Dynamic Rational Inattention and the Phillips Curve*

Hassan Afrouzi†
Columbia University

Choongryul Yang‡
Federal Reserve Board

First Draft: April, 2017
This Draft: April, 2021

[Code Repositories: Matlab, Julia] [Software Documentation] [Jupyter Notebooks]

Abstract

We study and fully characterize the dynamics of belief distributions in linear-quadratic-Gaussian rational inattention models. Building on these results, we propose a novel solution method that is orders of magnitude faster than alternative methods and is efficient enough for quantitative work. As an application, we develop an attention-driven theory of pricing where the Phillips curve slope responds endogenously to the conduct monetary policy. While more hawkish monetary policy flattens the Phillips curve, a more dovish monetary policy flattens it in the short run but leads to a steeper Phillips curve in the long run.

JEL Classification: E71, E32, D83
Keywords: Rational Inattention, Dynamic Information Acquisition, Phillips Curve

*We are grateful to Saroj Bhattarai and Olivier Coibion for their guidance and support. We are also thankful to George-Marios Angeletos and Mirko Wiederholt for their insightful discussions of this paper as well as to Eric Sims for a thoughtful discussion of an earlier version. We would also like to thank Miguel Acosta, Mark Dean, Chad Fulton, 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, Michael Woodford, Yeji Sung as well as seminar participants at 2018 and 2020 ASSA Meetings, Columbia University, SED Mexico City Conference, UT Austin, Federal Reserve Bank of Chicago, Cornell University, City University of New York, University of Wisconsin, 2020 NBER Summer Institute, VMACS Junior Conference, and the Fall 2020 NBER Monetary Economics meetings 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, the 2018 SED Meeting in Mexico City and 19th KEA International Conference in Seoul, South Korea.

†Columbia University, Department of Economics. Email: hassan.afrouzi@columbia.edu.
‡Federal Reserve Board of Governors. Email: choongryul.yang@frb.gov.

1 Introduction

Since Muth (1961), full-information rational expectations theory has grown to be an essential part of macroeconomic modeling. Nonetheless, a growing body of evidence on costly information has called for modifications to the theory that take such costs into account. To this end, the rational inattention theory (Sims, 2003, 2006, 2010) provides an appealing alternative by preserving the consistency of expectations within an optimizing framework. However, these models are notoriously complex to solve, as a result of which their implications for dynamics of belief distributions remain largely unexplored.

A central question that emerges from these models is how do distributions of optimal beliefs evolve over time from an arbitrary prior? The answer to this question has immediate practical and broad implications, from interpreting the consequences of information provision experiments that perturb initial priors to designing optimal communication policies with the public.

In this paper, we make three contributions. First, we provide an analytical *information Euler equation* that characterizes the dynamics of optimal belief distributions from any arbitrary initial prior in linear-quadratic-Gaussian (LQG) rational inattention models.[1] Second, we use our theoretical results to develop a novel method for solving these models that is several orders of magnitude faster and more accurate in characterizing the *dynamics* and the *steady-state* of belief distributions than alternative algorithms that rely on approximate solutions or value function iteration. Third, as an application of our general framework, we develop an attention-based theory of pricing, in which the slope of the Phillips curve is endogenous to the conduct of monetary policy and is occasionally flat due to transition dynamics of attention.

Dynamic rational inattention problems (DRIPs) are notoriously complex because both the state and choice variables are distributions with endogenous supports. LQG settings reduce these choice and state variables to covariance matrices of Gaussian distributions. However, even in these cases, for an n-dimensional Gaussian Markov process, the corresponding rational inattention problem has $\frac{n(n+1)}{2}$ state variables. In particular, one major complication is a set of n "no-forgetting" constraints that bind if the agent does not acquire information in the corresponding dimension.

Our first contribution is that we derive an analytical *information Euler equation* that fully characterizes the dynamics of optimal beliefs. In particular, our Euler equation provides a simple and intuitive rule for which no-forgetting constraints bind in each period by characterizing the marginal value of information in all dimensions of the state. The corresponding constraint then binds if this value is less than the marginal cost of one bit of information—which is the only parameter of the cost function for information acquisition.[2]

[1] In LQG rational inattention problems, the payoffs are quadratic, shocks are Gaussian, and the cost of information is linear in Shannon's mutual information function.

[2] In this sense, the extensive margin of rational inattention models connect with the notion of sparsity, as in Gabaix

As our second contribution, we propose a novel solution method for characterizing the transition dynamics and the steady-state of DRIPs based on iterating over our analytical information Euler equation. Our solution method is fast and can be implemented in quantitative work (e.g., our quantitative model or Song and Stern, 2020, who utilize our algorithm). Moreover, to demonstrate our toolbox's accuracy and efficiency for the steady-state information structure, we have replicated three canonical papers (Maćkowiak and Wiederholt, 2009a; Sims, 2010; Maćkowiak, Matějka, and Wiederholt, 2018) that use three different solution methods.[3] A summary of our computing times is reported in Table 1. Our computational toolbox is available for public use as the `DRIPs.m` repository for Matlab and the `DRIPs.jl` Julia package.[4] All examples and replications are available as interactive Jupyter notebooks that are accessible online with no software requirements.[5]

Application to Phillips Curve. Our third contribution is to apply our analytical framework to propose an attention-based theory of the Phillips curve. A recent growing literature documents that the slope of the Phillips curve has flattened during the last few decades.[6] While benchmark New Keynesian models would relate this flattening to changes in the model's structural parameters, in an analytical general equilibrium model 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 (i.e., when monetary policy is more hawkish), firms endogenously choose to pay less attention to changes in their input costs. Accordingly, when monetary policy is more hawkish, prices are less sensitive to the economy's slack, the Phillips curve is flatter, and firms' inflation expectations are more anchored. Therefore, our theory suggests that the decline in the slope of the Phillips curve is related to the more *hawkish* monetary policy adopted at the beginning of the Great Moderation.[7]

This effect, however, is not symmetric. While more hawkish monetary policy flattens the Phillips curve, a more dovish monetary policy completely flattens the Phillips curve in the short-run but steepens it 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

---

(2014), and provide a new perspective and microfoundation for how inattentive agents completely ignore certain dimensions of their environment.

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

[4]Link for Matlab GitHub Repository: `https://github.com/choongryulyang/DRIPs.m`
Link for Julia Package: `https://www.afrouzi.com/DRIPs.jl/dev`

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

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

[7]See, e.g., Clarida, Galí, and Gertler (2000) for evidence on more hawkish monetary policy in the post-Volcker era.

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 grows and, eventually, they restart paying attention to their costs. In this new regime, firms have a lower stock of knowledge, but they acquire information at a higher rate. The higher rate of information acquisition makes prices more sensitive to changes in input costs and leads to a *steeper* Phillips curve and less anchored inflation expectations relative to the previous regime.

Finally, we examine the quantitative relevance of our proposed mechanism for the change in the Phillips curve slope. Using our computational toolbox, we solve and calibrate a dynamic general equilibrium rational inattention model with monetary policy and supply shocks to the post-Volcker U.S. data (1983-2007). In the spirit of Maćkowiak and Wiederholt (2015), we examine the out-of-sample fit of our model by replacing the post-Volcker Taylor rule with an estimated Taylor rule for the pre-Volcker period. We find that our model quantitatively matches the higher variance of inflation and GDP in the pre-Volcker era as non-targeted moments. As our main empirical exercise, we simulate data from our calibrated model using our estimated pre- and post-Volcker monetary policy rules 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. We contribute to the literature that has laid the ground for solving DRIPs in LQG settings (Sims, 2003; Maćkowiak and Wiederholt, 2009a; Maćkowiak, Matějka, and Wiederholt, 2018; Fulton, 2018).[8] These papers make two assumptions that we depart from: (1) they abstract away from transition dynamics, and (2) they solve for the long-run steady-state information structure without discounting. A notable exception is Sims (2010), who studies the special case when solutions are interior, along with an example with two shocks and one action. More recently, Miao, Wu, and Young (2020) propose a value function iteration method but they do not characterize optimality conditions. We contribute to this literature by fully characterizing the optimality conditions, which serve as the foundation of our Euler equation and solution method. We provide a more thorough and detailed description of our contribution to this literature in Section 2.5.

We also contribute to the literature that considers how rational inattention affects the pricing decisions of firms in dynamic environments (Maćkowiak and Wiederholt, 2009a, 2015; Paciello and Wiederholt, 2014; Woodford, 2009; Matějka, 2015; Stevens, 2019) and connect this literature to the set of papers that study the implications of imperfect information for Phillips curves (Lucas, 1972; Woodford, 2003; Nimark, 2008; Angeletos and Lian, 2018; Angeletos and Huo, 2021; Gabaix, 2020).[9] Our main departure is to derive a Phillips curve in a model with rational inatten-

[8]See Maćkowiak, Matějka, and Wiederholt (2020) for a recent survey of the rational inattention literature.
[9]See, also, Mankiw and Reis (2002); Angeletos and La'O (2009); Angeletos and Lian (2016).

tion and study how *monetary policy* shapes and alters the incentives in information acquisition of firms, which in turn determines the slope of the Phillips curve. 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, our attention-based theory of the Phillips curve is motivated by the recent literature that studies the apparent flattening of the Phillips curve in the last few decades. Very broadly, this literature discusses two different explanations for this flattening. The first explanation is that the structural slope of the Phillips curve is stable over time, but the reduced form relationship between inflation and the output gap has changed either due to changes in the cyclical behavior of inflation (Stock and Watson, 2020) or changes in the conduct of monetary policy (e.g., Hooper, Mishkin, and Sufi, 2020; McLeay and Tenreyro, 2020) with a particular focus on the role of inflation expectations (e.g., Coibion and Gorodnichenko, 2015b; Jorgensen and Lansing, 2019; Hazell, Herreño, Nakamura, and Steinsson, 2020).[10] A second explanation is the slope hypothesis that argues for a structural change in the slope of the Phillips curve (see, e.g., Forbes, 2019; Del Negro, Lenza, Primiceri, and Tambalotti, 2020; Rubbo, 2020).[11] We contribute to this literature by connecting these two explanations and developing a theory in which the *structural* slope of the Phillips curve is affected by the conduct of monetary policy.[12] In particular, in our theory, the structural slope of the Phillips curve responds asymmetrically to the dovishness of monetary policy over time; where it temporarily flattens after the policy becomes more dovish but eventually steepens once firms' uncertainty about their environment grows to a critical level. In that sense, our theory provides a cautionary tale against policies that treat the slope of the Phillips curve as exogenously determined.

Layout. The paper is organized as follows. Section 2 characterizes transition dynamics of DRIPs in LQG settings, outlines our solution algorithm, and provides a more thorough discussion of how our approach relates to the existing literature. Section 3 provides an attention-based theory of the Phillips curve in an analytical framework. Section 4 presents our quantitative model and results. Section 5 concludes.

2 Theoretical Framework

This section formalizes the problem of an agent who chooses her information structure endogenously over time. We start by setting up the general problem and deriving some properties for

[10] Hazell, Herreño, Nakamura, and Steinsson (2020) document 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. According to our model, the benchmark New Keynesian Phillips curve is (1) misspecified and (2) "too" forward-looking, which is why the slope implied by our model is larger in magnitude.

[11] For a more detailed review of the literature on the slope of the Phillips curve, see, for instance, the discussions in Hazell, Herreño, Nakamura, and Steinsson (2020) or Del Negro, Lenza, Primiceri, and Tambalotti (2020).

[12] See, also, L'Huillier and Zame (2020) who connect changes in price stickiness to the pursuit of price stability by the central bank.

its solution without making assumptions on payoffs and information structures. We then derive and solve the implied LQG problem and present our algorithm for solving DRIPs and compare its accuracy and efficiency by replicating results from previous literature. We conclude this section by discussing the properties of transition dynamics in DRIPs in the context of an extension of the pricing example in Sims (2010).

2.1 Environment

Preferences. Time is discrete and is indexed by $t \in \{0, 1, 2, \ldots\}$. At each time t, the agent chooses a vector of actions $\vec{a}_t \in \mathbb{R}^m$ and gains an instantaneous payoff of $v(\vec{a}_t; \vec{x}_t)$ where $\{\vec{x}_t \in \mathbb{R}^n\}_{t=0}^{\infty}$ is an exogenous stochastic process, and $v(.;.) : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ 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 the history of shocks, $X^t \equiv (\vec{x}_0, \ldots, \vec{x}_t)$. 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, commonly referred to as the "no-forgetting" condition. The agent understands that any choice of information will affect their priors in the future, and that information has a continuation value.[13] Formally, a sequence of information sets $\{S^t \subseteq \mathcal{S}^t\}_{t\geq 0}$ satisfy the no-forgetting condition 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 the cost of information is linear in Shannon's mutual information function.[14] 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

[13]Although we assume perfect memory in our benchmark, these dynamic incentives exist insofar as the agent carries a part of her memory with her over time. For a model with fading memory with exogenous information, see Nagel and Xu (2019). Furthermore, Azeredo da Silveira and Woodford (2019) endogenize noisy memory in a setting where carrying information over time is costly.

[14]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.

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 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 unit of information.[15]

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

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

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

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

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 follow directly from the linearity of the cost function in Shannon's mutual information function and 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, 2020). 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 RI problem in Equation* (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*.*

[15]This unit is either bits—if entropy is defined in terms of binary logarithm—or nats—if entropy is defined in terms of natural logarithm.

Proof. See Appendix A.1. ∎

In static environments, the sufficiency of actions for signals follows from optimality (Matějka and McKay, 2015). Since information is costly and only valuable in choosing the optimal action, an information set not revealed by the optimal action must be suboptimal (otherwise, there exists an information set that generates the same action but is less costly). In dynamic settings, however, the agent might find it optimal to acquire information about future actions before-hand. Lemma 2.1 rules out this case by showing that if the chain-rule of mutual information holds, then the agent has no incentives to acquire information about future actions.[16]

The result in 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 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:

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

Here, $\mathbf{H} \in \mathbb{R}^{n\times m}$ has full column rank and captures the interaction of the actions with the state.[17] The assumption of $rank(\mathbf{H}) = m$ is without loss of generality; in the case that any two columns

[16] 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 the information is measured altogether or part by part.

[17] 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.

of $\mathbf{H}$ are linearly dependent, we can reclassify the problem so that all co-linear actions are in one class. Moreover, we have normalized the Hessian matrix of v with respect to $\vec{a}$ to negative identity.[18]

Optimality of Gaussian Posteriors. We start by proving that optimal actions are Gaussian under quadratic payoff with a Gaussian initial prior. Maćkowiak and Wiederholt (2009b) prove 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 $\{\vec{x}_t\}_{t\geq 0}$ is a Gaussian process and that the initial conditional prior, $\vec{x}_0|S^{-1}$, has a Gaussian distribution. If $\{\vec{a}_t\}_{t\geq 0}$ is a solution to the RI problem with quadratic payoff, then 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 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 Gauss-Markov process with the following state-space representation:*

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

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

[18] 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. Finally, while we have abstracted away from endogenous state variables, such problems could be reformulated in a similar form by redefining the action and an adequate redefinition of the matrix $\mathbf{H}$ (see, e.g., Miao, Wu, and Young, 2020; Mackowiak and Wiederholt, 2020).

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

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

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

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

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

Proof. See Appendix A.4. ■

Lemma 2.4 reformulates the RI problem into an LQG problem in Equation (2.4) subject to the law of motion for the agent's priors in Equation (2.5) and a set of *no-forgetting constraints* in Equation (2.6) that follow directly from the no-forgetting condition and require the agent's posterior to be at least as precise as their prior in all dimensions of the state. Finally, Equation (2.7) specifies the initial condition for the problem as the covariance matrix of the agent's prior belief over $\vec{x}_0$ induced by the initial information set S^{-1}.

