

Dynamic Rational Inattention and the Phillips Curve^{*†}

Hassan Afrouzi[‡] Choongryul Yang[§]
Columbia U. UT Austin

First Draft: April, 2017

This Draft: November, 2019

Abstract

We develop a general equilibrium flexible price model with dynamic rational inattention in which the slope of the Phillips curve is endogenous to systematic aspects of monetary policy. This Phillips curve is flatter when the monetary policy is more hawkish: rationally inattentive firms find it optimal to ignore monetary policy shocks when the monetary authority commits to stabilize nominal variables. Moreover, an unexpectedly more dovish monetary policy leads to a *completely* flat Phillips curve in the short-run and a steeper Phillips curve in the long-run. We also develop a tractable method for solving general dynamic rational inattention models in linear quadratic Gaussian setups.

JEL Classification: D83, D84, E03, E58

Keywords: Rational Inattention, Dynamic Information Acquisition, Phillips Curve

*We are grateful to Saroj Bhattarai, Olivier Coibion, Mark Dean, Xavier Gabaix, Yuriy Gorodnichenko, Jennifer La'O, John Leahy, Yueran Ma, Filip Matějka, Emi Nakamura, Kris Nimark, Eric Sims, Jón Steinsson, Luminita Stevens, Laura Veldkamp, Venky Venkateswaran, Mirko Wiederhold, Mike Woodford and seminar participants at 2018 AEA Meetings, Columbia, SED Mexico City, UT Austin, Chicago Fed, Cornell and CUNY for helpful comments.

†Previous versions of this manuscript were presented under the title “Dynamic Inattention, the Phillips Curve and Forward Guidance” at the 2018 ASSA Annual Meeting in Philadelphia as well as the 2018 SED Meeting in New Mexico.

‡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

Since [Muth \(1961\)](#), full information rational expectations theory has grown to be an essential part of macroeconomic modeling. Nonetheless, a growing body of evidence on the costly nature of information acquisition, as well as its potential relevance in explaining economic phenomena, has led to an increasing demand for modifications of the theory that take such costs into account.

The rational inattention theory ([Sims, 2003, 2006, 2010](#)) has provided an appealing alternative by introducing a cost to information acquisition, while simultaneously preserving the consistency of expectations within an optimizing framework. However, rational inattention models tend to be enormously complex and applying them to broader contexts has proven to be a challenge. In this paper, we develop a tractable and portable method for solving dynamic multivariate rational inattention models in linear quadratic Gaussian (LQG) setups where the payoffs are quadratic and the cost of information is linear in Shannon’s mutual information function.

Our first contribution is to formulate and to characterize the full transition dynamics of LQG dynamic multivariate rational inattention (LQG DRI) problems. Similar to the solution to any other dynamic problem, we characterize the solution to the LQG DRI problems with (1) a forward-looking Euler equation for the marginal benefit of information with a terminal transversality condition, (2) a backward-looking law of motion for the state, which is the “stock” of uncertainty in these models, and (3) a policy function for the choice variable as a function of the state and the marginal benefit.

We show that the transition dynamics in LQG DRI problems is characterized by inaction regions for the second moment of the agent’s belief (uncertainty) in different dimensions of the state: while the marginal cost of attention is constant for a bit of information, the marginal benefit of attention increases in the prior uncertainty of the agent about a particular shock. Therefore, for small levels of prior uncertainty the marginal benefit falls below the marginal cost and the agent decides to pay no attention. Over time, as the agent’s uncertainty grows in such dimensions, one of two cases arises.

The first is when the agent’s uncertainty stabilizes at a level within her inaction region, in which case the agent never pays attention to such dimensions. For instance, uncertainty about an AR(1) process is bounded above by its unconditional variance. If this variance is within the inaction region imposed by the cost of information, the agent would choose to not pay attention to this process.

The second case is when the uncertainty grows to be larger than the bound of the inaction region. In this case, since the cost is linear, the agent immediately acquires enough

information to stabilize their uncertainty at the inaction bound. In the case of the AR(1) example above, for instance, if the agent's inaction bound is below the unconditional variance of the process, then in the steady state the agent's information acquisition is such that their uncertainty about the process is constant and equal to the inaction bound.

We apply our framework to propose an attention driven theory of the Phillips curve. Our first result is that under optimal information acquisition of firms, the slope of the Phillips curve is endogenous to how monetary policy is conducted. In economies where the monetary authority puts a larger weight on stabilizing the nominal variables – in other words, when monetary policy is more hawkish – firms endogenously choose to pay less attention to monetary policy shocks and the slope of the Phillips curve becomes flatter.

Our second result is that higher uncertainty about monetary policy shocks, stemming from a lower weight on stabilizing nominal variables – or in other words, a more dovish monetary policy – can lead to a completely flat Phillips curve in the short-run – by pushing the firms into the inaction region – while leading to a steeper slope of the Phillips curve in the long-run as firms' uncertainty grows to hit the inaction bound. When the monetary policy becomes more dovish, rationally inattentive firms find themselves in a more volatile environment where they need to acquire information at a higher rate to maintain the same level of uncertainty about monetary policy. Nevertheless, given that information is costly, such optimizing firms do not fully offset this effect. They find it optimal to abstain from information acquisition until their uncertainty rises to a level that is manageable given their new more uncertain environment. Once they start paying attention again, however, they do at a higher rate than before which leads to a steeper Phillips curve.

Our theory provides a new perspective on a growing empirical literature that documents a flattening of the Phillips curve in recent decades,¹ and provides an explanation for this phenomenon by suggesting that the flatter slope is due to the optimal response of firms to the onset of a more hawkish monetary policy in the post-Volcker era of the U.S. monetary policy: when information is costly, it is optimal to pay less attention to monetary policy once it commits more to stabilizing nominal variables.

Our theory also offers a new and a different viewpoint for the conduct of monetary policy relative to the New Keynesian models. While the slope of the Phillips curve in latter models is mainly pinned down by the frequency of price changes and is exogenous to how committed the monetary policy is to stabilizing the nominal variables, our model suggests a direct link between the two. Therefore, policy regimes that might seem optimal under an exogenously flat Phillips curve have completely different outlooks from the

¹Blanchard (2016) and Coibion and Gorodnichenko (2015b).

perspective of our model. For instance, from the perspective of a model where the slope of the Phillips curve is exogenous and flat, a more dovish monetary policy, or *running the economy hotter*, might seem appealing in order to reduce unemployment. Nonetheless, our model provides a different remedy: such policies would work in the short-run by pushing firms temporarily into their inaction region and paralyzing inflation. After this temporary period, however, firms would start paying more attention to monetary policy which would lead to a steeper Phillips curve and a higher sensitivity of inflation to monetary policy shocks.

Related Literature. Dynamic rational inattention models have been applied to different setups for years.² Most of this literature, however, relies on simplifying assumptions – such as independence of signals – and computational methods in characterizing the solution. We provide a tractable solution method by building on a subset of this literature that has laid the ground for solving dynamic rational inattention models in LQG setups (Sims, 2003; Maćkowiak, Matějka and Wiederholt, 2018; Fulton, 2018; Miao, Wu and Young, 2018). This literature makes two simplifying assumptions that we depart from: (1) they abstract away from transition dynamics by assuming that the cost of information is not discounted, and (2) they solve for the long-run steady-state information structure that is independent of time and state.

Our attention driven theory of the Phillips curve is motivated by two separate sets of empirical evidence. First, the literature that estimates³ and subsequently documents a flattening of the slope of the Phillips curve (Coibion and Gorodnichenko, 2015b; Blanchard, 2016; Bullard, 2018; Hooper, Mishkin and Sufi, 2019).⁴ Second, the empirical literature that documents the information rigidities that economic agents exhibit in forming their expectations.⁵

²See, for instance, Maćkowiak and Wiederholt (2009a); Paciello (2012); Melosi (2014); Pasten and Schoenle (2016); Matějka (2015); Afrouzi (2016); Yang (2019) for applications to pricing; Sims (2003); Luo (2008); Tutino (2013) for consumption; Luo et al. (2012) for current account; Zorn (2016) for investment; Woodford (2009); Stevens (2019); Khaw and Zorrilla (2018) for infrequent adjustments in decisions; Maćkowiak and Wiederholt (2015) for business cycles; Paciello and Wiederholt (2014) for optimal policy; Peng and Xiong (2006) for asset pricing; Mondria and Wu (2010) for home bias; and Ilut and Valchev (2017) for imperfect problem solving.

³See, for instance, Roberts (1995); Gali and Gertler (1999); Rudd and Whelan (2005); Coibion (2010) for estimation of Phillips curve.

⁴While we provide an attention based theory for this phenomena, an alternative explanation is nonlinearities in the slope of the Phillips curve. See, for instance, Kumar and Orrenius (2016); Babb and Detmeister (2017); Hooper et al. (2019).

⁵For recent progress in this literature, see for instance, Kumar et al. (2015); Coibion and Gorodnichenko (2015a); Ryngaert (2017); Coibion et al. (2018); Roth and Wohlfart (2018); Gaglianone et al. (2019) for survey evidence, and Khaw et al. (2017); Khaw and Zorrilla (2018); Landier et al. (2019) for experimental evidence.

Finally, we relate to the literature that considers how imperfect information affects the Phillips curve (Lucas, 1972; Mankiw and Reis, 2002; Woodford, 2003; Reis, 2006; Nimark, 2008; Angeletos and La'O, 2009). Our main departure is to derive a Phillips curve in a model with rational inattention and study the interaction of imperfect information and *monetary policy* in shaping the Phillips curve.

The paper is organized as follow. In Section 2 we start by setting up the dynamic rational inattention problem and then characterize the solution for the LQG case. In Section 3 we provide an attention driven theory of the Phillips curve. Section 4 concludes. All proofs are included in the Appendix.

2 Theoretical Framework

In this section we formalize the choice problem of an agent who chooses her information structure endogenously over time. We start by setting up the general problem without making assumptions on payoffs and information structures. We then derive and solve the implied LQG problem. This approach helps us (1) identify the necessary assumptions that are required for the solution and (2) discuss how our setup relates and differs from the cases considered in the previous literature.

2.1 Environment

Preferences. Time is discrete and is indexed by $t \in \{0, 1, 2, \dots\}$. At each time t the agent chooses a vector of actions $\vec{a}_t \in \mathbb{R}^m$ and gains an instantaneous payoff of $v(\vec{a}_t; \vec{x}_t)$ where $\{\vec{x}_t \in \mathbb{R}^n\}_{t=0}^\infty$ is an exogenous stochastic process, and $v(.,.) : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ is strictly concave and bounded above with respect to its first argument.

Set of Available Signals. We assume that at any time t , the agent has access to a set of available signals in the economy, which we call \mathcal{S}^t . Signals in \mathcal{S}^t are informative of $X^t \equiv (\vec{x}_0, \dots, \vec{x}_t)$. In particular, we assume:

1. \mathcal{S}^t is *rich*: for any posterior distribution on X^t , there is a set of signals $S^t \subset \mathcal{S}^t$ that generate that posterior.
2. Available signals do not expire over time: $S^t \subset S^{t+h}, \forall h \geq 0$.
3. Available signals at time t are not informative of future innovations to \vec{x}_t : $\forall S_t \in \mathcal{S}^t, \forall h \geq 1, S_t \perp \vec{x}_{t+h} | X^t$.

Information Sets and Dynamics of Beliefs. Our main assumption here is that the agent does not forget information over time, which is commonly referred to as the “no-forgetting constraint”. The agent understands that any choice of information will affect their priors in the future and that information has a continuation value.⁶ Formally, a sequence of information sets $\{S^t \subseteq S^t\}_{t \geq 0}$ satisfy the *no-forgetting* constraint for the agent if $S^t \subseteq S^{t+\tau}, \forall t \geq 0, \tau \geq 0$.

Cost of Information and the Attention Problem. We assume cost of information is linear in Shannon’s mutual information function.⁷ Formally, let $\{S^t\}_{t \geq 0}$ denote a set of information sets for the agent which satisfies the no-forgetting constraint. Then, the agent’s flow cost of information at time t is $\omega \mathbb{I}(X^t; S^t | S^{t-1})$, where

$$\mathbb{I}(X^t; S^t | S^{t-1}) \equiv h(X^t | S^{t-1}) - \mathbb{E}[h(X^t | S^t) | S^{t-1}]$$

is the reduction in the entropy of X^t that the agent experiences by expanding her knowledge from S^{t-1} to S^t , and ω is the marginal cost of a nat of information.

We can now formalize the inattention problem of the agent in our setup:

$$\begin{aligned} V_0(S^{-1}) &\equiv \sup_{\{S_t \subset S^t, \vec{a}_t: S^t \rightarrow \mathbb{R}^m\}_{t \geq 0}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}[v(\vec{a}_t; \vec{x}_t) - \omega \mathbb{I}(X^t; S^t | S^{t-1}) | S^{-1}] && \text{(RI Problem)} \\ \text{s.t. } S^t &= S^{t-1} \cup S_t, \forall t \geq 0, && \text{(evolution of information set + no-forgetting)} \\ S^{-1} &\text{ given.} && \text{(initial information set)} \end{aligned}$$

2.1.1 Two General Properties of the Solution

Solving the **RI Problem** is complicated by two issues: (1) the agent can choose any subset of signals in any period and (2) the cost of information depends on the whole history of actions and states, which increases the dimensionality of the problem with time. The following two lemmas present results that simplify these complications.

Sufficiency of Actions for Signals. An important consequence of assuming that the cost of information is linear in Shannon’s mutual information function is that it implies

⁶Although we assume perfect memory in our benchmark, these dynamic incentives exist as long as the agent can carry a part of her memory with her over time. For a model with fading memory with exogenous information, see Nagel and Xu (2019). Furthermore, da Silveira et al. (2019) endogenize noisy memory in a setting where carrying information over time is costly.

