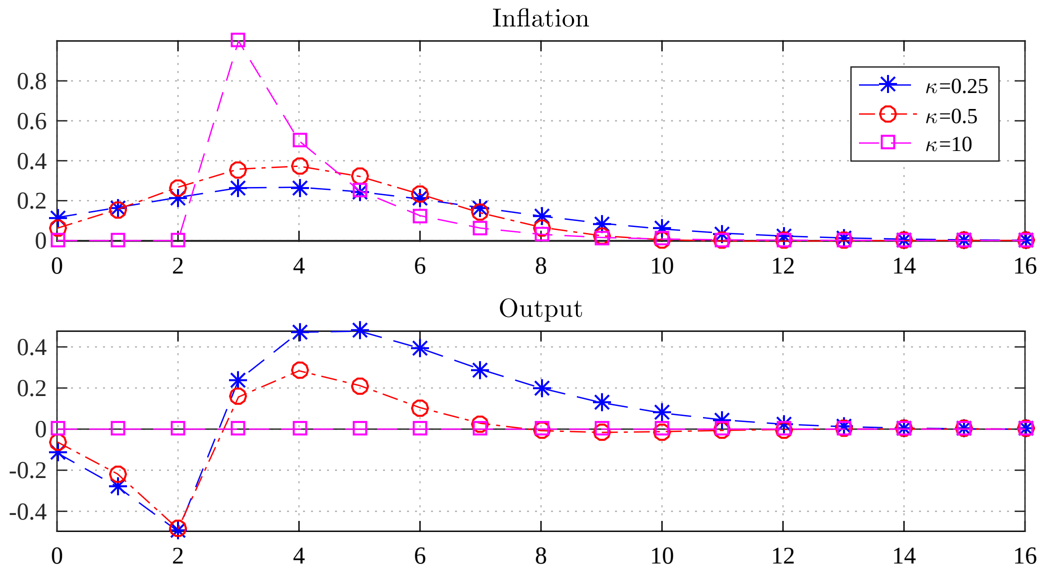


Figure 8: The figure shows impulse responses of three different models to a 1% news shock about aggregate demand that will take effect after three periods. Firms with a larger capacity are more confident that when the shock realizes, they will be able to recognize their fundamental and therefore choose to ignore the news shock. Thus, inflation does not respond when the shocks are announced.

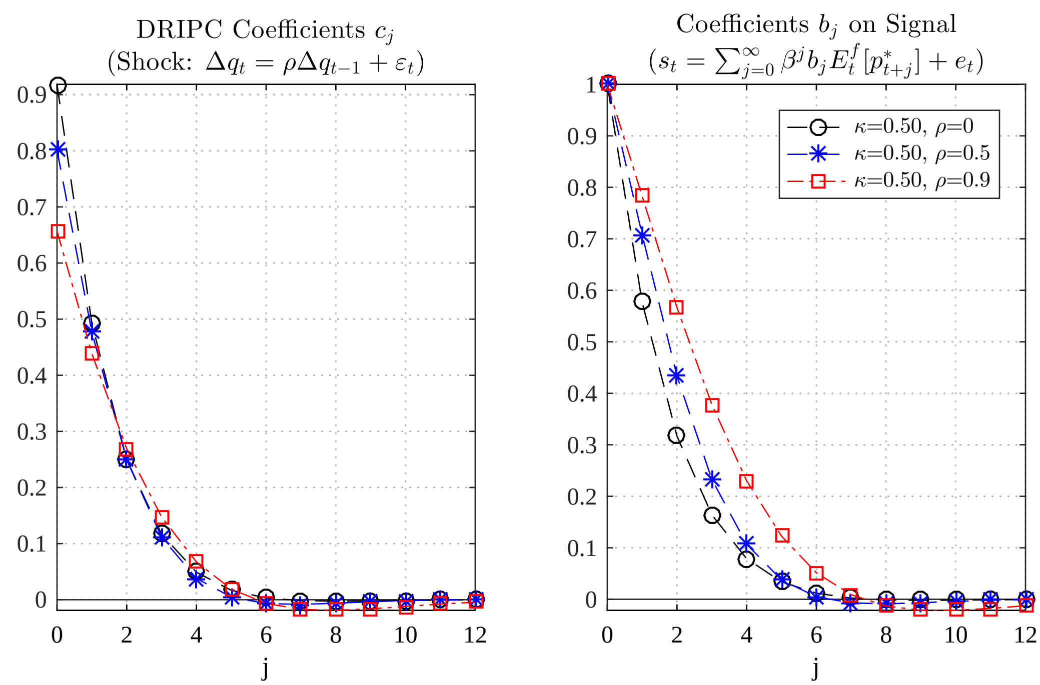


Figure 9: The figure shows the coefficients on the forward-looking terms in the dynamic rational inattention Phillips curve(left) and the coefficients on the forward-looking terms in the firms' optimal signal. The shock follows ARIMA(1) process with different persistence. We find that the coefficients quickly vanish to zero.