Solution. Sims (2010) derives a first-order condition for the solution to the LQG RI problem in Equation (2.4) when the no-forgetting constraints do not bind.[19] However, binding no-forgetting constraints arise frequently. In fact, for any $m < n$, at least $n - m$ constraints always bind by Lemma 2.2. Here, we provide a solution for the problem with arbitrary n and m characterizing when and which constraints bind at any given time. The following proposition takes these constraints into account and derives the following Karush-Kuhn-Tucker (KKT) conditions for the solution.

Proposition 2.1. *Suppose $\mathbf{\Sigma}_{0|-1}$ is strictly positive definite, and $\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'$ is of full rank. Then, all the future priors $\{\mathbf{\Sigma}_{t+1|t}\}_{t\geq 0}$ are invertible under the optimal solution to the LQG Problem in Equation (2.4), which is characterized by the following necessary conditions*

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

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

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

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

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

[19] He also provides a solution for a special case with $n = 2$ and $m = 1$ when these constraints do bind but does not extend that solution to the general problem with arbitrary m and n.

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. Since we allow for binding no-forgetting constraints, $\mathbf{\Lambda}_t$ 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 Equation (2.11) is the transversality condition on information acquisition of the agent.

An essential aspect of the first-order condition in Equation (2.8) is its forward-looking nature. In fact, the right-hand side of this equation captures the marginal benefit of reducing uncertainty for the agent. Intuitively, this benefit has two components: a contemporaneous benefit for information that captures how more information affects the agent's instantaneous utility and a continuation benefit that captures how information acquisition in period t will increase the agent's payoff in the future. Formally, defining $\mathbf{\Omega}_t$ as the right-hand side of the FOC in Equation (2.8), this decomposition can be stated as

$$\mathbf{\Omega}_t \equiv \underbrace{\mathbf{\Omega}}_{\text{contemporaneous benefit}} + \underbrace{\beta \mathbf{A}'(\omega \mathbf{\Sigma}_{t+1|t}^{-1} - \mathbf{\Lambda}_{t+1})\mathbf{A}}_{\text{continuation benefit}} \tag{2.12}$$

Moreover, while the novelty of Proposition 2.1 relative to the previous literature is to derive the necessary KKT conditions under potentially binding no-forgetting constraints, their sufficiency follows from two conditions: (1) affinity of the no-forgetting constraints in the sequence $\{\mathbf{\Sigma}_{t|t}\}_{t\geq 0}$ which is trivial, and (2) the concavity of the objective function in $\mathbf{\Sigma}_{t|t}, \forall t \geq 0$. Since $tr(\mathbf{\Omega}\mathbf{\Sigma}_{t|t})$ is linear in $\mathbf{\Sigma}_{t|t}$, it is straightforward to show that concavity of the objective function is equivalent to the convexity of the cost function, which was first shown to be true by Sims (2003) for an invertible $\mathbf{A}$, and later by Miao, Wu, and Young (2020) for an invertible $\mathbf{QQ}'$. Nonetheless, since we consider a more general parameter space that only requires $\mathbf{AA}' + \mathbf{QQ}'$ to be invertible, these proofs are no longer applicable to our setting. The following proposition proves this result for our case.

Proposition 2.2. *The necessary conditions in Proposition 2.1 are sufficient for optimality.*

Proof. See Appendix A.6. ■

With these necessary and sufficient conditions at hand, one can obtain the solution to the problem. Moving forward, we 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 introduce the following two matrix operators:

Definition 2.1. *For a symmetric matrix $\mathbf{X}$ with spectral decomposition $\mathbf{X} = \mathbf{UDU}'$, we define*

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

where $\max(\mathbf{D}, \omega)$ *and* $\min(\mathbf{D}, \omega)$ *operate on every element on the diagonal.*

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

Theorem 2.1. *Let $\mathbf{\Omega}_t$ be the forward-looking component of the FOC in Proposition 2.1, as defined in Equation (2.12). Then, $\mathbf{\Omega}_t$ is characterized by the following Euler equation:*

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

Furthermore, given $\mathbf{\Omega}_t$, the optimal posterior covariance matrix, $\mathbf{\Sigma}_{t|t}$, is characterized by the following policy function:

$$\mathbf{\Sigma}_{t|t} = \omega \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \left[\mathrm{Max}\left(\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \mathbf{\Omega}_t \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}, \omega \right) \right]^{-1} \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \tag{2.14}$$

Proof. See Appendix A.7. ∎

As $\mathbf{\Omega}_t$ is the *benefit* matrix that captures how information interacts with the agent's payoff, we refer to Equation (2.13) as the *information Euler equation* that captures how the agent encodes these benefits under her optimal information acquisition strategy. When $\beta = 0$, this benefit is simply given by $\mathbf{\Omega}_t = \mathbf{\Omega}$ which captures how $\vec{x}_t$ affects the agent's instantaneous payoff.[20] What is new here is that with $\beta > 0$, $\mathbf{\Omega}_t$ has an extra term that captures the continuation value of knowledge about $\vec{x}_t$, which depends on β itself, the persistence of the shocks $\mathbf{A}$, and the information acquisition policy of the agent in the next period.

Intuitively, information has marginal value *only* if (1) it generates higher payoff (captures by $\mathbf{\Omega}_t$), *and* (2) the agent is sufficiently uninformed (captured by $\mathbf{\Sigma}_{t|t-1}$). Based on this intuition, the *policy function* in Equation (2.14) shows that in acquiring information, the agent considers the orthogonalized dimensions of the matrix $\mathbf{\Sigma}_{t|t-1}^{1/2} \mathbf{\Omega}_t \mathbf{\Sigma}_{t|t-1}^{1/2}$. At the *extensive margin*, the agent ignores dimensions whose eigenvalues (marginal values of information) are less than ω—i.e., the agent is at a corner solution and her posterior uncertainty is the same as her prior uncertainty. On the *intensive margin*, the agent acquires information in dimensions whose eigenvalues are larger than ω, and her posterior uncertainty is lower than her prior uncertainty.

[20]The case of $\beta = 0$ collapses these results to the static cases studied in the literature prior to us (Fulton, 2018; Kőszegi and Matějka, 2020; Miao, Wu, and Young, 2020).

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), the information Euler equation in Equation (2.13) and the policy function in Equation (2.14) characterize the solution to the dynamic rational inattention problem.

Optimal Signals. While we have characterized the covariance matrix of the optimal posterior as a function of the agent's prior, the underlying assumption in Proposition 2.1 is that this posterior is generated by a sequence of signals about $\vec{x}_t$. It is important to note that both the number of these signals at a given time t, as well as how they load on different elements of the vector $\vec{x}_t$ are endogenous. Our next result characterizes these signals in a basis where the noise in these signals are independent.[21]

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 k_t is the number of the eigenvalues that are at least as large as ω, and for $i \leq k_t$, $\mathbf{y}_{i,t} \equiv \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{u}_{i,t}$ is the loading of signal i on $\vec{x}_t$, and $z_{i,t} \sim \mathcal{N}(0, \frac{\omega}{d_{i,t}-\omega})$ is the agent's rational inattention error in signal i that is orthogonal to $\vec{x}_t$ and all the other rational inattention errors.

Proof. See Appendix A.8. ■

In theory, an agent that learns about an n dimensional vector of shocks needs at most n signals, one for each dimension. However, in an environment with costly information acquisition, not all dimensions might hold enough value for information acquisition. To this end, we need a closer look at what determines these values. Information in a particular dimension is more valuable if (1) it provides more benefit to the agent (encoded in matrix $\mathbf{\Omega}_t$) *and* (2) if the agent is more uncertain about that dimension (encoded in matrix $\mathbf{\Sigma}_{t|t-1}$). The eigenvalues and eigenvectors of the matrix $\mathbf{\Sigma}_{t|t-1}^{1/2}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{1/2}$ capture these two forces. The eigenvalues are marginal values of information in a given dimension. Moreover, the corresponding eigenvectors are the dimensions along which the largest amount of information can be acquired for a given precision of a signal.

With this intuition in mind, Theorem 2.2 simply states that the agent's optimal posterior at a given time is generated by k_t signals, where k_t is the number of dimensions for which the marginal value of information, defined as the eigenvalues of the matrix was $\mathbf{\Sigma}_{t|t-1}^{1/2}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{1/2}$, is larger than its marginal cost, ω. Furthermore, for the dimensions that do have a higher marginal value than ω,

[21] Signals that generate a Gaussian posterior are not unique and, for instance, are equivalent up to linear transformations.

the loading of the corresponding signal on $\vec{x}_t$ is determined by the eigenvector associated with that marginal value.

Finally, the endogenous loadings of signals on the shock vector $\vec{x}_t$ captures the incentives of the agent in garbling information in different dimensions to reduce the cost of information acquisition. These incentives provide a microfoundation for *information spillovers* across different actions (Sims, 2010), where information about an action can affect others, either through a subjective correlated posterior ($\mathbf{\Sigma}_{t|t-1}$) or through complementarities or substitutabilities in actions captured by $\mathbf{\Omega}_t$.[22]

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 proposition characterizes how beliefs and actions evolve over time.

Proposition 2.3. *Let $\{(\mathbf{y}_{i,t}, d_{i,t}, z_{i,t})_{1\leq i\leq k_t}\}_{t\geq 0}$ be defined as in Theorem 2.2, 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, optimal actions is $\vec{a}_t = \mathbf{H}'\hat{x}_t$, where $\hat{x}_t$ evolve according to:*

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

Proof. See Appendix A.9. ∎

2.3 Solution Algorithm, Computational Accuracy and Efficiency

Given an initial prior $\mathbf{\Sigma}_{-1|0}$, the solution to the LQG dynamic rational inattention 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 LQG dynamic rational inattention problems based on this algorithm that is available as the `DRIPs.jl` package to the Julia programming language

[22]For instance, Kamdar (2018) documents that households have countercyclical inflation expectations—an observation that contradicts the negative comovement of inflation and unemployment in the data but is consistent with optimal information acquisition of households under substitutability of leisure and consumption. Similarly, Kőszegi and Matějka (2020) show that complementarities or substitutabilities in actions give rise to mental accounting in consumption behavior based on optimal information acquisition. While these two papers use static information acquisition, our framework allows for dynamic spillovers through information acquisition.

which is available at `https://github.com/afrouzi/DRIPs.jl` with a detailed software documentation available at `https://afrouzi.com/DRIPs.jl/dev`.[23]

Table 1: Summary of Computing Times

	Dimension	`DRIPs.jl`	Alternative Algorithms	
Computing time for:	n^2	Time (s)	Time (s)	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 (2018)				
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, Matějka, and Wiederholt (2018) (discussed in Section 2.4, Appendix C.1 and Appendix C.2 respectively). Tolerance level for convergence is 10^{-4} for the solution to rational inattention problem in all cases. Statistics from Miao, Wu, and Young (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).

We have also used our software package to replicate results from three canonical papers (Maćkowiak and Wiederholt, 2009a; Sims, 2010; Maćkowiak, Matějka, and Wiederholt, 2018) 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 available solution methods. Table 1 reports a summary of computing times for these replications. Moreover, all of our replication materials for these three papers are available along with our software documentations at the link above and are also accessible as executable Jupyter notebooks.

Our algorithm solves DRIPs in two stages. First, we solve for the steady-state of a problem that is independent of any initial prior belief. Second, we use a shooting algorithm on the information Euler equation (Equation 2.13) and the law of motion for the prior (Equation 2.5) to characterize the transition dynamics of the optimal information structure from an initial prior belief. In Appendix

[23] In addition to the Julia package, a Matlab code repository for the algorithm is also available at `https://github.com/choongryulyang/DRIPs.m`.

B, we describe these two stages in more detail.

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 solution to the rational inattention problem. Here we study an extension of that problem to investigate the *transition dynamics* of attention from an initial prior.

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 $v(a_t, \vec{x}_t) = -(a_t - \mathbf{H}'\vec{x}_t)^2$ where a_t is the agent's action (here the monopolist's price), and $\vec{x}_t = \mathbf{A}\vec{x}_{t-1} + \mathbf{Q}\vec{u}_t$ is an exogenous Gaussian process with two AR(1) shocks. Assuming the agent discounts future payoffs at an exponential rate β, Equation (10) in Sims (2010) derives the equivalent LQG rational inattention problem 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, Wu, and Young (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}:

$$\bar{\mathbf{\Sigma}} \equiv \lim_{t\to\infty} \mathbf{\Sigma}_{t|t} = \begin{bmatrix} 0.3592 & -0.1770 \\ -0.1770 & 0.7942 \end{bmatrix}, \quad \bar{\mathbf{\Sigma}}_{-1} \equiv \lim_{t\to\infty} \mathbf{\Sigma}_{t+1|t} = \begin{bmatrix} 0.4217 & -0.0673 \\ -0.0673 & 0.9871 \end{bmatrix} \tag{2.17}$$

This solution is close to the posterior covariance reported in Sims (2010).[24] Moreover, it is almost identical to the one reported in Miao, Wu, and Young (2020) who use conventional value function iteration methods to calculate this solution.[25]

Transition Dynamics of the Optimal Information Structure. This section reports 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

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

[25] In Miao, Wu, and Young (2020), the posterior covariance matrix, $\bar{\mathbf{\Sigma}}$, is $\begin{bmatrix} 0.3590 & -0.1769 \\ -0.1769 & 0.7945 \end{bmatrix}$.

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. Thus, at time 0, the agent's prior about $\vec{x}_0$ conditional their initial information set S^{-1} is

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

where $\bar{\mathbf{\Sigma}}_{-1}$ is the prior covariance matrix in the steady-state from Equation (2.17). By setting the mean of this prior to $\mathbf{0}$, we are implicitly assuming that both components of the monopolist's cost 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).

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 $\mathbf{\Sigma}^{\frac{1}{2}}_{t|t-1}\mathbf{\Omega}_t\mathbf{\Sigma}^{\frac{1}{2}}_{t|t-1}$ that are at least as large as ω. Since the dimension of the state in this problem is 2, there are two eigenvalues, representing the two dimensions to which the agent can pay attention.

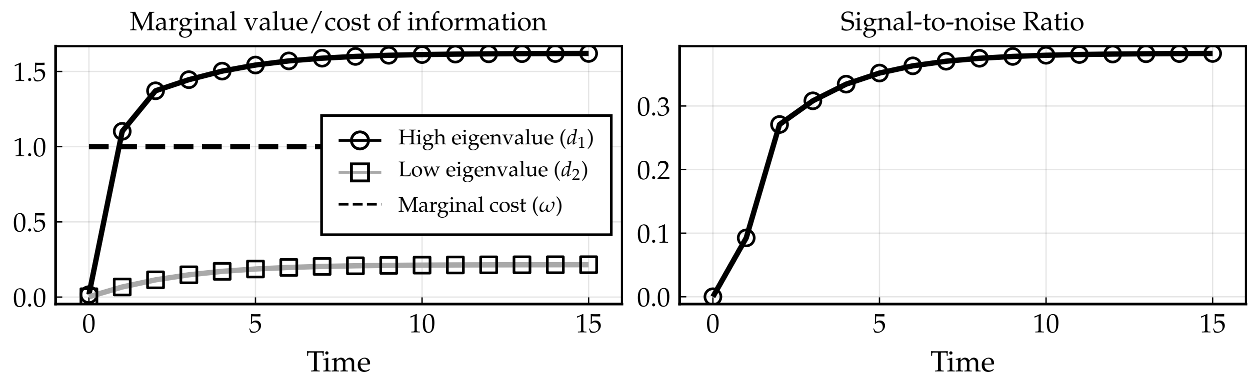

Figure 1: Transition Dynamics of Attention

Notes: The Figure depicts the transition dynamics of attention in our extension of the example in Sims (2010). The left panel shows the marginal values of information in orthogonal dimensions, and the right panel shows the transition path of the Kalman gain for the optimal signal. 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 (2.17). All values are constant in the steady-state.

