Dynamic Rational Inattention and the Phillips Curve[*†]

Hassan Afrouzi[‡]
Columbia University

Choongryul Yang[§]
Federal Reserve Board

First Draft: April, 2017
This Draft: November, 2020

[Software Documentation] [Code Repository] [Interactive Notebooks]

Abstract

We develop a fast, tractable, and robust method for solving the transition path of dynamic rational inattention problems (DRIPs) in LQG settings. As an application of our general framework, we develop an attention-driven analytical theory of dynamic pricing in which the Phillips curve slope is endogenous to systematic aspects of monetary policy. In our model, when the monetary authority is more committed to stabilizing nominal variables, rationally inattentive firms pay less attention to changes in their input costs, which leads to a flatter Phillips curve and more anchored inflation expectations. This effect, however, is not symmetric. A more dovish monetary policy flattens the Phillips curve in the short-run but generates a steeper Phillips curve in the long-run. In a quantitative exercise, we calibrate our general equilibrium model with TFP and monetary policy shocks to post-Volcker U.S. data and find that our mechanism quantifies a 75% decline in the slope of the Phillips curve in the post-Volcker period, relative to the period before it.

JEL Classification: D83, D84, E03, E58
Keywords: Rational Inattention, Dynamic Information Acquisition, Phillips Curve

[*]We are grateful to Saroj Bhattarai and Olivier Coibion for their guidance and support as well as to George-Marios Angeletos and Eric Sims for thoughtful discussions of this paper. We would also like to thank Miguel Acosta, Mark Dean, Xavier Gabaix, Yuriy Gorodnichenko, Jennifer La'O, John Leahy, Yueran Ma, Filip Matějka, Emi Nakamura, Kris Nimark, Jón Steinsson, Luminita Stevens, Laura Veldkamp, Venky Venkateswaran, Mirko Wiederholt, Michael Woodford and seminar participants at 2018 and 2020 ASSA Meetings, Columbia University, SED Mexico City Conference, University of Texas at Austin, Federal Reserve Bank of Chicago, Cornell University, City University of New York, University of Wisconsin and NBER Summer Institute for helpful comments. The views expressed herein are those of the authors, and do not reflect the views of the Federal Reserve Board or any person associated with the Federal Reserve System.

[†]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 Mexico City.

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

[§]Federal Reserve Board of Governors, 20th Street and Constitution Avenue NW, Washington, DC 20551, U.S.A. Email: choongryul.yang@frb.gov.

"The relationship between the slack in the economy or unemployment and inflation was a strong one 50 years ago...and has gone away" –Jerome Powell (2019)

1 Introduction

A recent growing literature documents that the slope of the Phillips curve has flattened during the last few decades.[1] Since the trade-off between inflation and unemployment is at the core of the monetary theory, understanding the sources of this change is important for studying the impact of monetary policy.

While benchmark New Keynesian models relate this flattening to changes in the model's structural parameters, in an analytical framework with rationally inattentive firms, we show that the Phillips curve slope is endogenous to the conduct of monetary policy. In our model, when monetary authority puts a larger weight on stabilizing the nominal variables, firms endogenously choose to pay less attention to changes in their input costs. Accordingly, when monetary policy is more stabilizing, prices are less sensitive to the slack in the economy, and the Phillips curve is flatter. Therefore, our theory suggests that the decline in the slope of the Phillips curve can be explained, at least partially, by the more *hawkish* monetary policy adopted at the beginning of the Great Moderation.[2]

This effect, however, is not symmetric in our model. While more hawkish monetary policy flattens the Phillips curve, a more dovish monetary policy completely flattens the Phillips curve in the short-run but eventually leads to a steeper Phillips curve in the long-run. The key to this asymmetry lies in the dynamic incentives in information acquisition. In our model, forward-looking firms learn about their input costs' persistent changes and invest in a stock of knowledge about these processes. When monetary policy becomes more dovish, firms suddenly find themselves in a more uncertain environment where their stock of knowledge depreciates faster. Hence, a more dovish monetary policy decreases the net present value of knowledge, and *crowds out* firms' information acquisition in the short-run, a period during which prices are not sensitive to changes in input costs and the Phillips curve is *completely flat*. However, this effect dissipates as firms' uncertainty about their input costs grow, and they eventually restart paying attention to their costs once their knowledge has depreciated enough. In this new regime, firms have a lower stock of knowledge, but they acquire information at a higher rate. The higher rate

[1] See, for instance, Coibion and Gorodnichenko (2015b); Blanchard (2016); Bullard (2018); Hooper, Mishkin and Sufi (2019); Del Negro, Lenza, Primiceri and Tambalotti (2020).

[2] See Clarida, Galí and Gertler (2000); Coibion and Gorodnichenko (2011) for evidence on more hawkish monetary policy in the post-Volcker period.

of information acquisition makes prices more sensitive to changes in input costs and leads to a *steeper* Phillips curve relative to the previous regime.

Furthermore, our model provides an endogenous explanation for how monetary policy's conduct affects the anchoring of inflation expectations. Since attention is endogenous, firms' expectations of inflation are less sensitive to short-run fluctuations and comove less with the output gap when monetary policy is more hawkish. Moreover, the asymmetries outlined above also appear in how inflation expectations are formed. When monetary policy becomes more dovish, the pass-through of changes in input costs to firms' inflation expectations is zero as long as the Phillips curve is flat. It is only in the long-run, when firms restart paying attention to input costs, that inflation expectations become more sensitive to changes in those costs.

Methodological Contributions. Our theory of the Phillips curve is an application of a tractable method that we develop for solving dynamic rational inattention problems (DRIPs) with multiple shocks and actions in linear quadratic Gaussian (LQG) settings. Our first methodological contribution in this area is to formulate and analytically characterize the full transition dynamics of DRIPs. We show that the transition dynamics in DRIPs are characterized by inaction regions for the decision maker's uncertainty in different state dimensions.

Our second contribution is that we use our theoretical results to develop a computational toolbox that decreases the solution times for these problems by several orders of magnitude. Furthermore, as far as we are aware, we provide the first solution method that characterizes the transition dynamics of DRIPs. To demonstrate our computational toolbox's accuracy and efficiency, we have replicated three canonical papers (Maćkowiak and Wiederholt, 2009a; Sims, 2010; Maćkowiak, Matějka and Wiederholt, 2018a) that use three different solution methods. Our algorithm provides equally accurate solutions but is significantly faster than other available methods because it utilizes our theoretical findings.[3] A summary of our computing times are reported in Table 1. Our computational toolbox is available for public use as the `DRIPs.jl` Julia package.[4] Finally, all examples and replications are also available as interactive Jupyter notebooks that are accessible online with no software requirements.[5]

[3] Our replications of Sims (2010); Maćkowiak and Wiederholt (2009a); Maćkowiak et al. (2018a) is described in Section 2.4, Appendix B.1 and Appendix B.2, respectively.

[4] Link: `https://www.afrouzi.com/DRIPs.jl/dev/`

[5] Link: `https://mybinder.org/v2/gh/afrouzi/DRIPs.jl/binder?filepath=examples`

Quantitative Results. Our final contribution is to test the quantitative relevance of our proposed mechanism for the change in the slope of the Phillips curve. Our quantitative exercise uses our computational toolbox to solve and calibrate a dynamic general equilibrium version of our rational inattention model with monetary policy and supply shocks. To calibrate our model, we estimate a Taylor rule for the post-Volcker era and target the variance-covariance structure of output and inflation in this period.

To assess the out-of-sample fit of our model, we perform an exercise in the spirit of Maćkowiak and Wiederholt (2015). In particular, in our calibrated model, we replace the post-Volcker Taylor rule of monetary policy with an estimated Taylor rule for the pre-Volcker period and find that our model quantitatively matches the higher variance of inflation and GDP in the pre-Volcker era as non-targeted moments.

Our main empirical exercise directly assesses whether our proposed mechanism can explain the decline in the Phillips curve slope. To do so, we simulate data from our calibrated model using our pre- and post-Volcker monetary policy rule estimates and estimate the implied slope of the Phillips curve in both samples. We find that our model can explain up to a 75% decline in the Phillips curve slope in the post-Volcker period.

Related Literature. Dynamic rational inattention models have long been applied to different settings in macroeconomics.[6] We contribute to a subset of this literature that has laid the ground for solving dynamic rational inattention problems in LQG settings (Sims, 2003; Maćkowiak, Matějka and Wiederholt, 2018a; Fulton, 2018). These papers make two simplifying assumptions that we depart from: (1) they abstract away from transition dynamics, and (2) they solve for the long-run steady-state information structure that is independent of time and state. However, two papers study the same problem as we do and are the closest to our study. The first paper is Sims (2010) who formulates the dynamic rational inattention problem on the transition path, but only provides solutions for two special cases: first, an example with two shocks and one action, and second, a first-order condition to the general problem assuming that the solution is interior. The second paper is Miao, Wu and Young (2020), who also study the same problem as we do, but provide an approximate solution to the steady-state information structure, which they call the

[6]See, 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, Nie and Young (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); Van Nieuwerburgh and Veldkamp (2010) for asset pricing; Mondria and Wu (2010) for home bias; and Ilut and Valchev (2017) for imperfect problem solving. See also Angeletos and Lian (2016); Maćkowiak, Matějka and Wiederholt (2018b).

Golden rule approximation.

We make two main contributions relative to this literature. (1) We provide an analytical characterization of the solution to the general problem, both on the transition path and in the steady-state. Our solution goes beyond the method in Sims (2010) in that we fully characterize the optimal information structure, taking corner solutions into account. We show that these cases are not rare and arise under fairly general circumstances—most notably when the number of actions is smaller than the number of shocks.[7] Relative to Miao et al. (2020), we go beyond studying the steady-state information structure and characterize transition dynamics as well. Moreover, our solution does not restrict us to the Golden rule approximation but embeds it as a special case. (2) Our second contribution is to provide a fast and robust algorithm and a software package that utilize our theoretical results. As far as we are aware, our solution method is the first one that allows for solving the transition dynamics of these problems without value function iteration, which significantly improves computing times. For instance, Miao et al. (2020) report that it takes 18 minutes to solve for Sims (2010)'s example using value function iteration and 3 seconds to solve for the Golden rule approximation of the solution. It takes our algorithm 7.9×10^{-4} seconds to solve for the steady-state as well as the transition dynamics and 1.6×10^{-4} seconds to solve for the Golden rule approximation of the solution.

Our attention-driven theory of the Phillips curve is motivated by the evidence for the flattening of the Phillips curve in the last few decades (Blanchard, 2016; Bullard, 2018; Hooper et al., 2019; Del Negro et al., 2020).[8] Another literature that motivates our model provides evidence for the information rigidities that economic agents exhibit in forming their expectations.[9]

We also contribute to the literature that considers how imperfect information affects the Phillips curve (Lucas, 1972; Mankiw and Reis, 2002; Woodford, 2003; Nimark, 2008;

[7]These corner solutions arise when the agent values information in a particular dimension of the state less than the cost of acquiring it. Similar corner solutions arise in Van Nieuwerburgh and Veldkamp (2010) who refer to these as no-forgetting constraints.

[8]A recent study by Hazell, Herreño, Nakamura and Steinsson (2020) also documents that the Phillips curve is flatter in the post-Volcker era, but once viewed through the lens of benchmark New Keynesian models, the slope implied by their estimates is so small that the flattening is irrelevant. Our model has different implications for their estimates as we do not rely on the benchmark New Keynesian models. According to our model, the benchmark New Keynesian Phillips curve is "too" forward-looking, which is why the implied slope in Hazell et al. (2020) is small.

[9]For recent progress in this literature, see for instance, Kumar, Afrouzi, Coibion and Gorodnichenko (2015); Coibion and Gorodnichenko (2015a); Ryngaert (2017); Coibion, Gorodnichenko and Ropele (2018); Roth and Wohlfart (2018); Gaglianone, Giacomini, Issler and Skreta (2019); Angeletos, Huo and Sastry (2020) for survey evidence, and Khaw, Stevens and Woodford (2017); Khaw and Zorrilla (2018); Landier, Ma and Thesmar (2019) for experimental evidence.

Angeletos and La'O, 2009; Angeletos and Huo, 2018).[10] Our main departure is to derive a Phillips curve in a model with rational inattention and study how *monetary policy* shapes and alters the incentives in information acquisition of firms. Specifically, a notable implication of our model is the different short-run and long-run implications of changes in monetary policy for the slope of the Phillips curve.

Finally, while we provide an attention-based theory for the Phillips curve slope, a series of alternative explanations have been proposed by other recent studies. These explanations include non-linearities in the slope of the Phillips curve (Kumar and Orrenius, 2016; Babb and Detmeister, 2017; Hooper et al., 2019), identification issues due to optimal monetary policy (McLeay and Tenreyro, 2020), changes in the input-output structure of the economy (Rubbo, 2020), or changes in price stickiness due to pursuit of price stability by the central bank (L'Huillier and Zame, 2020). Instead, our Phillips curve model focuses on how the conduct of monetary policy affects the attention allocation of firms.

Layout. 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. We conclude Section 2 by outlining our solution method and studying the transition dynamics of attention in an extension of the example from Sims (2010). In Section 3, we outline our attention driven theory of the Phillips curve with analytical solutions. In Section 4, we present our quantitative model and results. Section 5 concludes. Proofs for Sections 2 and 3 are included in Appendices A and C respectively.

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. We then present an algorithm for solving DRIPs and compare the accuracy and efficiency of our approach with other available methods by replicating three canonical papers in the rational inattention 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

[10] See, also, Reis (2006); Angeletos and Lian (2016, 2018); Gabaix (2016).

$\{\vec{x}_t \in \mathbb{R}^n\}_{t=0}^{\infty}$ is an exogenous stochastic process, and $v(.;.) : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ 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 $X^t \equiv (\vec{x}_0, \ldots, \vec{x}_t)$. 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: $\mathcal{S}^t \subset \mathcal{S}^{t+h}, \forall h \geq 0$.

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.[11] Formally, a sequence of information sets $\{S^t \subseteq \mathcal{S}^t\}_{t\geq 0}$ 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.[12] 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.

[11] Although 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, Sung and Woodford (2019) endogenize noisy memory in a setting where carrying information over time is costly.

[12] For a discussion of Shannon's mutual information function and generalizations see Caplin, Dean and Leahy (2017). See also Hébert and Woodford (2018) for an alternative cost function.

We can now formalize the rational inattention problem (henceforth RI Problem) of the agent in our setup:

$$V_0(S^{-1}) \equiv \sup_{\{S_t \subset \mathcal{S}^t, \vec{a}_t : S^t \to \mathbb{R}^m\}_{t \geq 0}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}[v(\vec{a}_t; \vec{x}_t) - \omega \mathbb{I}(X^t; S^t | S^{t-1}) | S^{-1}] \tag{2.1}$$

$$s.t.\ S^t = S^{t-1} \cup S_t, \forall t \geq 0, \tag{2.2}$$

$$S^{-1} \text{ given.} \tag{2.3}$$

where Equation (2.1) is the RI Problem in which the agent maximizes the net present value of her payoffs minus the cost of attention; Equation (2.2) captures the evolution of the agent's information set over time and Equation (2.3) specifies the initial condition for the dynamic problem.

It is important to note that this problem is a dynamic problem *only* because of information acquisition: any information acquired in a given period potentially reduces the expected costs of information acquisition in the future by expanding the agent's information set.

2.1.1 Two General Properties of the Solution

Solving the RI problem in Equation (2.1) 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 actions are sufficient statistics for signals over time (Steiner, Stewart and Matějka, 2017; Ravid, 2019). The following lemma formalizes this result in our setting.

Lemma 2.1. *Suppose $\{(S^t \subset \mathcal{S}^t, \vec{a}_t : S^t \to \mathbb{R}^m\}_{t=0}^{\infty} \cup S^{-1}$ is a solution to the 2.1. $\forall t \geq 0$, define $a^t \equiv \{\vec{a}_\tau\}_{0 \leq \tau \leq t} \cup S^{-1}$. Then, $X^t \to a^t \to S^t$ forms a Markov chain, i.e. a^t is a sufficient statistic for S^t with respect to X^t.*

Proof. See Appendix A.1. ∎

Lemma 2.1 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 2.1 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 2.2. *Suppose $\{\vec{x}_t\}_{t\geq 0}$ is a Markov process and $\{\vec{a}_t\}_{t\geq 0}$ is a solution to the 2.1 given an initial information set S^{-1}. Then $\forall t \geq 0$:*

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

Proof. See Appendix A.2. ∎

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:

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

Here, $\mathbf{H} \in \mathbb{R}^{n\times m}$ has full column rank and captures the interaction of the actions with the state.[13] The assumption of $rank(\mathbf{H}) = m$ 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.

Moreover, we have normalized the Hessian matrix of v with respect to $\vec{a}$ to negative identity.[14]

Optimality of Gaussian Posteriors. We start by proving that optimal actions are Gaussian under quadratic payoff with a Gaussian initial prior. Maćkowiak and Wiederholt

[13]While we take this as an assumption, this payoff function can also be derived as a second order approximation to a twice differentiable function $v(.;.)$ around the non-stochastic optimal action and disregarding the terms that are independent of the agent's choices.

[14]This 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.

(2009b) prove this result in their setup where the cost of information is given by

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{I}(X^T; a^T)$$

Our setup is marginally 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 2.1. The following lemma presents a modified proof that applies to our specification.

Lemma 2.3. *Suppose the initial conditional prior, $\vec{x}_0|S^{-1}$, is Gaussian. If $\{\vec{a}_t\}_{t\geq 0}$ is a solution to the 2.1 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.*

Proof. See Appendix A.3. ∎

The Equivalent LQG Problem. Lemma 2.3 simplifies the structure of the problem in that it allows us to re-write the RI problem in Equation (2.1) in terms of choosing a set of Gaussian joint distributions between the actions and the state. This is a canonical formulation of the rational inattention problems in LQG settings and it appears in different forms throughout the literature. For completeness, the following Lemma derives the LQG problem in our setting that follows from the RI problem in Equation (2.1). A similar formulation appears in Equation (27) in Sims (2010).