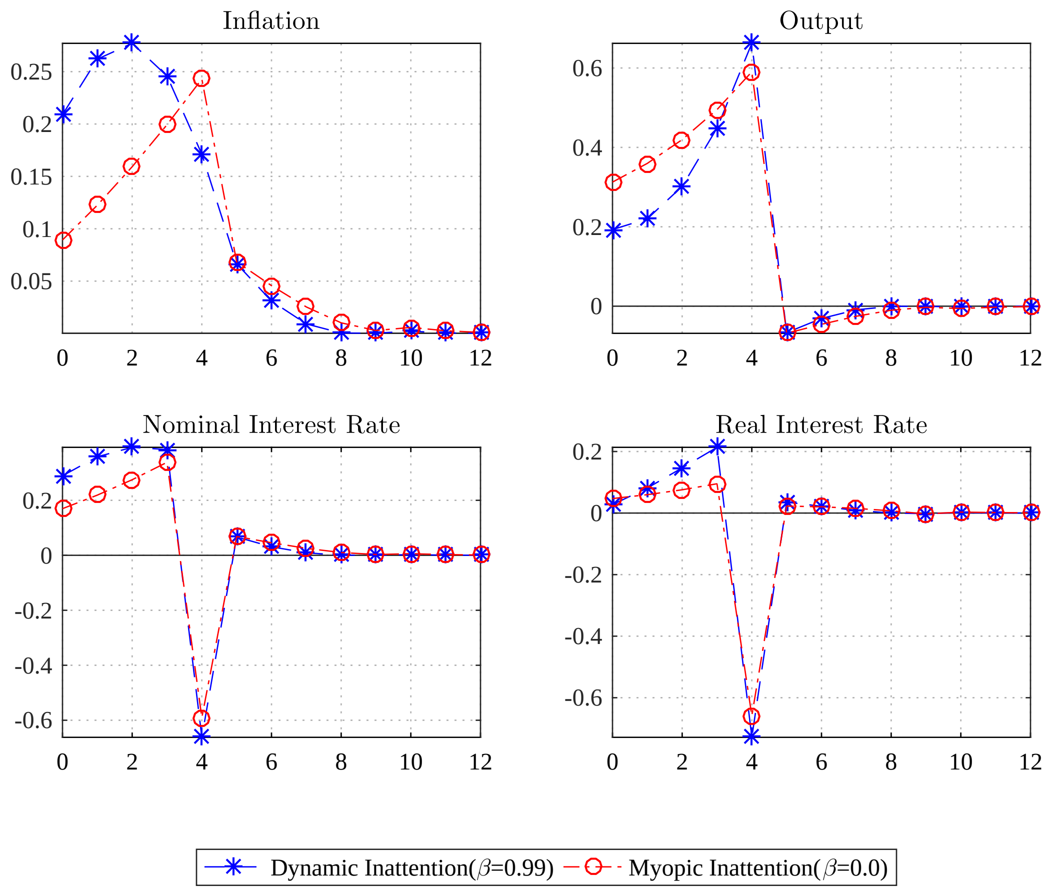


Figure 10: The figure shows the impulse responses of variables to a 4-period ahead forward guidance shock. We compare the IRFs of dynamic inattention model with those of myopic inattention model.

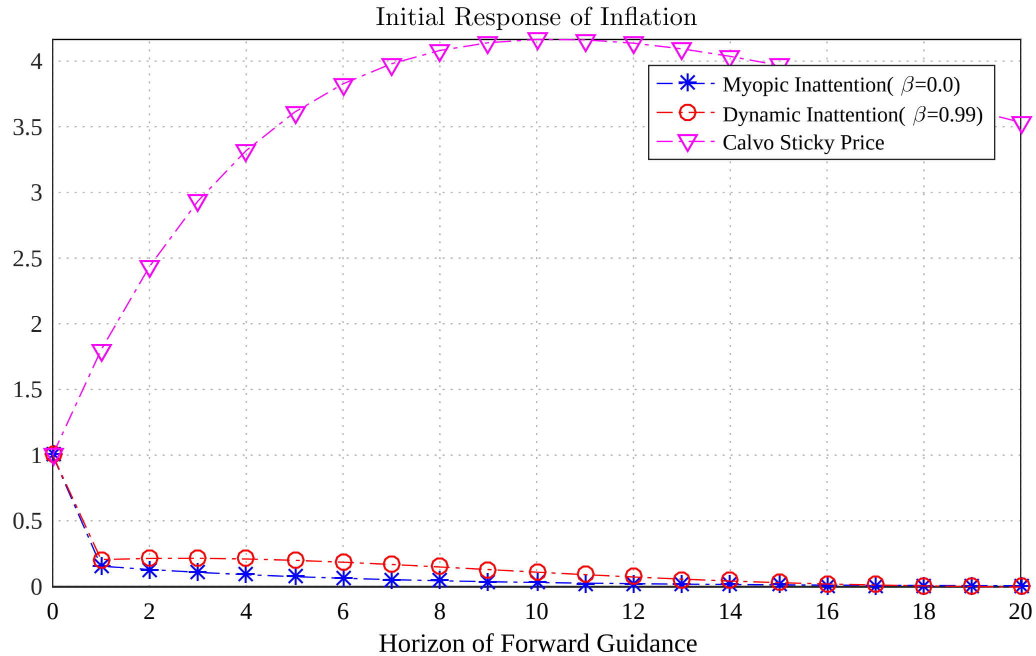


Figure 11: The figure shows the initial responses of inflation to the forward guidance shocks of different horizons. Unlike the Calvo sticky price model, the initial response of inflation in the rational inattention models decreases with the horizons of forward guidance.