The left panel of 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 the eigenvalues is larger than 1, which implies that the agent receives one signal starting at $t = 1$. It takes approximately ten periods for these eigenvalues to reach their steady-state, at which point only one of them remains above ω. Therefore, even in the steady-state, the agent receives only one signal.[26]

[26]This is consistent with Lemma 2.2 which specifies that the number of signals should be bounded above by the

Moreover, in contrast to the steady-state of the information structure, the signal-to-noise ratio of the agent's signal varies over time on the transition path. According to Proposition 2.3, this ratio is given by $1 - \frac{\omega}{d_{1,t}}$. The right panel of Figure 1 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. However, starting at $t = 1$, this quantity is positive and, in approximately ten periods, converges to its steady-state from below. Accordingly, the knowledge shock at $t = -1$ has dynamic consequences by *crowding out* information acquisition in later periods.

Impulse Response Functions. How important is the transition dynamics of attention? To answer this question, we compare the impulse responses of the monopolist's price to both shocks between the steady-state and the transition path of the optimal information structure, and show they are significantly different.

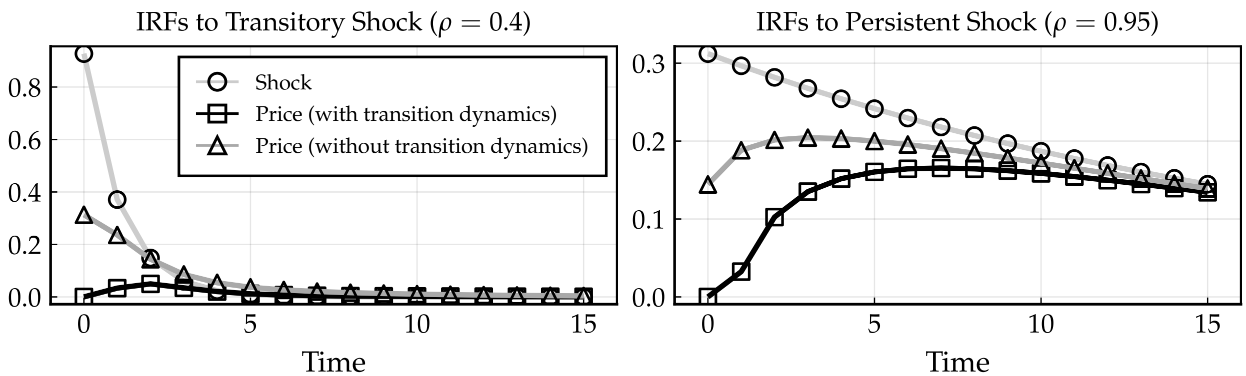

Figure 2: Impulse Response Functions in Steady State versus on the Transition Path

Notes: This figure the impulse response functions of the price with both the steady-state information structure and 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 (2.17). 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 one as the agent receives no signals in that period.

Figure 2 plots these impulse response functions for a one standard deviation innovation to both components of the monopolist's cost, under both information structures. The main observation is that the impulse responses under the transition dynamics of the information structure are significantly muted. Being highly certain after the knowledge shock at $t = -1$, the monopolist temporarily substitutes away from information acquisition in later periods and pays little attention to costs on the transition path. These muted responses mirror the smaller signal-to-noise ratio of the monopolist's signal on the transition path, and show that the monopolist is significantly

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.

less responsive to shocks under the transition path of the information structure compared to the steady-state one.

2.5 Relation to Alternative Solution Methods

The literature has focused on two different versions of the DRIPs in LQG settings in the past. The problem that we study, as posed in Lemma 2.4, was introduced in Sims (2010) who derived optimality conditions for the general problem when the solution is interior.[27] In related work, Miao, Wu, and Young (2020) also propose a solution method for transition dynamics of LQG control problems based on value function iteration, but they do not derive any first-order optimality conditions. Our contribution relative to these papers is that we fully characterize the optimality conditions for transition dynamics, taking corner solutions into account. These corner solutions are significant in economics because they micro-found why an agent might completely ignore specific shocks (for instance, in our application in Section 3, they lead to an occasionally *flat* Phillips curve). Our characterization of these optimality conditions is essential to our second contribution: by directly iterating over our *information Euler equation*, we propose a novel solution method that is significantly faster than value function iteration.

Another approach in formulating DRIPs in LQG settings abstracts away from transition dynamics and focuses on the steady-state information structure (Sims 2003; Maćkowiak, Matějka, and Wiederholt 2018; Fulton 2018 and the "Golden rule approximation" in Miao, Wu, and Young 2020):

$$\max_{\mathbf{\Sigma}\succeq 0} -tr(\mathbf{\Sigma\Omega}) - \omega \ln\left(\frac{|\mathbf{\Sigma}_{-1}|}{|\mathbf{\Sigma}|}\right) \; s.t. \, \mathbf{\Sigma}_{-1} = \mathbf{A\Sigma A}' + \mathbf{QQ}', \, \mathbf{\Sigma}_{-1} \succeq \mathbf{\Sigma}. \tag{2.18}$$

This problem does not capture transition dynamics, and its solution does not depend on β (the discount factor), implicitly assuming that the agent is perfectly patient (in Appendix A.10 we show that the solution to this problem collapses to the steady-state information structure of the dynamic problem when $\beta = 1$). By assuming the discount factor is 1, this method also implies a different steady-state information structure than the case when $\beta < 1$: (1) perfectly patient agents acquire more information in general because they do not discount the dynamic benefits of information, and (2) keeping the amount of information fixed, perfectly patient agents acquire more information about dimensions that are more persistent or, in general, have higher continuation value. Although this problem ignores transition dynamics and discounting, its appeal to the literature has been its simplicity, as pointed out by Miao, Wu, and Young (2020), who show that it is significantly faster

[27] This proved to be a restrictive assumption. For instance, Maćkowiak, Matějka, and Wiederholt (2018) showed that when the agent has one action, all but one of the no-forgetting constraints always bind (i.e., a firm that only chooses one price would never pay attention to anything but their marginal cost). More generally, per Lemma 2.2, as long as the number of actions is fewer than the number of shocks ($m < n$), the solution is not interior.

to solve. However, our proposed solution method does not discriminate between the two problems in terms of solution times and is equally fast in characterizing the steady-state solutions to both problems (see Table 1 for a summary of computing times).

3 An Attention-Based Theory of the Phillips Curve

This section introduces an analytical general equilibrium model with rationally inattentive firms and provides an attention-based theory of the Phillips curve.

3.1 Environment

Households. Consider a fully attentive representative household who supplies labor L_t in a competitive labor market at the nominal wage W_t, trades nominal bonds with the net 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

$$\max_{\{(C_{i,t})_{i\in[0,1]}, B_t, L_t\}_{t=0}^{\infty}} \mathbb{E}_0^f \left[\sum_{t=0}^{\infty} \beta^t (\log(C_t) - L_t)\right] \tag{3.1}$$

$$s.t. \int_0^1 P_{i,t}C_{i,t}di + B_t \leq W_tL_t + R_{t-1}B_{t-1} + \Pi_t + T_t, \quad C_t = \left[\int_0^1 C_{i,t}^{\frac{\theta-1}{\theta}} di\right]^{\frac{\theta}{\theta-1}}$$

where $\mathbb{E}_t^f[.]$ is the expectation operator of a fully informed household at time t, $C_{i,t}$ is the demand for variety i given its price $P_{i,t}$, B_t is the demand for nominal bonds at t that yield a nominal return of R_t at $t + 1$, Π_t is the aggregated profits of firms, and T_t is the net lump-sum transfers. Finally, C_t is the final consumption good aggregated with a constant elasticity of substitution $\theta > 1$ across varieties.

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

$$C_{i,t} = C_t \left(\frac{P_{i,t}}{P_t}\right)^{-\theta}, \qquad \forall i \in [0, 1], \forall t \geq 0, \tag{3.2}$$

$$1 = \beta R_t \mathbb{E}_t^f \left[\frac{Q_t}{Q_{t+1}}\right], \qquad \forall t \geq 0, \tag{3.3}$$

$$W_t = Q_t, \qquad \forall t \geq 0. \tag{3.4}$$

Here, Equation (3.2) is the household's demand for variety i at time t, Equation (3.3) is the consumption Euler equation, and Equation (3.4) specifies the equilibrium relationship between the nominal wage and the nominal aggregate demand.[28]

[28]This is the household's labor supply condition given that the Frisch elasticity of labor supply is assumed to be

Monetary Policy. For analytical tractability, we assume that the monetary authority targets the growth of the nominal aggregate demand, which can be interpreted as targeting inflation and output growth similarly:

$$\log(R_t) = \log(\bar{R}) + \phi\Delta q_t - \sigma_u u_t,\ u_t \sim \mathcal{N}(0,1)$$

where $\bar{R} \equiv \beta^{-1}$ is the steady-state nominal rate at zero trend inflation, $q_t \equiv \log(Q_t)$ is the log of the nominal aggregate demand, and u_t is an exogenous shock to monetary policy that affects the nominal rates with a standard deviation of σ_u. We consider a more standard Taylor rule in our quantitative model in Section 4.

Lemma 3.1. *Suppose $\phi > 1$. Then, in the log-linearized version of this economy, the nominal aggregate demand is uniquely determined by the history of monetary policy shocks, and is characterized by the random walk process, $q_t = q_{t-1} + \frac{\sigma_u}{\phi} u_t$.*

Proof. See Appendix D.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.[29] 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 $L_{i,t}$ 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 (Galí, 2015, p. 73). Firms produce their product with a linear technology in labor, $Y_{i,t} = L_{i,t}$. Therefore, for a given history $\{P_t, Q_t\}_{t\geq 0}$ and a set of prices $\{P_{i,t}\}_{t\geq 0}$, the net present value of the firm's profits, discounted by the household's marginal utility of consumption is

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

where the second line is a second-order approximation with small letters denoting the logs of corresponding variables.[30] This expression provides a quadratic approximation of a monopolistic

infinite. This is a common assumption in monetary models (Golosov and Lucas, 2007). We relax this assumption in our quantitative model in Section 4.

[29] See, for instance, Mankiw and Reis (2002), Woodford (2003), Golosov and Lucas (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).

[30] For a detailed derivation of this second-order approximation see, for instance, Maćkowiak and Wiederholt (2009a).

firm's losses from not matching its marginal cost (q_t in this setting.) Moreover, the approximation shows that the magnitude of these losses is proportional to how elastic the firm's demand is ($\theta - 1$). Firms with more elastic demand lose more profits by charging a suboptimal price.

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

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

where $p_i^t \equiv (p_{i,\tau})_{\tau \leq t}$ denotes the history of firm's prices up to time t. It is important to note that $\{p_{i,t}\}_{t\geq 0}$ is a stochastic process that is a sufficient statistic for the underlying signals that the firm receives—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 Gauss-Markov process, this problem satisfies the assumptions of Lemma 2.4. Formally, let $\sigma^2_{i,t|t-1} \equiv var(q_t | p_i^{t-1}), \sigma^2_{i,t|t} \equiv var(q_t | p_i^t)$ denote the prior and posterior variances of firm i's belief about q_t at time t. Then, the corresponding LQG problem to the one in Lemma 2.4 is

$$V(\sigma^2_{i,0|-1}) = \max_{\{\sigma_{i,t|t}, \sigma_{i,t+1|t}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \left[-(\theta-1)\sigma^2_{i,t|t} - \omega \ln \left(\frac{\sigma^2_{i,t|t-1}}{\sigma^2_{i,t|t}} \right) \right]$$

$$s.t. \quad \sigma^2_{i,t+1|t} = \sigma^2_{i,t|t} + \frac{\sigma_u^2}{\phi^2} \tag{3.5}$$

$$0 \leq \sigma_{i,t|t} \leq \sigma_{i,t|t-1}. \tag{3.6}$$

Here Equation (3.5) is the law of motion for the prior and Equation (3.6) is the no-forgetting constraint that correspond to this problem.

3.2 Characterization of Solution

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

Proposition 3.1. *Firms pay attention to monetary policy shocks only if their prior uncertainty is above a reservation level, $\underline{\sigma}^2$. Formally,*

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

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

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

Proof. See Appendix D.2. ■

The first part of Proposition 3.1 shows that firms pay attention to nominal demand only when they are sufficiently uncertain about it. In particular, for small enough levels of prior uncertainty—where the marginal benefit of acquiring a bit of information falls below its marginal cost—the no-forgetting constraint binds, and the firm receives no information. However, if the firm's prior uncertainty is higher than a reservation level, it acquires enough information to restore and maintain that uncertainty level. The second part of Proposition 3.1 shows that in the region where the firm does not pay attention to the nominal demand, its price is entirely insensitive to monetary shocks as the implied Kalman-gain is zero.

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

$$\kappa_{i,t} = \kappa \equiv \frac{\sigma_u^2}{\phi^2\underline{\sigma}^2 + \sigma_u^2}. \tag{3.7}$$

Comparative Statics. Here, we study how the reservation uncertainty $\underline{\sigma}^2$, and the steady-state Kalman-gain κ, change with the model parameters.

Corollary 3.1. *The reservation uncertainty of firms increases with* ω *and* σ_u, *and decreases with* ϕ, θ *as well as* β. *Moreover, the steady-state Kalman-gain of firms increases with* σ_u, θ *and* β, *and decreases with* ϕ *and* ω.

Proof. See Appendix D.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:[31]

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

[31] This approximation becomes the exact solution to the analogous problem in continuous time as the variance of the innovation is proportional to the length of time between two consecutive decisions.

3.3 Aggregation

For aggregation, we make two assumptions: (1) firms all start from the same initial condition, $\sigma^2_{i,0|-1} = \sigma^2_{0|-1}, \forall i \in [0,1]$—which is without loss of generality if all firms start with the steady-state prior as their initial prior—and (2) firms' rational inattention errors are independently distributed.[32]

Notation-wise, we let $\pi_t \equiv \log(P_t) - \log(P_{t-1})$ denote the aggregate inflation rate and $y_t \equiv \log(Q_t) - \log(P_t)$ be the log of aggregate output. The following proposition derives the Phillips curve of this economy.

Proposition 3.2. *Suppose all firms have the same initial condition, $\sigma^2_{0|-1} \geq 0$. Then,*

1. *the Phillips curve of this economy is*

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

where $\sigma^2_{t+1|t} = \min\{\underline{\sigma}^2, \sigma^2_{t|t-1}\} + \frac{\sigma_u^2}{\phi^2}$ for all $t \geq 0$.

2. *For any given $T \geq 0$, if $\sigma^2_{T|T-1} < \underline{\sigma}^2$, then $\pi_t = 0$ and $y_t = y_{t-1} + \frac{\sigma_u}{\phi} u_t$.*

3. *For any given $T \geq 0$, if $\sigma^2_{T|T-1} \geq \underline{\sigma}^2$, then for all $t \geq T+1$,*

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

where $\kappa \equiv \frac{\sigma_u^2}{\phi^2 \underline{\sigma}^2 + \sigma_u^2}$ is the steady-state Kalman-gain of firms in Equation (3.7).