⁷For a discussion of Shannon’s mutual information function and generalizations see Caplin et al. (2017). See also Hébert and Woodford (2018) for an alternative cost function.

actions are sufficient statistics for signals over time (Steiner et al., 2017; Ravid, 2019). The following lemma formalizes this result in our setting.

Lemma 1. *Suppose $\{(S^t \subset \mathcal{S}^t, \vec{a}_t : S^t \rightarrow \mathbb{R}^m)\}_{t=0}^\infty \cup S^{-1}$ is a solution to the **RI Problem**. $\forall t \geq 0$, define $a^t \equiv \{\vec{a}_\tau\}_{0 \leq \tau \leq t} \cup S^{-1}$. Then, $X^t \rightarrow a^t \rightarrow S^t$ forms a Markov chain.*

Lemma 1 allows us to directly substitute actions for signals. In particular, we can impose that the agent directly chooses $\{\vec{a}_t \in \mathcal{S}^t\}_{t \geq 0}$ without any loss of generality.

Conditional Independence of Actions from Past Shocks. It follows from Lemma 1 that if an optimal information structure exists, then $\forall t \geq 0 : \mathbb{I}(X^t; S^t | S^{t-1}) = \mathbb{I}(X^t; a^t | a^{t-1})$. Here we show this can be simplified if $\{\vec{x}_t\}_{t \geq 0}$ follows a Markov process.

Lemma 2. *Suppose $\{\vec{x}_t\}_{t \geq 0}$ is a Markov process and $\{\vec{a}_t\}_{t \geq 0}$ is a solution to the **RI Problem** given an initial information set S^{-1} . Then $\forall t \geq 0$:*

$$\mathbb{I}(X^t; a^t | a^{t-1}) = \mathbb{I}(\vec{x}_t; \vec{a}_t | a^{t-1}), \quad a^{-1} \equiv S^{-1}. \quad (2.1)$$

When $\{\vec{x}_t\}_{t \geq 0}$ is Markov, at any time t , \vec{x}_t is all the agent needs to know to predict the future states. Therefore, it is suboptimal to acquire information about previous realizations of the state.

2.2 The Linear-Quadratic-Gaussian Problem

In this section, we characterize the necessary and sufficient conditions for the optimal information structure in a Linear-Quadratic-Gaussian (LQG) setting. In particular, we assume that $\{\vec{x}_t \in \mathbb{R}^n : t \geq 0\}$ is a Gaussian process and the payoff function of the agent is quadratic and given by:

$$v(\vec{a}_t; \vec{x}_t) = -\frac{1}{2}(\vec{a}'_t - \vec{x}'_t \mathbf{H})(\vec{a}_t - \mathbf{H}' \vec{x}_t) \quad (2.2)$$

Here, $\mathbf{H} \in \mathbb{R}^{n \times m}$ has full column rank and captures the interaction of the actions with the state.⁸ The assumption of $\text{rank}(\mathbf{H}) = m$ is without loss of generality; in the case that any two column of \mathbf{H} are linearly dependent, we can reclassify the problem so that all colinear actions are in one class.

⁸While we take this as an assumption, this payoff function can also be derived as a second order approximation to a twice differentiable function $v(\cdot; \cdot)$ around the non-stochastic optimal action.

Moreover, we have normalized the Hessian matrix of v with respect to \vec{a} to negative identity.⁹

Optimality of Gaussian Posteriors. We start by proving that optimal actions are Gaussian under quadratic payoff with a Gaussian initial prior. Maćkowiak and Wiederholt (2009b) prove a version of this result in their setup where the cost of information is given by $\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{I}(X^T; a^T)$. Our setup is slightly different as in our case the cost of information is discounted at rate β and is equal to $(1 - \beta) \sum_{t=0}^{\infty} \beta^t \mathbb{I}(X^t; a^t)$, as derived in the proof of Lemma 1 for the derivation.

Lemma 3. *Suppose the initial conditional prior, $\vec{x}_0 | S^{-1}$, is Gaussian. If $\{\vec{a}_t\}_{t \geq 0}$ is a solution to the **RI Problem** with quadratic payoff given S^{-1} , then $\forall t \geq 0$, the posterior distribution $\vec{x}_t | \{\vec{a}_\tau\}_{0 \leq \tau \leq t} \cup S^{-1}$ is also Gaussian.*

The Equivalent LQG Problem. Lemma 3 simplifies the structure of the problem in that it allows us to re-write the **RI Problem** in terms of choosing a set of Gaussian joint distributions between the actions and the state.

Proposition 1. *Suppose the initial prior $\vec{x}_0 | S^{-1}$ is Gaussian and that $\{\vec{x}_t\}_{t \geq 0}$ is a Markov process with the following minimal state-space representation:*

$$\begin{aligned} \vec{x}_t &= \mathbf{A}\vec{x}_{t-1} + \mathbf{Q}\vec{u}_t, \\ \vec{u}_t &\perp \vec{x}_{t-1}, \quad \vec{u}_t \sim \mathcal{N}(0, \mathbf{I}^{k \times k}), k \in \mathbb{N}, \end{aligned} \tag{2.3}$$

Then, the **RI Problem** with quadratic payoff is equivalent to choosing a set of symmetric positive semidefinite matrices $\{\Sigma_{t|t}\}_{t \geq 0}$:

$$\begin{aligned} V_0(\Sigma_{0|-1}) &= \max_{\{\Sigma_{t|t} \in \mathbf{S}_+^n\}_{t \geq 0}} -\frac{1}{2} \sum_{t=0}^{\infty} \beta^t \left[\text{tr}(\Sigma_{t|t} \mathbf{\Omega}) + \omega \ln \left(\frac{|\Sigma_{t|t-1}|}{|\Sigma_{t|t}|} \right) \right] && \text{(LQG Problem)} \\ \text{s.t.} \quad \Sigma_{t+1|t} &= \mathbf{A}\Sigma_{t|t}\mathbf{A}' + \mathbf{Q}\mathbf{Q}', \quad \forall t \geq 0, && \text{(law of motion for priors)} \\ \Sigma_{t|t-1} - \Sigma_{t|t} &\succeq 0, \quad \forall t \geq 0 && \text{(no-forgetting)} \\ 0 &\prec \Sigma_{0|-1} \prec \infty \quad \text{given.} && \text{(initial prior)} \end{aligned}$$

Here, $\Sigma_{t|t} \equiv \text{var}(\vec{x}_t | a^t)$, $\Sigma_{t|t-1} \equiv \text{var}(\vec{x}_t | a^{t-1})$, $\mathbf{\Omega} \equiv \mathbf{H}\mathbf{H}'$ and \mathbf{S}_+^n is the n -dimensional symmetric positive semidefinite cone.

⁹This is without loss of generality; for any negative definite Hessian matrix $-\mathbf{H}_{aa} \prec 0$, normalize the action vectors by $\mathbf{H}_{aa}^{-\frac{1}{2}}$ to transform the payoff function to our original formulation.

This characterization of the problem matches the formulation in Sims (2010) but differs from the one in Sims (2003) and Miao, Wu and Young (2018) which solve the problem by optimizing at the steady-state.¹⁰

Solution Sims (2010) derives a first order condition for the solution to this problem when the no-forgetting constraint does not bind. Nonetheless, this constraint plays a key role in the solution of the **LQG Problem**. We extensively discuss the significance of this constraint for the economics of inattention in sub-section 2.3 as well as in the context of our application to Phillips curve.

Proposition 2. *Suppose $\Sigma_{0|-1}$ is strictly positive definite, and $\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'$ is of full rank. Then, all the future priors $\{\Sigma_{t+1|t}\}_{t \geq 0}$ are invertible under the optimal solution to the **LQG Problem**, which is characterized by*

$$\begin{aligned} \omega \Sigma_{t|t}^{-1} - \Lambda_t &= \Omega + \beta \mathbf{A}' (\omega \Sigma_{t+1|t}^{-1} - \Lambda_{t+1}) \mathbf{A}, & \forall t \geq 0, & \quad (\text{FOC}) \\ \Lambda_t (\Sigma_{t|t-1} - \Sigma_{t|t}) &= \mathbf{0}, \Lambda_t \succeq \mathbf{0}, \Sigma_{t|t-1} - \Sigma_{t|t} \succeq \mathbf{0}, & \forall t \geq 0, & \quad (\text{complementary slackness}) \\ \Sigma_{t+1|t} &= \mathbf{A} \Sigma_{t|t} \mathbf{A}' + \mathbf{Q}\mathbf{Q}', & \forall t \geq 0, & \quad (\text{law of motion for priors}) \\ \lim_{T \rightarrow \infty} \beta^{T+1} \text{tr}(\Lambda_{T+1} \Sigma_{T+1|T}) &= \mathbf{0} & & \quad (\text{transversality condition}) \end{aligned}$$

where Λ_t and $\Sigma_{t|t-1} - \Sigma_{t|t}$ are *simultaneously diagonalizable*.

The eigenvalues of Λ_t are in fact the shadow costs of the no-forgetting constraint. Therefore, when the no-forgetting constraint is not binding, $\Lambda_t = \mathbf{0}$ and the FOC is equivalent to the one derived in Sims (2010).

For the remainder of this section we rely on two matrix operators that are defined as following.

Definition 1. For a diagonal matrix $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ let

$$\text{Max}(\mathbf{D}, \omega) \equiv \text{diag}(\max(d_1, \omega), \dots, \max(d_n, \omega)) \quad (2.4)$$

$$\text{Min}(\mathbf{D}, \omega) \equiv \text{diag}(\min(d_1, \omega), \dots, \min(d_n, \omega)) \quad (2.5)$$

¹⁰The implied problem under the second approach is

$$\max_{\Sigma \succeq \mathbf{0}} -\text{tr}(\Sigma \Omega) - \omega \ln \left(\frac{|\Sigma_{-1}|}{|\Sigma|} \right) \text{ s.t. } \Sigma_{-1} = \mathbf{A} \Sigma \mathbf{A}' + \mathbf{Q}\mathbf{Q}', \Sigma_{-1} \succeq \Sigma.$$

Moreover, for a symmetric matrix \mathbf{X} with spectral decomposition $\mathbf{X} = \mathbf{U}'\mathbf{D}\mathbf{U}$, we define

$$\text{Max}(\mathbf{X}, \omega) \equiv \mathbf{U}' \text{Max}(\mathbf{D}, \omega) \mathbf{U}, \quad \text{Min}(\mathbf{X}, \omega) \equiv \mathbf{U}' \text{Min}(\mathbf{D}, \omega) \mathbf{U}. \quad (2.6)$$

Theorem 1. Let $\Omega_t \equiv \Omega + \beta \mathbf{A}' (\omega \Sigma_{t+1|t}^{-1} - \Lambda_{t+1}) \mathbf{A}$ denote the forward-looking component of the FOC in Proposition 2, which represents the marginal benefit of information. Then,

$$\Sigma_{t|t} = \omega \Sigma_{t|t-1}^{\frac{1}{2}} \left[\text{Max} \left(\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}}, \omega \right) \right]^{-1} \Sigma_{t|t-1}^{\frac{1}{2}} \quad (\text{policy function})$$

$$\Omega_t = \Omega + \beta \mathbf{A}' \Sigma_{t+1|t}^{-\frac{1}{2}} \text{Min} \left(\Sigma_{t+1|t}^{\frac{1}{2}} \Omega_{t+1} \Sigma_{t+1|t}^{\frac{1}{2}}, \omega \right) \Sigma_{t+1|t}^{-\frac{1}{2}} \mathbf{A} \quad (\text{Euler equation})$$

The **policy function** characterizes the optimal posterior given the state $\Sigma_{t|t-1}$ and the benefit matrix Ω_t . The **Euler equation** then characterizes Ω_t through a forward-looking difference equation that captures the dynamics of attention. Together with the **law of motion for priors** and **transversality condition**, these equations characterize the solution to the dynamic rational inattention problem.

While we have characterized the optimal posterior as a function of the agent's prior, the underlying assumption is that this posterior is generated by a vector of signals about \vec{x}_t . Both the number of these signals as well as how they load on different elements of the vector \vec{x}_t are endogenous. Our next result characterizes these signals.

Theorem 2. $\forall t \geq 0$, let $\{d_{i,t}\}_{1 \leq i \leq n}$ be the set of eigenvalues of the matrix $\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}}$ indexed in descending order. Moreover, let $\{\mathbf{u}_{i,t}\}_{1 \leq i \leq n}$ be orthonormal eigenvectors that correspond to those eigenvalues. Then, the agent's posterior belief at t is spanned by the following $0 \leq k_t \leq m$ signals

$$s_{i,t} = \mathbf{y}'_{i,t} \vec{x}_t + z_{i,t}, \quad 1 \leq i \leq k_t, \quad (2.7)$$

where

1. k_t is the number of the eigenvalues that are at least as large as ω : $k_t = \max\{i | d_{i,t} \geq \omega\}$.
2. $\forall i \in \{1, \dots, k_t\}$, $\mathbf{y}_{i,t} \equiv \Sigma_{t|t-1}^{-\frac{1}{2}} \mathbf{u}_{i,t}$.
3. $\forall i \in \{1, \dots, k_t\}$, $z_{i,t} \sim \mathcal{N}(0, \frac{\omega}{d_{i,t} - \omega})$, $z_{i,t} \perp (\vec{x}_t, z_{j,t})_{j \neq i}$.

Evolution of Optimal Beliefs and Actions. While Theorems 1 and 2 provide a representation for the optimal posteriors, we are often interested in the evolution of the agents' beliefs and actions. Our next theorem characterizes how beliefs and actions evolve over time.