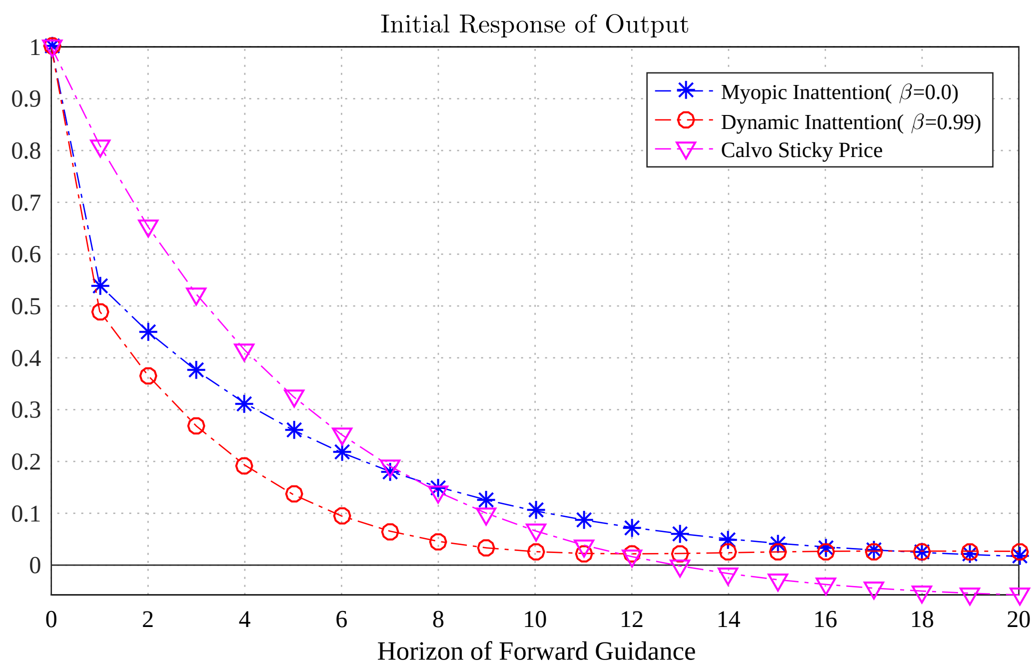


Figure 12: The figure shows the initial responses of output to the forward guidance shocks of different horizons. Compared to the Calvo sticky price mode, the initial response of output is small in the rational inattention models.

A Proofs

Proof of Theorem 1.

Proof. Recall that the agent's problem is

$$\begin{aligned}
\mathcal{L}_0(\Sigma_{0|-1}) &= \min_{\{\mathbf{y}_t\in\hat{\mathcal{S}}_t^F,\kappa_t\}_{t=0}^\infty} \sum_{t=0}^{\infty}\beta^t\left(\mathbf{w}'\Sigma_{t|t}\mathbf{w}+\lambda\kappa_t\right)\\
s.t. \quad & \Sigma_{t|t}=\Sigma_{t|t-1}-\Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1}\\
& \Sigma_{t+1|t}=\mathbf{M}\Sigma_{t|t}\mathbf{M}'+\mathbf{e}_1\mathbf{e}_1'\\
& \frac{1}{2}\log_2\frac{1}{1-\mathbf{y}_t'\Sigma_{t|t-1}\mathbf{y}_t}\le\kappa_t\\
& \Sigma_{0|-1}\text{ given.}
\end{aligned}$$

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:

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

Notice that since the objective function is convex, the information flow constraint is binding. 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 $\mathbf{y}_t$.[32]

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.)

$$\begin{aligned}
L &= \sum_{t=0}^{\infty}\beta^t\left(-\mathbf{w}'\Sigma_t\mathbf{w}+\mathbf{w}'\mathbf{z}_t\mathbf{z}_t'\mathbf{w}+\lambda\frac{1}{2}\log_2\left(1-\mathbf{z}_t'\Sigma_t^{-1}\mathbf{z}_t\right)\right)\\
&+ \sum_{t=0}^{\infty}\beta^t\left(\sum_{j=1}^{T}\eta_{j,t}'[\Sigma_{t+1}-\mathbf{M}(\Sigma_t-\mathbf{z}_t\mathbf{z}_t')\mathbf{M}'-\mathbf{e}_1\mathbf{e}_1']\mathbf{e}_j\right)
\end{aligned}$$

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. We start with the first order condition with respect

[32] 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.

to Σ_{t+1}:[33]

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

$$
\begin{aligned}
L &= \sum_{t=0}^{\infty}\beta^{t}\left(-\mathbf{w}'\Sigma_t\mathbf{w} + \mathbf{w}'\mathbf{z}_t\mathbf{z}_t'\mathbf{w} + \lambda\frac{1}{2}\log_2\left(1-\mathbf{z}_t'\Sigma_t^{-1}\mathbf{z}_t\right)\right)\right) \\
&+ \sum_{t=0}^{\infty}\beta^{t}\left(\sum_{j=1}^{T}\eta_{j,t}'[\Sigma_{t+1} - \mathbf{M}(\Sigma_t - \mathbf{z}_t\mathbf{z}_t')\mathbf{M}' - \mathbf{e}_1\mathbf{e}_1']\mathbf{e}_j\right)
\end{aligned}
$$

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

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

where $\mathbf{X}_t \equiv \frac{1}{2}\sum_{j=1}^{T}\mathbf{M}'\left[\mathbf{e}_j\eta_{j,t}' + \eta_{j,t}\mathbf{e}_j'\right]\mathbf{M}$ and $\phi_{t+1} = \frac{1}{2\ln 2}\frac{\lambda}{\left(1-\mathbf{y}_{t+1}'\Sigma_{t+1}\mathbf{y}_{t+1}\right)}$.