Lemma 2.4. *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 state-space representation:*

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

Then, the RI problem in Equation (2.1) with quadratic payoff is equivalent to choosing a set of symmetric positive semidefinite matrices $\{\mathbf{\Sigma}_{t|t}\}_{t\geq 0}$:

$$V_0(\mathbf{\Sigma}_{0|-1}) = \max_{\{\mathbf{\Sigma}_{t|t}\in\mathbb{S}_+^n\}_{t\geq 0}} -\frac{1}{2}\sum_{t=0}^{\infty}\beta^t\left[tr(\mathbf{\Sigma}_{t|t}\mathbf{\Omega}) + \omega\ln\left(\frac{|\mathbf{\Sigma}_{t|t-1}|}{|\mathbf{\Sigma}_{t|t}|}\right)\right] \tag{2.4}$$

$$s.t. \qquad \mathbf{\Sigma}_{t+1|t} = \mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{Q}\mathbf{Q}', \quad \forall t\geq 0, \tag{2.5}$$

$$\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq 0, \quad \forall t \geq 0 \tag{2.6}$$

$$0 \prec \mathbf{\Sigma}_{0|-1} \prec \infty \quad given. \tag{2.7}$$

Here, $|.|$ is the determinant operator, $\succeq$ denotes positive semidefiniteness, $\mathbf{\Sigma}_{t|t} \equiv var(\vec{x}_t|a^t)$, $\mathbf{\Sigma}_{t|t-1} \equiv var(\vec{x}_t|a^{t-1})$, $\mathbf{\Omega} \equiv \mathbf{HH}'$ and $\mathbb{S}^n_+$ is the n-dimensional symmetric positive semidefinite cone.

Proof. See Appendix A.4. ■

Lemma 2.4 reformulates the RI problem in Equation (2.1) into an LQG problem in Equation (2.4) subject to the law of motion for the agent's priors in Equation (2.5), a set of no-forgetting constraints in Equation (2.6) that require agent's posterior to be at least as precise as their prior, and an initial condition in Equation (2.7).

This characterization of the problem matches the formulation in Sims (2010) but differs from the one in Sims (2003) and the "Golden rule approximation" in Miao, Wu and Young (2020) who solve a problem in which the cost of attention is not discounted.[15]

Solution. Sims (2010) derives a first order condition for the solution to this problem when the no-forgetting constraints do not bind, and also provides a solution for $n = 2$ and $m = 1$ when these constraints do bind. It follows from 2.2 that binding no-forgetting constraints are cases that should arise frequently. In fact, for any $m < n$, at least $n - m$ constraints always bind since actions are sufficient statistics for signals. Thus, any approach to solve the general case of arbitrary n and m should characterize when and which constraints bind at any given time.

The problem in Proposition 2.1 can be solved using the standard Karush-Kuhn-Tucker (KKT) conditions for elements of the posterior matrix $\mathbf{\Sigma}_{t|t}$. The one potential complication in doing so is that the no-forgetting constraints in Equation (2.6) are a set of n non-negativity constraints that operate only on the eigenvalues of the matrix $\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}$, which are themselves affected by the choice of $\mathbf{\Sigma}_{t|t}$. One of our theoretical contributions is to transform these constraints to a form that makes the application of KKT conditions straightforward. The core idea here is to take advantage of the fact that positive semi-definiteness is independent of rotation. Hence, the constraints on the eigenvalues of $\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}$ can be written in any basis with no loss of generality. We formalize this idea in the proof of the following proposition and derive the following KKT conditions for the solution.

[15]The implied problem under the second approach is

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

Proposition 2.1. *Suppose $\mathbf{\Sigma}_{0|-1}$ is strictly positive definite, and $\mathbf{AA}' + \mathbf{QQ}'$ is of full rank. Then, all the future priors $\{\mathbf{\Sigma}_{t+1|t}\}_{t\geq 0}$ are invertible under the optimal solution to the 2.4, which is characterized by*

$$\omega\mathbf{\Sigma}_{t|t}^{-1} - \mathbf{\Lambda}_t = \mathbf{\Omega} + \beta\mathbf{A}'(\omega\mathbf{\Sigma}_{t+1|t}^{-1} - \mathbf{\Lambda}_{t+1})\mathbf{A}, \qquad \forall t \geq 0, \tag{2.8}$$

$$\mathbf{\Lambda}_t(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}) = \mathbf{0}, \mathbf{\Lambda}_t \succeq \mathbf{0}, \mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq \mathbf{0}, \qquad \forall t \geq 0, \tag{2.9}$$

$$\mathbf{\Sigma}_{t+1|t} = \mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{QQ}', \qquad \forall t \geq 0, \tag{2.10}$$

$$\lim_{T\to\infty} \beta^{T+1} tr(\mathbf{\Lambda}_{T+1}\mathbf{\Sigma}_{T+1|T}) = \mathbf{0} \tag{2.11}$$

*where $\mathbf{\Lambda}_t$ and $\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}$ are **simultaneously diagonalizable**.*

Proof. See Appendix A.5. ■

Here, Equation (2.8) is the first order condition for the problem with eigenvalues of $\mathbf{\Lambda}_t$ being the Lagrange multipliers on the no-forgetting constraints. If none of these constraints are binding under the optimal solution, then $\mathbf{\Lambda}_t = \mathbf{0}$ and this first order condition collapses to Equation (31) in Sims (2010). Here we allow for the possibility of binding no-forgetting constraints, so $\mathbf{\Lambda}_t$ is possibly non-zero and characterized by the complementarity slackness condition in Equation (2.9). Furthermore, Equation (2.10) is the law of motion for the agent's prior and finally, Equation (2.11) is the transversality condition on information acquisition of the agent.

With these equations at hand, one can obtain the solution to the problem. However, we can go further in interpreting and adjusting these equations to fit the standard characterizations of dynamic problems in economics. In particular, we can reformulate these conditions to derive a forward-looking *Euler equation* that captures the contemporaneous and continuation value of information and a *policy function* that given the value of information, maps the state variable of the agent at time t (prior uncertainty denote by $\mathbf{\Sigma}_{t|t-1}$) to a choice variable (posterior uncertainty denoted by $\mathbf{\Sigma}_{t|t}$). To present these two equations as concisely as possible, we start by introducing the following two matrix operators:

Definition 2.1. *For a diagonal matrix $\mathbf{D} = \text{diag}(d_1, \ldots, d_n)$ let*

$$\text{Max}(\mathbf{D}, \omega) \equiv \text{diag}(\max(d_1, \omega), \ldots, \max(d_n, \omega))$$
$$\text{Min}(\mathbf{D}, \omega) \equiv \text{diag}(\min(d_1, \omega), \ldots, \min(d_n, \omega))$$

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

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

In short, $\text{Max}(\mathbf{X}, \omega)$ preserves the $\mathbf{X}'s$ eigenvectors but replaces its eigenvalues with ω if they are smaller than ω. Similarly, $\text{Min}(\mathbf{X}, \omega)$ preserves $\mathbf{X}$'s eigenvectors but replaces its eigenvalues with ω if they are larger than ω.

Theorem 2.1. *Let* $\mathbf{\Omega}_t \equiv \mathbf{\Omega} + \beta\mathbf{A}'(\omega\mathbf{\Sigma}_{t+1|t}^{-1} - \mathbf{\Lambda}_{t+1})\mathbf{A}$ *denote the forward-looking component of the FOC in Proposition 2.1. Then,*

$$\mathbf{\Sigma}_{t|t} = \omega\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \left[\text{Max}\left(\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}, \omega\right)\right]^{-1} \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \tag{2.12}$$

$$\mathbf{\Omega}_t = \mathbf{\Omega} + \beta\mathbf{A}'\mathbf{\Sigma}_{t+1|t}^{-\frac{1}{2}} \,\text{Min}\left(\mathbf{\Sigma}_{t+1|t}^{\frac{1}{2}}\mathbf{\Omega}_{t+1}\mathbf{\Sigma}_{t+1|t}^{\frac{1}{2}}, \omega\right) \mathbf{\Sigma}_{t+1|t}^{-\frac{1}{2}}\mathbf{A} \tag{2.13}$$

Proof. See Appendix A.6. ∎

Equation (2.12) is the policy function that characterizes the optimal posterior of the agent given the state $\mathbf{\Sigma}_{t|t-1}$ and the benefit matrix $\mathbf{\Omega}_t$. Furthermore, Equation (2.13) is the forward-looking Euler equation that characterizes $\mathbf{\Omega}_t$ that captures the dynamics of attention. Together with the law of motion for the agent's prior in Equation (2.10) as well as the transversality condition in Equation (2.11), 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 2.2. $\forall t \geq 0$, *let* $\{d_{i,t}\}_{1\leq i\leq n}$ *be the set of eigenvalues of the matrix* $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$ *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*

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

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. $\forall i \in \{1,\ldots,k_t\}$, $\mathbf{y}_{i,t} \equiv \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{u}_{i,t}$.

3. $\forall i \in \{1,\ldots,k_t\}$, $z_{i,t} \sim \mathcal{N}(0, \frac{\omega}{d_{i,t}-\omega})$, $z_{i,t} \perp (\vec{x}_t, z_{j,t})_{j\neq i}$.

Proof. See Appendix A.7. ∎

Evolution of Optimal Beliefs and Actions. While Theorems 2.1 and 2.2 provide a representation for the optimal posteriors and signals, 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 2.2. *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 2.2, and let $\hat{x}_t \equiv \mathbb{E}[\vec{x}_t|a^t]$ be the mean of agent's posterior about $\vec{x}_t$ at time t. Then, $\hat{x}_t$ and optimal actions evolve according to:*

$$\hat{x}_t = \underbrace{\mathbf{A}\hat{x}_{t-1}}_{\text{prior belief}} + \sum_{i=1}^{k_t} \overbrace{\underbrace{(1-\frac{\omega}{d_{i,t}})}_{\substack{\text{signal-to-noise}\\ \text{ratio of } i}} \mathbf{\Sigma}_{t|t-1}\mathbf{y}_{i,t}}^{\text{Kalman gain vector of } i} \times \underbrace{[\mathbf{y}_{i,t}'(\vec{x}_t - \mathbf{A}\hat{x}_{t-1}) + z_{i,t}]}_{\text{surprise in signal } i}$$

$$\vec{a}_t = \mathbf{H}'\hat{x}_t$$

Proof. See Appendix A.8. ∎

2.3 Solution Algorithm, Computational Accuracy and Efficiency

Given an initial prior $\mathbf{\Sigma}_{-1|0}$, the solution to the LQG problem in Equation (2.4) is characterized by a sequence of matrices $\{\mathbf{\Sigma}_{t|t}, \mathbf{\Sigma}_{t+1|t}, \mathbf{\Omega}_t\}_{t\geq 0}$ that satisfy the policy function and Euler equation in Theorem 2.1, the law of motion for the priors in Equation (2.5) as well as the transversality condition in Equation (2.11).

Our main methodological contribution here is that, based on our theoretical findings in Theorems 2.1 and 2.2, we provide a new algorithm for characterizing the sequence of these matrices. We also provide a software package for solving dynamic rational inattention problems (DRIPs) based on this algorithm that is available as the `DRIPs.jl` package to the Julia programming language.[16]

[16]See `https://afrouzi.com/DRIPs.jl/dev/` for installation and usage instructions. A Matlab code for the algorithm is also available at `https://github.com/choongryulyang/dynamic_multivariate_RI`.

We have also used our software package to replicate results from three canonical papers (Maćkowiak and Wiederholt, 2009a; Sims, 2010; Maćkowiak et al., 2018a) that use different methods for solving DRIPs and assess the accuracy and the efficiency of our algorithm. Our algorithm produces identical results to each of these papers but is considerably faster than alternative solution methods. A summary of computing times for these papers are reported in Table 1. Moreover, the replication files for these three papers are available at the link above and are also accessible as executable Jupyter notebooks that accompany this manuscript.

Table 1: Summary of computing times

Computing time for:	Dimension n^2	`DRIPs.jl` Time (s)	Alternative Algorithms Time (s)	Alternative Algorithms Source
Sims (2010)				
Benchmark parameterization:				
steady state	2^2	1.6×10^{-4}		
transition dynamics	2^2	6.3×10^{-4}	1.08×10^{3}	Miao, Wu and Young (2020)
"Golden rule" approximation	2^2	1.6×10^{-4}	3.00×10^{0}	Miao, Wu and Young (2020)
Maćkowiak and Wiederholt (2009a)				
Benchmark parameterization:				
problem without feedback	20^2	1.83×10^{-1}	4.58×10^{1}	original (published)
problem with feedback	20^2	3.97×10^{0}	1.72×10^{2}	replication files
Maćkowiak, Matějka and Wiederholt (2018a)				
Price setting with rational inattention				
without feedback	2^2	0.45×10^{-3}		
with feedback	40^2	4.42×10^{-1}		
Business cycle model with news shocks	40^2	9.40×10^{-1}		

Notes: This table shows the summary of computing times for our replication of Sims (2010), Maćkowiak and Wiederholt (2009a) and Maćkowiak et al. (2018a) (discussed in Section 2.4, Appendix B.1 and Appendix B.2 respectively). Tolerance level for convergence is 10^{-4} for the solution to rational inattention problem in all cases. Statistics from Miao et al. (2020) are taken directly from their manuscript. All other calculations were performed on a 2019 MacBook Pro with 16GB of memory, a 2.3 GHz processor and 8 cores (but no multi-core functionality was used).

The general outline of our algorithm is to solve the problem in two stages: first, we solve for the steady state of the problem that is independent of the initial prior, and second, we use a shooting algorithm on the Euler equation (Equation 2.13) and the law of motion for the prior (Equation 2.5) to characterize the transition path. We now describe these two stages in more detail.

Solving for the Steady State Information Structure. By the "steady state" information structure, we mean a triple $(\bar{\boldsymbol{\Sigma}}_{-1}, \bar{\boldsymbol{\Sigma}}, \bar{\boldsymbol{\Omega}})$ that satisfy the stationary versions of the policy function, the law of motion for the prior and the Euler equation (Equations 2.12, 2.5 and 2.13 respectively):

$$\bar{\mathbf{\Sigma}} = \omega \bar{\mathbf{\Sigma}}_{-1}^{\frac{1}{2}} \left[\text{Max} \left(\bar{\mathbf{\Sigma}}_{-1}^{\frac{1}{2}} \bar{\mathbf{\Omega}} \bar{\mathbf{\Sigma}}_{-1}^{\frac{1}{2}}, \omega \right) \right]^{-1} \bar{\mathbf{\Sigma}}_{-1}^{\frac{1}{2}} \tag{2.16}$$

$$\bar{\mathbf{\Sigma}}_{-1} = \mathbf{A} \bar{\mathbf{\Sigma}} \mathbf{A}' + \mathbf{Q}\mathbf{Q}' \tag{2.17}$$

$$\bar{\mathbf{\Omega}} = \mathbf{\Omega} + \beta \mathbf{A}' \bar{\mathbf{\Sigma}}_{-1}^{-\frac{1}{2}} \text{Min} \left(\bar{\mathbf{\Sigma}}_{-1}^{\frac{1}{2}} \bar{\mathbf{\Omega}} \bar{\mathbf{\Sigma}}_{-1}^{\frac{1}{2}}, \omega \right) \bar{\mathbf{\Sigma}}_{-1}^{-\frac{1}{2}} \mathbf{A}$$

One can then solve for this steady state triple by the following iterative algorithm, starting with an initial guess for $\bar{\mathbf{\Sigma}}_{-1} = \bar{\mathbf{\Sigma}}_{-1,(0)}$ and $\bar{\mathbf{\Omega}} = \bar{\mathbf{\Omega}}_{(0)}$. Then, in any iteration $j \geq 1$:

1. Obtain the eigenvalue and eigenvector decomposition of

$$\mathbf{X}_{(j)} \equiv \bar{\mathbf{\Sigma}}_{-1,(j-1)}^{\frac{1}{2}} \bar{\mathbf{\Omega}}_{(j-1)} \bar{\mathbf{\Sigma}}_{-1,(j-1)}^{\frac{1}{2}}$$

2. Use the steady state policy function in Equation (2.16) to form a guess about the steady state posterior covariance matrix:

$$\bar{\mathbf{\Sigma}}_{(j)} = \omega \bar{\mathbf{\Sigma}}_{-1,(j-1)}^{\frac{1}{2}} \left[\text{Max} \left(\mathbf{X}_{(j)}, \omega \right) \right]^{-1} \bar{\mathbf{\Sigma}}_{-1,(j-1)}^{\frac{1}{2}}$$

3. Use the steady state law of motion for the prior in Equation (2.17) and the steady state Euler equation in 2.3 to update the guesses for $\bar{\mathbf{\Sigma}}_{-1}$ and $\bar{\mathbf{\Omega}}$:

$$\bar{\mathbf{\Omega}}_{(j)} = \mathbf{\Omega} + \beta \mathbf{A}' \bar{\mathbf{\Sigma}}_{-1,(j-1)}^{-\frac{1}{2}} \text{Min} \left(\mathbf{X}_{(j)}, \omega \right) \bar{\mathbf{\Sigma}}_{-1,(j-1)}^{-\frac{1}{2}} \mathbf{A}$$

$$\bar{\mathbf{\Sigma}}_{-1,(j)} = \mathbf{A} \bar{\mathbf{\Sigma}}_{(j)} \mathbf{A}' + \mathbf{Q}\mathbf{Q}$$

4. Repeat with j+=1 if $\|\mathbf{\Sigma}_{-1,(j)} - \mathbf{\Sigma}_{-1,(j-1)}\| / \|\mathbf{\Sigma}_{-1,(j-1)}\| >$ tolerance.