Proposition 3. *Let $\{(\mathbf{y}_{i,t}, d_{i,t}, z_{i,t})_{1 \leq i \leq k_t}\}_{t \geq 0}$ be defined as in Theorem 2, and let $\hat{x}_t \equiv \mathbb{E}[\bar{x}_t | a^t]$ be the mean of agent's posterior about \bar{x}_t at time t . Then, \hat{x}_t and optimal actions evolve according to:*

$$\hat{x}_t = \mathbf{A}\hat{x}_{t-1} + \sum_{i=1}^{k_t} \left(1 - \frac{\omega}{d_{i,t}}\right) \boldsymbol{\Sigma}_{t|t-1} \mathbf{y}_{i,t} [\mathbf{y}'_{i,t}(\bar{x}_t - \mathbf{A}\hat{x}_{t-1}) + z_{i,t}] \quad (\text{evolution of beliefs})$$

$$\bar{a}_t = \mathbf{H}'\hat{x}_t \quad (\text{optimal actions})$$

Transition Dynamics and the Steady State. A key property of the **LQG Problem** is that it is deterministic. Additionally, as it is evident from the FOC in Proposition 2, eigenvectors of $\boldsymbol{\Sigma}_{t|t}$ are jump variables except for when the no-forgetting constraint binds. Thus, on the transition path, the agent has the desire to move on to their "steady state" posterior in each orthogonalized dimension unless the no-forgetting constraint binds, in which case they have to wait until their uncertainty stabilizes, either by climbing out of the inaction region or by reaching a steady state level within that region. Using the results of Proposition 2 and Theorem 1 we can represent the steady state of the problem with three equations that characterize a triple $(\bar{\boldsymbol{\Sigma}}_{-1}, \bar{\boldsymbol{\Sigma}}, \bar{\boldsymbol{\Omega}})$:

$$\bar{\boldsymbol{\Sigma}} = \omega \bar{\boldsymbol{\Sigma}}_{-1}^{\frac{1}{2}} \left[\text{Max} \left(\bar{\boldsymbol{\Sigma}}_{-1}^{\frac{1}{2}} \bar{\boldsymbol{\Omega}} \bar{\boldsymbol{\Sigma}}_{-1}^{\frac{1}{2}}, \omega \right) \right]^{-1} \bar{\boldsymbol{\Sigma}}_{-1}^{\frac{1}{2}} \quad (\text{policy function in steady state})$$

$$\bar{\boldsymbol{\Omega}} = \boldsymbol{\Omega} + \beta \mathbf{A}' \bar{\boldsymbol{\Sigma}}_{-1}^{-\frac{1}{2}} \text{Min} \left(\bar{\boldsymbol{\Sigma}}_{-1}^{\frac{1}{2}} \bar{\boldsymbol{\Omega}} \bar{\boldsymbol{\Sigma}}_{-1}^{\frac{1}{2}}, \omega \right) \bar{\boldsymbol{\Sigma}}_{-1}^{-\frac{1}{2}} \mathbf{A} \quad (\text{Euler equation in steady state})$$

$$\bar{\boldsymbol{\Sigma}}_{-1} = \mathbf{A} \bar{\boldsymbol{\Sigma}} \mathbf{A}' + \mathbf{Q} \mathbf{Q}' \quad (\text{prior variance in steady state})$$

The reduction of the problem to these three equations makes the problem computationally trivial. A toolbox to solve this system is available online. Once a solution is obtained, the impulse response functions for actions can be constructed using classic tools for solving Kalman filters.

2.3 Discussion and the Economics of Dynamic Rational Inattention

In this section we discuss the economic properties of the solution to the dynamic rational inattention problem.

Incentives. An important property of the **RI Problem** is that the marginal benefit of a bit of information is increasing in the prior uncertainty of the agent, while marginal cost of a bit is assumed to be a constant ω . Accordingly, for a large enough ω , the marginal benefit of acquiring a bit of information in different dimensions (eigenspaces) of the state might fall below its marginal cost, in which case the agent will decide not to pay attention to that dimension at all. This is clear from Theorem 1 which shows that the optimal **policy function** ignores eigenvalues that are less than ω .

Underneath its technical representation, Theorem 1 encodes an intuitive economic result. It shows that in acquiring information, the agent first decomposes the matrix $\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}}$, which captures the marginal benefit of information, into its orthogonal eigenspaces. At the *extensive margin*, the agent ignores eigenspaces whose eigenvalues are less than ω : the marginal benefit of acquiring information in these dimensions is outweighed by its marginal cost. On the *intensive margin*, the agent acquires signals for eigenspaces whose eigenvalues are larger than ω . Moreover, Theorem 2 shows that the loading of each of these signals on the state \vec{x}_t is given by the eigenvector associated with the signal's eigenspace.

Endogenous Sparsity. The extensive margin of information acquisition under dynamic rational inattention provides a microfoundation for why an agent might decide to *completely* ignore certain shocks or dimensions of the state in acquiring information and constitutes a microfoundation for sparsity of attention as in Gabaix (2014). This microfoundation endogenizes two objects relative to previous models of sparsity: (1) the dimensions of sparsity – which are pinned down by the eigenvectors of $\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}}$ with eigenvalues *less* than ω , and (2) the size of the information inaction region that is generated by the extensive margin as a function of the marginal benefit of information.

In our framework, sparsity is governed by the **no-forgetting** constraints. The most obvious and likely case for a binding **no-forgetting** constraint is when the number of actions m is strictly less than the dimension of the state n . This follows directly from Lemma 1 which states that the agent's actions at any given period are sufficient statistics for the underlying signals that she receives under the optimal solution.¹¹

¹¹Therefore, $\text{rank}(\Sigma_{t|t-1} - \Sigma_{t|t}) \leq m < n$ and the constraint binds as its nullity is at least $n - m > 0$.

In static environments, the fact that actions are sufficient statistics for the underlying signals follows directly from optimality (Matějka and McKay, 2015). If the agent’s action does not reveal the underlying signal, then he must have received information that was not used in choosing the action. Nonetheless, such a strategy is suboptimal given that information is costly. In dynamic settings, however, this is not necessarily true due to smoothing incentives. The agent might find it optimal to acquire signals about future actions before-hand in which case the history of actions at a given time is no longer sufficient for the information set of the agent. Lemma 1 rules this out by showing that if the chain-rule of mutual information holds, then the agent has no smoothing incentives. Thus, upon acquiring signals for every given action, the discounting of the cost induces the agent to postpone acquiring information to the period in which that action is taken.¹²

The economic consequence of this result is that rationally inattentive agents are not concerned about identification: independent of how many shocks they face, they are only interested in how those shocks affect their actions. An important reference for why this matters in an economic sense is Hellwig and Venkateswaran (2009) which shows that when firms receive signals about a sufficient statistic for their prices, they charge the right prices even though they cannot tell aggregate and idiosyncratic shocks apart.¹³

Information Spillovers. The intensive margin of information acquisition under dynamic rational inattention provides a microfoundation for *information spillovers* across different actions. These spillover effects are uniquely identified by the eigenvectors of $\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}}$ with eigenvalues *larger* than ω . Therefore, information about an action can effect other actions either through a subjective correlated posterior ($\Sigma_{t|t-1}$) or through complementarities or substitutabilities in actions captured by Ω_t .¹⁴

¹²The chain-rule of mutual information implies that for every three random variables:

$$\mathbb{I}(X; (Y, Z)) = \mathbb{I}(X; Y) + \mathbb{I}(X; Z|Y).$$

Intuitively, it imposes a certain type of linearity: mutual information is independent of whether information is measured altogether or part by part.

¹³Hellwig and Venkateswaran (2009) do not endogenize information choice, but the exogenous signal structure that they consider is optimal under our model with a particular parametrization.

¹⁴For instance, Kamdar (2018) documents that households have countercyclical inflation expectations – an observation that is contradictory to the negative comovement of inflation and unemployment in the data but is consistent of optimal information acquisition of households under substitutability of leisure and consumption. Similarly, Kőszegi and Matějka (2019) show that complementarities or substitutabilities in actions give rise to mental accounting in consumption behavior through optimal information acquisition. While these two papers use static information acquisition, our framework allows for dynamic spillovers through information acquisition.

3 An Attention Driven Phillips Curve

In this section we introduce a tractable general equilibrium model with rationally inattentive firms and provide an attention driven theory of the Phillips curve.

3.1 Environment

Households. Consider a fully attentive representative household who supplies labor N_t in a competitive labor market with nominal wage W_t , trades nominal bonds with net interest rate of i_t , and forms demand over a continuum of varieties indexed by $i \in [0, 1]$. Furthermore, the household's flow utility is $u(C_t, N_t) = \log(C_t) - N_t$. Formally, the representative household's problem is

$$\begin{aligned} & \max_{\{(C_{i,t})_{i \in [0,1]}, N_t\}_{t=0}^{\infty}} \mathbb{E}_0^f \left[\sum_{t=0}^{\infty} \beta^t (\log(C_t) - N_t) \right] \\ \text{s.t. } & \int_0^1 P_{i,t} C_{i,t} di + B_t \leq W_t N_t + (1 + i_{t-1}) B_{t-1} + T_t \\ & C_t = \left[\int_0^1 C_{i,t}^{\frac{\theta-1}{\theta}} di \right]^{\frac{\theta}{\theta-1}} \end{aligned}$$

where $\mathbb{E}_t^f[\cdot]$ is the expectation operator of this fully informed agent at time t , and T_t is the net lump-sum transfers to the household at t .

For ease of notation, let $P_t \equiv \left[\int_0^1 P_{i,t}^{1-\theta} \right]^{\frac{1}{1-\theta}}$ denote the aggregate price index and $Q_t \equiv P_t C_t$ be the nominal aggregate demand in this economy. The solution to the household's problem is summarized by:

$$C_{i,t} = C_t P_t^\theta P_{i,t}^{-\theta}, \quad \forall i \in [0, 1], \forall t \geq 0, \quad (3.1)$$

$$1 = \beta(1 + i_t) \mathbb{E}_t^f \left[\frac{Q_t}{Q_{t+1}} \right], \quad \forall t \geq 0, \quad (3.2)$$

$$W_t = Q_t, \quad \forall t \geq 0. \quad (3.3)$$

Monetary Policy. We assume that the monetary authority targets the growth of the nominal aggregate demand. This can be interpreted as targeting inflation and output growth similarly:

$$i_t = \rho + \phi \Delta q_t - \sigma_u u_t, \quad u_t \sim \mathcal{N}(0, 1)$$

where $\rho \equiv -\log(\beta)$ is the natural rate of interest, $q_t \equiv \log(P_t C_t)$ is the log of the nominal aggregate demand, and u_t is an exogenous shock to monetary policy that affects the

nominal interest rates with a standard deviation of σ_u .

Lemma 4. *Suppose $\phi > 1$. Then, in the log-linearized version of this economy, the aggregate demand is uniquely determined by the history of monetary policy shocks, and is characterized by the following random walk process:*

$$q_t = q_{t-1} + \frac{\sigma_u}{\phi} u_t. \quad (3.4)$$

Assuming that the monetary authority directly controls the nominal aggregate demand is a popular framework in the literature to study the effects of monetary policy on pricing.¹⁵ We derive this as an equilibrium outcome in Lemma 4 in order to relate the variance of the innovations to the nominal demand to the *strength* with which the monetary authority targets its growth: a larger ϕ stabilizes the nominal demand while a larger σ_u increases its variance.

Firms. Every variety $i \in [0, 1]$ is produced by a price-setting firm. Firm i hires labor $N_{i,t}$ from a competitive labor market at a subsidized wage $W_t = (1 - \theta^{-1})Q_t$ where the subsidy θ^{-1} is paid per unit of worker to eliminate steady state distortions introduced by monopolistic competition. Firms produce their product with a linear technology in labor, $Y_{i,t} = N_{i,t}$. Therefore, for a particular history $\{(P_t, Q_t)\}_{t \geq 0}$ and set of prices $\{P_{i,t}\}_{t \geq 0}$, the net present value of the firms' profits, discounted by the marginal utility of the household is given by

$$\begin{aligned} & \sum_{t=0}^{\infty} \beta^t \frac{1}{P_t C_t} (P_{i,t} - (1 - \theta^{-1})Q_t) C_t P_t^\theta P_{i,t}^{-\theta} \\ &= -(\theta - 1) \sum_{t=0}^{\infty} \beta^t (p_{i,t} - q_t)^2 + \mathcal{O}(\|(p_{i,t}, q_t)_{t \geq 0}\|^3) + \text{terms independent of } \{p_{i,t}\}_{t \geq 0} \end{aligned} \quad (3.5)$$

where the second line is a second order approximation with small letters denoting the logs of corresponding variables.¹⁶ This approximation states that for a monopolistic competitive firms, their loss from not matching their marginal cost in pricing, which is this setting is the nominal demand, is quadratic and proportional to $\theta - 1$, with θ denoting the elasticity of demand.

¹⁵See, for instance, Mankiw and Reis (2002), Woodford (2003), Golosov and Lucas Jr (2007), Maćkowiak and Wiederholt (2009a) and Nakamura and Steinsson (2010). This is also analogous to formulating monetary policy in terms of an exogenous rule for money supply as in, for instance, Caplin and Spulber (1987) or Gertler and Leahy (2008).

¹⁶For a detailed derivation of this second order approximation see, for instance, Maćkowiak and Wiederholt (2009a) or Afrouzi (2016).

We assume prices are perfectly flexible but firms are rationally inattentive and set their prices based on imperfect information about the underlying shocks in the economy. The rational inattention problem of firm i in the notation of the previous section is then given by

$$V(p_i^{-1}) = \max_{\{p_{i,t} \in \mathcal{S}^t\}_{t \geq 0}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}[-(\theta - 1)(p_{i,t} - q_t)^2 - \omega \mathbb{I}(p_i^t, q^t) | p_i^{-1}] \quad (3.6)$$

where $p_i^t \equiv (p_{i,\tau})_{\tau \leq t}$ denotes the history of firm's prices over up to time t . It is important to note that $\{p_{i,t}\}_{t \geq 0}$ is a stochastic process that proxies for the underlying signals that the firm receives over time – a result that follows from Lemma 2.