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

$$
\begin{aligned}
& (\mathbf{w}'\mathbf{z}_t)\,\mathbf{w} - \lambda\tfrac{1}{2\ln 2}\tfrac{1}{(1-\mathbf{y}_t'\Sigma_t\mathbf{y}_t)}\Sigma_t^{-1}\mathbf{z}_t + \mathbf{X}_t\mathbf{z}_t = 0 \\
\Rightarrow \quad & (\mathbf{w}'\Sigma_t\mathbf{y}_t)\mathbf{w} - \phi_t\mathbf{y}_t + \mathbf{X}_t\Sigma_t\mathbf{y}_t = 0
\end{aligned}
$$

Hence the FOCs reduce to

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

Let $\widehat{\mathbf{X}}_t = \mathbf{w}\mathbf{w}' + \mathbf{X}_t$ be a symmetric matrix. Then the FO(N)C for $\mathbf{y}_t$ is

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

Note that any ϕ_t satisfying the above equation is an eigenvalue of $\widehat{\mathbf{X}}_t\Sigma_t$ and $\mathbf{y}_t$ is the corresponding eigenvector.

[33] For a guide to taking the derivative of symmetric matrices, see for example Petersen and Pedersen (2012).

We want to find the (second-order) sufficient and necessary conditions for the optimal solutions of the problem. First, we consider the necessary condition. Let Φ_t be the set of all eigenvalues of $\widehat{\mathbf{X}}_t\Sigma_t$. Let $\mathbf{y}_t^*$ be a maximizer for the problem and ϕ_t^* be the corresponding eigenvalue which satisfies the FOCs together with $\mathbf{y}_t^*$.[34] Then, ϕ_t^* should satisfy the following second-order necessary condition:

$$\widehat{\mathbf{X}}_t\Sigma_t - \phi_t^*\mathbf{I} \preceq 0$$

Thus, for every non-zero column vector $\mathbf{s}$ of T real numbers, ϕ_t^* should satisfy

$$\begin{aligned}\mathbf{s}'\left(\widehat{\mathbf{X}}_t\Sigma_t - \phi_t^*\mathbf{I}\right)\mathbf{s} &= \mathbf{s}'\left(\mathbf{V}D\mathbf{V}' - \phi_t^*\mathbf{I}\right)\mathbf{s}\\ &= \tilde{\mathbf{s}}'\left(D - \phi_t^*\mathbf{I}\right)\tilde{\mathbf{s}}\\ &\leq 0\end{aligned}$$

where D is a diagonal matrix formed from the eigenvalues of $\widehat{\mathbf{X}}_t\Sigma_t$[35], the columns of $\mathbf{V}$ are the corresponding eigenvectors with $\mathbf{V}\mathbf{V}' = \mathbf{V}'\mathbf{V} = \mathbf{I}$ and $\tilde{\mathbf{s}} = \mathbf{V}'\mathbf{s}$. Since ϕ_t^* is an eigenvalue of $\widehat{\mathbf{X}}_t\Sigma_t$, the last inequality holds when $\phi_t^* = \max_{\{\phi_t^j \in \Phi_t\}} \phi_t^j$. Thus, among the stationary points that satisfy the FOCs, the optimal signal $\mathbf{y}_t^*$ is the eigenvector which corresponds the largest eigenvalue ϕ_t^* of a symmetric matrix $\widehat{\mathbf{X}}_t\Sigma_t$.

Now, we show that this condition is sufficient for the optimal solution. It is enough to show that and the largest eigenvalue ϕ_t^* of $\widehat{\mathbf{X}}_t\Sigma_t$ and the corresponding eigenvector $\mathbf{y}_t^*$ satisfy the second-order sufficient condition. Let $\mathcal{S} = \left\{\mathbf{s} \neq \mathbf{0} \mid \mathbf{s}'\nabla_{(\mathbf{y}_t = \mathbf{y}_t^*)}\left(1 - 2^{-2\kappa_t} - \mathbf{y}_t'\Sigma_t\mathbf{y}_t\right) = \mathbf{s}'\Sigma_t\mathbf{y}_t^* = 0\right\}$. Then for any $\mathbf{s} = \mathcal{S}$, we have

$$\begin{aligned}\mathbf{s}'\left(\nabla^2 L^*\left(\left\{\mathbf{y}_t^*, \phi_t^*, \{\eta_{j,t}\}_{j=1}^{\infty}\right\}_{t=0}^{\infty}\right)\right)\mathbf{s} &= \mathbf{s}'\left(\widehat{\mathbf{X}}_t\Sigma_t - \phi_t^*\mathbf{I}\right)\mathbf{s}\\ &\leq 0\end{aligned}$$

where the equality holds when $\mathbf{s} = \mathbf{y}_t^*$ due to the first-order condition. Note that $\mathbf{y}_t^* \notin \mathcal{S}$ as ${\mathbf{y}_t^*}'\Sigma_t\mathbf{y}_t^* \neq 0$. Thus, $\mathbf{y}_t^*$ and ϕ_t^* satisfy the second-order sufficient condition.