Proof. See Appendix D.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. When firms' uncertainty is below the reservation uncertainty, they pay no attention to changes in their input costs, and inflation does not respond to monetary policy shocks. However, when the no-forgetting constraint binds, firms' uncertainty about their cost grows linearly with time and eventually reaches its reservation level, at which point firms begin to pay attention to their costs again at a constant rate that is determined by the steady-state information structure. This section analyzes this steady-state and then considers the dynamic consequences of unanticipated disturbances (MIT shocks) to the model's parameters.

[32]This assumption follows as a result from assuming that the cost of attention is Shannon's mutual information (Denti, 2015; Afrouzi, 2020). With other classes of cost functions, however, non-fundamental volatility can be optimal—see Hébert and La'O (2020) for characterization of these cost functions.

3.4.1 The Long-run Slope of the Phillips Curve

It follows from Proposition 3.2 that once the rational 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.[33] Moreover, the last part of Proposition 3.2 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, κ determines the Phillips curve slope and the magnitude and the persistence of the real effects of monetary policy shocks in this economy. A lower value of κ leads to a flatter Phillips curve, more persistent inflation and output processes, and higher monetary non-neutrality. Firms with a lower κ are more inattentive and acquire information at a slower pace. It takes longer for such firms to learn about changes in their input costs and respond to them. Meanwhile, firms' output responds to compensate for the partial pass-through of input costs to prices. Hence, more inattention leads to a larger and longer output response.

While comparative statics of κ with respect to the model parameters are derived in Corollary 3.1, in this section, we are particularly interested in how the rule of monetary policy affects the Phillips curve slope and consequently the transmission of monetary policy to output and inflation.

3.4.2 The Aftermath of an Unexpectedly More Hawkish Monetary Policy

According to Corollary 3.1, κ increases with $\frac{\sigma_u}{\phi}$. We interpret this ratio as a measure for the hawkishness of monetary policy. A higher value for ϕ (or a smaller value for $\frac{\sigma_u}{\phi}$) corresponds to a more stabilized process for nominal demand and constitutes a more hawkish monetary policy.

What happens to the Phillips curve slope when the monetary policy is more hawkish? To answer this question, we consider an economy where all firms are in the steady-state of their information acquisition and study the transition dynamics of their attention when monetary policy unexpectedly becomes more hawkish ($\frac{\sigma_u}{\phi}$ *decreases*).

Corollary 3.2. *Suppose the economy is in the steady state of its attention problem, and consider an unexpected and permanent decrease in $\frac{\sigma_u}{\phi}$. Then, the economy immediately jumps to a new steady state of the attention problem, in which (a) the Phillips curve is flatter and (b) output and inflation responses are more persistent.*

Proof. See Appendix D.5. ■

When monetary becomes unexpectedly more hawkish, nominal demand becomes more stable because innovations to its process are less volatile. This lower volatility affects firms' information

[33] In the steady-state of the rational inattention problem, $\sigma^2_{t|t-1} = \underline{\sigma}^2 + \frac{\sigma_u^2}{\phi^2}$. Plugging this into the expression for the Phillips curve in Part (1) of Proposition 3.2, we get $\pi_t = \frac{\sigma_u^2}{\phi^2 \underline{\sigma}^2} y_t$. It is straightforward from Equation (3.7) to verify that $\frac{\kappa}{1-\kappa} = \frac{\sigma_u^2}{\phi^2 \underline{\sigma}^2}$.

acquisition in two different ways. First, it increases the value of information because knowledge about nominal demand depreciates less over time when its process is more stable. Second, firms can afford a lower level of steady-state uncertainty with the same information acquisition rate (κ). These two forces manifest themselves in a desire for lower $\underline{\sigma}^2$ and a lower steady-state information acquisition rate (lower κ).

Therefore, when the shock happens, firms acquire enough information to jump to the new steady-state because their uncertainty from the previous steady-state is above their new reservation uncertainty. After that, firms also acquire information with a lower κ because their uncertainty grows more slowly. Thus, when monetary policy becomes more hawkish, output and inflation are more persistent, and the Phillips curve is flatter. These results are consistent with the flattening of the Phillips curve since the onset of the Great Moderation.[34] Our theory provides a new perspective on this issue. Firms do not need to be attentive to monetary policy in an environment where the policymakers 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 and permanent 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 and (b) output and inflation responses are less persistent.*

Proof. See Appendix D.6. ■

An increase in $\frac{\sigma_u}{\phi}$ makes the nominal demand more volatile and raises the reservation uncertainty of firms. Hence, immediately after an unexpected increase in $\frac{\sigma_u}{\phi}$, firms find themselves with an uncertainty below their new reservation level. The no-forgetting constraint begins to bind, and firms temporarily stop paying attention to the shocks until their uncertainty grows to its new reservation level. In the meantime, the Phillips curve is entirely flat; inflation is non-responsive to shocks, and output responds one to one to changes in nominal demand.

[34]For evidence on the flattening of the Phillips curve, see e.g., Coibion and Gorodnichenko (2015b); Blanchard (2016); Bullard (2018); Hooper, Mishkin, and Sufi (2020); Del Negro, Lenza, Primiceri, and Tambalotti (2020).

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.

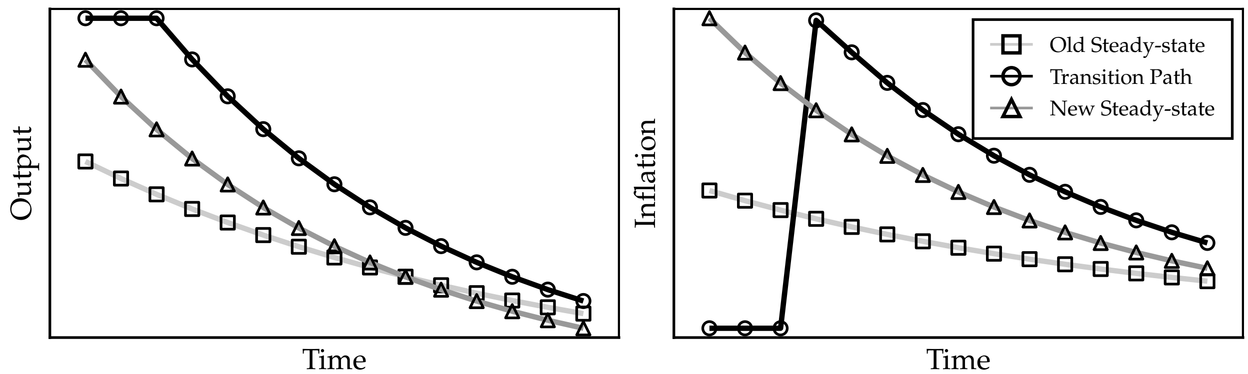

Figure 3: IRFs to a 1 S.D. Expansionary Shock When Policy Becomes More Dovish

Notes: This figure plots a numerical example for impulse responses of inflation and output to a one standard deviation expansionary shock to monetary policy. The lines with square markers are under the steady-state information structure. The lines with circle markers are the impulse responses after the policy becomes unexpectedly more dovish at time 0, with the transition dynamics of the information structure. The lines with triangle markers are the responses after the information structure converges to new steady-state with the more dovish policy. See Proposition 3.2 and Corollary 3.3 for details.

Figure 3 illustrates this point in a numerical example by plotting three different sets of impulse response functions. The light gray lines with square markers show the impulse responses under an initial steady-state information structure for firms. The black lines show these impulse responses after the monetary policy become unexpectedly more dovish in period zero. The main observation is the temporary flatness of the Phillips curve that occurs in the first three periods. The shock to dovishness of policy increases firms' reservation uncertainty and crowds out their information acquisition for three periods. After that, they begin to pay attention to shocks again, and inflation sharply picks up to match nominal demand shock. Finally, the dark gray lines with triangle markers show the impulse responses of output and inflation in the long-run after firms' information structure converges to its new steady-state. In this new steady-state, both inflation and output are more volatile, and their impact responses are larger because nominal demand is more volatile than the previous regime. However, these responses are less persistent as firms acquire information with a higher κ.

These findings provide a new perspective on the recently perceived disconnect between inflation and monetary policy. Our model offers an attention-based rationale for this disconnect, assuming that the Great Recession was followed by a period of higher uncertainty and more lenient monetary policy.

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 monetary policy success as they imply that the publics' inflation expectations are not very sensitive to temporary disturbances in economic variables. Moreover, inflation expectations have become more anchored in the U.S. since the onset of the Great Moderation: inflation expectations are more stable and 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. When monetary policy becomes more hawkish, firms pay less attention to shocks. Hence, their beliefs become less sensitive to short-run fluctuations in economic data, and their expectations become more anchored. 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,*

$$\hat{\pi}_t = (1-\kappa)\hat{\pi}_{t-1} + \frac{\kappa^2}{(2-\kappa)(1-\kappa)} y_t \tag{3.8}$$

$$= 2(1-\kappa)\hat{\pi}_{t-1} - (1-\kappa)^2 \hat{\pi}_{t-2} + \frac{\kappa^2}{2-\kappa}\frac{\sigma_u}{\phi} u_t \tag{3.9}$$

where κ is the steady-state Kalman-gain of firms in Equation (3.7).

Proof. See Appendix D.7. ■

Proposition 3.3 illustrates the degree of anchoring in firms' inflation expectations from two perspectives. Equation (3.8) derives the relationship between inflation expectations and output gap and shows that inflation expectations' sensitivity to output gap depends positively on κ. Equation (3.9) 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 indicates the 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.[35]

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

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

shocks depend on the conduct of monetary policy through κ. The following Corollary formalizes this relationship.[36]

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 D.8. ■

4 Quantitative Analysis

In this section, we extend our simple model in Section 3 to a more quantitatively plausible setup. Our objective is to assess whether our mechanism can generate a quantitatively relevant change in the Phillips curve slope in a calibrated model.

Our exercise in this section is in the spirit of the literature that interprets the Great Moderation, at least partially, through the lens of a shift in monetary policy in the post-Volcker era (Clarida, Galí, and Gertler, 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 Phillips curve slope, 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 Model

We extend our simple model from Section 3 in three dimensions. First, we introduce two new parameters on the household side for the inverse of the intertemporal elasticity of substitution (σ) and the inverse of the Frisch elasticity of labor supply (ψ). Second, we allow for strategic complementarities in pricing, which we excluded from the simple model but are quantitatively important for inflation dynamics. Third, we relax our Taylor rule specification to allow for interest rate smoothing and different central bank responses to inflation, output gap, and output growth. Appendix F provides a detailed explanation of this setup and the definition of general equilibrium. Here we present the log-linearized equilibrium conditions that characterize that equilibrium:

[36]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, 2021).

$$x_t = \mathbb{E}_t^f \left[x_{t+1} - \sigma^{-1} \left(i_t - \pi_{t+1} \right) \right] + \mathbb{E}_t^f \left[\Delta y_{t+1}^n \right] \qquad \forall t \geq 0 \qquad (4.1)$$

$$i_t = \rho i_{t-1} + (1-\rho) \left(\phi_\pi \pi_t + \phi_x x_t + \phi_{\Delta y} \Delta y_t \right) - \sigma_u u_t \qquad \forall t \geq 0 \qquad (4.2)$$

$$p_{i,t} = \mathbb{E}_{i,t} \left[p_t + \alpha x_t \right] \qquad \forall t \geq 0, \forall i \in [0,1]. \qquad (4.3)$$

Equation (4.1) is the standard log-linearized Euler equation for the household with full-information rational expectations. Here, σ^{-1} is the intertemporal elasticity of substitution, $y_t^n \equiv \frac{\sigma+\psi}{1+\psi} a_t$ is the log of the natural level of output with no frictions (which is uniquely determined by the productivity shock), x_t is the output gap defined as the log difference between output and its natural level, and π_t is inflation.

Equation (4.2) is the log-linearized Taylor rule, where ρ is the degree of interest rate smoothing, y_t is the log output, $u_t \sim \mathcal{N}(0,1)$ is the monetary policy shock, and ϕ_π, ϕ_x and $\phi_{\Delta y}$ are the responses of the central bank to inflation, output gap and output growth respectively.

Equation (4.3) shows that firm i tracks its nominal marginal cost, $p_t + \alpha x_t$, where $p_{i,t}$ is the firm's log price at t, p_t is the log of the aggregate price level, and $\alpha \equiv \frac{\sigma+\psi}{1+\psi\theta}$ is the degree of strategic complementarity. Moreover, $\mathbb{E}_{i,t}[\cdot]$ is firm i's expectation operator conditional on its time t information set under the solution to its rational inattention problem.

4.2 Computing the Equilibrium

The main computational challenge is solving for firms' rational inattention problem. This has two stages: first, given a Markov state-space representation for $p_t + \alpha x_t$ we can use our algorithm from Section 2. However, the process for $p_t + \alpha x_t$ is endogenous to the equilibrium decisions of firms and households. Therefore, a second step is to find the equilibrium process for $p_t + \alpha x_t$. It is important to note that for these two steps to be consistent, we need to choose a state-space representation for $p_t + \alpha x_t$ that is Markov.

We start by guessing for the MA representation of $p_t + \alpha x_t$ as a function of the productivity (ε_t) and monetary policy (u_t) shocks, which gives us a Markov representation for the process. We then approximate the process with a truncated MA process and use this truncated process as the input to our algorithm for DRIPs. We then solve for the implied state-space representations of the output gap and aggregate price and update our guess for the MA process of $p_t + \alpha x_t$. We repeat until convergence. When truncating the MA process of $p_t + \alpha x_t$, we approximate this process with an MA(160) process. Therefore, the corresponding rational inattention problem has $12880 = \frac{160 \times 161}{2}$ state variables. On average, for a given guess for this process, it takes 0.30 seconds for our DRIPs algorithm to solve for the implied steady-state covariance matrix (that is of dimension 160^2).[37]

[37] Truncated MA process are not necessarily the most efficient guesses for endogenous variables and one can gain efficiency in the second step of this iterative process by using ARMA guesses, which reduce the number of state variables (Maćkowiak, Matějka, and Wiederholt, 2018). However, MA processes assume a higher flexibility for the

Appendix F.3 provides a detailed description of matrix representations, and our solution algorithm.

4.3 Calibration

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

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	Quarterly frequency
Elasticity of substitution across firms (θ)	10	Firms' average markup
Elasticity of intertemporal substitution ($1/\sigma$)	0.4	Aruoba, Bocola, and Schorfheide (2017)
Inverse of Frisch elasticity (ψ)	2.5	Aruoba, Bocola, and Schorfheide (2017)
Taylor rule: smoothing (ρ)	0.946	Estimates 1983–2007 (Table G.1)
Taylor rule: response to inflation (ϕ_π)	2.028	Estimates 1983–2007 (Table G.1)
Taylor rule: response to output gap (ϕ_x)	0.168	Estimates 1983–2007 (Table G.1)
Taylor rule: response to output growth ($\phi_{\Delta y}$)	3.122	Estimates 1983–2007 (Table G.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 G.1)
Taylor rule: response to inflation (ϕ_π)	1.589	Estimates 1969–1978 (Table G.1)
Taylor rule: response to output gap (ϕ_x)	0.292	Estimates 1969–1978 (Table G.1)
Taylor rule: response to output growth ($\phi_{\Delta y}$)	1.028	Estimates 1969–1978 (Table G.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 quantitative 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.

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, Bocola, and Schorfheide (2017).

solution when the length of truncation is large. Since our algorithm is fast enough to be able to handle a large number of state variables, we use this structure in our solution method.