Assuming that the distribution of q_0 conditional on p_i^{-1} is a Gaussian process, and noting that $\{q_t\}_{t \geq 0}$ is itself a Markov Gaussian process, this problem satisfies the assumptions of Proposition 1. Formally, let $\sigma_{i,t|t-1} \equiv \sqrt{\text{var}(q_t | p_i^{t-1})}$, $\sigma_{i,t|t} \equiv \sqrt{\text{var}(q_t | p_i^t)}$ denote the prior and posterior standard deviations of firm i belief about q_t at time t . Then, the corresponding LQG problem to the one in Proposition 1 is

$$\begin{aligned} V(\sigma_{i,0|-1}) &= \max_{\{\sigma_{i,t|t}, \sigma_{i,t+1|t}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left[-(\theta - 1)\sigma_{i,t|t}^2 - \omega \ln \left(\frac{\sigma_{i,t|t-1}^2}{\sigma_{i,t|t}^2} \right) \right] \\ \text{s.t. } \sigma_{i,t+1|t}^2 &= \sigma_{i,t|t}^2 + \frac{\sigma_u^2}{\phi^2} \\ 0 &\leq \sigma_{i,t|t} \leq \sigma_{i,t|t-1} \end{aligned}$$

3.2 Characterization of Solution

The solution to this problem follows from Proposition 2, and is characterized by the following proposition.

Proposition 4. *Firms only pay attention to the monetary policy shocks if their prior uncertainty is above a reservation prior uncertainty. Formally,*

1. *the policy function of a firm for choosing their posterior uncertainty is*

$$\sigma_{i,t|t}^2 = \min\{\underline{\sigma}^2, \sigma_{i,t|t-1}^2\}, \quad \forall t \geq 0 \quad (3.7)$$

where $\underline{\sigma}^2$ is the positive root of the following quadratic equation:

$$\underline{\sigma}^4 + \left[\frac{\sigma_u^2}{\phi^2} - (1 - \beta) \frac{\omega}{\theta - 1} \right] \underline{\sigma}^2 - \frac{\omega}{\theta - 1} \frac{\sigma_u^2}{\phi^2} = 0 \quad (3.8)$$

2. the firm's price evolves according to:

$$p_{i,t} = p_{i,t-1} + \kappa_{i,t}(q_t - p_{i,t-1} + e_{i,t}) \quad (3.9)$$

where $\kappa_{i,t} \equiv \max\{0, 1 - \frac{\sigma^2}{\sigma_{i,t|t-1}^2}\}$ is the Kalman-gain of the firm under optimal solution and $e_{i,t}$ is the firm's rational inattention error.

The first part of Proposition 4 shows that firms pay attention to nominal demand only when they are sufficiently uncertain about it. The result follows from the fact that the marginal benefit of a bit of information is increasing in the prior uncertainty of a firm but the marginal cost is constant. Thus, for small levels of prior uncertainty where the marginal benefit of acquiring a bit of information falls below the marginal cost, the firm pays no attention to the nominal demand. However, once the prior uncertainty is at least as large as the reservation uncertainty, the firm always acquires enough information to maintain that level of uncertainty.

The second part of Proposition 1 shows that in the region where the firm does not pay attention to the nominal demand, their price does not respond to monetary policy shocks as the implied Kalman-gain is zero and the price is constant: $p_{i,t} = p_{i,t-1}$.

Nonetheless, as the nominal demand follows a random walk, it cannot be that the firm stays in the no-attention region forever. The variance of a random walk grows linearly with time, and it would only be below the reservation uncertainty for a finite amount of time. Once the firm's uncertainty reaches this level, the problem enters its steady state and the Kalman-gain is

$$\kappa_{i,t} = \kappa \equiv \frac{\sigma_u^2}{\phi^2 \underline{\sigma}^2 + \sigma_u^2}. \quad (3.10)$$

Comparative Statics. It is useful to study how the reservation uncertainty, $\underline{\sigma}^2$ and the steady state Kalman-gain κ change with the underlying parameters of the model.

Corollary 1. *The following hold:*

1. *The reservation uncertainty of firms increases with ω and σ_u , and decreases with ϕ, θ as well as β .*

2. The steady state Kalman-gain of firms increases with σ_u, θ and β , and decreases with ϕ and ω .

While Corollary 1 holds for all values of the underlying parameters, a simple first order approximation to the reservation uncertainty and steady state Kalman-gain can be derived when firms are perfectly patient ($\beta \rightarrow 1$) and σ_u^2 is small relative to the cost of information ω :¹⁷

$$[\underline{\sigma}^2]_{\beta=1, \sigma_u^2 \ll \omega} \approx \frac{\sigma_u}{\phi} \sqrt{\frac{\omega}{\theta - 1}} \quad (3.11)$$

$$[\kappa]_{\beta=1, \sigma_u^2 \ll \omega} \approx \frac{\sigma_u}{\phi} \sqrt{\frac{\theta - 1}{\omega}} \quad (3.12)$$

3.3 Aggregation

For aggregation, we make two assumptions: (1) firms all start from the same initial prior uncertainty, $\sigma_{i,0|0}^2 = \sigma_{0|0}^2, \forall i \in [0, 1]$, and (2) firms' rational inattention errors are independently distributed.¹⁸

Notation-wise, we define the log-linearized aggregate price as the average price of all firms, $p_t \equiv \int_0^1 p_{i,t} di$, the log-linearized inflation as $\pi_t = p_t - p_{t-1}$ and log-linearized aggregate output as the difference between the nominal demand and aggregate price, $y_t \equiv q_t - p_t$.

Proposition 5. *Suppose all firms start from the same prior uncertainty. Then,*

1. *the Phillips curve of this economy is*

$$\pi_t = \max\left\{0, \frac{\sigma_{t|t-1}^2 - \underline{\sigma}^2}{\sigma_{t|t}^2}\right\} y_t \quad (3.13)$$

2. *Suppose $\sigma_{T|T-1}^2 \leq \underline{\sigma}^2$, then $\forall t \leq T$:*

$$\pi_t = 0, \quad y_t = y_{t-1} + \frac{\sigma_u}{\phi} u_t. \quad (3.14)$$

¹⁷This approximation becomes the exact solution to the analogous problem in continuous time. This follows from the fact that in continuous time the variance of the innovation is arbitrarily small because it is proportional to the time between consecutive decisions.

¹⁸Our second assumption is not without loss of generality once we assume that the cost of information is Shannon's mutual information (Denti, 2015; Afrouzi, 2016). With other classes of cost functions, however, non-fundamental volatility can be optimal – see Hébert and La'O (2019) for characterization of these cost functions.

3. Suppose $\sigma_{T|T-1}^2 > \underline{\sigma}^2$, then for $t \geq T + 1$:

$$\pi_t = (1 - \kappa)\pi_{t-1} + \frac{\kappa\sigma_u}{\phi}u_t \quad (3.15)$$

$$y_t = (1 - \kappa)y_{t-1} + \frac{(1 - \kappa)\sigma_u}{\phi}u_t \quad (3.16)$$

where $\kappa \equiv \frac{\sigma_u^2}{\phi^2\sigma^2 + \sigma_u^2}$ is the Kalman-gain of firms in the steady state of the attention problem.

3.4 Discussion of Results

Proposition 5 shows that this economy has a Phillips curve with a time-varying slope, *which is flat* if and when the no-forgetting constraint binds. At a time when firm's uncertainty is below the reservation uncertainty, firms pay no attention to the monetary policy and the inflation does not respond to monetary policy shocks.

Nonetheless, since nominal demand follows a random walk process and the attention problem is deterministic, Proposition 5 also shows that the rational inattention problem will eventually enter and remain at its steady state where firms do pay attention to the nominal demand. In this section, we start by analyzing this steady state, and then consider the dynamic consequences of unanticipated disturbances (MIT shocks) to the parameters of the model.

3.4.1 The Long-run Slope of the Phillips Curve

It follows from Proposition 5 that once the inattention problem settles in its state, the Phillips curve is given by

$$\pi_t = \frac{\kappa}{1 - \kappa}y_t \quad (3.17)$$

where κ is the steady state Kalman gain. Moreover, the last part of the Proposition also shows that in this steady state, both output and inflation follow AR(1) processes whose persistence are given by $1 - \kappa$.

Thus, in the long-run, the parameter κ is sufficient for determining the slope of the Phillips curve as well as the magnitude and persistence of the real effects of monetary policy shocks in this economy: a lower value for κ leads to a flatter Phillips curve, a more persistent process for inflation and output, and larger monetary non-neutrality. The intuition behind all of these is that a lower value for κ is equivalent to lower attention to monetary policy shocks on the part of firms. It takes longer for less attentive firms to

learn about monetary policy shocks and respond to them. In the meantime, since firms are not adjusting their prices one to one with the shock, their output has to compensate. Thus, less attention, leads to a longer half-life for π and a larger degree of π – monetary non-neutrality.

Comparative statics of κ with respect to the underlying parameters of the model are derived in Corollary 1. In particular, we would like to focus on how the rule of monetary policy affects the slope of the Phillips curve and consequently the persistence and the magnitude of the real effect so of monetary policy shocks.

Corollary 1 shows that κ is increasing with $\frac{\sigma_u}{\phi}$. We interpret this ratio as a measure for how dovish the monetary policy is in this economy since a larger $\frac{\sigma_u}{\phi}$ corresponds to a lower relative weight on stabilizing inflation. It follows that in the long-run, the Phillips curve is steeper in more dovish economies. If the monetary authority opts for a lower weight on the stabilization of the nominal variables, the firms face a more volatile process for their marginal cost and optimally choose to pay more attention to monetary policy shocks in the steady state of their attention problem. As a result, such firms are more responsive to monetary policy shocks and are quicker in adjusting their prices.

3.4.2 The Aftermath of An Unexpectedly More Hawkish Monetary Policy

An interesting exercise is to consider an unexpected *decrease* in $\frac{\sigma_u}{\phi}$. This can correspond to lower variance of monetary policy shocks or a higher weight on stabilizing inflation in the rule of monetary policy.

Corollary 2. *Suppose the economy is in the steady state of its attention problem, and consider an unexpected decrease in $\frac{\sigma_u}{\phi}$. Then, the economy immediately jumps to a new steady state of the attention problem, in which:*

1. *The Phillips curve is flatter.*
2. *Output and inflation responses are more persistent.*

The comparative statics follow directly from Corollary 1 and are straight forward; however, the reason that the economy jumps to its new steady state needs some intuition. The reason for this jump is that a more hawkish economy has a less volatile nominal demand process and firms have lower reservation uncertainties in less volatile environments. Therefore, once the monetary policy rule becomes more hawkish, firms find themselves with a prior uncertainty that is higher than their new reservation uncertainty. Consequently, they acquire enough information to immediately reduce their uncertainty to the new reservation level. The key observation is that once they reach this new lower

level of uncertainty they need a lower rate of information acquisition to maintain that level of uncertainty. Hence, while the reservation uncertainty decreases with a more hawkish rule, the steady state Kalman-gain also decreases and leads to flatter Phillips curve and a higher persistence in responses of output and inflation.

Conceptually, our results speak to, and are consistent with, the post-Volcker era in the U.S. monetary policy. A large strand of the literature has documented that the slope of the Phillips curve has become flatter in the last few decades.¹⁹ Our theory provides a new perspective on this issue. Firms do not need to be attentive to monetary policy in an environment where the policy makers follow a hawkish rule.

3.4.3 The Aftermath of An Unexpectedly More Dovish Monetary Policy

The model is non-symmetric in response to changes in the rule of monetary policy. While the economy jumps to the new steady state of the attention problem after a decreases in $\frac{\sigma_u}{\phi}$, as shown in Corollary 2, the reverse is not true. An unexpected increase in $\frac{\sigma_u}{\phi}$ has different short-run implications due to its effect on reservation uncertainty.

Corollary 3. *Suppose the economy is in the steady state of its attention problem, and consider an unexpected increase in $\frac{\sigma_u}{\phi}$. Then,*

1. *The Phillips curve becomes temporarily flat until firms' uncertainty increases to its new reservation level.*
2. *Once firms' uncertainty reaches to its new reservation level, the economy enters its new steady state in which:*
 - (a) *the Phillips curve is steeper.*
 - (b) *output and inflation responses are less persistent.*

The intuition follows from Corollary 1. An increase in $\frac{\sigma_u}{\phi}$ makes the nominal demand more volatile and raises the reservation uncertainty of firms. Hence, immediately after such a shock, firms find themselves with an uncertainty that is below this reservation level; the no-forgetting constraint binds and they temporarily stop paying attention to the monetary policy shocks until their uncertainty grows to its new reservation level. In the meantime, the Phillips curve is flat and inflation is non-responsive to monetary policy shocks. The duration of this temporary phase depends

¹⁹See Coibion and Gorodnichenko (2015b) who do separate estimations for the pre- and post-Volcker period and document a decrease in the slope. See also, for instance, Blanchard (2016); Bullard (2018); Hooper et al. (2019).

Once firms' uncertainty reaches its new reservation level, however, they start paying attention at a higher rate to maintain this new level as the process is now more volatile. Thus, while a more dovish policy leads to a temporarily flat Phillips curve, it eventually leads to a steeper Phillips curve once firms adapt to their new environment.

These findings provide a new perspective on the recent perceived disconnect between inflation and monetary policy. If the Great Recession was followed by a period of higher uncertainty about monetary policy shocks or more lenient policy, then our model predicts that it would be optimal for firms to stop paying attention to monetary policy in the transition period to the new steady state.

4 Concluding Remarks

We characterize and solve dynamic multivariate rational inattention models and apply our findings to derive an attention driven Phillips curve.

Our theory of the Phillips curve puts forth a new perspective on the flattening of the slope of the Phillips curve in recent decades, and suggests that this was an endogenous response of the private sector to a more disciplined monetary policy in the post-Volcker era which put a larger weight on stabilizing nominal variables.