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

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

[34] Since our objective function is continuous and the constraint is a compact set, the problem attains a maximum by Weierstrass theorem.

[35] The matrix $\widehat{\mathbf{X}}_t\Sigma_t$ is diagonalizable.

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

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

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

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

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

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

Finally, assuming that $\phi_t > 0$, 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.

Proof. 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 $\exists \left\{ \left(\alpha_i^j\right)_{i=0}^{p-1}, \left(\gamma_i^j\right)_{i=0}^{q-1} \right\}_{j=0}^{\infty}$ 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

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

Let $c_{i,t}=\sum_{j=0}^{\infty}\beta^j b_{t,j}\alpha_i^j, \forall i\in\{0,1,\ldots,p-1\}, \forall t\geq 0$ and $d_{i,t}=\sum_{j=0}^{\infty}\beta^j b_{t,j}\gamma_i^j$, $\forall i\in\{0,1,\ldots,q-1\}$, $\forall t\geq 0$. 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. Letκ_t be the information flow that the agent chooses. Then we can write the optimal information problem:

$$\begin{aligned}\min_{\{\mathbf{y}_t\in\hat{\mathcal{S}}_t^F\}_{t=0}^{\infty}} \quad & \sum_{t=0}^{\infty}\beta^t\left(\mathbf{w}'\left(\Sigma_{t|t-1}-\Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1}\right)\mathbf{w}+\lambda\kappa_t\right)\\ s.t. \quad & \Sigma_{t+1|t}=\mathbf{M}\left(\Sigma_{t|t-1}-\Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1}\right)\mathbf{M}'+\mathbf{e}_1\mathbf{e}_1'\\ & \frac{1}{2}\log_2\left(\frac{1}{1-\mathbf{y}_t'\Sigma_{t|t-1}\mathbf{y}_t}\right)\leq\kappa_t\end{aligned}$$

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 $\mathbf{y}_t=\alpha_t\mathbf{w}$. Notice that at the optimum, the information flow constraint binds. Let's define $\mathbf{z}_t=\left(\Sigma_{t|t-1}\right)^{\frac{1}{2}}\mathbf{w}$. Then the agent's problem boils down to choose the optimal information flow, κ_t, and $\mathbf{z}_{t+1}$:

$$\begin{aligned}\max_{\kappa_t,z_{t+1}} \quad & \sum_{t=0}^{\infty}\beta^t\left(-\mathbf{z}_t'\mathbf{z}_t 2^{-2\kappa_t}-\lambda\kappa_t\right)\\ s.t. \quad & \mathbf{z}_{t+1}'\mathbf{z}_{t+1}=\rho^2 2^{-2\kappa_t}\mathbf{z}_t'\mathbf{z}_t+1\end{aligned}$$

The Lagrangian is:

$$\begin{aligned}\mathcal{L}\left(\mathbf{z}_t\right) &= \sum_{t=0}^{\infty}\beta^t\left(-\mathbf{z}_t'\mathbf{z}_t 2^{-2\kappa_t}-\lambda\kappa_t+\psi_t\left(\mathbf{z}_{t+1}'\mathbf{z}_{t+1}-\rho^2 2^{-2\kappa_t}\mathbf{z}_t'\mathbf{z}_t-1\right)\right)\\ &= \sum_{t=0}^{\infty}\beta^t\left(-\mathbf{z}_t'\mathbf{z}_t 2^{-2\kappa_t}\left(1+\rho^2\psi_t\right)-\lambda\kappa_t+\psi_t\left(\mathbf{z}_{t+1}'\mathbf{z}_{t+1}-1\right)\right)\end{aligned}$$

The first-order conditions give

$$\begin{aligned}
\frac{\lambda}{2\ln 2} &= 2^{-2\kappa_t}\mathbf{z}_t'\mathbf{z}_t\left(1+\rho^2\psi_t\right)\\
\psi_t &= \beta 2^{-2\kappa_{t+1}}\left(1+\rho^2\psi_{t+1}\right)\\
&= \beta\sum_{j=1}^{\infty}\left(\beta\rho^2\right)^{j-1}2^{-2\left(\sum_{i=1}^{j}\kappa_{t+i}\right)}+\beta\lim_{k\to\infty}\left(\beta\rho^2\right)^{k}2^{-2\left(\sum_{i=1}^{k}\kappa_{t+i}\right)}\psi_{t+k}\\
&= \beta\sum_{j=1}^{\infty}\left(\beta\rho^2\right)^{j-1}2^{-2\left(\sum_{i=1}^{j}\kappa_{t+i}\right)}
\end{aligned}$$

By combining two equations, we have

$$\left(\sum_{j=0}^{\infty}\left(\beta\rho^2\right)^{j}2^{-2\left(\sum_{i=0}^{j}\kappa_{t+i}\right)}\right)=\frac{\lambda}{2\ln 2}\left(\frac{1}{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}\right).$$