Solving for the Transition Dynamics. The objective here is to solve for the transition path of the triple $(\mathbf{\Sigma}_{t|t}, \mathbf{\Sigma}_{t+1|t}, \mathbf{\Omega}_t)$ to the steady state conditional on a given initial prior, $\mathbf{\Sigma}_{-1|0}$. We use a shooting algorithm to solve for this. In particular, we start with the guess that after some large T, the sequence has converged to the steady state solution from the previous step so that, up to some tolerance,

$$\mathbf{\Sigma}_{t+1|t} = \bar{\mathbf{\Sigma}}_{-1}, \quad \mathbf{\Sigma}_{t|t} = \bar{\mathbf{\Sigma}}, \quad \mathbf{\Omega}_t = \bar{\mathbf{\Omega}}, \quad \forall t \geq T+1$$

Therefore, conditional on this guess, we only need to solve for a finite sequence

$$\{\boldsymbol{\Sigma}_{t|t}, \boldsymbol{\Sigma}_{t+1|t}, \boldsymbol{\Omega}_t\}_{0 \le t \le T}, \quad \boldsymbol{\Sigma}_{-1|0} \quad \text{given.}$$

We find this sequence using the following iterative procedure:

1. Start with the guess that $\forall 0 \le t \le T$, $\boldsymbol{\Omega}_{t,(0)} = \bar{\boldsymbol{\Omega}}$.

2. At iteration $j \ge 1$, given the sequence $\{\boldsymbol{\Omega}_{t,(j-1)}\}_{0 \le t \le T}$ and $\boldsymbol{\Sigma}_{-1|0,(j)} \equiv \boldsymbol{\Sigma}_{-1|0}$, iterate forward in time using the policy function from Theorem 2.1 and the law of motion for priors:

$$\begin{aligned}
&\text{for } t = 0 \uparrow T: \\
&\boldsymbol{\Sigma}_{t|t,(j)} \equiv \omega \boldsymbol{\Sigma}^{\frac{1}{2}}_{t|t-1,(j)} \left[\text{Max} \left(\boldsymbol{\Sigma}^{\frac{1}{2}}_{t|t-1,(j)} \boldsymbol{\Omega}_{t,(j-1)} \boldsymbol{\Sigma}^{\frac{1}{2}}_{t|t-1,(j)}, \omega \right) \right]^{-1} \boldsymbol{\Sigma}^{\frac{1}{2}}_{t|t-1,(j)} \\
&\boldsymbol{\Sigma}_{t+1|t,(j)} \equiv \mathbf{A} \boldsymbol{\Sigma}_{t|t,(j)} \mathbf{A}' + \mathbf{Q}\mathbf{Q}'
\end{aligned}$$

3. At iteration $j \ge 1$, given the sequence $\{\boldsymbol{\Sigma}_{t+1|t,(j)}\}_{0 \le t \le T} \cup \{\Sigma_{T+1|T,(j)} \equiv \bar{\boldsymbol{\Sigma}}_{-1}\}$ and $\boldsymbol{\Omega}_{T+1,(j)} \equiv \bar{\boldsymbol{\Omega}}$, iterate backward in time using the Euler equation from Theorem 2.1:

$$\begin{aligned}
&\text{for } t = T \downarrow 0: \\
&\boldsymbol{\Omega}_{t,(j)} \equiv \boldsymbol{\Omega} + \beta \mathbf{A}' \boldsymbol{\Sigma}^{-\frac{1}{2}}_{t+1|t,(j)} \text{Min} \left(\boldsymbol{\Sigma}^{\frac{1}{2}}_{t+1|t,(j)} \boldsymbol{\Omega}_{t+1,(j)} \boldsymbol{\Sigma}^{\frac{1}{2}}_{t+1|t,(j)}, \omega \right) \boldsymbol{\Sigma}^{-\frac{1}{2}}_{t+1|t,(j)} \mathbf{A}
\end{aligned}$$

4. Repeat Steps 2 to 4 with j+=1 if

$$\|(\boldsymbol{\Sigma}_{t+1|t,(j)})_{t=0}^T - (\boldsymbol{\Sigma}_{t+1|t,(j-1)})_{t=0}^T\| / \|(\boldsymbol{\Sigma}_{t+1|t,(j-1)})_{t=0}^T\| > \text{tolerance}.$$

5. Finally, check if T was large enough for convergence to the steady state. If not, repeat starting from Step 1 with larger T.

2.4 Example: Transition Dynamics in Sims (2010)

In his Handbook of Monetary Economics chapter, Sims (2010) provides an example with two shocks ($n = 2$) and one action ($m = 1$). He then characterizes the *steady state* posterior covariance matrix under the rational inattention problem. A discussion of the steady state solution to this problem also appears in the recent paper by Miao et al. (2020) based on the solution method proposed by them. What is novel here is that our solution method allows us to extend the solution and study its *transition dynamics* to that steady state from any initial prior.[17]

Background. The example in Sims (2010) is of a monopolist who chooses its price to match the sum of two AR(1) processes, where one is more persistent than the other. The contemporaneous profit of the monopolist is decreasing in the distance of its price from this linear sum and is given by $v(a_t, x_{1,t}, x_{2,t}) = -(a_t - x_{1,t} - x_{2,t})^2$ where a_t is the agent's action (here the monopolist's price). Moreover, $x_{1,t}$ and $x_{2,t}$ are both shocks with AR(1) processes. Assuming the agent discounts future payoffs at an exponential rate β, Equation (10) in Sims (2010) derives the equivalent LQG rational inattention problem as:

$$\min_{\{\mathbf{\Sigma}_{t|t}\succeq 0\}_{t\geq 0}} \sum_{t=0}^{\infty} \beta^t \left[tr(\mathbf{\Sigma}_{t|t}\mathbf{H}\mathbf{H}') + \omega \log\left(\frac{|\mathbf{\Sigma}_{t|t-1}|}{|\mathbf{\Sigma}_{t|t}|}\right)\right]$$
$$s.t. \quad \mathbf{\Sigma}_{t+1|t} = \mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{Q}\mathbf{Q}', \quad \mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq \mathbf{0}, \quad \mathbf{\Sigma}_{-1|0} \succeq \mathbf{0} \quad \text{given.}$$

with the following parameterization:

$$\beta = 0.9, \quad \omega = 1, \quad \mathbf{H} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0.95 & 0 \\ 0 & 0.4 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \sqrt{0.0975} & 0 \\ 0 & \sqrt{0.86} \end{bmatrix}$$

Here, we have renamed the parameters so that the problem directly maps to our formulation in Equation (2.4). Otherwise, the problem is the same as in Sims (2010).

Steady State Solution. The steady state information structure has appeared prior to our paper in Sims (2010) and Miao et al. (2020). Our objective here is to compare the solution based on our algorithm with these benchmarks. Our solution method yields the following posterior and prior covariance matrices for the steady state information structure up to a tolerance of 10^{-4}:

$$\bar{\mathbf{\Sigma}} \equiv \lim_{t\to\infty} \mathbf{\Sigma}_{t|t} = \begin{bmatrix} 0.3592 & -0.1770 \\ -0.1770 & 0.7942 \end{bmatrix}, \quad \bar{\mathbf{\Sigma}}_{-1} \equiv \lim_{t\to\infty} \mathbf{\Sigma}_{t+1|t} = \begin{bmatrix} 0.4217 & -0.0673 \\ -0.0673 & 0.9871 \end{bmatrix} \tag{2.18}$$

This solution is close to the posterior covariance reported in Sims (2010).[18] Moreover, it is almost identical to the one reported in Miao et al. (2020) who use conventional value

[17] See Table 1 for how our algorithm compares to the one proposed by Miao et al. (2020) in computing times.

[18] Sims (2010) reports the following posterior covariance matrix: $\bar{\mathbf{\Sigma}} = \begin{bmatrix} 0.373 & -0.174 \\ -0.174 & 0.774 \end{bmatrix}$.

function iteration methods to calculate this solution.[19]

Transition Dynamics of the Optimal Information Structure. In this section we report results for the transition path of the optimal information structure from a highly certain prior. In particular, we assume that in the steady state of the information acquisition problem, the agent's prior is affected by a one time "knowledge shock" that reduces their prior uncertainty to 1 percent of its long-run value. We refer to period -1 as the period in which this knowledge shock happens. This implies that at time 0, the agent's prior about $\vec{x}_0$ is given by

$$\vec{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{0|-1}), \quad \mathbf{\Sigma}_{0|-1} = 0.01 \times \bar{\mathbf{\Sigma}}_{-1}$$

where $\vec{x}_0 = (x_{1,0}, x_{2,0})$ is the vector of the transitory and persistence shocks at time 0 and $\bar{\mathbf{\Sigma}}_{-1}$ is the prior covariance matrix in the steady state from Equation (2.18). By assuming that the mean of this prior is $\mathbf{0}$, we are implicitly assuming that both shocks were at their steady state values when the knowledge shock happened. We use the shooting algorithm outlined in Section 2.3 to solve for this transition path. It takes our code 630 microseconds to obtain the solution (See Table 1 for details).

Number of Signals Over Time. We start by characterizing the number of signals that the agent observes over time. It follows from Theorem 2.2 that this number is equal to the number of the eigenvalues of the matrix $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$ that are at least as large as ω. Since the dimension of the state in this problem is 2, there are two eigenvalues—or in other words, two dimensions to which the agent can pay attention.

Figure 1 plots these eigenvalues over time. At time 0, none of these eigenvalues are larger than ω, which implies that the agent acquires no information right after the knowledge shock. Starting at time 1, one of eigenvalues is larger than 1, which implies that the agent receives one signal starting at $t = 1$. It takes approximately 10 periods for these eigenvalues to reach their steady state, at which point only one of them remains above ω. This implies that even in the steady state the agent receives only one signal.[20]

[19] Miao et al. (2020) report the following posterior covariance matrix: $\bar{\mathbf{\Sigma}} = \begin{bmatrix} 0.3590 & -0.1769 \\ -0.1769 & 0.7945 \end{bmatrix}$.

[20] This is consistent with Lemma 2.2 which specifies that the number of signals should be bounded above by the agent's number of actions. Since the number of actions in this example is 1, the number of signals received by the agent should always be less than or equal to 1.

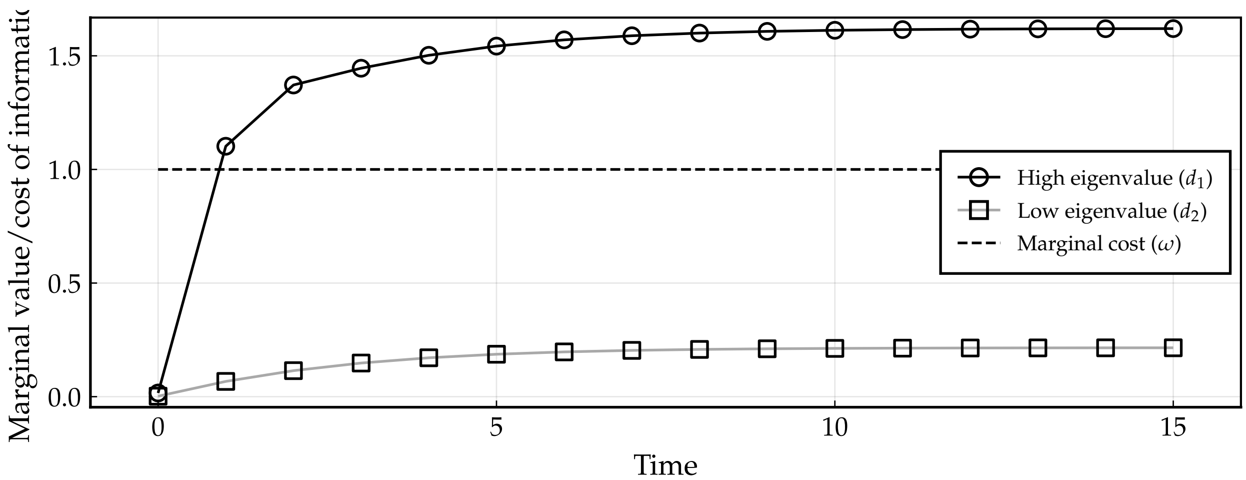

Figure 1: Marginal Value of Information on the Transition Path

Notes: This figure shows the marginal values of information in orthogonal dimensions for our extension of the example in Sims (2010). The transition dynamics are from an initial prior $\mathbf{\Sigma}_{0|-1} = 0.01 \times \mathbf{\Sigma}_{-1}$, where $\mathbf{\Sigma}_{-1}$ is the steady state prior covariance matrix reported in Equation (2.18). Following Theorem 2.2, these marginal values are defined as the eigenvalues of the matrix $\mathbf{\Sigma}_{t|t-1}^{1/2}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{1/2}$. The agent receives a signal in a particular dimension if the corresponding marginal value is larger than ω. In this example, the agent observes no signals in period 0, and only one signal after that, including in the steady state.

Signal-to-noise Ratio and Pass-through of Shocks on the Transition Path. Although the agent receives one signal starting at time 1, the signal-to-noise ratio and the loading of this signal on each of the two shocks still varies over the transition path. This implies that, in contrast to the steady state information structure, there is time variation in how the monopolist's beliefs and price responds to shocks. We use our results in Proposition 2.2 to study these time varying responses. In particular, we focus on two different set of statistics in this section:

1. The first quantity is the signal-to-noise itself, which according to Proposition 2.2 is defined as $1 - \frac{\omega}{d_{1,t}}$. The left panel of Figure 2 plots this quantity after the knowledge shock happens at time -1. At time 0, the signal-to-noise ratio is zero since the agent is not receiving any signals, but starting at time 1, this quantity is positive and converges to its steady state value of around 0.4 from below in approximately 10 periods. This shows that the initial knowledge shock at $t = -1$ has dynamic consequences and *crowds out* information acquisition in later periods.

2. The second set of quantities are the pass-through of shocks to the monopolist's price, which we define as the elasticity of price to innovations to each of the persistent and

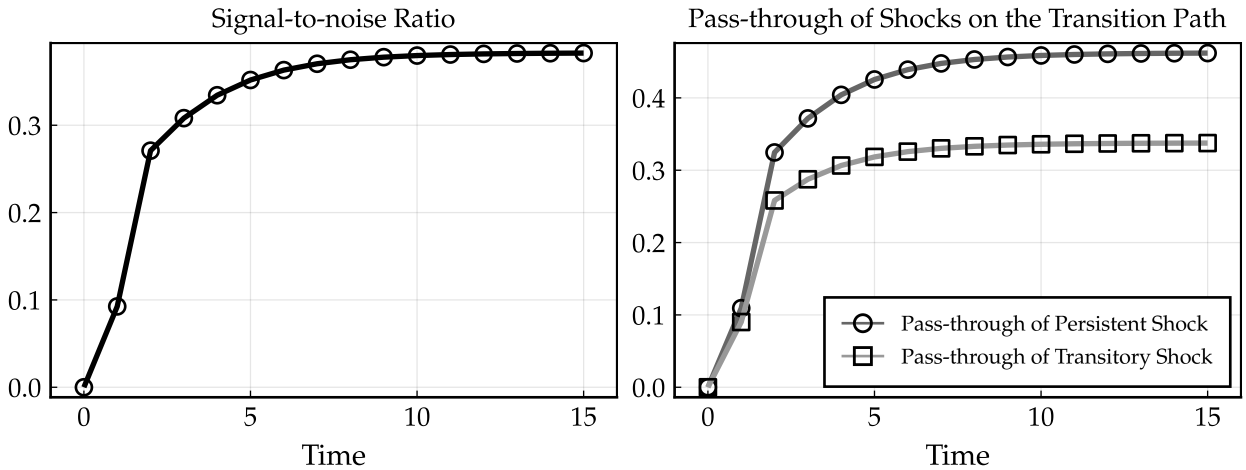

Figure 2: Signal Informativeness and Price Responsiveness on the Transition Path

Notes: The left panel shows the transition path of the Kalman gain for the optimal signal for our extension of the example from Sims (2010). Moreover, the right panel shows the instantaneous pass-through of the persistence and transitory shocks to the agent's action on the transition path. All values are constant in the steady state. Transition dynamics are from an initial prior $\boldsymbol{\Sigma}_{0|-1} = 0.01 \times \boldsymbol{\Sigma}_{-1}$, where $\boldsymbol{\Sigma}_{-1}$ is the steady state prior covariance matrix reported in Equation (2.18).

transitory components of the cost:

$$\text{Pass-through}_{s,t} \equiv \frac{\partial a_t}{\partial u_{s,t}}, \quad s \in \{1,2\}.$$

Here, a_t is the action (price) of the monopolist and $u_{s,t}$ is an innovation to the AR(1) process $x_{s,t}$, $s \in \{1,2\}$, that is realized at time t. The right panel of Figure 2 plots these two quantities over time. At time 0, because the monopolist receives no signals, both shocks have a pass-through of zero. Starting at time 1, both pass-throughs are positive and converge to their steady state from below—which directly follows from the crowding out effect of the initial knowledge shock. Moreover, on the transition path, as well as in the steady state, the more persistent shock has a higher pass-through relative to the transitory shock. This is due to the fact that the acquiring more information about the persistent shock is more valuable for a forward-looking agent.

Finally, it is important to note that all three of these quantities are constant under the steady state information structure, and the time-varying responses shown in Figure 2 are all due to the transition dynamics of attention.

Impulse Response Functions. How important can the transition dynamics of attention be? One way of answering this question is to compare how the impulse response functions of the monopolist's price to shocks vary between the steady state and the transition path of the optimal information structure.