On the policy front, our results speak to an ongoing debate on the trade-off between stabilizing inflation and maintaining a lower unemployment rate. Our theory suggests that while a dovish policy might seem appealing in the current climate where inflation seems hardly responsive to monetary policy, such a policy might have an adverse effect once implemented.

References

- Afrouzi, Hassan**, "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money," 2016. Manuscript.
- Angeletos, George-Marios and Jennifer La'O**, "Incomplete Information, Higher-Order Beliefs and Price Inertia," *Journal of Monetary Economics*, 2009, 56, S19–S37.
- Babb, Nathan and Alan K Detmeister**, "Nonlinearities in the Phillips Curve for the United States: Evidence Using Metropolitan Data," FEDS Working Paper 2017-070 2017.
- Blanchard, Olivier**, "The Phillips Curve: Back to the '60s?," *American Economic Review*, 2016, 106 (5), 31–34.

- Bullard, James**, “The Case of the Disappearing Phillips Curve,” in “The 2018 ECB Forum on Central Banking, Sintra, Portugal, June,” Vol. 19 2018.
- Caplin, Andrew, Mark Dean, and John Leahy**, “Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy,” Working Paper 23652, National Bureau of Economic Research 2017.
- Caplin, Andrew S and Daniel F Spulber**, “Menu Costs and the Neutrality of Money,” *The Quarterly Journal of Economics*, 1987, 102 (4), 703–725.
- Coibion, Olivier**, “Testing the Sticky Information Phillips Curve,” *The Review of Economics and Statistics*, 2010, 92 (1), 87–101.
- **and Yuriy Gorodnichenko**, “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts,” *American Economic Review*, August 2015, 105 (8), 2644–78.
- **and –**, “Is the Phillips Curve Alive and Well After All? Inflation Expectations and the Missing Disinflation,” *American Economic Journal: Macroeconomics*, January 2015, 7 (1), 197–232.
- , – , **and Tiziano Ropele**, “Inflation Expectations and Firm Decisions: New Causal Evidence,” Working Paper 25412, National Bureau of Economic Research 2018.
- Cover, Thomas M and Joy A Thomas**, *Elements of Information Theory*, John Wiley & Sons, 2012.
- da Silveira, Rava Azeredo, Yeji Sung, and Michael Woodford**, “Noisy Memory and Over-Reaction to News,” 2019. Manuscript.
- Denti, Tommaso**, “Unrestricted Information Acquisition,” 2015. Manuscript.
- Fulton, Chad**, “Mechanics of Static Quadratic Gaussian Rational Inattention Tracking Problems,” Working Paper, Board of Governors of the Federal Reserve System 2018.
- Gabaix, Xavier**, “A Sparsity-Based Model of Bounded Rationality,” *The Quarterly Journal of Economics*, 2014, 129 (4), 1661–1710.
- Gaglianone, Wagner Piazza, Raffaella Giacomini, Joao Issler, and Vasiliki Skreta**, “Incentive-driven Inattention,” Discussion Paper 13619, CEPR 2019.
- Gali, Jordi and Mark Gertler**, “Inflation Dynamics: A Structural Econometric Analysis,” *Journal of Monetary Economics*, 1999, 44 (2), 195–222.
- Gertler, Mark and John Leahy**, “A Phillips Curve with an Ss Foundation,” *Journal of Political Economy*, 2008, 116 (3), 533–572.
- Golosov, Mikhail and Robert E Lucas Jr**, “Menu Costs and Phillips curves,” *Journal of Political Economy*, 2007, 115 (2), 171–199.

- Hébert, Benjamin and Jennifer La'O**, "Information Acquisition, Efficiency, and Non-Fundamental Volatility," 2019. Manuscript.
- **and Michael Woodford**, "Information Costs and Sequential Information Sampling," Working Paper 25316, National Bureau of Economic Research 2018.
- Hellwig, Christian and Venky Venkateswaran**, "Setting the Right Prices for the Wrong Reasons," *Journal of Monetary Economics*, 2009, 56, S57–S77.
- Hooper, Peter, Frederic S Mishkin, and Amir Sufi**, "Prospects for Inflation in a High Pressure Economy: Is the Phillips Curve Dead or is It Just Hibernating?," Working Paper 25792, National Bureau of Economic Research 2019.
- Ilut, Cosmin and Rosen Valchev**, "Economic Agents As Imperfect Problem Solvers," 2017. Manuscript.
- Kamdar, Rupal**, "The Inattentive Consumer: Sentiment and Expectations," 2018. Manuscript.
- Khaw, Mel Win and Oskar Zorrilla**, "Deeper Habits," 2018. Manuscript.
- **, Luminita Stevens, and Michael Woodford**, "Discrete Adjustment to a Changing Environment: Experimental Evidence," *Journal of Monetary Economics*, 2017, 91, 88–103.
- Kőszegi, Botond and Filip Matějka**, "An Attention-Based Theory of Mental Accounting," 2019. Manuscript.
- Kumar, Anil and Pia M Orrenius**, "A Closer Look at the Phillips Curve Using State-Level Data," *Journal of Macroeconomics*, 2016, 47, 84–102.
- Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko**, "Inflation Targeting Does Not Anchor Inflation Expectations: Evidence from Firms in New Zealand," *Brookings Papers on Economic Activity*, 2015, pp. 187–226.
- Landier, Augustin, Yueran Ma, and David Thesmar**, "Biases in Expectations: Experimental Evidence," 2019. Manuscript.
- Lucas, Robert E**, "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 1972, 4 (2), 103–124.
- Luo, Yulei**, "Consumption Dynamics under Information Processing Constraints," *Review of Economic Dynamics*, 2008, 11 (2), 366–385.
- **, Jun Nie, and Eric R Young**, "Robustness, Information Processing Constraints, and the Current Account in Small Open Economies," *Journal of International Economics*, 2012, 88 (1), 104–120.
- Maćkowiak, Bartosz and Mirko Wiederholt**, "Optimal Sticky Prices under Rational Inattention," *The American Economic Review*, 2009, 99 (3), 769–803.

- **and** –, “Optimal Sticky Prices under Rational Inattention,” Working Paper Series 1009, European Central Bank February 2009.
- Maćkowiak, Bartosz and Mirko Wiederholt**, “Business Cycle Dynamics under Rational Inattention,” *The Review of Economic Studies*, 2015, 82 (4), 1502–1532.
- Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt**, “Dynamic Rational Inattention: Analytical Results,” *Journal of Economic Theory*, 2018, 176, 650 – 692.
- Mankiw, N Gregory and Ricardo Reis**, “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve,” *The Quarterly Journal of Economics*, 2002, 117 (4), 1295–1328.
- Matějka, Filip**, “Rationally Inattentive Seller: Sales and Discrete Pricing,” *The Review of Economic Studies*, 2015, 83 (3), 1125–1155.
- Matějka, Filip and Alisdair McKay**, “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model,” *American Economic Review*, January 2015, 105 (1), 272–98.
- Melosi, Leonardo**, “Estimating Models with Dispersed Information,” *American Economic Journal: Macroeconomics*, 2014, 6 (1), 1–31.
- Miao, Jianjun, Jieran Wu, and Eric Young**, “Multivariate Rational Inattention,” December 2018. Boston University - Department of Economics.
- Mondria, Jordi and Thomas Wu**, “The Puzzling Evolution of the Home Bias, Information Processing and Financial Openness,” *Journal of Economic Dynamics and Control*, 2010, 34 (5), 875–896.
- Muth, John F.**, “Rational Expectations and the Theory of Price Movements,” *Econometrica*, 1961, pp. 315–335.
- Nagel, Stefan and Zhengyang Xu**, “Asset Pricing with Fading Memory,” Working Paper 26255, National Bureau of Economic Research 2019.
- Nakamura, Emi and Jon Steinsson**, “Monetary Non-Neutrality in a Multisector Menu Cost Model,” *The Quarterly journal of economics*, 2010, 125 (3), 961–1013.
- Nimark, Kristoffer**, “Dynamic Pricing and Imperfect Common Knowledge,” *Journal of Monetary Economics*, 2008, 55 (2), 365–382.
- Paciello, Luigi**, “Monetary Policy and Price Responsiveness to Aggregate Shocks under Rational Inattention,” *Journal of Money, Credit and Banking*, 2012, 44 (7), 1375–1399.
- **and Mirko Wiederholt**, “Exogenous Information, Endogenous Information, and Optimal Monetary Policy,” *The Review of Economic Studies*, 2014, 81 (1), 356–388.
- Pasten, Ernesto and Raphael Schoenle**, “Rational Inattention, Multi-Product Firms and the Neutrality of Money,” *Journal of Monetary Economics*, 2016, 80, 1–16.

- Peng, Lin and Wei Xiong**, "Investor Attention, Overconfidence and Category Learning," *Journal of Financial Economics*, 2006, 80 (3), 563–602.
- Ravid, Doron**, "Bargaining with Rational Inattention," 2019. Manuscript.
- Reis, Ricardo**, "Inattentive Producers," *The Review of Economic Studies*, 2006, 73 (3), 793–821.
- Roberts, John M**, "New Keynesian Economics and the Phillips Curve," *Journal of Money, Credit and Banking*, 1995, 27 (4), 975–984.
- Roth, Christopher and Johannes Wohlfart**, "How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?," 2018.
- Rudd, Jeremy and Karl Whelan**, "New Tests of the New-Keynesian Phillips Curve," *Journal of Monetary Economics*, 2005, 52 (6), 1167 – 1181.
- Ryngaert, Jane**, "What do (and Don't) Forecasters Know About US Inflation," 2017. Manuscript.
- Sims, Christopher A**, "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, 50 (3), 665–690.
- , "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 2006, 96 (2), 158–163.
- , "Rational Inattention and Monetary Economics," in "Handbook of Monetary Economics," Vol. 3, Elsevier, 2010, pp. 155–181.
- Steiner, Jakub, Colin Stewart, and Filip Matějka**, "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 2017, 85 (2), 521–553.
- Stevens, Luminita**, "Coarse Pricing Policies," *The Review of Economic Studies*, 07 2019.
- Tutino, Antonella**, "Rationally Inattentive Consumption Choices," *Review of Economic Dynamics*, 2013, 16 (3), 421–439.
- Woodford, Michael**, "Imperfect Common Knowledge and the Effects of Monetary Policy," *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 2003.
- , "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics*, 2009, 56, S100–S124.
- Yang, Choongryul**, "Rational Inattention, Menu Costs, and Multi-Product Firms: Evidence and Monetary Non-Neutrality," 2019. Manuscript.
- Zorn, Peter**, "Investment under Rational Inattention: Evidence from US Sectoral Data," 2016. Manuscript.

APPENDIX

A Proofs

Proof of Lemma 1. First, note that observing $\{a^t\}_{t=0}^\infty$ induces the same action payoffs over time as $\{S^t\}_{t=0}^\infty$ because at any time t and for every possible realization of S^t , the agent gets $a(S^t)$ – the optimal action induced by that realization – as a direct signal. Suppose now that a^t is not a sufficient statistic for S^t relative to X^t . Then, we can show that $\{a^t\}_{t=0}^\infty$ costs less in terms of information than $\{S^t\}_{t=0}^\infty$. To see this, note that for any $t \geq 1$ and S^t , consecutive applications of the chain-rule of mutual information imply

$$\mathbb{I}(X^t; S^t) = \mathbb{I}(X^t; S^t | S^{t-1}) + \mathbb{I}(X^t; S^{t-1}) = \mathbb{I}(X^t; S^t | S^{t-1}) + \mathbb{I}(X^{t-1}; S^{t-1}) + \underbrace{\mathbb{I}(X^t; S^{t-1} | X^{t-1})}_{=0},$$

where the third term is zero by availability of information at time $t - 1$; $S^{t-1} \perp X^t | X^{t-1}$. Moreover, for $t = 0$ applying the chain-rule implies:

$$\mathbb{I}(X^0; S^0) = \mathbb{I}(X^0; S^0 | S^{-1}) + \mathbb{I}(X^0; S^{-1})$$

Thus,

$$\sum_{t=0}^{\infty} \beta^t \mathbb{I}(X^t; S^t | S^{t-1}) = \sum_{t=0}^{\infty} \beta^t (\mathbb{I}(X^t; S^t) - \mathbb{I}(X^{t-1}; S^{t-1})) = \mathbb{I}(X^0; S^{-1}) + (1 - \beta) \sum_{t=0}^{\infty} \beta^t \mathbb{I}(X^t; S^t).$$

Similarly, noting that a^{-1} is equal to S^{-1} by definition, we can show

$$\sum_{t=0}^{\infty} \beta^t \mathbb{I}(X^t; a^t | a^{t-1}) = \mathbb{I}(X^0; S^{-1}) + (1 - \beta) \sum_{t=0}^{\infty} \beta^t \mathbb{I}(X^t; a^t).$$

Finally, note that $X^t \rightarrow S^t \rightarrow a^t$ form a Markov chain so that $X^t \perp a^t | S^t$. A final application of the chain-rule for mutual information implies

$$\mathbb{I}(X^t; a^t, S^t) = \mathbb{I}(X^t; a^t) + \mathbb{I}(X^t; S^t | a^t) = \mathbb{I}(X^t; S^t) + \underbrace{\mathbb{I}(X^t; a^t | S^t)}_{=0}.$$

Therefore,

$$\begin{aligned} \sum_{t=0}^{\infty} \beta^t \mathbb{I}(X^t; S^t | S^{t-1}) - \sum_{t=0}^{\infty} \beta^t \mathbb{I}(X^t; a^t | a^{t-1}) &= (1 - \beta) \sum_{t=0}^{\infty} \beta^t [\mathbb{I}(X^t; S^t) - \mathbb{I}(X^t; a^t)] \\ &= \sum_{t=0}^{\infty} \beta^t \mathbb{I}(X^t; S^t | a^t) \geq 0. \end{aligned}$$