Consider the steady-state prior of the problem. Then the optimal information flows are constant and this gives:

$$\begin{aligned}
\frac{2^{-2\kappa}}{1-\beta\rho^2 2^{-2\kappa}} &= \frac{\lambda}{2\ln 2}\left(\frac{1}{\mathbf{w}'\Sigma\mathbf{w}}\right).\\
\kappa &= \frac{1}{2}\log_2\left(\left(\frac{2\ln 2}{\lambda}\right)\mathbf{w}'\Sigma\mathbf{w}+\beta\rho^2\right)
\end{aligned}$$

Notice that now the optimal choice of information flows depends on the discount factor and the persistence of the AR(1) shock process. Since $1-\alpha_t^2\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}=2^{-2\kappa}$,

$$\begin{aligned}
\alpha^2 &= \frac{1-2^{-2\kappa}}{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}\\
&= \frac{1}{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}\left(1-\frac{\frac{\lambda}{2\ln 2}\left(\frac{1}{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}\right)}{1+\beta\rho^2\frac{\lambda}{2\ln 2}\left(\frac{1}{\mathbf{w}'\Sigma_{t|t-1}\mathbf{w}}\right)}\right)
\end{aligned}$$

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

Moreover, by $\Sigma_{t+1|t}=\mathbf{M}\left(\Sigma_{t|t-1}-\Sigma_{t|t-1}\mathbf{y}_t\mathbf{y}_t'\Sigma_{t|t-1}\right)\mathbf{M}'+\mathbf{e}_1\mathbf{e}_1'$, and by the fact that $x_t=$

$\rho x_{t-1} + u_t$ implies $\mathbf{w} = \left(1, \rho, \rho^2, \dots\right)' \Rightarrow \mathbf{M}'\mathbf{w} = \rho\mathbf{w}$, $\forall t \geq 1$, we have

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

the last inequality holds with the steady-state prior at the optimum. Finally, to get the law of motion for the optimal action, by Kalman filter

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

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

Q.E.D. □

Proof of Lemma 3.

Proof. Recall that the first-order conditions are

$$\phi_t\mathbf{dy}_t = \left(\mathbf{dwdw}' + \mathbf{X}_t\right)\Sigma_{t|t-1}\mathbf{dy}_t$$
$$\mathbf{X}_t = \beta\left(\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1'\right)\left(\mathbf{dwdw}' + \mathbf{X}_{t+1} - \phi_{t+1}\mathbf{dy}_{t+1}\mathbf{dy}_{t+1}'\right)\left(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1'\right).$$

Let $\mathbf{M}_p = \mathbf{M} + \mathbf{e}_1\mathbf{e}_1'$. The second equation of the FOCs gives us

$$\mathbf{X}_t = \textstyle\sum_{j=1}^{\infty}\beta^j\left(\mathbf{M}_p'\right)^j\left(\mathbf{dwdw}' - \phi_{t+j}\mathbf{dy}_{t+j}\mathbf{dy}_{t+j}'\right)\mathbf{M}_p^j$$
$$\mathbf{X}_t\Sigma_t\mathbf{dy}_t = \textstyle\sum_{j=1}^{\infty}\beta^j\left(\mathbf{M}_p'\right)^j\left(\mathbf{dwdw}' - \phi_{t+j}\mathbf{dy}_{t+j}\mathbf{dy}_{t+j}'\right)\mathbf{M}_p^j\Sigma_t\mathbf{dy}_t$$
$$\Rightarrow \quad \mathbf{X}_t\Sigma_t\mathbf{dy}_t = \textstyle\sum_{j=1}^{\infty}\beta^j\left(\mathbf{dw}'\mathbf{M}_p^j\Sigma_t\mathbf{dy}_t\right)\left(\mathbf{M}_p'\right)^j\mathbf{dw} - \sum_{j=1}^{\infty}\beta^j\left(\mathbf{dy}_{t+j}'\mathbf{M}_p^j\Sigma_t\mathbf{dy}_t\right)\left(\mathbf{M}_p'\right)^j\left(\phi_{t+j}\mathbf{dy}_{t+j}\right)$$

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

$$\phi_t \mathbf{dy}_t = \sum_{j=0}^{\infty} \beta^j \left(\mathbf{dw}'\mathbf{M}_p^j \Sigma_t \mathbf{dy}_t\right) \left(\mathbf{M}_p^{'}\right)^j \mathbf{dw} - \sum_{j=1}^{\infty} \beta^j \left(\mathbf{dy}_{t+j}' \mathbf{M}_p^j \Sigma_t \mathbf{dy}_t\right) \left(\mathbf{M}_p^{'}\right)^j \left(\phi_{t+j}\mathbf{dy}_{t+j}\right)$$

Now guess that $\phi_{t+j}\mathbf{dy}_{t+j} = \sum_{k=0}^{\infty} \beta^k a_{t+j,k} \left(\mathbf{M}_p^{'}\right)^k \mathbf{dw}$. Plugging in this guess in the above equation