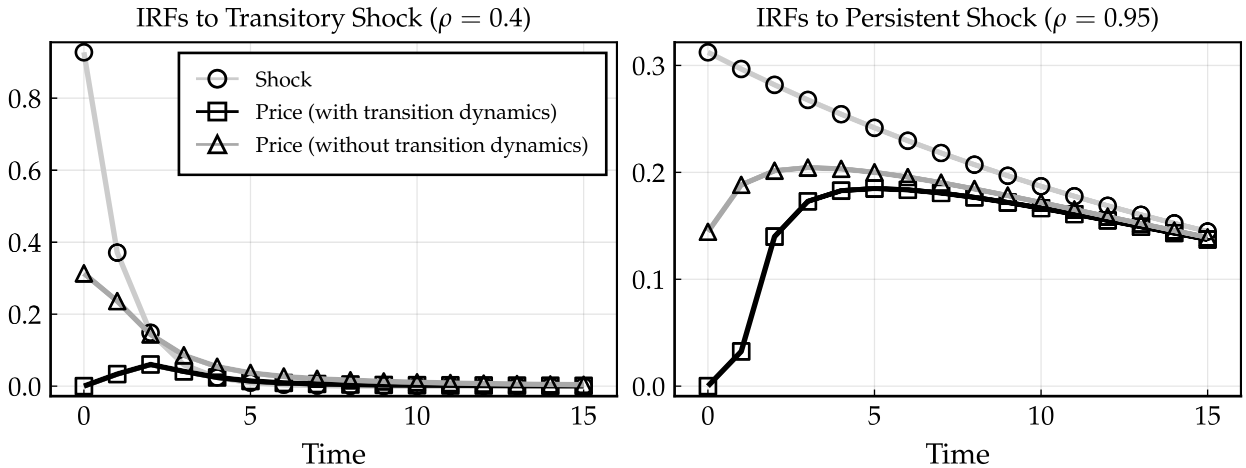

Figure 3: Impulse Response Functions in Steady State versus on the Transition Path

Notes: This figure the impulse response functions of the price with both the steady state information structure as well as the information structure on the transition path in our extension of the example from Sims (2010). Transition dynamics are from an initial prior $\mathbf{\Sigma}_{0|-1} = 0.01 \times \mathbf{\Sigma}_{-1}$, where $\mathbf{\Sigma}_{-1}$ is the steady state prior covariance matrix reported in Equation (2.18). The agent consistently acquires less information relative to the steady state on the transition path and the impulse responses are more muted. In particular, price does not respond to shocks at all at time 1 as the agent receives no signals in that period.

Figure 3 plots these impulse response functions for a one standard deviation innovation at time 0 to both components of the cost and under both information structures. The main observation is that the impulse responses are significantly muted under the information structure in the transition path. The monopolist, being certain about both shocks at $t = -1$, substitutes temporarily substitutes away from information acquisition and pays little attention to costs after the initial knowledge shock. At time 0, it receives no signal about any of the two AR(1) processes and has no reaction to their innovations. Starting at time 1, the reaction is larger than zero but significantly below what it is under the steady state information structure. This directly follows from the fact that the signal-to-noise ratio on the transition path is significantly smaller than its steady state value.

2.5 Further Discussion of Dynamic Rational Inattention Problems

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

Incentives. Underneath its technical representation, Theorem 2.1 encodes an intuitive economic result. It shows that in acquiring information, the agent first decomposes the matrix $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$, 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 2.2 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 $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$ 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 corner solutions generated through the no-forgetting constraints in Equation (2.6). The most obvious and likely case for binding no-forgetting constraints is when the number of actions m is strictly less than the dimension of the state n—as was the case in the example from Sims (2010) in Section 2.4. This follows directly from Lemma 2.1 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.[21]

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 2.1 rules out this case by showing that if

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

the chain-rule of mutual information holds, then the agent has no smoothing incentives.[22]

The economic consequence of this result is that independent of how many shocks they face, rationally inattentive agents 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.[23]

Information Spillovers. Sufficiency of actions for signals also generates endogenous correlation across exogenously independent shocks and provides a microfoundation for *information spillovers* across different actions (Sims, 2010). These effects are uniquely characterized by the eigenvectors of $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$ with eigenvalues *larger* than ω. Therefore, information about an action can effect other actions either through a subjective correlated posterior ($\mathbf{\Sigma}_{t|t-1}$) or through complementarities or substitutabilies in actions captured by $\mathbf{\Omega}_t$.[24]

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

[22]The 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.

[23]Hellwig 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.

[24]For 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 (2020) show that complementarities or substitutabilies 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.

interest rate of i_t, and forms demand over a continuum of varieties indexed by $i \in [0,1]$. Formally, the representative household's problem is

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

where $\mathbb{E}_t^f[.]$ is the expectation operator of this fully informed agent at time t, Π_t is the aggregated profits of firms, 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:

$$\begin{aligned}
C_{i,t} &= C_t P_t^{\theta} P_{i,t}^{-\theta}, && \forall i \in [0,1], \forall t \ge 0, \\
1 &= \beta(1+i_t)\mathbb{E}_t^f \left[\frac{Q_t}{Q_{t+1}} \right], && \forall t \ge 0, \\
W_t &= Q_t, && \forall t \ge 0.
\end{aligned}$$

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:

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

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 3.1. *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:*

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

Proof. See Appendix C.1. ■

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.[25] We derive this as an equilibrium outcome in Lemma 3.1 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

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

where the second line is a second order approximation with small letters denoting the logs of corresponding variables.[26] 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.

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

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

[25] See, 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).

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

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 2.2.

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 2.4. Formally, let $\sigma_{i,t|t-1} \equiv \sqrt{var(q_t|p_i^{t-1})}, \sigma_{i,t|t} \equiv \sqrt{var(q_t|p_i^t)}$ 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 2.4 is

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

3.2 Characterization of Solution

The solution to this problem follows from Proposition 2.1, and is characterized by the following proposition.

Proposition 3.1. *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*

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

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

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

2. *the firm's price evolves according to:*

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

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

Proof. See Appendix C.2. ∎

The first part of Proposition 3.1 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 2.4 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

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

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 3.1. *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.*

Proof. See Appendix C.3. ∎

While Corollary 3.1 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 ($\beta \to 1$) and σ_u^2 is small relative to the cost of information ω:[27]

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

3.3 Aggregation

For aggregation, we make two assumptions: (1) firms all start from the same initial prior uncertainty, $\sigma_{i,0|-1}^2 = \sigma_{0|-1}^2, \forall i \in [0,1]$, and (2) firms' rational inattention errors are independently distributed.[28]

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

1. *the Phillips curve of this economy is*

$$\pi_t = \max\{0, \frac{\sigma_{t|t-1}^2 - \underline{\sigma}^2}{\sigma_{t|t}^2}\} y_t$$

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

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

[27]This 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.

[28]Our 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 $\sigma^2_{T|T-1} > \underline{\sigma}^2$, then for $t \geq T+1$:*

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

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

Proof. See Appendix C.4. ■

3.4 Implications for the Slope of the Phillips Curve

Proposition 3.2 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 3.2 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 3.2 that once the inattention problem settles in its steady-state, the Phillips curve is given by

$$\pi_t = \frac{\kappa}{1-\kappa}y_t$$

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 – and a larger degree of – monetary non-neutrality.

Comparative statics of κ with respect to the underlying parameters of the model are derived in Corollary 3.1. 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 3.1 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 3.2. *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.*

Proof. See Appendix C.5. ■

The comparative statics follow directly from Corollary 3.1 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.[29] 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 3.2, 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 3.3. *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.

Proof. See Appendix C.6. ∎

The intuition follows from Corollary 3.1. An increase in $\frac{\sigma_u}{\phi}$ makes the nominal demand more volatile and raises the reservation uncertainty of firms. Hence, immediately after

[29] See 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).

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.

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.

3.5 Implications for Anchoring of Inflation Expectations

One of the most salient indicators to which monetary policymakers pay specific attention, especially under inflation targeting regimes, is the *anchoring of inflation expectations*. "Well-anchored" inflation expectations are considered a sign of success for monetary policy as they imply that the publics' inflation expectations are not very sensitive to temporary disturbances in economic variables. Moreover, the extent to which inflation expectations are anchored in the U.S. economy seem to have increased over time. Since the onset of the Great Moderation, inflation expectations are more stable and seem to have lower sensitivity to short-run fluctuations in the economic data (Bernanke, 2007; Mishkin, 2007).

The dependence of firms' information acquisition incentives on the rule of monetary policy in our framework provides a natural explanation for this trend. Intuitively, when monetary policy becomes more Hawkish in stabilizing prices, firms pay less attention to shocks that affect their nominal marginal costs and hence their beliefs become less sensitive to short-run fluctuations in economic data. The following proposition characterizes the dynamics of firms' inflation expectations in our simple model.

Proposition 3.3. *Let $\hat{\pi}_t \equiv \int_0^1 \mathbb{E}_{i,t}[\pi_t]di$ denote the average expectation of firms about aggregate inflation at time t. Then, in the steady state of the attention problem,*

1. *the relationship between inflation expectations, $\hat{\pi}_t$, and output gap, y_t, is given by*

$$\hat{\pi}_t = (1-\kappa)\hat{\pi}_{t-1} + \frac{\kappa^2}{(2-\kappa)(1-\kappa)} y_t$$

2. *dynamics of $\hat{\pi}_t$ is captured by the following AR(2) process:*

$$\hat{\pi}_t = 2(1-\kappa)\hat{\pi}_{t-1} - (1-\kappa)^2\hat{\pi}_{t-2} + \frac{\kappa^2}{2-\kappa}\frac{\sigma_u}{\phi}u_t$$

where κ is the steady-state Kalman-gain of firms.

Proof. See Appendix C.7. ■

Proposition 3.3 illustrates the degree of anchoring in firms' inflation expectations from two perspectives. The first part of the Proposition, derives relationship between inflation expectations and output gap and shows that the sensitivity of inflation expectations with respect to the output gap depends positively on κ. The second part of the proposition then recasts this relationship in terms of the exogenous monetary policy shocks, which are the sole drivers of short-run fluctuations in this economy.

The AR(2) nature of these expectations indicate the inherent inertia that expectations inherit from firms' imperfect information – the counterfactual being full-information rational expectations, in which case both inflation and inflation expectations are i.i.d. over time.[30]

Moreover, both the degree of the inertia in firms' inflation expectations, which is determined by $1-\kappa$, as well as the sensitivity of firms' inflation expectations to output gap or monetary policy shocks depend on the conduct of monetary policy through κ. The following Corollary formalizes this relationship.

Corollary 3.4. *Firms' inflation expectations are less sensitive to both output gap and short-run monetary policy shocks (are more "anchored") and are more persistent when monetary policy is more hawkish – i.e. $\frac{\sigma_u}{\phi}$ is smaller.*

Proof. See Appendix C.8. ■

The intuition behind the result in Corollary 3.4 is the same as for the slope of the Phillips curve. With more Hawkish monetary policy, firms pay lower attention to monetary policy shocks which decreases the sensitivity of their beliefs to these shocks and increases their persistence.

[30] With full-information rational expectations, $\int_0^1 \mathbb{E}_{i,t}[\pi_t] = \pi_t = \Delta q_t = \sigma_u \phi^{-1} u_t$.

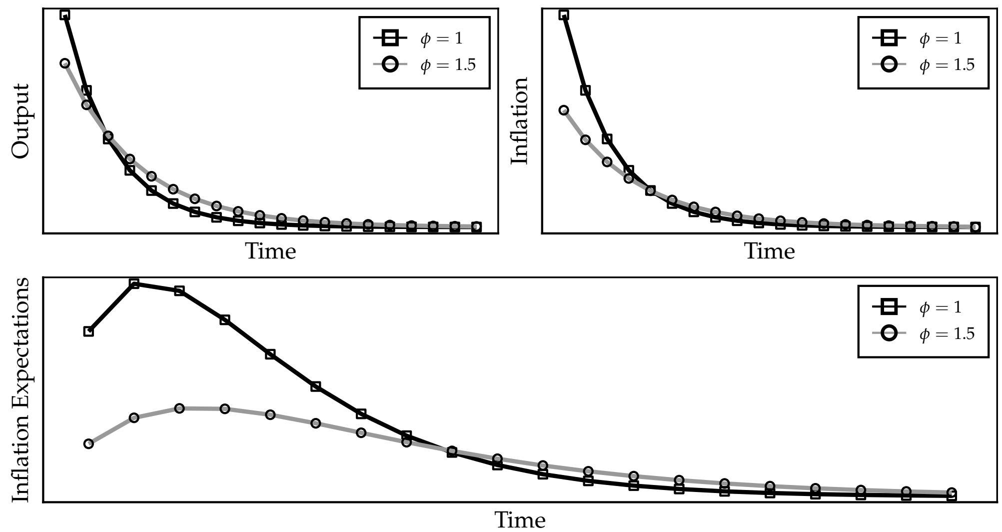

Figure 4: Impulse Responses to a 1 Std. Dev. Expansionary Monetary Policy Shock

Notes: This figure plots a numerical example for impulse responses of inflation, output, and firms' inflation expectations to a one standard deviation expansionary shock to monetary policy under two different values for $\phi \in \{1, 1.5\}$.

Figure 4 illustrates these results. The top two panels in the Figure show the impulse responses of output and inflation to a one standard deviation expansionary monetary policy shock under two different values for $\phi \in \{1, 1.5\}$. Moreover, the bottom panel of Figure 4 shows the impulse responses of firms average inflation expectations under these two parameters: with more Hawkish monetary policy, expectations are less sensitive to monetary policy shocks, but their responses are more persistent.[31]

4 Quantitative Analysis

Our simple model from the previous section illustrates the mechanism of how the slope of the Phillips curve depends on the rule of monetary policy. In this section, we relax the simplifying assumptions and extend that model to a general equilibrium model with rational inattention to assess whether our mechanism is quantitatively valid.

Our exercise in this section is very much in the spirit of the literature that interprets the Great Moderation, at least partially, through the lens of a shift in monetary policy in

[31]While in our setup higher anchoring of the expectation are generated by a combination of higher order beliefs and lower information acquisition on the part of firms, it is also important to note that higher persistence and anchoring can be generated in a context that takes the role of strategic interactions into account (see, e.g., Angeletos and Huo, 2018).

the post-Volcker era (Clarida et al., 2000; Coibion and Gorodnichenko, 2011; Maćkowiak and Wiederholt, 2015). In particular, we are interested in the following question: can the shift in the rule of monetary policy in the post-Volcker era explain the decline in the slope of the Phillips curve, and if so by how much? To answer this question, we calibrate a quantitative version of our model with TFP and monetary policy shocks to the U.S. inflation and output data in the post-Volcker era and examine whether the model can generate a quantitatively relevant shift in the slope of the Phillips curve.

4.1 Setup of the Quantitative Model

Household. The representative household forms demand over a unit measure of weakly substitutable varieties, indexed by $i \in [0,1]$, supplies labor in segmented but competitive labor markets for each variety, and has access to a risk-free nominal bond. For simplicity, we assume that the representative household is fully informed about all prices and wages as the main purpose of this paper is to study the effects of rational inattention among firms on price setting decisions. Formally, the household solves:

$$\max_{\left\{C_t,\left\{C_{i,t},L_{i,t}\right\}_{i\in[0,1]},B_t\right\}_{t\geq 0}} \mathbb{E}_t^f\left[\sum_{t=0}^{\infty}\beta^t\left(\frac{C_t^{1-\sigma}}{1-\sigma}-\frac{\int_0^1 L_{i,t}^{1+\psi}di}{1+\psi}\right)\right] \tag{4.0}$$

$$s.t. \quad \int_0^1 P_{i,t}C_{i,t}di + B_t \leq R_{t-1}B_{t-1} + \int_0^1 W_{i,t}L_{i,t}di + \Pi_t$$

$$C_t = \left[\int_0^1 C_{i,t}^{\frac{\theta-1}{\theta}}di\right]^{\frac{\theta}{\theta-1}}$$

Here $\mathbb{E}_t^f[\cdot]$ is the full-information rational expectation operator at time t, B_t is the demand for nominal bonds and R_{t-1} is the nominal interest rate. $L_{i,t}$ is the firm-specific labor supply of the household, $W_{i,t}$ is the firm-specific nominal wage, and Π_t is the aggregate profit of firms. C_t is the aggregator over the consumption of goods produced by firms. θ is the constant elasticity of substitution across different firms.

Firms. There is a measure one of firms, indexed by i, that operate in monopolistically competitive markets. Firms take wages and demands for their goods as given, and choose their prices $P_{i,t}$ based on their information set, S_i^t, at that time. After setting their prices, firms hire labor from a competitive labor market and produce the realized level of demand that their prices induce with a production function, $Y_{i,t} = A_t L_{i,t}^d$, where $L_{i,t}^d$ is firm

i's demand for labor and A_t is an aggregate TFP shock. We assume $a_t \equiv \log(A_t)$, follows a AR(1) process:

$$a_t = \rho_a a_{t-1} + \varepsilon_{a,t}, \;\; \varepsilon_{a,t} \sim N(0, \sigma_a^2)$$

It follows that firm i's nominal profit at time t is given by

$$\Pi(P_{i,t}, A_t, W_{i,t}, P_t, Y_t) = \left(P_{i,t} - \frac{W_{i,t}}{A_t}\right)\left(\frac{P_{i,t}}{P_t}\right)^{-\theta} Y_t,$$

where Y_t is the aggregate output, P_t is the aggregate price index and $P_{i,t}$ is the firm's price.

Firms are rationally inattentive and choose their prices subject to a cost that is linear in Shannon's mutual information function, as in the RI problem in Equation (2.1). Firm i's dynamic rational inattention problem is given by:

$$\begin{aligned} \max_{\{S_{i,t} \subseteq \mathcal{S}_{i,t}, P_{i,t}(S_i^t)\}_{t\geq 0}} \quad & \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t C_t^{-\sigma} \left\{ \Pi(P_{i,t}, A_t, W_{i,t}, P_t, Y_t) \right.\right. \\ & \left.\left. - \omega \mathbb{I}(S_i^t; (A_\tau, W_{i,\tau}, P_\tau, Y_\tau)_{\tau \leq t} | S_i^{t-1}) \right\} \middle| S_i^{-1} \right] \\ \text{s.t.} \quad & S_i^t = S_i^{t-1} \cup S_{i,t}, \end{aligned} \tag{4.1}$$

where $\mathcal{S}_{i,t}$ is the set of available signals for the firm that satisfies the assumptions specified in Section 2.1.