Hence, while $\{a^t\}_{t=0}^{\infty}$ induces the same action payoffs as $\{S^t\}_{t=0}^{\infty}$, it costs less in terms of information costs, and induce higher total utility for the agent. Therefore, if $\{S^t\}_{t \geq 0}$ is optimal, it has to be that

$$\mathbb{I}(X^t; S^t | a^t) = 0, \forall t \geq 0 \quad (\text{A.1})$$

which implies $S^t \perp X^t | a^t$ and $X^t \rightarrow a^t \rightarrow S^t$ forms a Markov chain $\forall t \geq 0$. \blacksquare

Proof of Lemma 2. The chain-rule implies $\mathbb{I}(X^t; a^t | a^{t-1}) = \mathbb{I}(X^t; a_t, a^{t-1} | a^{t-1}) = \mathbb{I}(X^t; a_t | a^{t-1})$. Moreover, it also implies

$$\mathbb{I}(X^t; \vec{a}_t | a^{t-1}) = \mathbb{I}(\vec{x}_t; \vec{a}_t | a^{t-1}) + \mathbb{I}(X^{t-1}; \vec{a}_t | a^{t-1}, \vec{x}_t).$$

Since $a_t = \arg \max_a \mathbb{E}[u(a; X_t) | S^t]$ and given that a^t is a sufficient statistic for S^t , then optimality requires that $\mathbb{I}(X^{t-1}; a_t | a^{t-1}, \vec{x}_t) = 0$. To see why, suppose not. Then, we can construct a an information structure that costs less but implies the same expected payoff. Thus, for the optimal information structure, this mutual information is zero, which implies

$$\mathbb{I}(X^t; a^t | a^{t-1}) = \mathbb{I}(\vec{x}_t; \vec{a}_t | a^{t-1}), \quad \vec{a}_t \perp X^{t-1} | (\vec{x}_t, a^{t-1}).$$

\blacksquare

Proof of Lemma 3. We prove this Proposition by showing that for any sequence of actions, we can construct a Gaussian process that costs less in terms of information costs, but generates the exact same payoff sequence. To see this, take an action sequence $\{\vec{a}_t\}_{t \geq 0}$, and let $a^t \equiv \{\vec{a}_\tau : 0 \leq \tau \leq t\} \cup S^{-1}$ denote the information set implied by this action sequence. Now define a sequence of Gaussian variables $\{\hat{a}_t\}_{t \geq 0}$ such that for $t \geq 0$,

$$\text{var}(X^t | \hat{a}^t) = \mathbb{E}[\text{var}(X^t | a^t) | S^{-1}].$$

Note that both these sequence of actions imply the same sequence of utilities for the agent since they have the same covariance matrix by construction. So we just need to show that

the Gaussian sequence costs less. To see this note:

$$\begin{aligned}
& \mathbb{E} \left[\sum_{t=0}^{\infty} \beta^t \left(\mathbb{I}(X^t; a^t | a^{t-1}) - \mathbb{I}(X^t; \hat{a}^t | \hat{a}^{t-1}) \right) | S^{-1} \right] \\
&= (1 - \beta) \mathbb{E} \left[\sum_{t=0}^{\infty} \beta^t \left(\mathbb{I}(X^t; a^t) - \mathbb{I}(X^t; \hat{a}^t) \right) | S^{-1} \right] \\
&= (1 - \beta) \mathbb{E} \left[\sum_{t=0}^{\infty} \beta^t \left(h(X^t | \hat{a}^t) - h(X^t | a^t) \right) | S^{-1} \right] \geq 0,
\end{aligned}$$

where the last inequality is followed from the fact that among the random variables with the same expected covariance matrix, the Gaussian variable has maximal entropy.²⁰ ■

Proof of Proposition 1. We know from Lemma 3 that optimal posteriors, if the problem attains its maximum, are Gaussian. So without loss of generality we can restrict our attention to Gaussian signals. Moreover, since $\{\vec{x}_t\}_{t \geq 0}$ is Markov, we know from Lemma 2 that optimal actions should satisfy $\vec{a}_t \perp X^{t-1} | (a^{t-1}, \vec{x}_t)$ where $a^t = \{\vec{a}_\tau\}_{0 \leq \tau \leq t} \cup S^{-1}$. Thus, we can decompose:

$$\vec{a}_t - \mathbb{E}[\vec{a}_t | a^{t-1}] = \mathbf{Y}'_t (\vec{x}_t - \mathbb{E}[\vec{x}_t | a^{t-1}]) + \vec{z}_t, \quad \vec{z}_t \perp (a^{t-1}, X^t), \quad \vec{z}_t \sim \mathcal{N}(0, \Sigma_{z,t}),$$

for some $\mathbf{Y}_t \in \mathbb{R}^{n \times m}$. Now, note that choosing actions is equivalent to choosing a sequence of $\{(\mathbf{Y}_t \in \mathbb{R}^{n \times m}, \Sigma_{z,t} \succeq 0)\}_{t \geq 0}$.

Now, let $\vec{x}_t | a^{t-1} \sim \mathcal{N}(\vec{x}_{t|t-1}, \Sigma_{t|t-1})$ and $\vec{x}_t | a^t \sim \mathcal{N}(\vec{x}_{t|t}, \Sigma_{t|t})$ denote the prior and posterior beliefs of the agent at time t . Kalman filtering implies $\forall t \geq 0$:

$$\begin{aligned}
\vec{x}_{t|t} &= \vec{x}_{t|t-1} + \Sigma_{t|t-1} \mathbf{Y}_t (\mathbf{Y}'_t \Sigma_{t|t-1} \mathbf{Y}_t + \Sigma_{z,t})^{-1} (\vec{a}_t - \vec{a}_{t|t-1}), \quad \vec{x}_{t+1|t} = \mathbf{A} \vec{x}_{t|t} \\
\Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1} \mathbf{Y}_t (\mathbf{Y}'_t \Sigma_{t|t-1} \mathbf{Y}_t + \Sigma_{z,t})^{-1} \mathbf{Y}'_t \Sigma_{t|t-1}, \\
\Sigma_{t+1|t} &= \mathbf{A} \Sigma_{t|t} \mathbf{A}' + \mathbf{Q} \mathbf{Q}'.
\end{aligned}$$

Note that positive semi-definiteness of $\Sigma_{z,t}$ implies that $\Sigma_{t|t} \preceq \Sigma_{t|t-1}$. Furthermore, note that for any posterior $\Sigma_{t|t} \preceq \Sigma_{t|t-1}$ that is generated by fewer than or equal to m signals, there exists at least one set of $\mathbf{Y}_t \in \mathbb{R}$ and $\Sigma_{v,t} \in \mathbb{S}_+^m$ that generates it. Moreover, note that any linear map of \vec{a}_t , as long as it is of rank m , is sufficient for $\vec{x}_{t|t}$ by sufficiency of action for signals. So we normalize $\vec{a}_t = \mathbf{H}' \vec{x}_{t|t}$ which is allowed as \mathbf{H} has full column rank.

²⁰See Chapter 12 in Cover and Thomas (2012).

Additionally, observe that given a^t :

$$\mathbb{E}[(\vec{a}_t - \vec{x}_t' \mathbf{H})(\vec{a}_t - \mathbf{H}' \vec{x}_t') | a^t] = \mathbb{E}[(\vec{x}_t - \vec{x}_{t|t})' \mathbf{H} \mathbf{H}' (\vec{x}_t - \vec{x}_{t|t}) | a^t] = \text{tr}(\mathbf{\Omega} \mathbf{\Sigma}_{t|t}), \mathbf{\Omega} \equiv \mathbf{H} \mathbf{H}'.$$

Thus, the **RI Problem** becomes:

$$\begin{aligned} & \sup_{\{\mathbf{\Sigma}_{t|t} \in \mathbf{S}_+^n\}_{t \geq 0}} -\frac{1}{2} \sum_{t=0}^{\infty} \beta^t \left[\text{tr}(\mathbf{\Sigma}_{t|t} \mathbf{\Omega}) + \omega \ln \left(\frac{|\mathbf{\Sigma}_{t|t-1}|}{|\mathbf{\Sigma}_{t|t}|} \right) \right] && \text{(LQG Problem)} \\ \text{s.t.} \quad & \mathbf{\Sigma}_{t+1|t} = \mathbf{A} \mathbf{\Sigma}_{t|t} \mathbf{A}' + \mathbf{Q} \mathbf{Q}', \quad \forall t \geq 0, && \text{(law of motion for priors)} \\ & \mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq 0, \quad \forall t \geq 0 && \text{(no-forgetting)} \\ & 0 \prec \mathbf{\Sigma}_{0|-1} = \text{var}(\vec{x}_0 | S^{-1}) \prec \infty \quad \text{given.} && \text{(initial prior)} \end{aligned}$$

Finally, note that we can replace the sup operator with max because $\forall t \geq 0$ the objective function is continuous as a function of $\mathbf{\Sigma}_{t|t}$ and the set $\{\mathbf{\Sigma}_{t|t} \in \mathbf{S}_+^n | 0 \preceq \mathbf{\Sigma}_{t|t} \preceq \mathbf{\Sigma}_{t|t-1}\}$ is a compact subset of the positive semidefinite cone. ■

Proof of Proposition 2. We start by writing the Lagrangian. Let $\mathbf{\Gamma}_t$ be a symmetric matrix whose k 'th row is the vector of shadow costs on the k 'th column of the evolution of prior at time t . Moreover, let λ_t be the vector of shadow costs on the no-forgetting constraint which can be written as $\text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}) \geq 0$ where $\text{eig}(\cdot)$ denotes the vector of eigenvalues of a matrix.

$$\begin{aligned} L_0 = & \max_{\{\mathbf{\Sigma}_{t|t} \in \mathbf{S}_+^n\}_{t \geq 0}} \frac{1}{2} \sum_{t=0}^{\infty} \beta^t [-\text{tr}(\mathbf{\Sigma}_{t|t} \mathbf{\Omega}) - \omega \ln(|\mathbf{\Sigma}_{t|t-1}|) + \omega \ln(|\mathbf{\Sigma}_{t|t}|) \\ & - \text{tr}(\mathbf{\Gamma}_t (\mathbf{A} \mathbf{\Sigma}_{t|t} \mathbf{A}' + \mathbf{Q} \mathbf{Q}' - \mathbf{\Sigma}_{t+1|t})) + \lambda_t' \text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t})] \end{aligned}$$

But notice that

$$\lambda_t' \text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}) = \text{tr}(\text{diag}(\lambda_t) \text{diag}(\text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}))).$$

where $\text{diag}(\cdot)$ is the operator that places a vector on the diagonal of a square matrix with zeros elsewhere. Finally notice that for $\mathbf{\Sigma}_{t|t}$ such that $\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}$ is symmetric and positive semidefinite, there exists an orthonormal basis \mathbf{U}_t such that

$$\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} = \mathbf{U}_t \text{diag}(\text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t})) \mathbf{U}_t'$$

Now, let $\mathbf{\Lambda}_t \equiv \mathbf{U}_t \text{diag}(\lambda_t) \mathbf{U}'_t$ and observe that

$$\text{tr}(\text{diag}(\lambda_t) \text{diag}(\text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}))) = \text{tr}(\mathbf{\Lambda}_t(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t})).$$

Moreover, note that complementary slackness for this constraint requires:

$$\begin{aligned} \lambda'_t \text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t-1}) &= 0, \lambda_t \geq 0, \text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t-1}) \geq 0 \\ \Leftrightarrow \text{diag}(\lambda_t) \text{diag}(\text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t})) &= 0, \text{diag}(\lambda_t) \succeq 0, \mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq 0 \\ \Leftrightarrow \mathbf{\Lambda}_t(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}) &= 0, \mathbf{\Lambda}_t \succeq 0, \mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq 0 \end{aligned}$$

re-writing the Lagrangian we get:

$$\begin{aligned} L_0 &= \max_{\{\mathbf{\Sigma}_{t|t} \in \mathcal{S}_+^n\}_{t \geq 0}} \frac{1}{2} \sum_{t=0}^{\infty} \beta^t [-\text{tr}(\mathbf{\Sigma}_{t|t} \mathbf{\Omega}) - \omega \ln(|\mathbf{\Sigma}_{t|t-1}|) + \omega \ln(|\mathbf{\Sigma}_{t|t}|) \\ &\quad - \text{tr}(\mathbf{\Gamma}_t(\mathbf{A} \mathbf{\Sigma}_{t|t} \mathbf{A}' + \mathbf{Q} \mathbf{Q}' - \mathbf{\Sigma}_{t+1|t})) + \text{tr}(\mathbf{\Lambda}_t(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}))] \end{aligned}$$

Differentiating with respect to $\mathbf{\Sigma}_{t|t}$ and $\mathbf{\Sigma}_{t|t-1}$ and imposing symmetry we have

$$\mathbf{\Omega} - \omega \mathbf{\Sigma}_{t|t}^{-1} + \mathbf{A}' \mathbf{\Gamma}_t \mathbf{A} + \mathbf{\Lambda}_t = 0 \quad (\text{w.r.t. } \mathbf{\Sigma}_{t|t})$$

$$\omega \beta \mathbf{\Sigma}_{t+1|t}^{-1} - \mathbf{\Gamma}_t - \beta \mathbf{\Lambda}_{t+1} = 0 \quad (\text{w.r.t. } \mathbf{\Sigma}_{t+1|t})$$

Notice that the assumptions of the Theorem imply that we can invert the prior matrices because:

$$\mathbf{\Sigma}_{t|t-1} \succ 0 \Rightarrow \mathbf{\Sigma}_{t+1|t} = \mathbf{A} \mathbf{\Sigma}_{t|t} \mathbf{A} + \mathbf{Q} \mathbf{Q}' \succ 0, \forall t \geq 0$$

To see why, suppose otherwise, then $\exists \mathbf{w} \neq 0$ such that

$$\mathbf{w}'(\mathbf{A} \mathbf{\Sigma}_{t|t} \mathbf{A}' + \mathbf{Q} \mathbf{Q}') \mathbf{w} = 0 \Leftrightarrow \mathbf{w}' \mathbf{A} \mathbf{\Sigma}_{t|t} \mathbf{A}' \mathbf{w} = \mathbf{w}' \mathbf{Q} \mathbf{Q}' \mathbf{w} = 0$$

Thus,

$$(\mathbf{\Sigma}_{t|t}^{\frac{1}{2}} \mathbf{A}' \mathbf{w} = 0) \wedge (\mathbf{Q}' \mathbf{w} = 0) \quad (\text{A.2})$$

Moreover, note that $\mathbf{\Sigma}_{t|t}$ is invertible because the cost of attention has to be finite:

$$\ln \left(\frac{\det(\mathbf{\Sigma}_{t|t-1})}{\det(\mathbf{\Sigma}_{t|t})} \right) < \infty \Rightarrow \det(\mathbf{\Sigma}_{t|t}) > 0 \quad (\text{A.3})$$

Hence, $\Sigma_{t|t}^{\frac{1}{2}}$ is invertible, and we can write the above equations as:

$$(\mathbf{A}\mathbf{A}'\mathbf{w} = 0) \wedge (\mathbf{Q}\mathbf{Q}'\mathbf{w} = 0) \Rightarrow (\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}')\mathbf{w} = 0 \quad (\text{A.4})$$

but since $\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'$ is invertible by assumption, this implies that $\mathbf{w} = 0$ which is a contradiction with $\mathbf{w} \neq 0$. Thus, $\Sigma_{t+1|t}$ has to be invertible as well.