Monetary Policy Rule(s). In our benchmark calibration, we choose the standard deviation of monetary policy shocks (σ_u) to match the size of these shocks, as identified by Romer and Romer (2004), for the period 1983–2007.[38]

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. For our benchmark calibration, we perform this estimation for the the post-Volcker (1983–2007) sample.[39] The point estimates are reported in Panel B of Table 2, and more detailed results including standard errors are reported in Appendix Table G.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).[40]

Finally, for our counterfactual analysis in later sections, we do a similar estimation of these parameters for the pre-Volcker sample (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 G.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.[41] The three moments (variances of inflation and GDP with their covariance) 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 larger than the standard deviation of the monetary policy shock (σ_u) for the post-Volcker period.[42] Moreover, the calibrated cost of information processing, $\omega\mathbb{I}(.,.)$, is 0.1 percent

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

[39] 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.

[40] 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.

[41] We detrend the log of real GDP and core CPI inflation data using quadratic trends to measure the covariance matrix of output and inflation.

[42] Since we assume that only two shocks (productivity and monetary shocks) drive business cycles in our model, the calibrated size of the productivity shocks is quite large compared to the previous business cycle literature. For example, mean estimate of standard deviation of TFP shocks in Smets and Wouters (2007) is 0.45 percent.

of firms' steady-state real revenue.[43] This small calibrated cost implies that imperfect information models do not require large information costs to match the data.

4.4 Model Fit

Targeted Moments. Columns (1) and (2) of Table 3 reports our targeted moments both in the data and as implied by the model. All three targeted moments are matched by the model.

Table 3: Targeted and Non-targeted Moments

	Targeted moments (Post-Volcker: 1983–2007)		Non-targeted moments (Pre-Volcker: 1969–1978)	
Moment	(1) Data	(2) Model	(3) Data	(4) Model
Standard deviation of inflation	0.015	0.015	0.025	0.025
Standard deviation of real GDP	0.018	0.018	0.022	0.020
Correlation(inflation, real GDP)	0.209	0.209	0.242	0.245

Notes: Columns (1) and (2) present moments of the data and simulated series from the model parameterized at the baseline values in Table 2. Columns (3) and (4) 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

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). Our estimated values for these parameters are reported in Panel C of Table 2, and indicate that monetary policy was less responsive to inflation and output growth in the pre-Volcker period, and the monetary shocks were more volatile.

We then simulate the model under the calibrated values for the cost of attention and the process for the TFP shocks and calculate the implied covariance matrix for GDP and inflation. Columns (3) and (4) of Table 3 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 matches the high volatility of inflation and GDP in the pre-Volcker period.

[43]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.

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

Because the Phillips curve slope is endogenous in the model, the change in the rule of monetary policy in the post-Volcker period would imply a potential change in this slope as well. In this section, we study whether these changes are consistent with a flatter Phillips curve in the post-Volcker period within the model. If so, is the mechanism quantitatively relevant?

The main challenge here is to constitute the right comparison between the model and the empirical evidence on the Phillips curve slope. While the empirical literature uses the New Keynesian Phillips curve (NKPC) as the equation guiding their estimation strategy, according to our model the NKPC is misspecified. 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 goes back long enough in time to cover both periods. Such a dataset does not exist for the U.S. to the best of our knowledge.[44] The alternative strategy that we employ here is to simulate data from our model under the two specifications of monetary policy and run similar regressions as in the empirical literature. These regressions are misspecified from our model's perspective and provide biased estimates due to an omitted variables bias issue. However, they constitute a fair comparison to the evidence on the Phillips curve slope.

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.

$$\pi_t = constant + \gamma \mathbb{E}_t[\pi_{t+1}] + (1-\gamma)\pi_{t-1} + \kappa x_t + \varepsilon_t. \tag{4.4}$$

We use four lags of both inflation and output gap as instruments. Column (1) in Table 4 shows the estimates of the NKPC. 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.

It worth noting that while we are interested in the relative slope in the two periods, the magnitude of these estimates slopes are well above the estimates in the empirical literature.[45] However, this is not necessarily inconsistent with our model. One challenge that the the empirical literature faces is controlling for supply shocks that confound the estimates of the Phillips curve slope and

[44] An alternative dataset for testing our model's implications might be the Survey of Professional Forecasters. In fact, using this dataset, Coibion and Gorodnichenko (2015a) find that the estimated degree of informational rigidity for professional forecasters reached its minimum level in 1983–84 and since then it has consistently increased, which is consistent with the implications of our model. Moreover, as shown in Angeletos, Huo, and Sastry (2020), the estimated informational rigidities are also higher in the post-Volcker period than in the pre-Volcker period when they use unemployment as a running variable. Nonetheless, we are cautious against using the Survey of Professional Forecasters data since previous literature has found that firms' expectations formation process is very different from that of professional forecasters' (e.g., Coibion and Gorodnichenko, 2012, 2015a; Coibion, Gorodnichenko, and Kumar, 2018).

[45] For example, the estimated slope of the hybrid NKPC in Galí, Gertler, and López-Salido (2005) is 0.002. In Del Negro, Lenza, Primiceri, and Tambalotti (2020), the slope estimates for the post-Volcker period range from 0 to 0.01.

Table 4: 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 NKPC using simulated data from the baseline model presented in Section 4.2. Column (1) and (2) show the estimates of the NKPC in Equation (4.4) 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.

introduces a downward bias (McLeay and Tenreyro, 2020). To see whether this type of downward bias can bring us closer to these estimates we repeat our estimation exercise with imperfect measures of output gap in our model. Column (2) in Table 4 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, in Column (3) we partially control for the supply shocks by subtracting a moving average of the natural level of output from realized output 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.[46]

5 Concluding Remarks

We derive an information Euler equation that fully characterizes the transition path of dynamic rational inattention problems in LQG settings and use our theoretical results to propose a novel and fast solution method that significantly reduces the computing times for solving these problems. We 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 Phillips curve slope in recent decades. It suggests that this was an endogenous response of the private sector to a more disciplined monetary policy in the post-Volcker era, putting a larger weight on stabilizing nominal variables.

Our results also speak to an ongoing debate on the trade-off between stabilizing inflation and

[46] Appendix Table G.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.

maintaining a lower unemployment rate on the policy front. Our theory suggests that while a dovish policy might seem appealing in the current climate where inflation seems hardly responsive to monetary policy, once implemented, such a policy might have an adverse effect by steepening the Phillips curve in the long-run.

References

AFROUZI, H. (2020): "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money," Working Paper 8218, CESifo.

ANGELETOS, G.-M., AND Z. HUO (2021): "Myopia and Anchoring," *American Economic Review*, 111(4), 1166–1200.

ANGELETOS, G.-M., Z. HUO, AND K. A. SASTRY (2020): "Imperfect Macroeconomic Expectations: Evidence and Theory," in *NBER Macroeconomics Annual 2020, volume 35*, NBER Chapters. National Bureau of Economic Research, Inc.

ANGELETOS, G.-M., AND J. LA'O (2009): "Incomplete Information, Higher-Order Beliefs and Price Inertia," *Journal of Monetary Economics*, 56, S19–S37.

ANGELETOS, G.-M., AND C. LIAN (2016): "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," vol. 2 of *Handbook of Macroeconomics*, pp. 1065–1240. Elsevier.

——— (2018): "Forward Guidance without Common Knowledge," *American Economic Review*, 108(9), 2477–2512.

ARUOBA, B., L. BOCOLA, AND F. SCHORFHEIDE (2017): "Assessing DSGE model nonlinearities," *Journal of Economic Dynamics and Control*, 83(C), 34–54.

AZEREDO DA SILVEIRA, R., AND M. WOODFORD (2019): "Noisy Memory and Over-Reaction to News," *AEA Papers and Proceedings*, 109, 557–61.

BERNANKE, B. S. (2007): "Inflation Expectations and Inflation Forecasting," Speech 306, Board of Governors of the Federal Reserve System.

BHATIA, R. (2002): "Eigenvalues of AB and BA," *Resonance*, 7(1), 88–93.

BLANCHARD, O. (2016): "The Phillips Curve: Back to the'60s?," *American Economic Review*, 106(5), 31–34.

BOYD, S., AND L. VANDENBERGHE (2004): *Convex optimization*. Cambridge university press.

BULLARD, J. B. (2018): "The Case of the Disappearing Phillips Curve," Speech 314, Federal Reserve Bank of St. Louis.

CAPLIN, A., M. DEAN, AND J. LEAHY (2017): "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy," Working Paper 23652, National Bureau of Economic Research.

CAPLIN, A., AND D. SPULBER (1987): "Menu Costs and the Neutrality of Money," *The Quarterly Journal of Economics*, 102(4), 703–25.

CLARIDA, R., J. GALÍ, AND M. GERTLER (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, 115(1), 147–80.

COIBION, O., AND Y. GORODNICHENKO (2011): "Monetary Policy, Trend Inflation, and the Great Moderation: An Alternative Interpretation," *American Economic Review*, 101(1), 341–70.

——— (2012): "What Can Survey Forecasts Tell Us about Information Rigidities?," *Journal of Political Economy*, 120(1), 116–159.

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

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

COIBION, O., Y. GORODNICHENKO, L. KUENG, AND J. SILVIA (2017): "Innocent Bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 88, 70–89.

COIBION, O., Y. GORODNICHENKO, AND S. KUMAR (2018): "How Do Firms Form Their Expectations? New Survey Evidence," *American Economic Review*, 108(9), 2671–2713.

COVER, T., AND J. THOMAS (2012): *Elements of Information Theory*. John Wiley & Sons.

DEL NEGRO, M., M. LENZA, G. E. PRIMICERI, AND A. TAMBALOTTI (2020): "What's up with the Phillips Curve?," Working Paper 27003, National Bureau of Economic Research.

DENTI, T. (2015): "Unrestricted Information Acquisition," Manuscript.

FORBES, K. (2019): "Inflation Dynamics: Dead, Dormant, or Determined Abroad?," Discussion paper, National Bureau of Economic Research.

FULTON, C. (2018): "Mechanics of Static Quadratic Gaussian Rational Inattention Tracking Problems," Working paper, Federal Reserve Board of Governors.

GABAIX, X. (2014): "A Sparsity-Based Model of Bounded Rationality," *The Quarterly Journal of Economics*, 129(4), 1661–1710.

——— (2020): "A behavioral New Keynesian model," *American Economic Review*, 110(8), 2271–2327.

GALÍ, J. (2015): *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.

GALÍ, J., M. GERTLER, AND J. D. LÓPEZ-SALIDO (2005): "Robustness of the estimates of the hybrid New Keynesian Phillips curve," *Journal of Monetary Economics*, 52(6), 1107 – 1118, The econometrics of the New Keynesian price equation.

GERTLER, M., AND J. LEAHY (2008): "A Phillips Curve with an Ss Foundation," *Journal of Political Economy*, 116(3), 533–72.

GOLOSOV, M., AND R. LUCAS (2007): "Menu Costs and Phillips curves," *Journal of Political Economy*, 115(2), 171–99.

HAZELL, J., J. HERREÑO, E. NAKAMURA, AND J. STEINSSON (2020): "The Slope of the Phillips Curve: Evidence from U.S. States," Working Paper 28005, National Bureau of Economic Research.

HÉBERT, B., AND J. LA'O (2020): "Information Acquisition, Efficiency, and Non-Fundamental Volatility," Working Paper 26771, National Bureau of Economic Research.

HÉBERT, B., AND M. WOODFORD (2018): "Information Costs and Sequential Information Sampling," Working Paper 25316, National Bureau of Economic Research.

HOOPER, P., F. S. MISHKIN, AND A. SUFI (2020): “Prospects for Inflation in a High Pressure Economy: Is the Phillips Curve Dead or is It Just Hibernating?,” *Research in Economics*, 74(1), 26–62.

JORGENSEN, P., AND K. LANSING (2019): “Anchored Inflation Expectations and the Flatter Phillips Curve,” Working Paper Series 2019-27, Federal Reserve Bank of San Francisco.

KAMDAR, R. (2018): “The Inattentive Consumer: Sentiment and Expectations,” Manuscript.

KIM, Y.-H., AND S.-J. KIM (2006): “On the Convexity of log det (I+ KX^{-1}),” *arXiv preprint cs/0611043*.

KŐSZEGI, B., AND F. MATĚJKA (2020): “Choice Simplification: A Theory of Mental Budgeting and Naive Diversification,” *The Quarterly Journal of Economics*, 135(2), 1153–1207.

LEVY, D., M. BERGEN, S. DUTTA, AND R. VENABLE (1997): “The Magnitude of Menu Costs: Direct Evidence From Large U. S. Supermarket Chains,” *The Quarterly Journal of Economics*, 112(3), 791–825.

L'HUILLIER, J.-P., AND W. ZAME (2020): “Flattening of the Phillips Curve and the Mandate of the Central Bank,” Discussion paper, Mimeo.

LUCAS, R. (1972): “Expectations and the Neutrality of Money,” *Journal of Economic Theory*, 4(2), 103–124.

MAĆKOWIAK, B., F. MATĚJKA, AND M. WIEDERHOLT (2018): “Dynamic Rational Inattention: Analytical Results,” *Journal of Economic Theory*, 176, 650 – 692.

——— (2020): “Rational Inattention: A Review,” CEPR Discussion Papers 15408, CEPR.

MAĆKOWIAK, B., AND M. WIEDERHOLT (2009a): “Optimal Sticky Prices under Rational Inattention,” *The American Economic Review*, 99(3), 769–803.

——— (2009b): “Optimal Sticky Prices under Rational Inattention,” Working Paper Series 1009, European Central Bank.

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

MACKOWIAK, B., AND M. WIEDERHOLT (2020): “Rational Inattention and the Business Cycle Effects of Productivity and News Shocks,” Manuscript.

MANKIW, G., AND R. REIS (2002): “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve,” *The Quarterly Journal of Economics*, 117(4), 1295–1328.

MATĚJKA, F. (2015): “Rationally Inattentive Seller: Sales and Discrete Pricing,” *The Review of Economic Studies*, 83(3), 1125–1155.

MATĚJKA, F., AND A. MCKAY (2015): “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model,” *American Economic Review*, 105(1), 272–98.

MCLEAY, M., AND S. TENREYRO (2020): “Optimal Inflation and the Identification of the Phillips Curve,” *NBER Macroeconomics Annual*, 34(1), 199–255.

MIAO, J., J. WU, AND E. YOUNG (2020): “Multivariate Rational Inattention,” Boston University - Department of Economics, February 2020.

MISHKIN, F. (2007): “Inflation dynamics,” *International Finance*, 10(3), 317–34.

MUTH, J. F. (1961): “Rational Expectations and the Theory of Price Movements,” *Econometrica*, pp. 315–35.

NAGEL, S., AND Z. XU (2019): "Asset Pricing with Fading Memory," Working Paper 26255, National Bureau of Economic Research.

NAKAMURA, E., AND J. STEINSSON (2010): "Monetary Non-Neutrality in a Multisector Menu Cost Model," *The Quarterly Journal of Economics*, 125(3), 961–1013.

NIMARK, K. (2008): "Dynamic Pricing and Imperfect Common Knowledge," *Journal of Monetary Economics*, 55(2), 365–82.

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

RAVID, D. (2020): "Ultimatum Bargaining with Rational Inattention," *American Economic Review*, 110(9), 2948–63.

ROMER, C., AND D. ROMER (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94(4), 1055–84.

RUBBO, E. (2020): "Networks, Phillips Curves and Monetary Policy," Discussion paper, Harvard University, Mimeo.

SIMS, C. A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665–690.

——— (2006): "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 96(2), 158–163.

——— (2010): "Rational Inattention and Monetary Economics," in *Handbook of Monetary Economics*, ed. by B. M. Friedman, and M. Woodford, vol. 3, pp. 155–81. Elsevier.

SMETS, F., AND R. WOUTERS (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97(3), 586–606.