Monetary Policy. Monetary policy is specified as a standard Taylor rule:

$$\frac{R_t}{\bar{R}} = \left(\frac{R_{t-1}}{\bar{R}}\right)^{\rho} \left(\left(\frac{P_t}{P_{t-1}}\right)^{\phi_\pi} \left(\frac{Y_t}{Y_t^n}\right)^{\phi_x} \left(\frac{Y_t}{Y_{t-1}}\right)^{\phi_{\Delta y}}\right)^{1-\rho} \exp(u_t)$$

where $\bar{R}$ is the steady-state nominal rate, Y_t^n is the natural level of output and $u_t \sim N(0, \sigma_u^2)$ is the monetary policy shock.

Definition of Equilibrium. Given exogenous processes for productivity and monetary policy shocks $\{a_t, u_t\}_{t\geq 0}$, a general equilibrium of this economy is an allocation for the representative household,

$$\Omega^H \equiv \left\{C_t, \{C_{i,t}, L_{i,t}\}_{i\in[0,1]}, B_t\right\}_{t=0}^{\infty},$$

an allocation for every firm $i \in [0,1]$ given their initial set of signals,

$$\Omega_i^F \equiv \left\{ s_{i,t} \in \mathcal{S}_{i,t}, P_{i,t}, L_{i,t}^d, Y_{i,t} \right\}_{t=0}^{\infty},$$

a set of prices $\left\{ P_t, R_t, \{W_{i,t}\}_{i \in [0,1]} \right\}_{t=0}^{\infty}$, and a stationary distribution over firms' states such that

1. given the set of prices and $\{\Omega_i^F\}_{i\in[0,1]}$, the household's allocation solves the problem in Equation (4.1),
2. given the set of prices and Ω^H, and the implied labor supply and output demand, firms' allocation solve their problem in Equation (4.1),
3. monetary policy satisfies the specified rule in Equation (4.1) ;
4. markets clear:

$$\begin{aligned} Y_{i,t} &= C_{i,t}, \qquad \forall i \in [0,1], \forall t \geq 0 \\ L_{i,t} &= L_{i,t}^d, \qquad \forall i \in [0,1], \forall t \geq 0 \\ Y_t &= C_t, \qquad \forall t \geq 0 \end{aligned}$$

4.2 Computing the Equilibrium

We use our theoretical framework for solving dynamic rational inattention problems to solve for the equilibrium by taking a second-order approximation to firms' profit function in Equation (4.1).[32] This problem is given by a similar expression in our simple model but now includes a role for strategic complementarities:

$$\min_{\{p_{i,t}\}_{t\geq 0}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}\left[(\theta-1)(p_{i,t} - p_t - \alpha x_t)^2 + \omega \mathbb{I}(p_{i,t}, \{p_{t-j} + \alpha x_{t-j}\}_{j=0}^{\infty} | p_i^{t-1}) | p_i^{-1} \right]$$

Here, $\alpha \equiv \frac{\sigma+\psi}{1+\theta\psi}$ is the degree of strategic complementarity that is pinned down by the underlying deep parameters of the model, $x_t \equiv y_t - y_t^n$ is the log output gap in the model defined as the log difference between output and its natural level in the economy with no frictions, and p_t is the log of aggregate price. Moreover, we have already incorporated the result from Lemma 2.2 that with Shannon's mutual information as the cost of attention, the history of prices is sufficient statistics for the firm's signals at any given time.

[32] This approach is commonly used to turn firms' problems to quadratic objectives – see, e.g., Maćkowiak and Wiederholt 2009a; Maćkowiak and Wiederholt 2015.

Consequently, our general equilibrium model, up to a first order approximation, is characterized by the following three equations with two stochastic processes of technology (y_t^n) and monetary policy shocks (u_t):

$$
\begin{aligned}
x_t &= \mathbb{E}_t^f \left[x_{t+1} - \frac{1}{\sigma} \left(i_t - \pi_{t+1} \right) \right] + \mathbb{E}_t^f \left[y_{t+1}^n \right] - y_t^n \\
p_{i,t} &= \mathbb{E}_{i,t} \left[p_t + \alpha x_t \right] \\
i_t &= \rho i_{t-1} + (1-\rho) \left(\phi_\pi \pi_t + \phi_x x_t + \phi_{\Delta y} \Delta y_t \right) + u_t
\end{aligned}
$$

where $\mathbb{E}_{i,t}[\cdot]$ is firm i's expectation operator conditional on its time t information set under the solution to its rational inattention problem.

Lastly, as firms' rational inattention problem depends on the state-space representation of $p_t + \alpha x_t$, which is itself an endogenous object to the model, we use the following iteration algorithm to solve for the equilibrium: we start by guessing for the MA representation of $p_t + \alpha x_t$ and solve the firms' rational inattention problem. Under the solution to that problem, we then solve for the implied state-space representations of the output gap and prices and then update our guess. The equilibrium is then characterized as a fixed point of this mapping. A detailed description of matrix representations and our solution algorithm are provided in Appendix D.1.

4.3 Calibration

Our benchmark model is calibrated at a quarterly frequency with a time discount factor of $\beta = 0.99$ to the post-Volcker U.S. data ending at the onset of the Great Recession (1983–2007). A summary of the calibrated values of the parameters is presented in Table 2. In the remainder of this section we go over the details of our calibration strategy.

Assigned parameters. We set the elasticity of substitution across firms to be ten ($\theta = 10$), which corresponds to a markup of 11 percent. We set the inverse of the Frisch elasticity (ψ) to be 2.5 and the elasticity of intertemporal substitution ($1/\sigma$) to be 0.4, which are consistent with estimates presented in Aruoba et al. (2017).

Monetary policy rule(s). We set the standard deviation of monetary policy shocks (σ_u) in our benchmark model to match the size of the identified monetary policy shocks constructed by Romer and Romer (2004) for the period 1983–2007.[33]

[33]Original data on monetary policy shocks in Romer and Romer (2004) are available until 1996, while we use extended data, which are available until 2007, from Coibion, Gorodnichenko, Kueng and Silvia (2017).

Table 2: Calibrated and Assigned Parameters

Parameter	Value	Moment Matched / Source
Panel A. Calibrated parameters		
Information cost (ω)	0.70×10^{-3}	Cov. matrix of GDP and inflation
Persistence of productivity shocks (ρ_a)	0.850	Cov. matrix of GDP and inflation
S.D. of productivity shocks (σ_a)	1.56×10^{-2}	Cov. matrix of GDP and inflation
Panel B. Assigned parameters		
Time discount factor (β)	0.99	
Elasticity of substitution across firms (θ)	10	Firms' average markup
Elasticity of intertemporal substitution ($1/\sigma$)	0.4	Aruoba et al. (2017)
Inverse of Frisch elasticity (ψ)	2.5	Aruoba et al. (2017)
Taylor rule: smoothing (ρ)	0.946	Estimates 1983–2007 (Table A.1)
Taylor rule: response to inflation (ϕ_π)	2.028	Estimates 1983–2007 (Table A.1)
Taylor rule: response to output gap (ϕ_x)	0.168	Estimates 1983–2007 (Table A.1)
Taylor rule: response to output growth ($\phi_{\Delta y}$)	3.122	Estimates 1983–2007 (Table A.1)
S.D. of monetary shocks (σ_u)	0.28×10^{-2}	Romer and Romer (2004)
Panel C. Counterfactual model parameters (Pre-Volcker: 1969–1978)		
Taylor rule: smoothing (ρ)	0.918	Estimates 1969–1978 (Table A.1)
Taylor rule: response to inflation (ϕ_π)	1.589	Estimates 1969–1978 (Table A.1)
Taylor rule: response to output gap (ϕ_x)	0.292	Estimates 1969–1978 (Table A.1)
Taylor rule: response to output growth ($\phi_{\Delta y}$)	1.028	Estimates 1969–1978 (Table A.1)
S.D. of monetary shocks (σ_u)	0.54×10^{-2}	Romer and Romer (2004)

Notes: The table presents the baseline parameters for the general equilibrium model. Panel A shows the calibrated parameters which match the three key moments shown in Table 3. Panel B shows values and the source of the assigned model parameters. Panel C shows the parameters for the counterfactual analysis in Section 4.5. See Section 4.3 for details.

Furthermore, for the parameters describing the monetary policy rule (ρ, ϕ_π, $\phi_{\Delta y}$, ϕ_x), we estimate the Taylor rule in Equation (4.2) using real-time U.S. data. Specifically, following Coibion and Gorodnichenko (2011), we use the Greenbook forecasts of inflation and real GDP growth. The measure of the output gap is also based on Greenbook forecasts. We consider two time samples: the pre-Volcker period (1969–1978) and the post-Volcker period (1983–2007).[34] The point estimates are reported in Panel B of Table 2, and more

[34] Coibion and Gorodnichenko (2011) use data from 1983 through 2002 for the post-Volcker period estimation. We extend the sample period until 2007. Another difference is that our specification allows for interest rate smoothing of order one, while they consider the smoothing of order two.

detailed results including standard errors are reported in Appendix Table A.1. These estimates point to strong long-run responses by the central bank to inflation and output growth (2.03 and 3.12, respectively) and a moderate response to the output gap (0.17).[35]

Finally, for our counterfactual analysis in later sections, we do a similar estimation of these parameters for the pre-Volcker era (1969–1978). The point estimates are reported in Panel C of Table 2, and more detailed results, including standard errors, are reported in Appendix Table A.1.

Calibrated Parameters. We calibrate the three remaining parameters of the model – marginal costs of information processing (ω) as well as the persistence (ρ_a) and the size (σ_a) of productivity shocks – jointly by targeting the covariance matrix of inflation and real GDP in post-Volcker U.S. data (1983–2007). The covariance matrix is measured after we detrend the CPI core inflation and real GDP data using log-quadratic trends. The three moments (variances of inflation and GDP along with their covariance) exactly identify the three model parameters, as reported in Table 2.

The standard deviation of the productivity shocks (σ_a) is around 1.56 percent per quarter, which is about six times bigger than the standard deviation of the monetary policy shock (σ_u) for the post-Volcker period.

Moreover, the calibrated cost of information processing, $\omega\mathbb{I}(.,.)$, is 0.1 percent of firms' steady-state real revenue.[36] This small calibrated cost implies that imperfect information models do not require large information costs to match the data. The cost is negligible compared to firms' revenue. One relevant measure that one could use to relate the degree of information acquisition to that of professional forecasters is the firms' Kalman gain on their signals under the optimal information structure. The implied Kalman gain for firms in the model is 0.8, which implies a large degree of information acquisition relative to professional forecasters – Coibion and Gorodnichenko (2015b) estimate professional forecasters' Kalman gain to be around 0.5.

[35]Because empirical Taylor rules are estimated using annualized rates while the Taylor rule in the model is expressed at quarterly rates, we rescale the coefficient on the output gap in the model such that $\phi_x = 0.673/4 = 0.168$. Also, because we use the Greenbook forecast data prepared by staff members of the Fed a few days before each FOMC meeting, the sample from 1969 through 1978 was monthly, whereas the sample from 1983 through 2007 was every six weeks. Thus, we convert the estimated AR(1) parameters from monthly or six-week frequency to quarterly and use the converted parameters for our model simulations.

[36]This number is on the lower end of the cost of pricing frictions that have been estimated in the literature. For instance, Levy, Bergen, Dutta and Venable (1997) estimate the cost of menu cost frictions as 0.7 percent of firms' steady state revenue.

4.4 Model Fit

Targeted Moments. Table 3 reports our targeted moments both in the data and as implied by the model. All three targeted moments, variances of GDP and inflation and their covariance, are matched by the model.

Table 3: Targeted Moments

Moment	**Data**	**Model**
Standard deviation of inflation (1983–2007)	0.015	0.015
Standard deviation of real GDP (1983–2007)	0.018	0.018
Correlation between inflation and real GDP (1983–2007)	0.209	0.209

Notes: The table presents moments of the data and simulated series from the model parameterized at the baseline values in Table 2. See Section 4.3 for details.

Non-targeted moments. To examine the model's ability to capture the out-of-sample behavior of GDP and inflation, following Maćkowiak and Wiederholt (2015), we compare the implied variance–covariance matrix of GDP and inflation for the pre-Volcker era with the one measured from the U.S. data.

To do so, we first replace the parameters related to monetary policy with the pre-Volcker era estimates. Specifically, we replace the estimates of the Taylor rule for the post-Volcker period with our estimates for the pre-Volcker period. Furthermore, we re-estimate the standard deviation of monetary policy shocks (σ_u) using the pre-Volcker period monetary policy shock series from Romer and Romer (2004). As shown in Panel C of Table 2, monetary policy is less responsive to inflation and output growth in the pre-Volcker period than in the post-Volcker period. Also, the monetary shock is more volatile in the pre-Volcker period than in the post-Volcker period.

We then simulate the model under the calibrated values for the cost of attention and the process for the TFP shocks and calculate the variance–covariance matrix for GDP and inflation. Table 4 reports the model-generated moments and their analogs in the data. While we only target the volatility of inflation and GDP for the post-Volcker period, our model is able to match the high volatility of inflation and GDP in the pre-Volcker period as a consequence of more dovish monetary policy during that period.

Furthermore, for both the pre- and post-Volcker parameterization of monetary policy, Figure 5 shows the impulse response functions of the variables of the model with respect to one standard-deviation TFP and monetary policy shocks. The main takeaway from

Table 4: Non-targeted Moments

Moment	**Data**	**Model**
Standard deviation of inflation (1969-1978)	0.025	0.025
Standard deviation of real GDP (1969-1978)	0.022	0.020
Correlation between inflation and real GDP (1969-1978)	0.242	0.245

Notes: The table compares the volatility of inflation and output gap and their correlation in the US data for the pre-Volcker era to the counterparts from the counterfactual model simulation. See Section 4.4 for details.

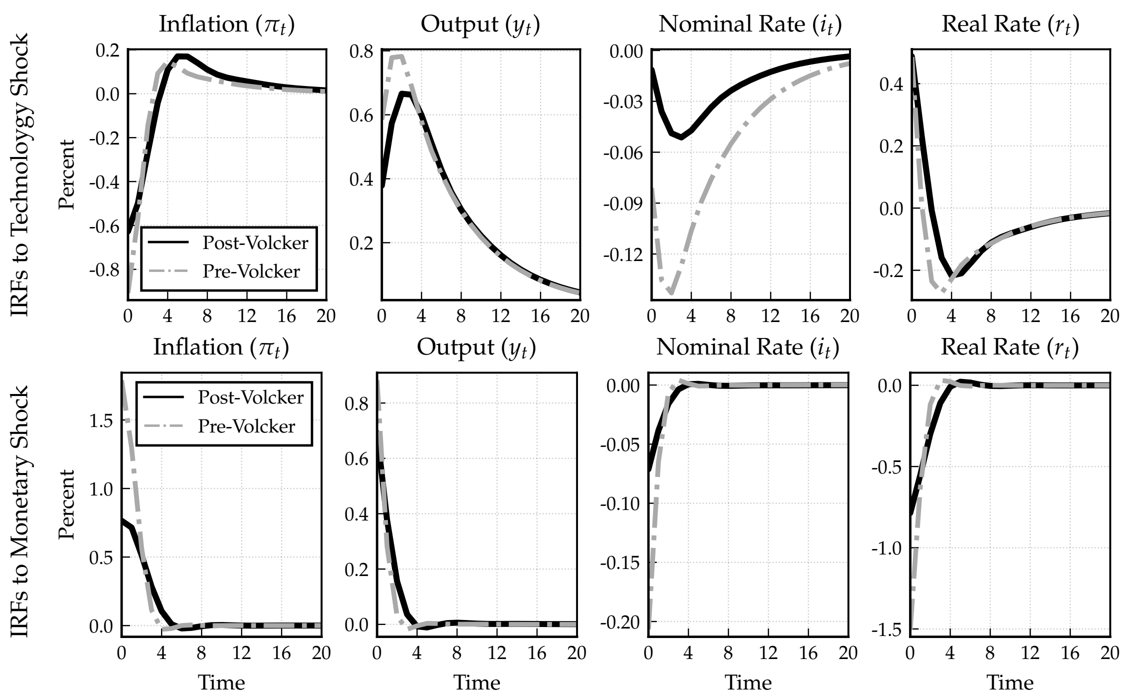

Figure 5: Impulse Responses to Technology and Monetary Shocks

Notes: This figure plots impulse responses of inflation, output, nominal rates, and real interest rates to a one standard deviation shock to technology (upper panels) and those to a one standard deviation shock to monetary policy (lower panels). Solid black lines are the responses in the model with the post-Volcker calibration while dashed gray lines are the responses in the model with the pre-Volcker calibration.

these IRFs is that inflation, output and nominal as well as real interest rates respond more to shocks under the pre-Volcker parameterization of monetary policy.

It is important to note that a change in the slope of the Phillips curve is neither a necessary condition for higher volatility of inflation and output under a more dovish monetary policy nor is it necessarily a consequence of it (Clarida et al., 2000). The takeaway from our exercise in matching these moments is to validate our model quantitatively. Whether the model can match the change in the slope of the Phillips curve is a different question that we investigate in the remainder of this section.

4.5 Quantification of the Change in the Slope of the Phillips Curve

Because the slope of the Phillips curve is endogenous in the model, the change in the rule of monetary policy in the post-Volcker period would lead to an endogenous change in the slope of the Phillips curve. The main question we try to answer in this section is, are the estimated monetary policy parameters for the pre- and post-Volcker periods consistent with a flatter Phillips curve in later periods within the model, and if so is the mechanism quantitatively relevant?