Now, replacing for Γ_t in the first order conditions we get the conditions in the theorem. Moreover, we have a terminal optimality condition that requires:

$$\lim_{T \rightarrow \infty} \beta^T \text{tr}(\Gamma_T \Sigma_{T+1|T}) \geq 0 \Leftrightarrow \lim_{T \rightarrow \infty} \beta^{T+1} \text{tr}(\Lambda_{T+1} \Sigma_{T+1|T}) \leq 0 \quad (\text{TVC})$$

Since both Λ_T and $\Sigma_{T+1|T}$ are positive semidefinite, we also have $\text{tr}(\Lambda_{T+1} \Sigma_{T+1|T}) \geq 0$. Thus, TVC becomes:

$$\lim_{T \rightarrow \infty} \beta^{T+1} \text{tr}(\Lambda_{T+1} \Sigma_{T+1|T}) = 0$$

■

Proof of Theorem 1. From the FOC in Proposition 2 observe that

$$\omega \Sigma_{t|t}^{-1} = \Omega_t + \Lambda_t \Rightarrow \Sigma_{t|t-1} - \Sigma_{t|t} = \Sigma_{t|t-1} - \omega(\Omega_t + \Lambda_t)^{-1}. \quad (\text{A.5})$$

For ease of notation let $\mathbf{X}_t \equiv \Sigma_{t|t-1} - \Sigma_{t|t}$. Multiplying the above equation by $\Omega_t + \Lambda_t$ from right we get

$$\mathbf{X}_t \Omega_t - \Sigma_{t|t-1} \Lambda_t = \Sigma_{t|t-1} \Omega_t - \omega \mathbf{I},$$

where we have imposed the complementarity slackness $\mathbf{X}_t \Lambda_t = 0$. Finally, multiply this equation by $\Sigma_{t|t-1}^{\frac{1}{2}}$ from right and $\Sigma_{t|t-1}^{-\frac{1}{2}}$ from left.²¹ We have

$$(\Sigma_{t|t-1}^{-\frac{1}{2}} \mathbf{X}_t \Sigma_{t|t-1}^{-\frac{1}{2}}) (\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}}) - \Sigma_{t|t-1}^{\frac{1}{2}} \Lambda_t \Sigma_{t|t-1}^{\frac{1}{2}} = \Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}} - \omega \mathbf{I}$$

Where $\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}} = \mathbf{U}_t \mathbf{D}_t \mathbf{U}_t'$ is the spectral decomposition stated in the Theorem.

²¹ $\Sigma_{t|t-1}^{\frac{1}{2}}$ exists since $\Sigma_{t|t-1}$ is positive semidefinite and $\Sigma_{t|t-1}^{-\frac{1}{2}}$ exists since we assumed that the initial prior is strictly positive definite.

Now, for ease of notation let

$$\hat{\mathbf{X}}_t \equiv \mathbf{U}'_t \boldsymbol{\Sigma}_{t|t-1}^{-\frac{1}{2}} \mathbf{X}_t \boldsymbol{\Sigma}_{t|t-1}^{-\frac{1}{2}} \mathbf{U}_t \quad (\text{A.6})$$

$$\hat{\boldsymbol{\Lambda}}_t \equiv \mathbf{U}'_t \boldsymbol{\Sigma}_{t|t-1}^{\frac{1}{2}} \boldsymbol{\Lambda}_t \boldsymbol{\Sigma}_{t|t-1}^{\frac{1}{2}} \mathbf{U}_t \quad (\text{A.7})$$

Plugging these in along with the spectral decomposition stated in the Theorem we have

$$\hat{\mathbf{X}}_t \mathbf{D}_t - \hat{\boldsymbol{\Lambda}}_t = \mathbf{D}_t - \omega \mathbf{I} \quad (\text{A.8})$$

Now, notice that \mathbf{X}_t and $\boldsymbol{\Lambda}_t$ are simultaneously diagonalizable if and only if $\hat{\mathbf{X}}_t$ and $\hat{\boldsymbol{\Lambda}}_t$ are simultaneously diagonalizable. Combined with complementarity slackness, this implies $\hat{\boldsymbol{\Lambda}}_t \hat{\mathbf{X}}_t = \hat{\mathbf{X}}_t \hat{\boldsymbol{\Lambda}}_t = \mathbf{0}$. Similarly, note that \mathbf{X}_t and $\boldsymbol{\Lambda}_t$ are positive semidefinite if and only if $\hat{\mathbf{X}}_t$ and $\hat{\boldsymbol{\Lambda}}_t$ are positive semidefinite, respectively. So we need for two simultaneously diagonalizable symmetric positive semidefinite matrices $\hat{\boldsymbol{\Lambda}}_t$ and $\hat{\mathbf{X}}_t$ that solve Equation A.8.

It follows from these that both these matrices are diagonal. To see this, re-write the above equation as

$$(\hat{\mathbf{X}}_t - \mathbf{I}) \mathbf{D}_t = \hat{\boldsymbol{\Lambda}}_t - \omega \mathbf{I} \quad (\text{A.9})$$

Now, notice that $\hat{\mathbf{X}}_t - \mathbf{I}$ and $\hat{\boldsymbol{\Lambda}}_t - \omega \mathbf{I}$ are simultaneously diagonalizable. Let α denote this basis. We have

$$[\hat{\mathbf{X}}_t - \mathbf{I}]_\alpha [\mathbf{D}_t]_\alpha = [\hat{\boldsymbol{\Lambda}}_t - \omega \mathbf{I}]_\alpha$$

Note that in this equation, the right hand side is diagonal and the left hand side is the product of a diagonal matrix with $[\mathbf{D}_t]_\alpha$. Thus, $[\mathbf{D}_t]_\alpha$ has to be diagonal as well. This implies α is the identity basis and that $\hat{\boldsymbol{\Lambda}}_t$ and $\hat{\mathbf{X}}_t$ are diagonal matrices. Using complementarity slackness $\hat{\boldsymbol{\Lambda}}_t \hat{\mathbf{X}}_t = \mathbf{0}$, feasibility constraint $\hat{\mathbf{X}}_t \succeq \mathbf{0}$, and dual feasibility constraint $\hat{\boldsymbol{\Lambda}}_t \succeq \mathbf{0}$ it is straight forward to show that $\boldsymbol{\Lambda}_t$ is strictly positive for the eigenvalues (entries on the diagonal) of \mathbf{D}_t that are smaller than ω .

$$\hat{\boldsymbol{\Lambda}}_t = \text{Max}(\omega \mathbf{I} - \mathbf{D}_t, \mathbf{0}) \quad (\text{A.10})$$

Now, using Equation A.7 we get:

$$\boldsymbol{\Lambda}_t = \boldsymbol{\Sigma}_{t|t-1}^{-\frac{1}{2}} \mathbf{U}_t \text{Max}(\omega \mathbf{I} - \mathbf{D}_t, \mathbf{0}) \mathbf{U}'_t \boldsymbol{\Sigma}_{t|t-1}^{-\frac{1}{2}} \quad (\text{A.11})$$

Moreover, recall that $\omega \boldsymbol{\Sigma}_{t|t}^{-1} = \boldsymbol{\Omega}_t + \boldsymbol{\Lambda}_t$. Hence, plugging in the spectral decomposition

and the solution for Λ_t :

$$\begin{aligned}
\omega \Sigma_{t|t}^{-1} &= \Sigma_{t|t-1}^{-\frac{1}{2}} \mathbf{U}_t \mathbf{D}_t \mathbf{U}_t' \Sigma_{t|t-1}^{-\frac{1}{2}} + \Sigma_{t|t-1}^{-\frac{1}{2}} \mathbf{U}_t \text{Max}(\omega \mathbf{I} - \mathbf{D}_t, \mathbf{0}) \mathbf{U}_t' \Sigma_{t|t-1}^{-\frac{1}{2}} \\
&= \Sigma_{t|t-1}^{-\frac{1}{2}} \mathbf{U}_t \text{Max}(\omega \mathbf{I}, \mathbf{D}_t) \mathbf{U}_t' \Sigma_{t|t-1}^{-\frac{1}{2}} \\
&= \Sigma_{t|t-1}^{-\frac{1}{2}} \text{Max}(\Sigma_{t|t-1}^{\frac{1}{2}} \mathbf{\Omega}_t \Sigma_{t|t-1}^{\frac{1}{2}}, \omega) \Sigma_{t|t-1}^{-\frac{1}{2}}
\end{aligned} \tag{A.12}$$

Inverting this gives us the expression in the Theorem – the matrix is invertible because all eigenvalues are bounded below by ω . Moreover, using the definition of $\mathbf{\Omega}_t$ in the statement of the Theorem, and the expression for Λ_t in Equation A.11 we have:

$$\begin{aligned}
\mathbf{\Omega}_t &= \mathbf{\Omega} + \beta \mathbf{A}' (\omega \Sigma_{t+1|t}^{-1} - \Lambda_{t+1}) \mathbf{A} \\
&= \mathbf{\Omega} + \beta \mathbf{A}' \Sigma_{t+1|t}^{-\frac{1}{2}} (\omega \mathbf{I} - \mathbf{U}_t \text{Max}(\omega \mathbf{I} - \mathbf{D}_t, \mathbf{0})) \Sigma_{t+1|t}^{-\frac{1}{2}} \mathbf{A} \\
&= \mathbf{\Omega} + \beta \mathbf{A}' \Sigma_{t+1|t}^{-\frac{1}{2}} \mathbf{U}_t \text{Min}(\mathbf{D}_t, \omega \mathbf{I}) \mathbf{U}_t' \Sigma_{t+1|t}^{-\frac{1}{2}} \mathbf{A} \\
&= \mathbf{\Omega} + \beta \mathbf{A}' \Sigma_{t+1|t}^{-\frac{1}{2}} \text{Min}(\Sigma_{t+1|t}^{\frac{1}{2}} \mathbf{\Omega}_{t+1} \Sigma_{t+1|t}^{\frac{1}{2}}, \omega) \Sigma_{t+1|t}^{-\frac{1}{2}} \mathbf{A}
\end{aligned} \tag{A.13}$$

■

Proof of Theorem 2. The upper bound m directly follows from Lemma 1. Recall from part 2 of Lemma 2 that when $\{\vec{x}_t\}$ is a Markov process, then $\vec{a}_t \perp X^{t-1} | (a^{t-1}, \vec{x}^t)$. Moreover, since actions are Gaussian in the LQG setting, we can then decompose the innovation to the action of the agent at time t as

$$\vec{a}_t - \mathbb{E}[\vec{a}_t | a^{t-1}] = \mathbf{Y}_t' (\vec{x}_t - \mathbb{E}[\vec{x}_t | a^{t-1}]) + \vec{z}_t, \vec{z}_t \perp (X^t, a^{t-1}) \tag{A.14}$$

where $\vec{z}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_{z,t})$ is the agent's rational inattention error – it is mean zero and Gaussian. It just remains to characterize \mathbf{Y}_t and the covariance matrix of \vec{z}_t . Now, since actions are sufficient for the signals of the agent at time t , we have

$$\begin{aligned}
\mathbb{E}[\vec{x}_t | a^t] &= \mathbb{E}[\vec{x}_t | a^{t-1}] + \mathbf{K}_t (\vec{a}_t - \mathbb{E}[\vec{a}_t | a^{t-1}]) \\
&= \mathbb{E}[\vec{x}_t | a^{t-1}] + \mathbf{K}_t \mathbf{Y}_t' (\vec{x}_t - \mathbb{E}[\vec{x}_t | a^{t-1}]) + \mathbf{K}_t \vec{z}_t
\end{aligned} \tag{A.15}$$

where $\mathbf{K}_t \equiv \Sigma_{t|t-1} \mathbf{Y}_t (\mathbf{Y}_t' \Sigma_{t|t-1} \mathbf{Y}_t + \Sigma_{z,t})^{-1}$ is the implied Kalman gain by the decomposition. The number of the signals that span the agent's posterior is therefore the rank of this Kalman gain matrix. Moreover, note that if the decomposition is of the optimal actions,

then the implied posterior covariance should coincide with the solution:

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \mathbf{K}_t \mathbf{Y}'_t \Sigma_{t|t-1} \Rightarrow \mathbf{K}_t \mathbf{Y}'_t = \mathbf{I} - \Sigma_{t|t} \Sigma_{t|t-1}^{-1} \quad (\text{A.16})$$