SONG, W., AND S. STERN (2020): "Firm Inattention and the Efficacy of Monetary Policy: A Text-Based Approach," Manuscript.

STEINER, J., C. STEWART, AND F. MATĚJKA (2017): "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 85(2), 521–553.

STEVENS, L. (2019): "Coarse Pricing Policies," *The Review of Economic Studies*, 87(1), 420–453.

STOCK, J. H., AND M. W. WATSON (2020): "Slack and cyclically sensitive inflation," *Journal of Money, Credit and Banking*, 52(S2), 393–428.

WOODFORD, M. (2003): "Imperfect Common Knowledge and the Effects of Monetary Policy," in *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, ed. by P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, pp. 25–58. Princeton University Press.

——— (2009): "Information-Constrained State-Dependent Pricing," *Journal of Monetary Economics*, 56, S100–S124.

APPENDIX
FOR ONLINE PUBLICATION ONLY

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

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

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 $X^t \to S^t \to a^t$ 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.[47] ■

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 $\forall t\geq 0$:

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

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

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

$\vec{a}_t = \mathbf{H}'\vec{x}_{t|t}$ 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 RI Problem in Equation (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. The problem has two set of constraints: (1) a set of $\frac{n(n+1)}{2}$ equality constraints that are introduced by the law of motion for the priors in Equation (2.5) (2) a set of n non-negativity constraints on the eigenvalues of the matrix $\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}$. We let $\mathbf{\Gamma}_t$ be a matrix whose k'th row is the vector of Lagrange multipliers on the k'th column of the evolution of prior at time t (note that in matrix notation, each constraint is introduced twice by the symmetry of the prior matrix, except for the ones on the diagonal. Hence, $\mathbf{\Gamma}_t$ is also symmetric by the symmetry of the constraints). Moreover, let λ_t be the vector of shadow costs on the vector of no-forgetting constraints, which we refer to as $\text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}) \geq 0$ where $\text{eig}(.)$ denotes the vector of eigenvalues of a matrix.

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

Our goal is to take the FO(N)C conditions with respect to the elements of the matrix $\mathbf{\Sigma}_{t|t}$. First, we transform the non-negativity constraints in terms of $\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}$ instead of its eigenvalues:

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

where $\text{diag}(.)$ 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

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

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

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

Moreover, note that complementary slackness for this constraint requires:

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

re-writing the Lagrangian we get:

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

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

$$\mathbf{\Omega} - \omega\mathbf{\Sigma}_{t|t}^{-1} + \mathbf{A}'\mathbf{\Gamma}_t\mathbf{A} + \mathbf{\Lambda}_t = 0, \;\text{ and }\; \omega\beta\mathbf{\Sigma}_{t+1|t}^{-1} - \mathbf{\Gamma}_t - \beta\mathbf{\Lambda}_{t+1} = 0.$$

Now, replacing for $\mathbf{\Gamma}_t$ in the first order conditions we get the conditions in the Proposition.

One result that we have assumed in writing this expression is that $\mathbf{\Sigma}_{t|t-1}$ is invertible, which follows from the assumptions of the Proposition. The claim is:

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

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

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

Thus,

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

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

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

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

$$(\mathbf{AA'w} = 0) \wedge (\mathbf{QQ'w} = 0) \Rightarrow (\mathbf{AA'} + \mathbf{QQ'})\mathbf{w} = 0$$

but since $\mathbf{AA'} + \mathbf{QQ'}$ 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.

Moreover, we have a terminal optimality condition that requires:

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

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

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

■

A.6 Proof of Proposition 2.2

Proof. Since the no-forgetting constraints are affine functions of the posterior variances, we only need to show that the objective function is concave in $(\mathbf{\Sigma}_{t|t})_{t\geq 0}$. Our first observation is that the objective is separable in the elements of this vector. In particular, substituting $\mathbf{\Sigma}_{t+1|t} = \mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{QQ}', \forall t \geq 0$ in Equation (2.4), we can write the objective of the agent as

$$\begin{aligned} V_0\left((\mathbf{\Sigma}_{t|t})_{t\geq 0}\right) = \; & -\frac{\omega}{2}\ln\left(|\mathbf{\Sigma}_{0|-1}|\right) \\ & +\frac{1}{2}\sum_{t=0}^{\infty}\beta^t\left[-tr(\mathbf{\Sigma}_{t|t}\mathbf{\Omega}) - \beta\omega\ln\left(|\mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{QQ}'|\right) + \omega\ln\left(|\mathbf{\Sigma}_{t|t}|\right)\right] \end{aligned}$$

Therefore, since all $\mathbf{\Sigma}_{t|t}$'s appear additively and independent of one another in the objective function, to show that the objective is concave in $(\mathbf{\Sigma}_{t|t})_{t\geq 0}$, it suffices to show that it is concave with respect to each $\mathbf{\Sigma}_{t|t}, \forall t \geq 0$.

Moreover, since $tr(\mathbf{\Sigma}_{t|t}\mathbf{\Omega})$ is linear in (and it second derivative is 0 with respect to) $\mathbf{\Sigma}_{t|t}$, we just need to show that for all $t \geq 0$,

$$\beta\ln\left(|\mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{QQ}'|\right) - \ln\left(|\mathbf{\Sigma}_{t|t}|\right)$$

is convex in $\mathbf{\Sigma}_{t|t}$. Now, observe that for any $\mathbf{\Sigma}_{t|t} = \mathbf{\Sigma} \succ \mathbf{0}$, this function can be rewritten in the following form:

$$
\begin{aligned}
&\beta \ln\left(|\mathbf{A}\mathbf{\Sigma}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'|\right) - \ln\left(|\mathbf{\Sigma}|\right) \\
=&\beta \ln(|\mathbf{A}(\mathbf{\Sigma} - \mathbf{I})\mathbf{A}' + \mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'|) - \ln(|\mathbf{\Sigma}|) && \text{(add and subtract } \mathbf{A}\mathbf{A}'\text{)} \\
=&\beta \ln\left(|\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'|\right) && \text{(factor out } \mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'\text{)} \\
+&\beta \ln\left(|\mathbf{I} + (\mathbf{\Sigma} - \mathbf{I})\mathbf{A}'(\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}')^{-1}\mathbf{A}|\right) - \ln\left(|\mathbf{\Sigma}|\right) && \text{(by Sylvester's identity)} \\
=&\beta \ln\left(|\mathbf{\Sigma}^{-1} + (\mathbf{I} - \mathbf{\Sigma}^{-1})\mathbf{A}'(\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}')^{-1}\mathbf{A}|\right) && \text{(factor out } \mathbf{\Sigma}\text{)} \\
-&(1-\beta) \ln\left(|\mathbf{\Sigma}|\right) + \beta \ln\left(|\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'|\right)
\end{aligned}
$$

Now, ignoring the constant $\ln\left(|\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'|\right)$ and defining $\Delta \equiv \mathbf{A}'(\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}')^{-1}\mathbf{A}$ we just need to show that

$$\beta \ln(|\mathbf{\Sigma}^{-1} + (\mathbf{I} - \mathbf{\Sigma}^{-1})\Delta|) - (1-\beta)\ln(|\mathbf{\Sigma}|) \tag{A.1}$$

is a convex function of $\mathbf{\Sigma}$. Notice that since for all $\beta \in [0,1]$, $-(1-\beta)\ln\left(|\mathbf{\Sigma}|\right)$ is itself a convex function of $\mathbf{\Sigma}$ (Boyd and Vandenberghe, 2004, , p. 74), a sufficient condition for the convexity of the function in Equation (A.1) is if the first term,

$$C(\mathbf{\Sigma}) \equiv \ln(|\Delta + \mathbf{\Sigma}^{-1}(\mathbf{I} - \Delta)|)$$

is convex (because it would imply that Equation (A.1) is the sum of two convex functions). To show this is indeed the case, in the remainder of this proof we proceed to show that $C(\mathbf{\Sigma})$ is the pointwise limit of a sequence of convex functions, and therefore is itself convex. To this end, we take advantage of the following two results. First, we rely on the following Lemma from information theory (see, e.g., Kim and Kim, 2006, for a short and elegant proof):

Lemma A.1. *For positive semi-definite matrices K and X, the function $\ln\left(|I + KX^{-1}|\right)$ is convex in X.*

Second, we rely on the following property of the matrix Δ:

Lemma A.2. *For real-valued matrices $\mathbf{A}$ and $\mathbf{Q}$, $\mathbf{0} \preceq \Delta = \mathbf{A}'(\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}')^{-1}\mathbf{A} \preceq \mathbf{I}$, meaning that all the eigenvalues of Δ are between 0 and 1.*

Proof. First, note that both $\mathbf{A}\mathbf{A}'$ and $\mathbf{Q}\mathbf{Q}'$ are positive semi-definite as they are the each the product of a real-valued matrix by its transpose. It immediately follows that Δ is positive semi-definite.

Suppose now that λ is an eigenvalue of Δ. Then $\lambda \geq 0$. Moreover, since eigenvalues of product of two square matrices are independent of the order of multiplication (see, e.g., Bhatia, 2002), λ is also an eigenvalue of $\mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}')^{-1}$, so there exists a vector $\vec{z}$ such that

$$\mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}')^{-1}\vec{z} = \lambda\vec{z}$$

summing this with $\mathbf{QQ}'(\mathbf{AA}'+\mathbf{QQ}')^{-1}\vec{z}$ on both sides, and moving terms around, we get that

$$\mathbf{QQ}'(\mathbf{AA}'+\mathbf{QQ}')^{-1}\vec{z}=(1-\lambda)\vec{z}$$

meaning that $1-\lambda$ is an eigenvalue of $\mathbf{QQ}'(\mathbf{AA}'+\mathbf{QQ}')^{-1}$. Therefore, $1-\lambda$ is also an eigenvalue of $\mathbf{Q}'(\mathbf{AA}'+\mathbf{QQ}')^{-1}\mathbf{Q}$, which is a symmetric, positive semi-definite matrix, so that $1-\lambda\geq 0$. Thus, we have established that $0\leq\lambda\leq 1$. ■

Now, consider the following spectral decomposition of $\Delta=\mathbf{UDU}'$ (where $\mathbf{D}$ is diagonal and $\mathbf{U}$ is an orthonormal matrix), and define the sequence $\Delta_n\equiv\mathbf{U}\max\{\mathbf{D},\frac{1}{n}\}\mathbf{U}'$, for $n\geq 1$. Notice this sequence has the following two properties:

1. it converges to Δ : $\lim_{n\to\infty}\Delta_n=\mathbf{U}\max\{\mathbf{D},0\}\mathbf{U}'=\mathbf{UDU}'=\Delta$ where the last equality follows from the fact that elements of $\mathbf{D}$ are non-negative.
2. $0\prec\frac{1}{n}\mathbf{I}\preceq\mathbf{\Delta}_n\preceq\mathbf{I}$ (since diagonal elements of $\max\{\mathbf{D},\frac{1}{n}\}$ are bounded between $1/n$ and 1). One notable implication of this property is that Δ_n is invertible (as all of its eigenvalues are strictly positive).

Lemma A.3. *For all $n\geq 1$, let $C_n(\mathbf{\Sigma})\equiv\ln(|\Delta_n+\mathbf{\Sigma}^{-1}(\mathbf{I}-\Delta_n)|)$. Then, (1) the function $C_n(\mathbf{\Sigma})$ is convex for all $n\geq 1$ and (2) $C_n(\mathbf{\Sigma})\to C(\mathbf{\Sigma})$ pointwise.*

Proof. **Part 1 (convexity).** Since Δ_n is invertible, factor out Δ_n to see

$$C_n(\mathbf{\Sigma})=\ln(|\mathbf{I}+\mathbf{\Sigma}^{-1}(\Delta_n^{-1}-\mathbf{I})|)+\ln(|\Delta_n|)$$

Therefore, applying Lemma A.1, we know that this is a convex function if $\Delta_n^{-1}-\mathbf{I}\succeq 0$, which is true because all the eigenvalues of Δ_n are between $1/n$ and 1 (so that all the eigenvalues of $\Delta_n^{-1}-\mathbf{I}$ are between 0 and $n-1\geq 0$).

Part 2 (pointwise convergence). Since $\ln(|X|)$ is continuous and $\Delta_n\to\Delta$, for any given $\mathbf{\Sigma}\succeq\mathbf{0}$, we have

$$\ln(|\Delta_n+\mathbf{\Sigma}^{-1}(\mathbf{I}-\Delta_n)|)\to\ln(|\Delta+\mathbf{\Sigma}^{-1}(\mathbf{I}-\Delta)|)\Rightarrow C_n(\mathbf{\Sigma})\to C(\mathbf{\Sigma}),\forall\mathbf{\Sigma}\succ\mathbf{0}$$

■

Now, since the pointwise limit of a sequence of convex functions is convex, $C(\mathbf{\Sigma})$ is convex on the positive semi-definite cone. ■

A.7 Proof of Theorem 2.1

Proof. From the FOC in Proposition 2.1 observe that

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

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 $\mathbf{X}_t\boldsymbol{\Lambda}_t = 0$. Finally, multiply this equation by $\boldsymbol{\Sigma}_{t|t-1}^{\frac{1}{2}}$ from right and $\boldsymbol{\Sigma}_{t|t-1}^{-\frac{1}{2}}$ from left.[48] We have

$$\hat{\mathbf{X}}_t\mathbf{D}_t - \hat{\boldsymbol{\Lambda}}_t = \mathbf{D}_t - \omega\mathbf{I} \tag{A.2}$$

where

$$\hat{\mathbf{X}}_t \equiv \boldsymbol{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{X}_t\boldsymbol{\Sigma}_{t|t-1}^{-\frac{1}{2}}, \quad \hat{\boldsymbol{\Lambda}}_t \equiv \boldsymbol{\Sigma}_{t|t-1}^{\frac{1}{2}}\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}_{t|t-1}^{\frac{1}{2}}, \quad \mathbf{D}_t \equiv \boldsymbol{\Sigma}_{t|t-1}^{\frac{1}{2}}\boldsymbol{\Omega}_t\boldsymbol{\Sigma}_{t|t-1}^{\frac{1}{2}} \tag{A.3}$$

Now, note that $\boldsymbol{\Lambda}_t\mathbf{X}_t = \mathbf{X}_t\boldsymbol{\Lambda}_t = 0$ implies $\hat{\boldsymbol{\Lambda}}_t\hat{\mathbf{X}}_t = \hat{\mathbf{X}}_t\hat{\boldsymbol{\Lambda}}_t = 0$. Similarly, note that $\mathbf{X}_t$ and $\boldsymbol{\Lambda}_t$ are positive semidefinite if and only if $\hat{\mathbf{X}}_t$ and $\hat{\boldsymbol{\Lambda}}_t$ are positive semidefinite, respectively. So we need for two simultaneously diagonalizable symmetric positive semidefinite matrices $\hat{\boldsymbol{\Lambda}}_t$ and $\hat{\mathbf{X}}_t$ that solve Equation (A.2).