The main challenge here is to constitute the right comparison between the model and the empirical estimates of the slope of the Phillips curve. While the empirical literature that documents the change in the slope of the Phillips curve uses the New Keynesian Phillips curve (NKPC) as the equation guiding their empirical strategy, our model has a different specification for the Phillips curve that does not necessarily comply with the NKPC formulation. In particular, our model subscribes that one should control for the forecast errors of firms regarding the output gap and inflation at different horizons, which stems from their endogenous information acquisition strategy.

While the ideal case would be to re-estimate the Phillips curve based on the specification subscribed by our model, such a strategy requires a time-series on firms' expectations that does not exist for the U.S. to our knowledge. Even in countries where a time-series of firms' expectations exists, such as in Italy for instance, the data do not go back in time enough to capture variations in the rule of monetary policy.

The alternative strategy that we employ here, which allows us to compare the predictions of our model to the empirical literature on the slope of the Phillips curve, is to simulate data from our model under the two specifications of monetary policy in the pre- and post-Volcker periods, and run similar regressions as in the empirical literature. While these regressions are mis-specified from the perspective of our model and are biased due to omitted variables, namely firms' expectations, they constitute a fair comparison for the estimates from the model and in the data.

Formally, we simulate the model for 50,000 periods for both the pre- and post-Volcker periods and estimate the following hybrid NKPC using GMM estimation.

$$\pi_t = constant + \gamma \mathbb{E}_t[\pi_{t+1}] + (1-\gamma)\pi_{t-1} + \kappa x_t + \varepsilon_t.$$

We use four lags of both inflation and output gap as instruments. Table 5 shows the parameter estimates of the NKPC using the simulated data from our model.

Table 5: Estimates of the New Keynesian Phillips Curve Using Simulated Data

	(1) Output gap		(2) Output		(3) Adj. output gap	
	Pre-Volcker	Post-Volcker	Pre-Volcker	Post-Volcker	Pre-Volcker	Post-Volcker
Slope of NKPC (κ)	1.160***	0.304***	0.035***	0.027***	0.024***	-0.012***
	(0.029)	(0.007)	(0.001)	(0.001)	(0.007)	(0.003)
Forward-looking (γ)	0.666***	0.612***	0.549***	0.499***	0.554***	0.512***
	(0.005)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)

Notes: This table shows the estimation results of the New Keynesian Phillips curve using simulated data from the baseline model presented in Section 4.2. Column (1) and (2) show the estimates of the New Keynesian Phillips curve (4.5) using the simulated output gap and output data, respectively. Column (3) shows the estimates using the simulated output gap data, which are adjusted by subtracting moving averages of natural level of output from actual output. Four lags of inflation and output gap (or output) are used as instruments for the GMM estimation. A constant term is included in the regressions but not reported. Newey-West standard errors are reported in parentheses. ***, **, * denotes statistical significance at 1%, 5%, and 10% levels respectively.

Another potential issue that the literature has noted and we need to address is the lack of refined measurements for the output gap in the data. In particular, as pointed out by McLeay and Tenreyro (2020), if one fails to fully control for the supply shocks in estimating the Phillips curve, the estimates for the slope are going to be downwardly biased. More importantly, a more hawkish monetary policy would induce a larger downward bias than a dovish monetary policy. In order to take this force into account, we estimate the equation above with three different measures of the output gap that vary the extent to which we control for the supply shocks.

Column (1) in Table 5 shows the estimates of the NKPC when we use the true output gap (fully controlling for supply shocks). In this scenario, the model predicts that the slope of the Phillips curve declined from 1.16 in the pre-Volcker era to 0.30 in the post-Volcker period – a 75% decline. The benefit of this specification is that we are controlling for the true output gap, which eliminates concerns regarding the downward bias induced by the omitted variables bias. Therefore, the entire decline in the slope that we observe in this specification is due to the change in the information acquisition incentives of firms.

Nonetheless, the potential issue with this specification is that it does not directly relate to the empirical estimates, as fully controlling for supply shocks is not feasible in the data. In fact, the large magnitude of the estimates and the significantly positive slope, even for the post-Volcker era, suggests that our model is not over-explaining the decline in the slope of the Phillips curve.

To illustrate this point further, Column (2) in Table 5 reports the estimated hybrid NKPC when we use the output minus the steady-state output as our measure of the output gap – fully omitting the supply shocks. In this case, the estimated slope for both periods is much smaller compared to the estimates in Column (1), but there is still a 25% decline in the slope of the Phillips curve from the pre- to post-Volcker period.

Finally, we consider an interim case in Column (3) where we partially control for the supply shocks by subtracting a moving average of the natural level of output from realized output in the model to construct the output gap. Again, the model predicts a decline in the slope of the Phillips curve from the pre- to post-Volcker period.[37]

5 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 Chen Lian, "Incomplete information in macroeconomics: Accommodating frictions in coordination," in "Handbook of macroeconomics," Vol. 2, Elsevier, 2016, pp. 1065–1240.

_ **and** _ , "Forward guidance without common knowledge," *American Economic Review*, 2018, *108* (9), 2477–2512.

[37] Appendix Table A.2 shows the estimates of both standard forward-looking NKPC and (unrestricted) hybrid NKPC using different measures of the output gap from the simulated data. In all cases, the slope of the NKPC declined from the pre- to post-Volcker era.

— **and Jennifer La'O**, "Incomplete Information, Higher-Order Beliefs and Price Inertia," *Journal of Monetary Economics*, 2009, *56*, S19–S37.

— **and Zhen Huo**, "Myopia and Anchoring," Working Paper 24545, National Bureau of Economic Research 2018.

—, —, **and Karthik A Sastry**, "Imperfect Macroeconomic Expectations: Evidence and Theory," Technical Report, National Bureau of Economic Research 2020.

Aruoba, S. Borağan, Luigi Bocola, and Frank Schorfheide, "Assessing DSGE model nonlinearities," *Journal of Economic Dynamics and Control*, 2017, *83* (C), 34–54.

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.

Bernanke, Ben, "Inflation expectations and inflation forecasting," Technical Report, Board of Governors of the Federal Reserve System (US) 2007.

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.

Clarida, Richard, Jordi Galí, and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, 02 2000, *115* (1), 147–180.

Coibion, Olivier and Yuriy Gorodnichenko, "Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation," *American Economic Review*, 2011, *101* (1), 341–70.

— **and** —, "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.

__, __, Lorenz Kueng, and John Silvia, "Innocent Bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 2017, *88*, 70 – 89.

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.

Del Negro, Marco, Michele Lenza, Giorgio E Primiceri, and Andrea Tambalotti, "What's up with the Phillips Curve?," Working Paper 27003, National Bureau of Economic Research April 2020.

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.

__, "A Behavioral New Keynesian Model," Working Paper 22954, National Bureau of Economic Research 2016.

Gaglianone, Wagner Piazza, Raffaella Giacomini, Joao Issler, and Vasiliki Skreta, "Incentive-driven Inattention," Discussion Paper 13619, CEPR 2019.

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.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson, "The Slope of the Phillips Curve: Evidence from US States," Technical Report, National Bureau of Economic Research 2020.

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, "Choice simplification: A theory of mental budgeting and naive diversification," *The Quarterly Journal of Economics*, 2020, *135* (2), 1153–1207.

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.

Levy, Daniel, Mark Bergen, Shantanu Dutta, and Robert Venable, "The Magnitude of Menu Costs: Direct Evidence From Large U. S. Supermarket Chains," *The Quarterly Journal of Economics*, 1997, *112* (3), 791–825.

L'Huillier, Jean-Paul and William R Zame, "Flattening of the Phillips Curve and the Mandate of the Central Bank," Technical Report 2020. Mimeo.

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.

_ , _ , and _ , "Survey: Rational inattention, a disciplined behavioral model," CEPR Discussion Paper DP13243, CEPR 2018.

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.

McLeay, Michael and Silvana Tenreyro, "Optimal inflation and the identification of the Phillips curve," *NBER Macroeconomics Annual*, 2020, *34* (1), 199–255.

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," January 2020. Boston University - Department of Economics.

Mishkin, Frederic S, "Inflation dynamics," *International Finance*, 2007, *10* (3), 317–334.

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.

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.

Nieuwerburgh, Stijn Van and Laura Veldkamp, "Information Acquisition and Under-Diversification," *The Review of Economic Studies*, 2010, *77* (2), 779–805.

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.

Powell, Jerome, "Semiannual Monetary Policy Report to the Congress," 7 2019. Before the Committee on Financial Services, U.S. House of Representatives, Washington, D.C.

Ravid, Doron, "Bargaining with Rational Inattention," 2019. Manuscript.

Reis, Ricardo, "Inattentive Producers," *The Review of Economic Studies*, 2006, *73* (3), 793–821.

Romer, Christina D. and David H. Romer, "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, September 2004, *94* (4), 1055–1084.

Roth, Christopher and Johannes Wohlfart, "How Do Expectations about the Macroeconomy Affect Personal Expectations and Behavior?," CESifo Working Paper 7154, CESifo 2018.

Rubbo, Elisa, "Networks, Phillips Curves and Monetary Policy," Technical Report, Harvard University 2020. Mimeo.

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 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: Micro Evidence and Aggregate Implications,” 2019. Manuscript.

Zorn, Peter, “Investment under Rational Inattention: Evidence from US Sectoral Data,” 2016. Manuscript.

APPENDIX

A Proofs for Section 2

A.1 Proof of Lemma 2.1

Proof. 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,

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

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

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

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

A.2 Proof of Lemma 2.2

Proof. 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

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

∎

A.3 Proof of Lemma 2.3

Proof. 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 $t \geq 0$,

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

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.[38] ■

A.4 Proof of Lemma 2.4

Proof. We know from Lemma 2.3 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 2.2 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:

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

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

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

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

[38] See Chapter 12 in Cover and Thomas (2012).

Note that positive semi-definiteness of $\mathbf{\Sigma}_{z,t}$ implies that $\mathbf{\Sigma}_{t|t} \preceq \mathbf{\Sigma}_{t|t-1}$. Furthermore, note that for any posterior $\mathbf{\Sigma}_{t|t} \preceq \mathbf{\Sigma}_{t|t-1}$ 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 $\mathbf{\Sigma}_{v,t} \in \mathbb{S}_+^m$ 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. Additionally, observe that given a^t:

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

Thus, the 2.1 becomes:

$$\begin{aligned}
&\sup_{\{\mathbf{\Sigma}_{t|t} \in \mathbb{S}_+^n\}_{t\geq 0}} -\frac{1}{2}\sum_{t=0}^{\infty}\beta^t \left[tr(\mathbf{\Sigma}_{t|t}\mathbf{\Omega}) + \omega \ln\left(\frac{|\mathbf{\Sigma}_{t|t-1}|}{|\mathbf{\Sigma}_{t|t}|}\right)\right] \\
s.t. \quad & \mathbf{\Sigma}_{t+1|t} = \mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{Q}\mathbf{Q}', \quad \forall t \geq 0, \\
& \mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq 0, \quad \forall t \geq 0 \\
& 0 \prec \mathbf{\Sigma}_{0|-1} = var(\vec{x}_0|S^{-1}) \prec \infty \quad \text{given.}
\end{aligned}$$

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 $\{\mathbf{\Sigma}_{t|t} \in \mathbb{S}_+^n | 0 \preceq \mathbf{\Sigma}_{t|t} \preceq \mathbf{\Sigma}_{t|t-1}\}$ is a compact subset of the positive semidefinite cone. ■

A.5 Proof of Proposition 2.1

Proof. 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 eig(.) denotes the vector of eigenvalues of a matrix.

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

But notice that

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

where diag(.) 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

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

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

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

Moreover, note that complementary slackness for this constraint requires:

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

re-writing the Lagrangian we get:

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

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

$$\begin{aligned}
\mathbf{\Omega} - \omega\mathbf{\Sigma}_{t|t}^{-1} + \mathbf{A}'\mathbf{\Gamma}_t\mathbf{A} + \mathbf{\Lambda}_t = 0 \\
\omega\beta\mathbf{\Sigma}_{t+1|t}^{-1} - \mathbf{\Gamma}_t - \beta\mathbf{\Lambda}_{t+1} = 0
\end{aligned}$$

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,

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

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

$$\ln\left(\frac{\det(\boldsymbol{\Sigma}_{t|t-1})}{\det(\boldsymbol{\Sigma}_{t|t})}\right) < \infty \Rightarrow \det(\boldsymbol{\Sigma}_{t|t}) > 0$$

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

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

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 $\boldsymbol{\Gamma}_t$ in the first order conditions we get the conditions in the theorem. Moreover, we have a terminal optimality condition that requires:

$$\lim_{T\to\infty} \beta^T tr(\boldsymbol{\Gamma}_T\boldsymbol{\Sigma}_{T+1|T}) \geq 0 \Leftrightarrow \lim_{T\to\infty} \beta^{T+1} tr(\boldsymbol{\Lambda}_{T+1}\boldsymbol{\Sigma}_{T+1|T}) \leq 0$$

Since both $\boldsymbol{\Lambda}_T$ and $\boldsymbol{\Sigma}_{T+1|T}$ are positive semidefinite, we also have $tr(\boldsymbol{\Lambda}_{T+1}\boldsymbol{\Sigma}_{T+1|T}) \geq 0$. Thus, TVC becomes:

$$\lim_{T\to\infty} \beta^{T+1} tr(\boldsymbol{\Lambda}_{T+1}\boldsymbol{\Sigma}_{T+1|T}) = 0$$

∎

A.6 Proof of Theorem 2.1

Proof. From the FOC in Proposition 2.1 observe that

$$\omega\boldsymbol{\Sigma}_{t|t}^{-1} = \boldsymbol{\Omega}_t + \boldsymbol{\Lambda}_t \Rightarrow \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t} = \boldsymbol{\Sigma}_{t|t-1} - \omega(\boldsymbol{\Omega}_t + \boldsymbol{\Lambda}_t)^{-1}.$$

For ease of notation let $\mathbf{X}_t \equiv \boldsymbol{\Sigma}_{t|t-1} - \boldsymbol{\Sigma}_{t|t}$. Multiplying the above equation by $\boldsymbol{\Omega}_t + \boldsymbol{\Lambda}_t$ from right we get

$$\mathbf{X}_t\boldsymbol{\Omega}_t - \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\Lambda}_t = \boldsymbol{\Sigma}_{t|t-1}\boldsymbol{\Omega}_t - \omega\mathbf{I},$$

where we have imposed the complementarity slackness $\mathbf{X}_t\mathbf{\Lambda}_t = 0$. Finally, multiply this equation by $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$ from right and $\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}$ from left.[39] We have

$$(\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{X}_t\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}})(\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}) - \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Lambda}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} = \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} - \omega\mathbf{I}$$

Where $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} = \mathbf{U}_t\mathbf{D}_t\mathbf{U}_t'$ is the spectral decomposition stated in the Theorem. Now, for ease of notation let

$$\hat{\mathbf{X}}_t \equiv \mathbf{U}_t'\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{X}_t\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{U}_t$$
$$\hat{\mathbf{\Lambda}}_t \equiv \mathbf{U}_t'\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Lambda}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{U}_t$$

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

$$\hat{\mathbf{X}}_t\mathbf{D}_t - \hat{\mathbf{\Lambda}}_t = \mathbf{D}_t - \omega\mathbf{I}$$

Now, notice that $\mathbf{X}_t$ and $\mathbf{\Lambda}_t$ are simultaneously diagonalizable if and only if $\hat{\mathbf{X}}_t$ and $\hat{\mathbf{\Lambda}}_t$ are simultaneously diagonalizable. Combined with complementarity slackness, this implies $\hat{\mathbf{\Lambda}}_t\hat{\mathbf{X}}_t = \hat{\mathbf{X}}_t\hat{\mathbf{\Lambda}}_t = 0$. Similarly, note that $\mathbf{X}_t$ and $\mathbf{\Lambda}_t$ are positive semidefinite if and only if $\hat{\mathbf{X}}_t$ and $\hat{\mathbf{\Lambda}}_t$ are positive semidefinite, respectively. So we need for two simultaneously diagonalizable symmetric positive semidefinite matrices $\hat{\mathbf{\Lambda}}_t$ and $\hat{\mathbf{X}}_t$ that solve Equation A.6.

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

$$(\hat{\mathbf{X}}_t - \mathbf{I})\mathbf{D}_t = \hat{\mathbf{\Lambda}}_t - \omega\mathbf{I}$$

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

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

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 $\hat{\mathbf{\Lambda}}_t$ and $\hat{\mathbf{X}}_t$ are diagonal matrices. Using complementarity slackness $\hat{\mathbf{\Lambda}}_t\hat{\mathbf{X}}_t = \mathbf{0}$, feasibility constraint $\hat{\mathbf{X}}_t \succeq \mathbf{0}$, and dual feasibility constraint

[39] $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$ exists since $\mathbf{\Sigma}_{t|t-1}$ is positive semidefinite and $\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}$ exists since we assumed that the initial prior is strictly positive definite.