Let $\mathbf{U}_t \mathbf{D}_t \mathbf{U}'_t$ denote the spectral decomposition of $\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}}$. Then, using Theorem 1, we have:

$$\begin{aligned} \mathbf{K}_t \mathbf{Y}'_t &= \Sigma_{t|t-1}^{\frac{1}{2}} \mathbf{U}_t (\mathbf{I} - \omega \text{Max}(\mathbf{D}_t, \omega)^{-1}) \mathbf{U}'_t \Sigma_{t|t-1}^{-\frac{1}{2}} \\ &= \sum_{i=1}^n \max(0, 1 - \frac{\omega}{d_{i,t}}) \Sigma_{t|t-1} \mathbf{y}_{i,t} \mathbf{y}'_{i,t} \end{aligned} \quad (\text{A.17})$$

where $d_{i,t}$ is the i 'th eigenvalue in \mathbf{D}_t and $\mathbf{y}_{i,t}$ is the i 'th column of the matrix $\Sigma_{t|t-1}^{-\frac{1}{2}} \mathbf{U}_t$. Notice that for any i , $\mathbf{y}_{i,t} = \Sigma_{t|t-1}^{-\frac{1}{2}} \mathbf{u}_{i,t}$ is an eigenvector for $\Omega_t \Sigma_{t|t-1}$:

$$\Omega_t \Sigma_{t|t-1} \mathbf{y}_{i,t} = \Sigma_{t|t-1}^{-\frac{1}{2}} (\Sigma_{t|t-1}^{\frac{1}{2}} \Omega_t \Sigma_{t|t-1}^{\frac{1}{2}}) \mathbf{u}_{i,t} = d_{i,t} \Sigma_{t|t-1}^{-\frac{1}{2}} \mathbf{u}_{i,t} = d_{i,t} \mathbf{y}_{i,t} \quad (\text{A.18})$$

Moreover, note that only eigenvectors with eigenvalue larger than ω get a positive weight in spanning $\mathbf{K}_t \mathbf{Y}'_t$, meaning that we can exclude eigenvectors associated with $d_{i,t} \leq \omega$. Formally, let \mathbf{Y}_t^+ be a matrix whose columns are columns of \mathbf{Y}_t whose eigenvalue is larger than ω . Let \mathbf{D}_t^+ be the diagonal matrix with these eigenvalues, and let $\Sigma_{z,t}^+$ be the corresponding principal minor of $\Sigma_{z,t}$. Then,

$$\begin{aligned} \mathbf{Y}_t (\mathbf{Y}'_t \Sigma_{t|t-1} \mathbf{Y}_t + \Sigma_{z,t})^{-1} \mathbf{Y}'_t &= \sum_{i=1}^n \max(0, 1 - \frac{\omega}{d_{i,t}}) \mathbf{y}_{i,t} \mathbf{y}'_{i,t} \\ &= \sum_{d_{i,t} > \omega} (1 - \frac{\omega}{d_{i,t}}) \mathbf{y}_{i,t} \mathbf{y}'_{i,t} \\ &= \mathbf{Y}_t^+ (\mathbf{Y}_t^{+\prime} \Sigma_{t|t-1} \mathbf{Y}_t^+ + \Sigma_{z,t}^+)^{-1} \mathbf{Y}_t^{+\prime} \end{aligned} \quad (\text{A.19})$$

Now we just need $\Sigma_{z,t}^+$ to fully characterize the signals. For this, note that $\forall i, j$:

$$\mathbf{y}'_{i,t} \Sigma_{t|t-1} \mathbf{y}_{j,t} = \begin{cases} \mathbf{u}'_{i,t} \mathbf{u}_{i,t} = 1 & \text{if } i = j \\ \mathbf{u}'_{i,t} \mathbf{u}_{j,t} = 0 & \text{if } i \neq j \end{cases} \quad (\text{A.20})$$

Thus, $\mathbf{Y}_t^{+\prime} \Sigma_{t|t-1} \mathbf{Y}_t^+ = \mathbf{I}_k$ where \mathbf{I}_k is the k -dimensional identity matrix with k being the number of eigenvalues in \mathbf{D}_t that are larger than ω . Combining this with Equation A.16

we have:

$$\begin{aligned}
\Sigma_{t|t-1} - \Sigma_{t|t} &= \Sigma_{t|t-1} \mathbf{Y}_t^+ (\mathbf{Y}_t^{+'} \Sigma_{t|t-1} \mathbf{Y}_t^+ + \Sigma_{z,t}^+)^{-1} \mathbf{Y}_t^{+'} \Sigma_{t|t-1} \\
&\Rightarrow \mathbf{Y}_t^{+'} (\Sigma_{t|t-1} - \Sigma_{t|t}) \mathbf{Y}_t^+ = \mathbf{Y}_t^{+'} \Sigma_{t|t-1} \mathbf{Y}_t^+ (\mathbf{Y}_t^{+'} \Sigma_{t|t-1} \mathbf{Y}_t^+ + \Sigma_{z,t}^+)^{-1} \mathbf{Y}_t^{+'} \Sigma_{t|t-1} \mathbf{Y}_t^+ \\
(\mathbf{Y}_t^{+'} \Sigma_{t|t-1} \mathbf{Y}_t^+) &\Rightarrow \Sigma_{z,t}^+ = (\mathbf{I}_k - \mathbf{Y}_t^{+'} \Sigma_{t|t} \mathbf{Y}_t^+)^{-1} - \mathbf{I}_k
\end{aligned} \tag{A.21}$$

Plugging in for $\Sigma_{t|t}$ from the **policy function** we have:

$$\Sigma_{z,t}^+ = (\mathbf{I}_k - \omega (\mathbf{D}_t^+)^{-1})^{-1} - \mathbf{I}_k = (\omega^{-1} \mathbf{D}_t^+ - \mathbf{I}_k)^{-1} \tag{A.22}$$

Note that $\Sigma_{z,t}^+$ is diagonal where the i 'th diagonal entry is $\frac{1}{\omega^{-1} d_{i,t} - 1}$.

Thus, the agent's posterior is spanned by the following k signals:

$$\vec{s}_t = \mathbf{Y}^+ \vec{x}_t + \vec{z}_t, \mathbf{Y}_t^{+'} \Sigma_{t|t-1} \mathbf{Y}_t^+ = \mathbf{I}_k, \vec{z}_t \sim \mathcal{N}(\mathbf{0}, (\omega^{-1} \mathbf{D}_t^+ - \mathbf{I}_k)^{-1}) \tag{A.23}$$

■

Proof of Proposition 3. From the proof of the last Theorem, recall that the Kalman gain for predicting the state is given by

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \mathbf{K}_t \mathbf{Y}_t^{+'} \Sigma_{t|t-1} \Rightarrow \mathbf{K}_t \mathbf{Y}_t^{+'} = \mathbf{I} - \Sigma_{t|t} \Sigma_{t|t-1}^{-1} \tag{A.24}$$

Plugging this into Equation A.15, multiplying it by \mathbf{H}' from left, and substituting $\vec{a}_t = \mathbf{H}' \mathbb{E}[\vec{x}|a^t]$ we have:

$$\vec{a}_t - \mathbb{E}[\vec{a}_t | a^{t-1}] = \mathbf{H}' (\mathbf{I} - \Sigma_{t|t} \Sigma_{t|t-1}^{-1}) (\vec{x}_t - \mathbb{E}[\vec{x}_t | a^{t-1}]) + \mathbf{H}' \mathbf{K}_t \vec{z}_t \tag{A.25}$$

Notice that this implies $(\mathbf{H}' \mathbf{K}_t - \mathbf{I}) \vec{z}_t = 0$. Now, taking the variance of the two sides we get

$$\begin{aligned}
\text{var}(\vec{a}_t | a^{t-1}) &= \mathbf{H}' (\Sigma_{t|t-1} - \Sigma_{t|t}) \mathbf{H} \\
&= \mathbf{H}' (\mathbf{I} - \Sigma_{t|t} \Sigma_{t|t-1}^{-1}) \Sigma_{t|t-1} (\mathbf{I} - \Sigma_{t|t-1}^{-1} \Sigma_{t|t}) \mathbf{H} + \Sigma_{z,t}.
\end{aligned} \tag{A.26}$$

where the first line follows from leaving $\mathbf{H}' \mathbf{K}_t$ as is, and the second line follows from plugging in $\mathbf{H}' \mathbf{K}_t \vec{z}_t = \vec{z}_t$. Solving for $\Sigma_{z,t}$ we get:

$$\Sigma_{z,t} = \mathbf{H}' (\Sigma_{t|t} - \Sigma_{t|t} \Sigma_{t|t-1}^{-1} \Sigma_{t|t}) \mathbf{H} \tag{A.27}$$

■

Proof of Lemma 4. The log-linearized Euler equation from the household side is

$$i_t = \rho + \mathbb{E}_t[\Delta q_{t+1}] \quad (\text{A.28})$$

Combining this with the monetary policy rule, we have

$$\Delta q_t = \phi^{-1} \mathbb{E}_t^f[\Delta q_{t+1}] + \frac{\sigma_u}{\phi} u_t \quad (\text{A.29})$$

Iterating this forward and noting that $\lim_{h \rightarrow \infty} \phi^{-h} \mathbb{E}_t^f[\Delta q_{t+h}] = 0$ due to $\phi > 1$, we get the result in the Lemma. ■

Proof of Proposition 4. Part 1. For ease of notation we drop the firm index i in the proof. The FOC in Proposition 2 in this case reduces to

$$\lambda_t = 1 - \theta + \frac{\omega}{\sigma_{t|t}^2} - \frac{\beta\omega}{\sigma_{t+1|t}^2} + \beta\lambda_{t+1} \quad (\text{A.30})$$

Since the problem is deterministic and the state variables grows with time when the constraint is binding, then there is a t after which the constraint does not bind. Given such a t , suppose $\lambda_t = \lambda_{t+1} = 0$, then noting that $\sigma_{t+1|t}^2 = \sigma_{t|t}^2 + \sigma_u^2 \phi^{-2}$, the FOC becomes:

$$\sigma_{t|t}^4 + \left[\frac{\sigma_u^2}{\phi^2} - (1 - \beta) \frac{\omega}{\theta - 1} \right] \sigma_{t|t}^2 - \frac{\omega}{\theta - 1} \frac{\sigma_u^2}{\phi^2} = 0 \quad (\text{A.31})$$

Note that given the values of parameters, this equation does not depend on any other variable than $\sigma_{t|t}^2$ (in particular it is independent of the state $\sigma_{t|t-1}^2$). Hence, for any t , if $\lambda_t = 0$, then the $\sigma_{t|t}^2 = \underline{\sigma}^2$, where $\underline{\sigma}^2$ is the positive root of the equation above. However, for this solution to be admissible it has to satisfy the no-forgetting constraint which holds only if $\underline{\sigma}^2 \leq \sigma_{t|t-1}^2$. Thus,

$$\sigma_{t|t}^2 = \min\{\sigma_{t|t-1}^2, \underline{\sigma}^2\}. \quad (\text{A.32})$$

Part 2. The Kalman-gain can be derived from the relationship between prior and posterior uncertainty:

$$\sigma_{i,t|t}^2 = (1 - \kappa_{i,t}) \sigma_{i,t|t-1}^2 \Rightarrow \kappa_{i,t} = 1 - \min\left\{1, \frac{\underline{\sigma}^2}{\sigma_{i,t|t-1}^2}\right\} = \max\left\{0, 1 - \frac{\underline{\sigma}^2}{\sigma_{i,t|t-1}^2}\right\}. \quad (\text{A.33})$$

■

Proof of Corollary 1. Follows from the characterization of $\underline{\sigma}^2$ in Proposition 4. ■

Proof of Proposition 5. Part 1. Recall from the proof of Proposition 4 that

$$p_{i,t} = p_{i,t-1} + \kappa_{i,t}(q_t - p_{i,t-1} + e_{i,t}) \quad (\text{A.34})$$

Aggregating this up and imposing $\kappa_{i,t} = \kappa_t$ since all firms start from the same uncertainty and solve the same problem, we get:

$$\pi_t = \frac{\kappa_t}{1 - \kappa_t} y_t. \quad (\text{A.35})$$

Plug in κ_t from Equation A.33 to get the expression for the slope of the Phillips curve.

Part 2. In this case the Phillips curve is flat so it immediately follows that $\pi_t = 0$. Moreover, since $\pi_t + \Delta y_t = \Delta q_t$, plugging in $\pi_t = 0$, we get $y_t = y_{t-1} + \Delta q_t$.

Part 3. If $\sigma_{T|T-1}^2 \geq \underline{\sigma}^2$, then $\forall t \geq T + 1$, $\sigma_{t|t}^2 = \underline{\sigma}^2$ and $\sigma_{t|t-1}^2 = \underline{\sigma}^2 + \sigma_u^2 \phi^{-2}$. Hence, for $t \geq T + 1$, the Phillips curve is given by $\pi_t = \frac{\kappa}{1 - \kappa} y_t$. Combining this with $\pi_t + \Delta y_t = \Delta q_t$ we get the dynamics stated in the Proposition. ■

Proof of Corollary 2. The jump to the new steady state follows from the result in Corollary 1 that $\underline{\sigma}^2$ increases with $\frac{\sigma_u}{\phi}$. The comparative statics follow from the fact that κ is the positive root of

$$\beta \kappa^2 + (1 - \beta + \xi) \kappa - \xi = 0 \quad (\text{A.36})$$

where $\xi \equiv \frac{\sigma_u^2(\theta-1)}{\phi^2 \omega}$. It suffices to observe that κ decreases with ξ , and ξ increases with $\frac{\sigma_u}{\phi}$. ■

Proof of Corollary 3. The transition to the new steady state follows from the fact that reservation uncertainty increases with a positive shock to $\underline{\sigma}^2$. The policy function of the firm in Proposition 4 that firms would wait until their uncertainty reaches this new level. Comparative statics in the steady state follow directly from Corollary 1. ■