$$\begin{aligned}
\left(1 - \mathbf{dy}_t'\Sigma_t \mathbf{dy}_t\right)\phi_t \mathbf{dy}_t &= \sum_{j=0}^{\infty} \beta^j \left(\mathbf{dw}'\mathbf{M}_p^j \Sigma_t \mathbf{dy}_t\right) \left(\mathbf{M}_p^{'}\right)^j \mathbf{dw} - \sum_{j=0}^{\infty} \beta^j \left(\mathbf{dy}_{t+j}' \mathbf{M}_p^j \Sigma_t \mathbf{dy}_t\right) \left(\mathbf{M}_p^{'}\right)^j \left(\phi_{t+j}\mathbf{dy}_{t+j}\right) \\
&= \sum_{j=0}^{\infty} \beta^j \left(\mathbf{dw}'\mathbf{M}_p^j \Sigma_t \mathbf{dy}_t\right) \left(\mathbf{M}_p^{'}\right)^j \mathbf{dw} - \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \beta^{j+k} a_{t+j,k} \left(\mathbf{dy}_{t+j}' \mathbf{M}_p^j \Sigma_t \mathbf{dy}_t\right) \left(\mathbf{M}_p^{'}\right)^{j+k} \mathbf{dw} \\
&= \sum_{j=0}^{\infty} \beta^j \left[\left(\mathbf{dw}'\mathbf{M}_p^j \Sigma_t \mathbf{dy}_t\right) - \sum_{k=0}^{j} a_{t+k,j-k} \left(\mathbf{dy}_{t+k}' \mathbf{M}_p^k \Sigma_t \mathbf{dy}_t\right)\right] \left(\mathbf{M}_p^{'}\right)^j \mathbf{dw}
\end{aligned}$$

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

$$a_{t,j} = 2^{2\kappa_t} \left[\mathbf{dw}'\mathbf{M}_p^j \Sigma_t \mathbf{dy}_t - \sum_{k=0}^{j} a_{t+k,j-k} \left(\mathbf{dy}_{t+k}' \mathbf{M}_p^k \Sigma_t \mathbf{dy}_t\right)\right].$$

Finally, assuming that $\phi_t > 0$, let $b_{t,j} \equiv \phi_t^{-1} a_{t,j}$, so that $\mathbf{dy}_t = \sum_{j=0}^{\infty} \beta^j b_{t,j} \left(\mathbf{M}_p^{'}\right)^j \mathbf{dw}$. Now, the optimal signal is

$$\begin{aligned}
s_t^* &= \mathbf{dy}_t' \tilde{\mathbf{u}}_t + e_t \\
&= \sum_{j=0}^{\infty} \beta^j b_{t,j} \mathbf{dw}' \left(\mathbf{M} + \mathbf{e}_1 \mathbf{e}_1'\right)^j \tilde{\mathbf{u}}_t + e_t.
\end{aligned}$$

but notice that $\left(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1'\right)^j \tilde{\mathbf{u}}_t = \mathbb{E}\left\{\tilde{\mathbf{u}}_{t+j}|\tilde{\mathbf{u}}_t\right\}$, and $\mathbf{dw}'\left(\mathbf{M} + \mathbf{e}_1\mathbf{e}_1'\right)^j \tilde{\mathbf{u}}_t = \mathbb{E}\left\{\mathbf{dw}'\tilde{\mathbf{u}}_{t+j}|\tilde{\mathbf{u}}_t\right\} = \mathbb{E}\left\{x_{t+j}|\tilde{\mathbf{u}}_t\right\} = \mathbb{E}_t^f\left\{x_{t+j}\right\}$. Hence,

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

Q.E.D. □

Proof of Corollary 2.

To be included. Basically the same as Corollary 1.

Proof of Lemma 4.

Proof. 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 $\mathbf{dy} = \sum_{j=0}^{\infty} \beta^j b_j \left(\mathbf{M}' + \mathbf{e}_1\mathbf{e}_1'\right)^j \mathbf{dw}_{p^*}$ be the representation of the optimal signal derived in that section. Notice that by Kalman filter

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

Moreover,

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

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

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

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

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

Therefore,

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

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

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

Now, we re-write the forward-looking term as follows:

$$
\begin{aligned}
\sum_{j=0}^{\infty}\beta^j b_j(\mathbb{E}_t^f\{p_{t+j}^*\}-\tilde{\mathbb{E}}_t\{p_{t+j}^*\}) &= b_0\alpha y_t+\sum_{j=1}^{\infty}\beta^j b_j(\mathbb{E}_t^f\{p_{t+j}^*-p_{t+j-1}^*\}-\tilde{\mathbb{E}}_t\{p_{t+j}^*-p_{t+j-1}^*\})\\
&+ \sum_{j=1}^{\infty}\beta^j b_j(\mathbb{E}_t^f\{p_{t+j-1}^*\}-\tilde{\mathbb{E}}_t\{p_{t+j-1}^*\})\\
&= b_0\alpha y_t+\sum_{j=1}^{\infty}\beta^j b_j(\mathbb{E}_t^f\{\pi_{t+j}+\alpha\Delta y_{t+j}\}-\tilde{\mathbb{E}}_t\{\pi_{t+j}+\alpha\Delta y_{t+j}\})\\
&+ \beta\sum_{j=0}^{\infty}\beta^j b_{j+1}(\mathbb{E}_t^f\{p_{t+j}^*\}-\tilde{\mathbb{E}}_t\{p_{t+j}^*\})
\end{aligned}
$$

Thus, by iteration,

$$
\begin{aligned}
\sum_{j=0}^{\infty}\beta^j b_j(\mathbb{E}_t^f\{p_{t+j}^*\}-\tilde{\mathbb{E}}_t\{p_{t+j}^*\}) &= (\sum_{j=0}^{\infty}\beta^j b_j)\alpha y_t\\
&+ \sum_{j=1}^{\infty}(\sum_{k=j}^{\infty}\beta^k b_k)(\mathbb{E}_t^f\{\pi_{t+j}+\alpha\Delta y_{t+j}\}-\tilde{\mathbb{E}}_t\{\pi_{t+j}+\alpha\Delta y_{t+j}\})\\
&+ \lim_{T\to\infty}\beta^T\sum_{j=0}^{\infty}\beta^j b_{j+T}(\mathbb{E}_t^f\{p_{t+j}^*\}-\tilde{\mathbb{E}}_t\{p_{t+j}^*\})
\end{aligned}
$$

Since $\beta<1$, b_{j+T} depends on the stationary part of the process of the fundamental, and the process for the difference of the expectations is stationary (they both are composed of a unit root and a stationary part, so their difference is the difference of two stationary processes, which is a stationary process[36]), the limit term is zero.