$\hat{\mathbf{\Lambda}}_t \succeq \mathbf{0}$ it is straight forward to show that $\mathbf{\Lambda}_t$ is strictly positive for the eigenvalues (entries on the diagonal) of $\mathbf{D}_t$ that are smaller than ω.

$$\hat{\mathbf{\Lambda}}_t = \mathrm{Max}(\omega\mathbf{I}-\mathbf{D}_t, \mathbf{0})$$

Now, using Equation A.6 we get:

$$\mathbf{\Lambda}_t = \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{U}_t\,\mathrm{Max}(\omega\mathbf{I} - \mathbf{D}_t, \mathbf{0})\mathbf{U}_t'\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}$$

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

$$\begin{aligned}
\omega\mathbf{\Sigma}_{t|t}^{-1} &= \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{U}_t\mathbf{D}_t\mathbf{U}_t'\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} + \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{U}_t\,\mathrm{Max}(\omega\mathbf{I} - \mathbf{D}_t, \mathbf{0})\mathbf{U}_t'\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \\
&= \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{U}_t\,\mathrm{Max}(\omega\mathbf{I}, \mathbf{D}_t)\mathbf{U}_t'\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \\
&= \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\,\mathrm{Max}(\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}, \omega)\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}
\end{aligned}$$

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 $\mathbf{\Lambda}_t$ in Equation A.6 we have:

$$\begin{aligned}
\mathbf{\Omega}_t &= \mathbf{\Omega} + \beta\mathbf{A}'(\omega\mathbf{\Sigma}_{t+1|t}^{-1} - \mathbf{\Lambda}_{t+1})\mathbf{A} \\
&= \mathbf{\Omega} + \beta\mathbf{A}'\mathbf{\Sigma}_{t+1|t}^{-\frac{1}{2}}(\omega\mathbf{I} - \mathbf{U}_t\,\mathrm{Max}(\omega\mathbf{I} - \mathbf{D}_t, \mathbf{0}))\mathbf{\Sigma}_{t+1|t}^{-\frac{1}{2}}\mathbf{A} \\
&= \mathbf{\Omega} + \beta\mathbf{A}'\mathbf{\Sigma}_{t+1|t}^{-\frac{1}{2}}\mathbf{U}_t\,\mathrm{Min}\,(\mathbf{D}_t, \omega\mathbf{I})\mathbf{U}_t'\mathbf{\Sigma}_{t+1|t}^{-\frac{1}{2}}\mathbf{A} \\
&= \mathbf{\Omega} + \beta\mathbf{A}'\mathbf{\Sigma}_{t+1|t}^{-\frac{1}{2}}\,\mathrm{Min}(\mathbf{\Sigma}_{t+1|t}^{\frac{1}{2}}\mathbf{\Omega}_{t+1}\mathbf{\Sigma}_{t+1|t}^{\frac{1}{2}}, \omega)\mathbf{\Sigma}_{t+1|t}^{-\frac{1}{2}}\mathbf{A}
\end{aligned}$$

■

A.7 Proof of Theorem 2.2

Proof. The upper bound m directly follows from Lemma 2.1. Recall from part 2 of Lemma 2.2 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

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

where $\vec{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{z,t})$ 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

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

where $\mathbf{K}_t \equiv \mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t(\mathbf{Y}_t'\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t + \mathbf{\Sigma}_{z,t})^{-1}$ 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:

$$\mathbf{\Sigma}_{t|t} = \mathbf{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{Y}_t'\mathbf{\Sigma}_{t|t-1} \Rightarrow \mathbf{K}_t\mathbf{Y}_t' = \mathbf{I} - \mathbf{\Sigma}_{t|t}\mathbf{\Sigma}_{t|t-1}^{-1}$$

Let $\mathbf{U}_t\mathbf{D}_t\mathbf{U}_t'$ denote the spectral decomposition of $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$. Then, using Theorem 2.1, we have:

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

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 $\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{U}_t$. Notice that for any i, $\mathbf{y}_{i,t} = \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{u}_{i,t}$ is an eigenvector for $\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}$:

$$\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}\mathbf{y}_{i,t} = \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}(\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}})\mathbf{u}_{i,t} = d_{i,t}\mathbf{\Sigma}_{i,t}^{-\frac{1}{2}}\mathbf{u}_{i,t} = d_{i,t}\mathbf{y}_{i,t}$$

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 $\mathbf{\Sigma}_{z,t}^+$ be the corresponding principal minor of $\mathbf{\Sigma}_{z,t}$. Then,

$$\begin{aligned}
\mathbf{Y}_t(\mathbf{Y}_t'\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t+\mathbf{\Sigma}_{z,t})^{-1}\mathbf{Y}_t' &= \sum_{i=1}^{n}\max(0,1-\frac{\omega}{d_{i,t}})\mathbf{y}_{i,t}\mathbf{y}_{i,t}' \\
&= \sum_{d_{i,t}\geq\omega}(1-\frac{\omega}{d_{i,t}})\mathbf{y}_{i,t}\mathbf{y}_{i,t}' \\
&= \mathbf{Y}_t^{+}(\mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t^{+}+\mathbf{\Sigma}_{z,t}^{+})^{-1}\mathbf{Y}_t^{+'}
\end{aligned}$$

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

$$\mathbf{y}_{i,t}'\mathbf{\Sigma}_{t|t-1}\mathbf{y}_{j,t} = \begin{cases}\mathbf{u}_{i,t}'\mathbf{u}_{i,t}=1 & \text{if } i=j\\ \mathbf{u}_{i,t}'\mathbf{u}_{j,t}=0 & \text{if } i\neq j\end{cases}$$

Thus, $\mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t^{+} = \mathbf{I}_k$ where $\mathbf{I}_k$ is the k-dimensioanl identity matrix with k being the number of eigenvalues in $\mathbf{D}_t$ that are larger than ω. Combining this with Equation A.7 we have:

$$\begin{aligned}
\mathbf{\Sigma}_{t|t-1}-\mathbf{\Sigma}_{t|t} &= \mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t^{+}(\mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t^{+}+\mathbf{\Sigma}_{z,t}^{+})^{-1}\mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t-1} \\
\Rightarrow\mathbf{Y}_t^{+'}(\mathbf{\Sigma}_{t|t-1}-\mathbf{\Sigma}_{t|t})\mathbf{Y}_t^{+} &= \mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t^{+}(\mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t^{+}+\mathbf{\Sigma}_{z,t}^{+})^{-1}\mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t^{+}
\end{aligned}$$
$$(\mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t-1}\mathbf{Y}_t^{+}) \Rightarrow\mathbf{\Sigma}_{z,t}^{+} = (\mathbf{I}_k-\mathbf{Y}_t^{+'}\mathbf{\Sigma}_{t|t}\mathbf{Y}_t^{+})^{-1}-\mathbf{I}_k$$

Plugging in for $\mathbf{\Sigma}_{t|t}$ from the 2.12 we have:

$$\mathbf{\Sigma}_{z,t}^{+} = (\mathbf{I}_k-\omega(\mathbf{D}_t^{+})^{-1})^{-1}-\mathbf{I}_k = (\omega^{-1}\mathbf{D}_t^{+}-\mathbf{I}_k)^{-1}$$

Note that $\mathbf{\Sigma}_{z,t}^{+}$ 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:

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

■

A.8 Proof of Proposition 2.2

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

$$\mathbf{\Sigma}_{t|t} = \mathbf{\Sigma}_{t|t-1} - \mathbf{K}_t \mathbf{Y}_t' \mathbf{\Sigma}_{t|t-1} \Rightarrow \mathbf{K}_t \mathbf{Y}_t' = \mathbf{I} - \mathbf{\Sigma}_{t|t} \mathbf{\Sigma}_{t|t-1}^{-1}.$$

Plugging this into Equation A.7, multiplying it by $\mathbf{H}'$ from left, and substituting $\vec{a}_t = \mathbf{H}'\mathbb{E}[\vec{x}|a^t]$ we have:

$$\vec{a}_t - \mathbb{E}[\vec{a}_t|a^{t-1}] = \mathbf{H}'(\mathbf{I} - \mathbf{\Sigma}_{t|t}\mathbf{\Sigma}_{t|t-1}^{-1})(\vec{x}_t - \mathbb{E}[\vec{x}_t|a^{t-1}]) + \mathbf{H}'\mathbf{K}_t\vec{z}_t$$

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

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

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 $\mathbf{\Sigma}_{z,t}$ we get:

$$\mathbf{\Sigma}_{z,t} = \mathbf{H}'(\mathbf{\Sigma}_{t|t} - \mathbf{\Sigma}_{t|t}\mathbf{\Sigma}_{t|t-1}^{-1}\mathbf{\Sigma}_{t|t})\mathbf{H}$$

∎

B Replications

In this appendix, we present briefly two models we replicate in Section 2.3.

B.1 Replication of Maćkowiak and Wiederholt (2009a)

The rational inattention problem in Maćkowiak and Wiederholt (2009a) is

$$\min_{\{\hat{\Delta}_{i,t},\hat{z}_{i,t}\}} \left\{ E\left[(\Delta_t - \hat{\Delta}_{i,t})^2\right] + \underbrace{\left(\frac{\hat{\pi}_{14}}{\hat{\pi}_{11}}\right)^2}_{\equiv\xi} E\left[(z_{i,t} - \hat{z}_{i,t})^2\right] \right\},$$

$$\begin{aligned} s.t.\quad & \mathcal{I}(\{\Delta_t\};\{\hat{\Delta}_{i,t}\}) + \mathcal{I}(\{z_{i,t}\};\{\hat{z}_{i,t}\}) \leq \kappa, \\ & \{\Delta_t, \hat{\Delta}_{i,t}\} \perp \{z_{i,t}, \hat{z}_{i,t}\} \end{aligned}$$

where $\Delta_t \equiv p_t + \left(\frac{|\hat{\pi}_{13}|}{|\hat{\pi}_{11}|}\right)(q_t - p_t)$ is the profit-maximizing response to aggregate conditions and $z_{i,t}$ is an idiosyncratic shock. Also, $\hat{\Delta}_{i,t} \equiv E_{i,t}[\Delta_t]$ and $\hat{z}_{i,t} \equiv E_{i,t}[z_{i,t}]$ are firm i's subjective expectation of Δ_t and $z_{i,t}$, respectively. $\mathcal{I}(\cdot;\cdot)$ is Shannon's mutual information and κ is a fixed capacity of processing information. Lastly, notice that aggregate price p_t and exogenous shock processes are defined:

$$\begin{aligned} p_t &= \int_0^1 \hat{\Delta}_{i,t} di \\ q_t &= \rho q_{t-1} + \nu_t, \nu_{q,t} \sim \mathcal{N}(0, \sigma_q^2) \\ z_{i,t} &= \rho z_{i,t-1} + \nu_{z,t}, \nu_{z,t} \sim \mathcal{N}(0, \sigma_z^2). \end{aligned}$$

To solve the model using our method, we translate the problem above into a `DRIPs` structure. The most efficient way, due to the independence assumption, is to write it as the sum of two `DRIPs`: one that solves the attention problem for the idiosyncratic shock, and one that solves the attention problem for the aggregate shock which also has endogenous feedback.

Moreover, since the problem above has a fixed capacity, instead of a fixed cost of attention (ω) as in `DRIPs` pacakge, we need to iterate over ω's to find the one that corresponds with κ. Lastly, the attention problem in this model coincides with our model when $\beta = 1$. The full documentation for replication is available in `https://afrouzi.com/DRIPs.jl/dev/examples/ex3_mw2009/ex3_Mackowiak_Wiederholt_2009/`.

B.2 Replication of Maćkowiak, Matějka and Wiederholt (2018a)

We first describe the model of price-setting in Maćkowiak et al. (2018a). We solve this model with and without endogenous feedback in firms' optimal prices.

B.2.1 A Model of Price-Setting

There is a measure of firms indexed by $i \in [0,1]$. Firm i chooses its price $p_{i,t}$ at time t to track its ideal price $p^*_{i,t}$. Formally, her flow profit is

$$-(p_{i,t} - p^*_{i,t})^2$$

Without Endogenous Feedback We first consider the case without endogenous feedback in the firm's optimal price by assuming that $p^*_{i,t} = q_t$ where

$$\Delta q_t = \rho \Delta q_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2)$$

Here q_t can be interpreted as money growth or the nominal aggregate demand. Therefore, the state-space representation of the problem is

$$\begin{aligned}
\vec{x}_t &= \begin{bmatrix} q_t \\ \Delta q_t \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \rho \\ 0 & \rho \end{bmatrix}}_{\mathbf{A}} \vec{x}_{t-1} + \underbrace{\begin{bmatrix} \sigma_u \\ \sigma_u \end{bmatrix}}_{\mathbf{Q}} u_t, \\
p^*_{i,t} &= \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}'}_{\mathbf{H}} \vec{x}_t
\end{aligned}$$

Endogenous Feedback with Strategic Complementarity Now we consider the case where there is general equilibrium feedback with the degree of strategic complementarity α:

$$p^*_{i,t} = (1-\alpha) q_t + \alpha p_t$$

where

$$\begin{aligned}
\Delta q_t &= \rho \Delta q_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2) \\
p_t &\equiv \int_0^1 p_{i,t} di
\end{aligned}$$

Note that now the state space representation for $p^*_{i,t}$ is no longer exogenous and is determined in the equilibrium. However, we know that this is a Guassian process and by Wold's theorem we can decompose it to its $MA(\infty)$ representation:

$$p^*_{i,t} = \Phi(L) u_t$$

where $\Phi(.)$ is a lag polynomial and u_t is the shock to nominal demand. Here, we have basically guessed that the process for $p^*_{i,t}$ is determined uniquely by the history of monetary shocks which requires that rational inattention errors of firms are orthogonal. Our objective is to find $\Phi(.)$.

Since we cannot put $MA(\infty)$ processes in the computer, we approximate them with truncation. In particular, we know for stationary processes, we can arbitrarily get close to the true process by truncating $MA(\infty)$ processes to $MA(T)$ processes. Our problem here is that $p^*_{i,t}$ has a unit root and is not stationary. We can bypass this issue by re-writing the state space in the following way:

$$p^*_{i,t} = \phi(L)\tilde{u}_t, \quad \tilde{u}_t = (1-L)^{-1}u_t = \sum_{j=0}^{\infty} u_{t-j}$$

here $\tilde{u}_{t-j}$ is the unit root of the process and basically we have differenced out the unit root from the lag polynomial, and $\phi(L) = (1-L)\Phi(L)$. Notice that since the original process was difference stationary, differencing out the unit root means that $\phi(L)$ is now in ℓ_2, and the process can now be approximated arbitrarily precisely with truncation.

B.2.2 A Business Cycle Model with News Shocks

In this subsection, we describe the business cycle model with news shocks in Section 7 in Maćkowiak et al. (2018a).

The techonology shock follows AR(1) process:

$$z_t = \rho z_{t-1} + \sigma \varepsilon_{t-k}$$

and the total labor input is:

$$n_t = \int_0^1 n_{i,t} di.$$

Under perfect information, the households chooses the utility-maximizing labor supply, all firms choose the profit-maximizing labor input, and the labor market clearing condition is:

$$\frac{1-\gamma}{\psi+\gamma} w_t = \frac{1}{\alpha}(z_t - w_t).$$

Then, the market clearing wages and the equilibrium labor input are:

$$w_t = \frac{\frac{1}{\alpha}}{\frac{1-\gamma}{\psi+\gamma} + \frac{1}{\alpha}} z_t \equiv \xi z_t$$
$$n_t = \frac{1}{\alpha}(1-\xi) z_t.$$

Firms are rationally inattentive and want to keep track of their ideal price,

$$n_t^* = \frac{1}{\alpha} z_t - \frac{1}{\alpha}\frac{\psi+\gamma}{1-\gamma} n_t$$

where $n_t = \int_0^1 n_{i,t} di$. Then, firm i's choice depends on its information set at time t:

$$n_{i,t} = E_{i,t}[n_t^*].$$

Note that now the state space representation for n_t^* is determined in the equilibrium. As we describe above, we can decompose it to its $MA(\infty)$ representation by Wold's theorem:

$$n_t^* = \Phi(L)\varepsilon_t$$

where $\Phi(.)$ is a lag polynomial and ε_t is the shock to technology. We have again guessed that the process for n_t^* is determined uniquely by the history of technology shocks. Then, we transform the problem to a state space representation. The full documentation for replication is available in https://afrouzi.com/DRIPs.jl/dev/examples/ex5_mmw2018/ex5_Mackowiak_Matejka_Wiederholt_2018/.