It follows from these that $\hat{\mathbf{X}}_t$, $\hat{\boldsymbol{\Lambda}}_t$ and $\mathbf{D}_t$ are simultaneously diagonalizable. To see this, consider the following revisions of Equation (A.2):

$$\hat{\mathbf{X}}_t\mathbf{D}_t = \mathbf{D}_t + \hat{\boldsymbol{\Lambda}}_t - \omega\mathbf{I}, \quad -\hat{\boldsymbol{\Lambda}}_t\mathbf{D}_t = \hat{\boldsymbol{\Lambda}}_t^2 - \omega\hat{\boldsymbol{\Lambda}}_t$$

where the equation on the left is a simple re-arrangement of Equation (A.2) and the equation on the right is one where we have multiplied Equation (A.2) with $\hat{\boldsymbol{\Lambda}}_t$ from left. Note that the right hand side of both of these equations are symmetric matrices. Therefore, the left hand side of the should also be symmetric, which implies that $\mathbf{D}_t$ commutes with both $\hat{\mathbf{X}}_t$ and $\hat{\boldsymbol{\Lambda}}_t$. Now, since all three matrices are diagonalizable (because they are symmetric) and any two of them commute with one another, they are simultaneously diagonalizable . Let α denote a basis in which these matrices are diagonal. Then, we have

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

Using complementarity slackness $[\hat{\boldsymbol{\Lambda}}_t]_\alpha[\hat{\mathbf{X}}_t]_\alpha = \mathbf{0}$, the constraint $[\hat{\mathbf{X}}_t]_\alpha \succeq \mathbf{0}$, and dual feasibility constraint $[\hat{\boldsymbol{\Lambda}}_t]_\alpha \succeq \mathbf{0}$ it is straight forward to show that $[\boldsymbol{\Lambda}_t]_\alpha$ is strictly positive for the eigenvalues (entries on the diagonal) of $[\mathbf{D}_t]_\alpha$ that are smaller than ω.

$$[\hat{\boldsymbol{\Lambda}}_t]_\alpha = \max(\omega\mathbf{I} - [\mathbf{D}_t]_\alpha, \mathbf{0}) \Leftrightarrow \hat{\boldsymbol{\Lambda}}_t = \mathrm{Max}(\omega\mathbf{I}{-}\mathbf{D}_t, \mathbf{0}).$$

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

Now, using Equation (A.3), we get:

$$\mathbf{\Lambda}_t = \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \operatorname{Max}(\omega \mathbf{I} - \mathbf{D}_t, \mathbf{0}) \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}. \tag{A.4}$$

Moreover, recall $\omega \mathbf{\Sigma}_{t|t}^{-1} = \mathbf{\Omega}_t + \mathbf{\Lambda}_t$. Using the solution for $\mathbf{\Lambda}_t$ and $\mathbf{\Omega}_t = \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \mathbf{D}_t \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}$:

$$\begin{aligned}
\omega \mathbf{\Sigma}_{t|t}^{-1} &= \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \left[\mathbf{D}_t + \operatorname{Max}(\omega \mathbf{I} - \mathbf{D}_t, \mathbf{0})\right] \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \\
&= \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \operatorname{Max}(\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \mathbf{\Omega}_t \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}, \omega) \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}.
\end{aligned}$$

Inverting this gives us the expression in the statement of 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.4) we have:

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

■

A.8 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

$$\begin{aligned}
\mathbb{E}[\vec{x}_t | a^t] &= \mathbb{E}[\vec{x}_t | a^{t-1}] + \mathbf{K}_t (\vec{a}_t - \mathbb{E}[\vec{a}_t | a^{t-1}]) \\
&= \mathbb{E}[\vec{x}_t | a^{t-1}] + \mathbf{K}_t \mathbf{Y}_t' (\vec{x}_t - \mathbb{E}[\vec{x}_t | a^{t-1}]) + \mathbf{K}_t \vec{z}_t
\end{aligned} \tag{A.5}$$

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:

$$\mathbf{\Sigma}_{t|t} = \mathbf{\Sigma}_{t|t-1} - \mathbf{K}_t \mathbf{Y}_t' \mathbf{\Sigma}_{t|t-1} \Rightarrow \mathbf{K}_t \mathbf{Y}_t' = \mathbf{I} - \mathbf{\Sigma}_{t|t} \mathbf{\Sigma}_{t|t-1}^{-1}. \tag{A.6}$$

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:

$$\begin{aligned}
\mathbf{K}_t \mathbf{Y}_t' &= \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \mathbf{U}_t (\mathbf{I} - \omega \operatorname{Max}(\mathbf{D}_t, \omega)^{-1}) \mathbf{U}_t' \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \\
&= \sum_{i=1}^{n} \max(0, 1 - \frac{\omega}{d_{i,t}}) \mathbf{\Sigma}_{t|t-1} \mathbf{y}_{i,t} \mathbf{y}_{i,t}'
\end{aligned} \tag{A.7}$$

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.6) we have:

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

Plugging in for $\mathbf{\Sigma}_{t|t}$ from Equation (2.14) 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.9 Proof of Proposition 2.3

Proof. Let $\hat{x}_t \equiv \mathbb{E}[\vec{x}_t|a^t]$. Combining Equation (A.5) and Equation (A.7), we have

$$\begin{aligned}\hat{x}_t &= \mathbb{E}[\vec{x}_t|a^{t-1}] + \sum_{i=1}^{n}\max(0, 1 - \frac{\omega}{d_{i,t}})\mathbf{\Sigma}_{t|t-1}\mathbf{y}_{i,t}\mathbf{y}_{i,t}'(\vec{x}_t - \mathbb{E}[\vec{x}_t|a^{t-1}]) + \mathbf{K}_t\vec{z}_t \\ &= \mathbf{A}\hat{x}_{t-1} + \sum_{i=1}^{k_t}(1 - \frac{\omega}{d_{i,t}})\mathbf{\Sigma}_{t|t-1}\mathbf{y}_{i,t}(\mathbf{y}_{i,t}'(\vec{x}_t - \mathbf{A}\hat{x}_{t-1}) + z_{i,t}),\end{aligned}$$

where k_t is the number of the eigenvalues that are at least as large as ω. Notice that, as shown in Appendix A.4, we normalize $\vec{a}_t = \mathbf{H}'\hat{x}_t$ since any linear map of $\vec{a}_t$, as long as it is of rank m, is sufficient for $\hat{x}_t$ by sufficiency of action for signals. ■

A.10 Steady-state Information Structure with $\beta = 1$

Here, we show that the solution to the problem in Equation (2.18) is characterized by the steady-state version of our conditions in Proposition 2.1 when $\beta = 1$. The Lagrangian for the problem in Equation (2.18) is:

$$\max_{\mathbf{\Sigma},\mathbf{\Sigma_{-1}}} -tr(\mathbf{\Omega\Sigma}) - \omega\ln(|\mathbf{\Sigma_{-1}}|) + \omega\ln(|\mathbf{\Sigma}|) - tr(\mathbf{\Gamma}(\mathbf{A\Sigma A}' + \mathbf{QQ}' - \mathbf{\Sigma}_{-1})) + \lambda' eig(\mathbf{\Sigma_{-1}} - \mathbf{\Sigma})$$

where $\mathbf{\Gamma}$ is a symmetric matrix whose k'th row is the set of Lagrange multipliers on the constraints imposed by the k'th column of $\mathbf{\Sigma}_{-1} = \mathbf{A\Sigma A}' + \mathbf{QQ}'$. Now, consider the following Spectral decomposition of $\mathbf{\Sigma}_{-1} - \mathbf{\Sigma}$:

$$\mathbf{\Sigma}_{-1} - \mathbf{\Sigma} = \mathbf{U}\,\mathrm{diag}(\mathrm{eig}(\mathbf{\Sigma}_{-1} - \mathbf{\Sigma}))\mathbf{U}'$$

where by the Spectral theorem, $\mathbf{U}$ is an orthonormal basis so that $\mathbf{UU}' = \mathbf{U}'\mathbf{U} = \mathbf{I}$, and $\mathrm{diag}(\mathrm{eig}(\mathbf{\Sigma}_{-1} - \mathbf{\Sigma}))$ denotes the diagonal matrix of the eigenvalues of $\mathbf{\Sigma}_{-1} - \mathbf{\Sigma}$.

Following the proof of Proposition 2.1, we can also do the following transformation: $\lambda'\,\mathrm{eig}(\mathbf{\Sigma}_{-1} -$

$\boldsymbol{\Sigma}) = tr(\boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{-1} - \boldsymbol{\Sigma}))$, where $\boldsymbol{\Lambda} \equiv \mathbf{U}\,\mathrm{diag}(\lambda)\mathbf{U}'$. Then, the KKT conditions are

$$\boldsymbol{\Omega} - \omega\boldsymbol{\Sigma}^{-1} + \mathbf{A}'\boldsymbol{\Gamma}\mathbf{A} + \boldsymbol{\Lambda} = \mathbf{0}, \quad \omega\boldsymbol{\Sigma}_{-1}^{-1} - \boldsymbol{\Gamma} - \boldsymbol{\Lambda} = \mathbf{0}.$$

Replacing for $\boldsymbol{\Gamma}$ we have $\omega\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Omega} = \boldsymbol{\Lambda} + \mathbf{A}'(\omega\boldsymbol{\Sigma}_{-1}^{-1} - \boldsymbol{\Lambda})\mathbf{A}$ with complementary slackness conditions that $\boldsymbol{\Lambda}(\boldsymbol{\Sigma}_{-1} - \boldsymbol{\Sigma}) = (\boldsymbol{\Sigma}_{-1} - \boldsymbol{\Sigma})\boldsymbol{\Lambda} = 0$ where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}_{-1} - \boldsymbol{\Sigma}$ are simultaneously diagonalizable. Notice that these conditions are identical to the conditions outlined in Proposition 2.1, when we impose the steady-state and set $\beta = 1$.

B Solution Algorithm for DRIPs

In this Appendix, we provide a detailed outline of our Euler equation algorithm for solving the transition dynamics and the steady state information structure of DRIPs.

Solving for the Steady-State Information Structure. The steady-state information structure is 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 2.14, 2.5 and 2.13 respectively). We solve for this triple using the following iterative algorithm, starting with initial guesses for $\bar{\boldsymbol{\Sigma}}_{-1} = \bar{\boldsymbol{\Sigma}}_{-1,(0)}$ and $\bar{\boldsymbol{\Omega}} = \bar{\boldsymbol{\Omega}}_{(0)}$.[49] Then, in any iteration $\mathtt{j} \geq 1$:

1. Obtain the eigenvalue and eigenvector decomposition of
$$\mathbf{X}_{(\mathtt{j})} \equiv \bar{\boldsymbol{\Sigma}}^{\frac{1}{2}}_{-1,(\mathtt{j}-1)}\bar{\boldsymbol{\Omega}}_{(\mathtt{j}-1)}\bar{\boldsymbol{\Sigma}}^{\frac{1}{2}}_{-1,(\mathtt{j}-1)}$$

2. Use Theorem 2.1 to update guesses:
$$\bar{\boldsymbol{\Omega}}_{(\mathtt{j})} = \boldsymbol{\Omega} + \beta\mathbf{A}'\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}_{-1,(\mathtt{j}-1)}\,\mathrm{Min}\left(\mathbf{X}_{(\mathtt{j})}, \omega\right)\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}_{-1,(\mathtt{j}-1)}\mathbf{A}$$
$$\bar{\boldsymbol{\Sigma}}_{-1,(\mathtt{j})} = \omega\mathbf{A}\bar{\boldsymbol{\Sigma}}^{\frac{1}{2}}_{-1,(\mathtt{j}-1)}\left[\mathrm{Max}\left(\mathbf{X}_{(\mathtt{j})}, \omega\right)\right]^{-1}\bar{\boldsymbol{\Sigma}}^{\frac{1}{2}}_{-1,(\mathtt{j}-1)}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'$$

3. Repeat with $\mathtt{j}$+=1 if $\|\boldsymbol{\Sigma}_{-1,(\mathtt{j})} - \boldsymbol{\Sigma}_{-1,(\mathtt{j}-1)}\| >$ tolerance.

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

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

[49] By default, our solution algorithm sets $\bar{\boldsymbol{\Omega}}_{(0)} = \mathbf{H}\mathbf{H}'$ and $\bar{\boldsymbol{\Sigma}}_{(0)} = \mathbf{A}\mathbf{A}' + \mathbf{Q}\mathbf{Q}'$. However, the user can specify alternative guesses, especially in iterative estimation exercises where a solution from a previous step might be closer to the solution.

We find this sequence using the following iterative procedure, starting from the initial guess that for all $t \in \{0, 1, \ldots, T\}$, $\mathbf{\Omega}_{t,(0)} = \bar{\mathbf{\Omega}}$:

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

$$\mathbf{\Sigma}_{t+1|t,(\mathtt{j})} \equiv \omega \mathbf{A} \mathbf{\Sigma}^{\frac{1}{2}}_{t|t-1,(\mathtt{j})} \left[\text{Max} \left(\mathbf{\Sigma}^{\frac{1}{2}}_{t|t-1,(\mathtt{j})} \mathbf{\Omega}_{t,(\mathtt{j}-1)} \mathbf{\Sigma}^{\frac{1}{2}}_{t|t-1,(\mathtt{j})}, \omega \right) \right]^{-1} \mathbf{\Sigma}^{\frac{1}{2}}_{t|t-1,(\mathtt{j})} \mathbf{A}' + \mathbf{Q}\mathbf{Q}'$$

2. At iteration $\mathtt{j} \geq 1$, given the sequence $\{\mathbf{\Sigma}_{t+1|t,(\mathtt{j})}\}_{0 \leq t \leq T} \cup \{\Sigma_{T+1|T,(\mathtt{j})} \equiv \bar{\mathbf{\Sigma}}_{-1}\}$ and $\mathbf{\Omega}_{T+1,(\mathtt{j})} \equiv \bar{\mathbf{\Omega}}$, iterate backward in time using the Euler equation from Theorem 2.1: for $t = T \downarrow 0$,

$$\mathbf{\Omega}_{t,(\mathtt{j})} \equiv \mathbf{\Omega} + \beta \mathbf{A}' \mathbf{\Sigma}^{-\frac{1}{2}}_{t+1|t,(\mathtt{j})} \, \text{Min} \left(\mathbf{\Sigma}^{\frac{1}{2}}_{t+1|t,(\mathtt{j})} \mathbf{\Omega}_{t+1,(\mathtt{j})} \mathbf{\Sigma}^{\frac{1}{2}}_{t+1|t,(\mathtt{j})}, \omega \right) \mathbf{\Sigma}^{-\frac{1}{2}}_{t+1|t,(\mathtt{j})} \mathbf{A}$$

3. Repeat Steps 2 to 4 with `j+=1` if $\|(\mathbf{\Sigma}_{t+1|t,(\mathtt{j})})_{t=0}^{T} - (\mathbf{\Sigma}_{t+1|t,(\mathtt{j}-1)})_{t=0}^{T}\| >$ tolerance.

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

C Replications

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

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

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

$$\min_{\{\hat{\Delta}_{i,t}, \hat{z}_{i,t}\}} E\left[(\Delta_t - \hat{\Delta}_{i,t})^2\right] + \left(\frac{\hat{\pi}_{14}}{\hat{\pi}_{11}}\right)^2 E\left[(z_{i,t} - \hat{z}_{i,t})^2\right],$$
$$s.t. \quad \mathcal{I}(\{\Delta_t\}; \{\hat{\Delta}_{i,t}\}) + \mathcal{I}(\{z_{i,t}\}; \{\hat{z}_{i,t}\}) \leq \kappa, \quad \{\Delta_t, \hat{\Delta}_{i,t}\} \perp \{z_{i,t}, \hat{z}_{i,t}\}$$

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 $p_t = \int_0^1 \hat{\Delta}_{i,t} di$ and exogenous shock processes are defined:

$$q_t = \rho_q q_{t-1} + \nu_{q,t}, \nu_{q,t} \sim \mathcal{N}(0, \sigma_q^2), \quad z_{i,t} = \rho_z z_{i,t-1} + \nu_{z,t}, \nu_{z,t} \sim \mathcal{N}(0, \sigma_z^2).$$

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 $\beta = 1$. We have included a complete replication of the figures and graphs from Maćkowiak and Wiederholt (2009a) online at https://afrouzi.com/DRIPs.jl/dev/examples/ex3_mw2009/ex3_Mackowiak_Wiederholt_2009/.