Let $c_j\equiv 2^{2\kappa}\delta_0(\sum_{k=j}^{\infty}\beta^k b_k)$, $\forall j\geq 0$. The Phillips curve is then:

$$
\pi_t=\tilde{\mathbb{E}}_{t-1}\left[\pi_t+\alpha\Delta y_t\right]+c_0\alpha y_t+\sum_{j=1}^{\infty}c_j\left(\mathbb{E}_t^f\left[\pi_{t+j}+\alpha\Delta y_{t+j}\right]-\tilde{\mathbb{E}}_t\left[\pi_{t+j}+\alpha\Delta y_{t+j}\right]\right).
$$

Q.E.D. □

Proof of example 4.

Proof. 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

[36]Proof: By Wold's Theorem any stationary process is the inner product of a summable sequence and its i.i.d innovations over time. Let ℓ^2 denote the space of summable sequences. We need to show that ℓ^2 is closed under addition, to show that the sum of two stationary processes is also stationary. Suppose $\{u,v\}\subset\ell^2$, then $||u+v||_2^2=\langle u+v,u+v\rangle\leq||u||_2^2+||v||_2^2+2|\langle u,v\rangle|$. By Cauchy-Shwarz inequality $|\langle u,v\rangle|\leq||u||.||v||$. Thus, $||u+v||_2^2\leq(||u||_2+||v||_2)^2<\infty$. Therefore, $u+v\in\ell^2$.

condition for the attention problem, we have

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

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

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

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

$$\begin{aligned} \phi \mathbf{dy} &= \left(\mathbf{e}_1' \Sigma \mathbf{dy}\right) \mathbf{e}_1 + \frac{\beta}{1-\beta}\left(1 - \phi\theta^2\right) \left(\mathbf{e}_1' \Sigma \mathbf{dy}\right) \mathbf{e}_1 \\ &= \theta \mathbf{e}_1 \left(\mathbf{e}_1' \Sigma \mathbf{e}_1\right) \left(\frac{1 - \beta\phi\theta^2}{1-\beta}\right) \\ \phi &= \left(\left(\mathbf{e}_1' \Sigma \mathbf{e}_1\right) \left(\frac{1}{1-\beta} - \frac{\beta\phi\theta^2}{1-\beta}\right)\right) \\ &= \frac{\mathbf{e}_1' \Sigma \mathbf{e}_1}{1 - \beta + \beta\theta^2 \left(\mathbf{e}_1' \Sigma \mathbf{e}_1\right)} \end{aligned}$$

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:

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

where the $\kappa = \frac{\phi}{\lambda} 2 \ln 2$. 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

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

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,

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

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

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

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

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

Thus,

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

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

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

which implies

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

Also,

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

Q.E.D. $\square$

Proof of example 5.

Proof. 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:

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

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:

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

where the last line is derived from $p_t^* = p_t + \alpha y_t$.

Q.E.D. $\square$

Proof of example 6.

Proof. 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 translates to

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

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

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

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

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

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

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

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

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

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 $\hat{\Sigma}$:

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

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

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

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

$$\begin{aligned}
\mathbf{e}_2'\hat{\Sigma}\mathbf{dy} &= \delta\left[(1-\gamma)\,\mathbf{e}_2'\hat{\Sigma}\mathbf{e}_2 + \gamma\mathbf{e}_2'\hat{\Sigma}\mathbf{e}_1\right] \\
&= \delta\mathbf{e}_1'\left(\hat{\Sigma}-\hat{\Sigma}\mathbf{dy}\mathbf{dy}'\hat{\Sigma}\right)\mathbf{e}_1 \\
&= 1-\gamma\delta
\end{aligned}$$

Thus,

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

where the second line was our guess. This implies

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

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

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

Multiply this once by $\mathbf{e}_1'$ and once by $\mathbf{e}_2'$ from left to get[37]

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

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

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

Thus,

$$\begin{aligned}\tilde{\mathbb{E}}_t\{\Delta q_{t+1}\} &= (1-\delta)\,\tilde{\mathbb{E}}_{t-1}\{\Delta q_t\}+\delta(y_t+\gamma u_t)-(1-\delta)\,\pi_t \\ \pi_t &= \tilde{\mathbb{E}}_{t-1}\{\Delta q_t\}+\tfrac{(1-\gamma\delta)\delta}{1-\delta(1-\gamma\delta)}(y_t+\gamma u_t)\end{aligned}\tag{18}$$

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

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

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 18 we get the following Phillip's curve:

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

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

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:

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

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

Q.E.D. $\square$