C Proofs for Section 3

C.1 Proof of Lemma 3.1

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

$$i_t = \rho + \mathbb{E}_t[\Delta q_{t+1}]$$

Combining this with the monetary policy rule, we have

$$\Delta q_t = \phi^{-1}\mathbb{E}_t^f[\Delta q_{t+1}] + \frac{\sigma_u}{\phi} u_t$$

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

C.2 Proof of Proposition 3.1

Proof. **Part 1.** For ease of notation we drop the firm index i in the proof. The FOC in Proposition 2.1 in this case reduces to

$$\lambda_t = 1 - \theta + \frac{\omega}{\sigma^2_{t|t}} - \frac{\beta\omega}{\sigma^2_{t+1|t}} + \beta\lambda_{t+1}$$

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 $\sigma^2_{t+1|t} = \sigma^2_{t|t} + \sigma^2_u\phi^{-2}$, the FOC becomes:

$$\sigma^4_{t|t} + \left[\frac{\sigma^2_u}{\phi^2} - (1-\beta)\frac{\omega}{\theta-1}\right]\sigma^2_{t|t} - \frac{\omega}{\theta-1}\frac{\sigma^2_u}{\phi^2} = 0$$

Note that given the values of parameters, this equation does not depend on any other variable than $\sigma^2_{t|t}$ (in particular it is independent of the state $\sigma^2_{t|t-1}$). Hence, for any t, if $\lambda_t = 0$, then the $\sigma^2_{t|t} = \underline{\sigma}^2$, 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 $\underline{\sigma}^2 \leq \sigma^2_{t|t-1}$. Thus,

$$\sigma^2_{t|t} = \min\{\sigma^2_{t|t-1}, \underline{\sigma}^2\}.$$

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

$$\sigma^2_{i,t|t} = (1-\kappa_{i,t})\sigma^2_{i,t|t-1} \Rightarrow \kappa_{i,t} = 1 - \min\{1, \frac{\underline{\sigma}^2}{\sigma^2_{i,t|t-1}}\} = \max\{0, 1 - \frac{\underline{\sigma}^2}{\sigma^2_{i,t|t-1}}\}.$$

■

C.3 Proof of Corollary 3.1

Proof. Follows from differentiating the expression for $\underline{\sigma}^2$ in Proposition 3.1. ■

C.4 Proof of Proposition 3.2

Proof. **Part 1.** Recall from the proof of Proposition 3.1 that

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

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:

$$\pi_t = \frac{\kappa_t}{1-\kappa_t} y_t.$$

Plug in κ_t from Equation C.2 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 $\sigma^2_{T|T-1} \geq \underline{\sigma}^2$, then $\forall t \geq T+1$, $\sigma^2_{t|t} = \underline{\sigma}^2$ and $\sigma^2_{t|t-1} = \underline{\sigma}^2 + \sigma_u^2 \phi^{-2}$. 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. ■

C.5 Proof of Corollary 3.2

Proof. The jump to the new steady state follows from the result in Corollary 3.1 that $\underline{\sigma}^2$ increases with $\frac{\sigma_u}{\phi}$. The comparative statics follow from the fact that κ is the positive root of

$$\beta\kappa^2 + (1 - \beta + \xi)\kappa - \xi = 0$$

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 $\frac{\sigma_u}{\phi}$. ■

C.6 Proof of Corollary 3.3

Proof. 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 3.1 that firms would wait until their uncertainty reaches this new level. Comparative statics in the steady state follow directly from Corollary 3.1. ■

C.7 Proof of Proposition 3.3

Proof. Note that in the steady state of the attention problem, inflation and nominal demand, $\vec{s}_t \equiv \begin{bmatrix} q_t \\ \pi_t \end{bmatrix}$, jointly evolve according to

$$\vec{s}_t = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1-\kappa \end{bmatrix}}_{\equiv \mathbf{A}_s} \vec{s}_{t-1} + \underbrace{\begin{bmatrix} \frac{\sigma_u}{\phi} \\ \frac{\kappa\sigma_u}{\phi} \end{bmatrix}}_{\equiv \mathbf{Q}_s} u_t$$

Moreover, given that we know that a firm's history of prices is a sufficient statistics for their information set at that time, we can solve for their belief about the vector $\vec{s}_t$ by applying the Kalman filtering:

$$\int_0^1 \mathbb{E}[\vec{s}_t | p_i^t] di = \int_0^1 \mathbb{E}[\vec{s}_t | p_i^{t-1}] di + \mathbf{K}_s (q_t - \mathbb{E}[q_t | p_i^{t-1}])$$

It follows that the steady-state covariance matrix, $\mathbf{\Sigma}_s \equiv \lim_{t\to\infty} var(\vec{s}_t | p_i^{t-1})$, solves the following Riccati equation:

$$\mathbf{\Sigma}_s = \mathbf{A}_s \mathbf{\Sigma}_s \mathbf{A}_s' - \kappa \frac{\mathbf{\Sigma}_s \mathbf{e}_1 \mathbf{e}_1' \mathbf{\Sigma}_s}{\mathbf{e}_1' \mathbf{\Sigma}_s \mathbf{e}_1}$$

where κ is the steady-state Kalman-gain of firms and $\mathbf{e}_1' \equiv (1, 0)$. The solution to this Riccati equation is given by

$$\mathbf{\Sigma}_s \equiv \begin{bmatrix} \frac{1}{\kappa} & \frac{1}{2-\kappa} \\ \frac{1}{2-\kappa} & \frac{(3-2\kappa)\kappa}{(2-\kappa)^3} \end{bmatrix} \frac{\sigma_u^2}{\phi^2}$$

which then implies that the Kalman-gain vector, $\mathbf{K}_s$ is given by

$$\begin{aligned} \mathbf{K}_s &= \kappa \frac{\mathbf{\Sigma}_s \mathbf{e}_1 \mathbf{e}_1'}{\mathbf{e}_1' \mathbf{\Sigma}_s \mathbf{e}_1} \\ &= \begin{bmatrix} \kappa \\ \frac{\kappa^2}{2-\kappa} \end{bmatrix} \mathbf{e}_1 \end{aligned}$$

Thus, noticing that the firms average inflation expectations is given by the second element of the vector $\int_0^1 \mathbb{E}[\vec{s}_t|p_i^t]di$, we have

$$\hat{\pi}_t = (1-\kappa)\hat{\pi}_{t-1} + \frac{\kappa^2}{2-\kappa}(q_t - p_{t-1}) = (1-\kappa)\hat{\pi}_{t-1} + \frac{\kappa^2}{(2-\kappa)(1-\kappa)}y_t$$

where in the second line we have plugged in $y_t \equiv q_t - p_t$ and the Phillips curve $\pi_t = \frac{\kappa}{1-\kappa}y_t$. Finally, multiplying the lag of the above equation by $1-\kappa$ and differencing them out we have

$$\begin{aligned}\hat{\pi}_t - (1-\kappa)\hat{\pi}_{t-1} &= (1-\kappa)\hat{\pi}_{t-1} - (1-\kappa)^2\hat{\pi}_{t-2} + \frac{\kappa^2}{(2-\kappa)(1-\kappa)}(y_t - (1-\kappa)y_{t-1}) \\ &= (1-\kappa)\hat{\pi}_{t-1} - (1-\kappa)^2\hat{\pi}_{t-2} + \frac{\kappa^2}{2-\kappa}\frac{\sigma_u}{\phi}u_t.\end{aligned}$$

■

C.8 Proof of Corollary 3.4

Proof. Note that the sensitivity of firms' inflation expectations to a one standard deviation shock to monetary policy ($\frac{\sigma_u}{\phi}u_t$) is

$$\frac{\partial \hat{\pi}_t}{\partial(\frac{\sigma_u}{\phi}u_t)} = \frac{\kappa^2}{2-\kappa}$$

Now, note that

$$\frac{\partial\left(\frac{\partial \hat{\pi}_t}{\partial(\frac{\sigma_u}{\phi}u_t)}\right)}{\partial\left(\frac{\sigma_u}{\phi}\right)} = \frac{4\kappa-\kappa^2}{(2-\kappa)^2} = \left[1 + \left(\frac{2}{2-\kappa}\right)^2\right]\frac{\partial\kappa}{\partial\left(\frac{\sigma_u}{\phi}\right)} < 0$$

where the negative sign follows from the fact that κ is decreasing in $\frac{\sigma_u}{\phi}$ (Corollary 3.1). ■

D Computing the Equilibrium

D.1 Matrix Representation and Solution Algorithm

Firms wants to keep track of their ideal price, $p_{i,t}^* = p_t + \alpha x_t$. Notice that the state space representation for $p_{i,t}^*$ is no longer exogenous and is determined in the equilibrium. However, we know that this is a Guassian process and by Wold's theorem we can decompose it to its $MA(\infty)$ representation:

$$p^*_{i,t} = \Phi_a(L)\varepsilon_{a,t} + \Phi_u(L)\varepsilon_{u,t}$$

where $\Phi_a(.)$ and $\Phi_u(.)$ are lag polynomials. Here, we have basically guessed that the process for $p^*_{i,t}$ is determined uniquely by the history of monetary shocks which requires that rational inattention errors of firms are orthogonal.

We cannot put $MA(\infty)$ processes in the computer and have to truncate them. However, we know that for stationary processes we can arbitrarily get close to the true process by truncating $MA(\infty)$ processes. Our problem here is that $p^*_{i,t}$ has a unit root and is not stationary. We can bypass this issue by re-writing the state space in the following way:

$$p^*_{i,t} = \Phi_a(L)\varepsilon_{a,t} + \phi_u(L)\tilde{\varepsilon}_{u,t}, \quad \tilde{\varepsilon}_{u,t} = (1-L)^{-1}\varepsilon_{u,t} = \sum_{j=0}^{\infty} \varepsilon_{u,t-j}$$

here $\tilde{\varepsilon}_{u,t}$ is the unit root of the process and basically we have differenced out the unit root from the lag polynomial, and $\phi_u(L) = (1-L)\Phi_u(L)$. Notice that since the original process was difference stationary, differencing out the unit root means that $\phi_u(L)$ is now in ℓ_2, and the process can now be approximated arbitrarily precisely with truncation.

For ease of notation, let $z_t = (\varepsilon_{a,t}, \varepsilon_{u,t})$ and $\tilde{z}_t = (\varepsilon_{a,t}, \tilde{\varepsilon}_{u,t})$. For a length of truncation L, let $\vec{x}'_t \equiv (z_t, z_{t-1}, \ldots, z_{t-(L+1)}) \in \mathbb{R}^{2L}$ and $\vec{\mathbf{x}}'_t \equiv (\tilde{z}_t, \tilde{z}_{t-1}, \ldots, \tilde{z}_{t-(L+1)}) \in \mathbb{R}^{2L}$. Notice that

$$\begin{aligned}
\vec{x}_t &= (\mathbf{I} - \mathbf{\Lambda}\mathbf{M}')\vec{\mathbf{x}}_t \\
\vec{\mathbf{x}}_t &= (\mathbf{I} - \mathbf{\Lambda}\mathbf{M}')^{-1}\vec{x}_t
\end{aligned}$$

where $\mathbf{I}$ is a $2L \times 2L$ identity matrix, $\mathbf{\Lambda}$ is a diagonal matrix where $\mathbf{\Lambda}_{(2i,2i)} = 1$ and $\mathbf{\Lambda}_{(2i-1,2i-1)} = 0$ for all $i = 1, 2, \cdots, L$, and $\mathbf{M}$ is a shift matrix:

$$\mathbf{M} = \begin{bmatrix} \mathbf{0}_{2\times(2L-2)} & \mathbf{0}_{2\times 2} \\ \mathbf{I}_{(2L-2)\times(2L-2)} & \mathbf{0}_{(2L-2)\times 2} \end{bmatrix}$$

Then, note that $p^*_{i,t} \approx \mathbf{H}'\vec{\mathbf{x}}_t$ where $\mathbf{H} \in \mathbb{R}^{2L}$ is the truncated matrix analog of the lag polynominal, and is endogenous to the problem. Our objective is to find the general equilibrium $\mathbf{H}$ along with the optimal information structure that it implies.

Moreover, note that

$$\begin{aligned}
a_t &= \mathbf{H}_a' \vec{x}_t, \quad \mathbf{H}_a' = (1, 0, \rho_a, 0, \rho_a^2, 0, \ldots, \rho_a^{L-1}, 0) \\
u_t &= \mathbf{H}_u' \vec{x}_t, \quad \mathbf{H}_u' = (0, 1, 0, \rho_u, 0, \rho_a^2, \ldots, 0, \rho_u^{L-1})
\end{aligned}$$

We will solve for $\mathbf{H}$ by iterating over the problem. In particular, in iteration $n \geq 1$, given the guess $\mathbf{H}_{(n-1)}$, we have the following state space representation for the firm's problem

$$\vec{\mathbf{x}}_t = \underbrace{\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 \end{bmatrix}}_{\mathbf{A}} \vec{\mathbf{x}}_{t-1} + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}}_{\mathbf{Q}} z_t,$$

$$p_{i,t}^* = \mathbf{H}_{(n-1)}' \vec{\mathbf{x}}_t$$

Now, note that

$$\begin{aligned}
p_t &= \int_0^1 p_{i,t} di = \mathbf{H}_{(n-1)}' \int_0^1 \mathbb{E}_{i,t}[\vec{\mathbf{x}}_t] di \\
&= \mathbf{H}_{(n-1)}' \sum_{j=0}^{\infty} [(\mathbf{I} - \mathbf{K}_{(n)} \mathbf{Y}_{(n)}') \mathbf{A}]^j \mathbf{K}_{(n)} \mathbf{Y}_{(n)}' \vec{\mathbf{x}}_{t-j} \\
&\approx \mathbf{H}_{(n-1)}' \underbrace{\left[\sum_{j=0}^{\infty} [(\mathbf{I} - \mathbf{K}_{(n)} \mathbf{Y}_{(n)}') \mathbf{A}]^j \mathbf{K}_{(n)} \mathbf{Y}_{(n)}' \mathbf{M}'^j \right]}_{\equiv \mathbf{X}_{(n)}} \vec{\mathbf{x}}_t \\
&= \mathbf{H}_{(n-1)}' \mathbf{X}_{(n)} \vec{\mathbf{x}}_t = \mathbf{H}_p' \vec{\mathbf{x}}_t
\end{aligned}$$

Let $x_t = \mathbf{H}_x' \vec{x}_t$, $i_t = \mathbf{H}_i' \vec{x}_t$, and $\pi_t = \mathbf{H}_\pi' \vec{x}_t = \mathbf{H}_p' (\mathbf{I} - \mathbf{\Lambda} \mathbf{M}')^{-1} (\mathbf{I} - \mathbf{M}') \vec{x}_t$. Then from the households Euler equation, we have:

$$\begin{aligned} x_t &= \mathbb{E}_t^f \left[x_{t+1} - \frac{1}{\sigma} \left(i_t - \pi_{t+1} \right) \right] + \mathbb{E}_t^f [y_{t+1}^n] - y_t^n \\ \implies \mathbf{H}_i &= \sigma \left(\mathbf{M}' - \mathbf{I} \right) \mathbf{H}_x + \frac{\sigma(1+\psi)}{\sigma+\psi} \left(\mathbf{M}' - \mathbf{I} \right) \mathbf{H}_a + \mathbf{M}' \mathbf{H}_\pi \end{aligned}$$

Also, the Taylor rule gives:

$$\begin{aligned} & i_t = \rho i_{t-1} + (1-\rho) \left(\phi_\pi \pi_t + \phi_x x_t + \phi_{\Delta y} \left(y_t - y_{t-1} \right) \right) + u_t \\ \implies & \left(\mathbf{I} - \rho \mathbf{M} \right) \mathbf{H}_i = (1-\rho) \phi_\pi \mathbf{H}_\pi + (1-\rho) \phi_x \mathbf{H}_x \\ & + (1-\rho) \phi_{\Delta y} \left(\mathbf{I} - \mathbf{M} \right) \left(\mathbf{H}_x + \frac{1+\psi}{\sigma+\psi} \mathbf{H}_a \right) + \mathbf{H}_u \end{aligned}$$

These give us $\mathbf{H}_x$ and $\mathbf{H}_i$ and we update new $\mathbf{H}_{(n)}$ using:

$$\mathbf{H}_{(n)} = \mathbf{H}_p + \alpha (\mathbf{I} - \mathbf{M} \mathbf{\Lambda}') \mathbf{H}_x$$

We iterate until convergence of $\mathbf{H}_{(n)}$.

E Appendix Tables

Table A.1: **Estimates of the Taylor Rule**

	constant	ρ	ϕ_π	$\phi_{\Delta y}$	ϕ_x
Pre-Volcker	0.096	0.957***	1.589*	1.028*	1.167**
(1969–1978)	(0.187)	(0.022)	(0.847)	(0.601)	(0.544)
Post-Volcker	-0.310***	0.961***	2.028***	3.122***	0.673***
(1983–2007)	(0.062)	(0.015)	(0.617)	(1.090)	(0.234)

Notes: This table reports least squares estimates of the Taylor rule. We use the Greenbook forecasts of current and future macroeconomic variables. The interest rate is the target federal funds rate set at each meeting from the Fed. The measure of the output gap is based on Greenbook forecasts. We consider two time samples: 1969–1978 and 1983–2007. Newey-West standard errors are reported in parentheses. ***, **, * denotes statistical significance at 1%, 5%, and 10% levels respectively.

Table A.2: Estimates of the New Keynesian Phillip Curve

	(1) Output gap		(2) Output		(3) Adj. output gap	
	Pre-Volcker	Post-Volcker	Pre-Volcker	Post-Volcker	Pre-Volcker	Post-Volcker
Panel A. Standard New Keynesian Phillips Curve						
Slope of NKPC (κ)	2.751***	0.846***	-0.347***	-0.231***	-0.278***	-0.057***
	(0.101)	(0.020)	(0.020)	(0.007)	(0.034)	(0.013)
Forward-looking (γ)	0.901***	0.894***	2.459***	1.649***	2.399***	1.592***
	(0.055)	(0.016)	(0.043)	(0.013)	(0.041)	(0.011)
Panel B. Hybrid New Keynesian Phillips Curve						
Slope of NKPC (κ)	1.020***	0.249***	-0.128***	-0.07***	-0.057***	-0.021***
	(0.063)	(0.012)	(0.013)	(0.004)	(0.016)	(0.005)
Forward-looking (γ_f)	0.738***	0.649***	1.420***	0.931***	1.299***	0.848***
	(0.027)	(0.006)	(0.049)	(0.016)	(0.038)	(0.010)
Backward-looking (γ_b)	0.335***	0.393***	0.304***	0.356***	0.332***	0.392***
	(0.005)	(0.003)	(0.011)	(0.007)	(0.009)	(0.004)
Panel C. Hybrid New Keynesian Phillips Curve ($\gamma_f + \gamma_b = 1$)						
Slope of NKPC (κ)	1.160***	0.304***	0.035***	0.027***	0.024***	-0.012***
	(0.029)	(0.007)	(0.001)	(0.001)	(0.007)	(0.003)
Forward-looking (γ_f)	0.666***	0.612***	0.549***	0.499***	0.554***	0.512***
	(0.005)	(0.003)	(0.002)	(0.001)	(0.002)	(0.001)

Notes: This table shows the estimates of the New Keynesian Phillips curves using simulated data from the baseline model presented in Section 4.2. Column (1) and (2) show the estimates of the New Keynesian Phillips curve using the simulated output gap and output data, respectively. Column (3) shows the estimates using the simulated output gap data, which are adjusted by subtracting moving averages of natural level of output from actual output. Panel A shows the estimates of the standard New Keynesian Phillips curve without backward-looking inflation and Panel B shows the estimates of the hybrid New Keynesian Phillips curve. Panel C shows the estimates of the hybrid New Keynesian Phillips curve with a coefficient restriction, $\gamma_f + \gamma_b = 1$. Four lags of inflation and output gap (or output) are used as instruments for the GMM estimation. A constant term is included in the regressions but not reported. Newey-West standard errors are reported in parentheses. ***, **, * denotes statistical significance at 1%, 5%, and 10% levels respectively.