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

We describe the model of price-setting in Maćkowiak, Matějka, and Wiederholt (2018) with and without endogenous feedback in firms' optimal prices.

C.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 $\Delta q_t = \rho \Delta q_{t-1} + u_t$ with $u_t \sim \mathcal{N}(0, \sigma_u^2)$. Then, the state-space representation of the problem is

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

Endogenous Feedback with Strategic Complementarity Now we consider the case where there is general equilibrium feedback with the degree of strategic complementarity α. Firm i's optimal price is $p^*_{i,t} = (1-\alpha) q_t + \alpha p_t$ where $p_t \equiv \int_0^1 p_{i,t} di$ and $\Delta q_t = \rho \Delta q_{t-1} + u_t$ with $u_t \sim \mathcal{N}(0, \sigma_u^2)$. 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 guessed that the process for $p^*_{i,t}$ is determined uniquely by the history of shocks which requires that rational inattention errors of firms are orthogonal. Our objective is to find $\Phi(.)$.

We approximate $MA(\infty)$ processes with truncation. In particular, 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: $p^*_{i,t} = \phi(L) \tilde{u}_t$ where $\tilde{u}_t = (1-L)^{-1} u_t = \sum_{j=0}^{\infty} u_{t-j}$. Here, $\tilde{u}_{t-j}$ is the unit root of the process and 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 in ℓ_2, and the process can now be approximated arbitrarily precisely with truncation.

C.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, Matějka, and Wiederholt (2018).

The techonology shock, z_t, 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:

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

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

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

Then, firm i's choice depends on its information set at time t and $n_{i,t} = E_{i,t}[n_t^*]$.

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. We have included a complete replication of the figures and graphs from Maćkowiak and Wiederholt (2009a) online at https://afrouzi.com/DRIPs.jl/dev/examples/ex5_mmw2018/ex5_Mackowiak_Matejka_Wiederholt_2018/.

D Proofs for Section 3

D.1 Proof of Lemma 3.1

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

$$\log(R_t) = \log(\beta^{-1}) + \mathbb{E}_t[\Delta q_{t+1}].$$

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. ■

D.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:

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

■

D.3 Proof of Corollary 3.1

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

D.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 (D.1) 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. ■

D.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 $\frac{\sigma_u}{\phi}$. ■

D.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. ■

D.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 in Equation (3.7) 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

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

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

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

■

D.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). ■

E Approximation of Firms' Profit Function

Consider a firm with the following net present value of its profits at time 0:

$$\sum_{t=0}^{\infty} \beta^t \Pi(P_t, W_t, X_t)$$

where $\Pi(P_t, W_t, X_t) = e^{\log(X_t)-\theta\log(P_t)}(e^{\log(P_t)} - (1-\theta^{-1})e^{\log(W_t)})$. Here, P_t is the firm's price X_t scales the profit function (in both our simple and quantitative models $X_t = \bar{P}_t^{\theta} Y_t (C_t/C_0)^{-\sigma}$ where $\bar{P}_t$ is the aggregate price, and Y_t is the aggregate output, $(C_t/C_0)^{-\sigma}$ is stochastic part of the discount factor). Moreover, W_t is the firm's marginal cost which corresponds to the nominal wage in the simple model, and nominal wage scaled by productivity in the quantitative model.

For any pair of W_t and X_t, note that

$$W_t = \arg\max_P \Pi(P_t, W_t, X_t) \Leftrightarrow \Pi_1(W_t, W_t, X_t) = 0$$

Therefore, in the non-stochastic steady-state $P = W$ for this firm. Now taking a second-order approximation to the function $L(P_t, W_t, X_t) \equiv \Pi(P_t, W_t, X_t) - \Pi(W_t, W_t, X_t)$ around these steady-state values, we have

$$\begin{aligned} L(P_t, W_t, X_t) &= \frac{1}{2}\Pi_{11}(\log(P_t)^2 - \log(W_t)^2) + \Pi_{12}\log(W_t)(\log(P_t) - \log(W_t)) \\ &+ \Pi_{13}\log(X_t)(\log(P_t) - \log(W_t)) + \mathcal{O}(\|\log(X_t), \log(W_t), \log(P_t)\|^3) \end{aligned} \tag{E.1}$$

where $\Pi_{1n}, n \in \{1,2,3\}$ denotes second order derivatives of the profit function with respect to log price, log price and log wage, and log price and $\log(X)$ around teh approximation point. Now, also note that since W_t maximizes the profit function for any W_t and X_t, we have

$$\Pi_1(W_t, W_t, X_t) = 0 \Rightarrow \Pi_{11}\log(W_t) + \Pi_{12}W_t + \Pi_{13}X_t + \mathcal{O}(\|\log(X_t), \log(W_t)\|^2) = 0$$

Combining this with Equation (E.1) we have

$$\begin{aligned} \Pi(P_t, W_t, X_t) = L(P_t, W_t, X_t) + \Pi(W_t, W_t, X_t) &= \frac{1}{2}\Pi_{11}(\log(P_t) - \log(W_t))^2 \\ &+ \mathcal{O}(\|\log(W_t), \log(X_t), \log(P_t)\|^3) + \text{terms independent of } P_t \end{aligned}$$

Finally, to calculate Π_{11}, note that

$$\begin{aligned} \Pi_1 &= -\theta e^{\log(X)-\theta\log(P)}(e^{\log(P)} - (1-\theta^{-1})e^{\log(W)}) + e^{\log(X)-(\theta-1)\log(P)}, \\ \Pi_{11} &= -\theta\Pi_1 - (\theta-1)e^{\log(X)-(\theta-1)\log(P)} \end{aligned}$$

Now assuming that the aggregate price is the same as the firm's individual price in the stochastic steady-state, $\log(X) - (\theta-1)\log(P) = \log(PYC^{-\sigma}) = \log(Q)$ where Q is the steady-state value of nominal demand. Moreover, since $\Pi_1 = 0$ in the steady-state, we have $\Pi_{11} = -(\theta-1)Q$. Hence, normalizing the steady-state value of nominal demand to 1:

$$\begin{aligned} \sum_{t=0}^{\infty}\beta^t \Pi(P_t, W_t, X_t) = -\frac{1}{2}\sum_{t=0}^{\infty}\beta^t \left[(\theta-1)(\log(P_t) - \log(W_t))^2 + \mathcal{O}(\|\log(P_t)\|^3)\right] \\ + \text{terms independent of } \{P_t\}_{t\geq 0} \end{aligned}$$

F Setup of the Quantitative Model

F.1 Environment

Household. The representative household's problem is similar to the one in Equation (3.1) with two extensions. First, we add two new parameters to preferences that capture the intertemporal elasticity of substitution and the Frisch elasticity of labor supply. Second, we assume segmented labor markets for different varieties, which is a known mechanism to generate quantitatively plausible strategic complementarities in pricing. Formally, the household solves:

$$\max_{\{C_t,B_t,(C_{i,t},L_{i,t})_{i\in[0,1]}\}_{t\geq 0}} \mathbb{E}_t^f \left[\sum_{t=0}^{\infty} \beta^t \left(\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{\int_0^1 L_{i,t}^{1+\psi} di}{1+\psi}\right)\right] \tag{F.1}$$
$$s.t. \quad \int_0^1 P_{i,t}C_{i,t}di + B_t \leq R_{t-1}B_{t-1} + \int_0^1 W_{i,t}L_{i,t}di + \Pi_t$$

with CES aggregator, $C_t = \left[\int_0^1 C_{i,t}^{\frac{\theta-1}{\theta}} di\right]^{\frac{\theta}{\theta-1}}$. Here all variables and notation are similarly defined as in Equation (3.1), with the addition that $L_{i,t}$ now represents the household's labor supply in the segmented labor market i given wage $W_{i,t}$. Moreover, σ is the inverse of the intertemporal elasticity of substitution, and ψ is the inverse of the Frisch elasticity of labor supply.

Monetary Policy. Monetary policy is specified as the following standard Taylor rule with interest smoothing that targets inflation, output gap and output growth:

$$\frac{R_t}{\bar{R}} = \left(\frac{R_{t-1}}{\bar{R}}\right)^{\rho} \left(\left(\frac{P_t}{P_{t-1}}\right)^{\phi_\pi} \left(\frac{Y_t}{Y_t^n}\right)^{\phi_x} \left(\frac{Y_t}{Y_{t-1}}\right)^{\phi_{\Delta y}}\right)^{1-\rho} \exp(-\sigma_u u_t) \tag{F.2}$$

where $\bar{R}$ is the steady-state nominal rate, $Y_t \equiv C_t$ is the aggregate output, Y_t^n is the natural-level of output in the economy with no frictions, and $u_t \sim \mathcal{N}(0,1)$ is the monetary policy shock.

Firms. There is a measure one of firms, indexed by i, that operate in monopolistically competitive markets and are price takers in their segmented labor market. 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 its prices, firm i hires labor $L_{i,t}^d$ to meet its demand with the production function $Y_{i,t} = A_t L_{i,t}^d$. Here, A_t is an aggregate productivity shock. We assume $a_t \equiv \log(A_t)$, follows a AR(1) process: $a_t = \rho_a a_{t-1} + \sigma_a \varepsilon_t, \ \ \varepsilon_t \sim \mathcal{N}(0,1)$.

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:

$$\max_{\{S_{i,t}\subseteq\mathcal{S}_{i,t},P_{i,t}(S_i^t)\}_{t\geq 0}} \quad \mathbb{E}\Bigg[\sum_{t=0}^{\infty}\beta^t C_t^{-\sigma}\Bigg\{\left(P_{i,t}-(1-\theta^{-1})\frac{W_{i,t}}{A_t}\right)\left(\frac{P_{i,t}}{P_t}\right)^{-\theta}Y_t \tag{F.3}$$

$$-\omega\mathbb{I}(S_i^t;(A_\tau,W_{i,\tau})_{\tau\leq t}|S_i^{t-1})\Bigg\}\Bigg|S_i^{-1}\Bigg]$$

$$s.t. \qquad S_i^t = S_i^{t-1}\cup S_{i,t},$$

where Y_t is the aggregate output, P_t is the aggregate price index, $P_{i,t}$ is the firm's price, $\theta^{-1}W_{i,t}$ is the optimal subsidy for hiring labor that eliminates the steady-state distortions from monopolistic pricing, and $\mathcal{S}_{i,t}$ is the set of available signals for the firm that satisfies the assumptions specified in Section 2.1.

Similar to our approach in the simple model, we derive a second-order approximation to the net present value of firms' profits (see Appendix E for detailed derivation) and define the firms' rational inattention problem as

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

Here, small letters denote log-deviations from the non-stochastic steady-state for their corresponding variables and $\alpha\equiv\frac{\sigma+\psi}{1+\theta\psi}$ is the degree of strategic complementarity. Moreover, $x_t\equiv y_t-y_t^n$ is the log output gap defined as the log difference between output and its natural-level in the economy with no frictions. The log natural-level of output is uniquely determined by the productivity shock and is given by $y_t^n\equiv\frac{1+\psi}{\psi+\sigma}a_t$. Finally, in stating this problem, we have already incorporated the result from Lemma 2.2, which states 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.

F.2 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, $\Omega^H\equiv\{C_t,B_t,(C_{i,t},L_{i,t})_{i\in[0,1]}\}_{t=0}^{\infty}$, 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 $\left\{\Omega_i^F\right\}_{i\in[0,1]}$, the household's allocation solves the problem in Equation (F.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 (F.3),

3. monetary policy satisfies the specified rule in Equation (F.2) ;

4. markets clear: $\forall i \in [0,1], \forall t \geq 0, Y_{i,t} = C_{i,t}, L_{i,t} = L^d_{i,t}$ and $Y_t = C_t$.

F.3 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. To bypass this issue, we re-write the state-space as: $p^*_{i,t} = \Phi_a(L)\varepsilon_{a,t} + \phi_u(L)\tilde{\varepsilon}_{u,t}$, $\tilde{\varepsilon}_{u,t} = (1-L)^{-1}\varepsilon_{u,t} = \sum_{j=0}^{\infty} \varepsilon_{u,t-j}$, where $\tilde{\varepsilon}_{u,t}$ 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 $\vec{x}'_t \equiv (z_t, z_{t-1}, \dots, z_{t-(L+1)}) \in \mathbb{R}^{2L}$ and $\vec{\mathbf{x}}'_t \equiv (\tilde{z}_t, \tilde{z}_{t-1}, \dots, \tilde{z}_{t-(L+1)}) \in \mathbb{R}^{2L}$. Notice that $\vec{x}_t = (\mathbf{I} - \mathbf{\Lambda M}')\vec{\mathbf{x}}_t$ and $\vec{\mathbf{x}}_t = (\mathbf{I} - \mathbf{\Lambda M}')^{-1}\vec{x}_t$ 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 polynomial, 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 $a_t = \mathbf{H}'_a\vec{x}_t$ and $u_t = \mathbf{H}'_u\vec{x}_t$ where$\mathbf{H}'_a = (1, 0, \rho_a, 0, \rho_a^2, 0, \dots, \rho_a^{L-1}, 0)$ and $\mathbf{H}'_u = (0, 1, 0, \rho_u, 0, \rho_a^2, \dots, 0, \rho_u^{L-1})$.

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

$$\vec{\mathbf{x}}_t = \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 \end{bmatrix}}_{\mathbf{A}} \vec{\mathbf{x}}_{t-1} + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}}_{\mathbf{Q}} z_t, \quad p^*_{i,t} = \mathbf{H}'_{(n-1)}\vec{\mathbf{x}}_t$$

Now, note that

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

$$= \mathbf{H}'_{(n-1)}\mathbf{X}_{(n)}\vec{\mathbf{x}}_t = \mathbf{H}'_p\vec{\mathbf{x}}_t$$

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 the households Euler equation, $x_t = \mathbb{E}^f_t \left[x_{t+1} - \frac{1}{\sigma}(i_t - \pi_{t+1})\right] + \mathbb{E}^f_t[y^n_{t+1}] - y^n_t$, gives:

$$\mathbf{H}_i = \sigma(\mathbf{M}' - \mathbf{I})\mathbf{H}_x + \frac{\sigma(1+\psi)}{\sigma+\psi}(\mathbf{M}' - \mathbf{I})\mathbf{H}_a + \mathbf{M}'\mathbf{H}_\pi$$

The Taylor rule, $i_t = \rho i_{t-1} + (1-\rho)(\phi_\pi \pi_t + \phi_x x_t + \phi_{\Delta y}(y_t - y_{t-1})) + u_t$, gives:

$$(\mathbf{I} - \rho\mathbf{M})\mathbf{H}_i = (1-\rho)\left(\phi_\pi \mathbf{H}_\pi + \phi_x \mathbf{H}_x + \phi_{\Delta y}(\mathbf{I} - \mathbf{M})\left(\mathbf{H}_x + \frac{1+\psi}{\sigma+\psi}\mathbf{H}_a\right)\right) + \mathbf{H}_u$$

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 $\mathbf{H}_{(n)}$.

F.4 Impulse Response Functions

For both the pre-Volcker and post-Volcker parameterization of monetary policy in Table 2, Figure G.1 shows the impulse responses of the model variables to one standard deviation TFP and monetary policy shocks. The main takeaway from these IRFs is that inflation, output, and both nominal and real interest rates respond more to shocks under the pre-Volcker parameterization of monetary policy.

G Appendix Figure and Tables

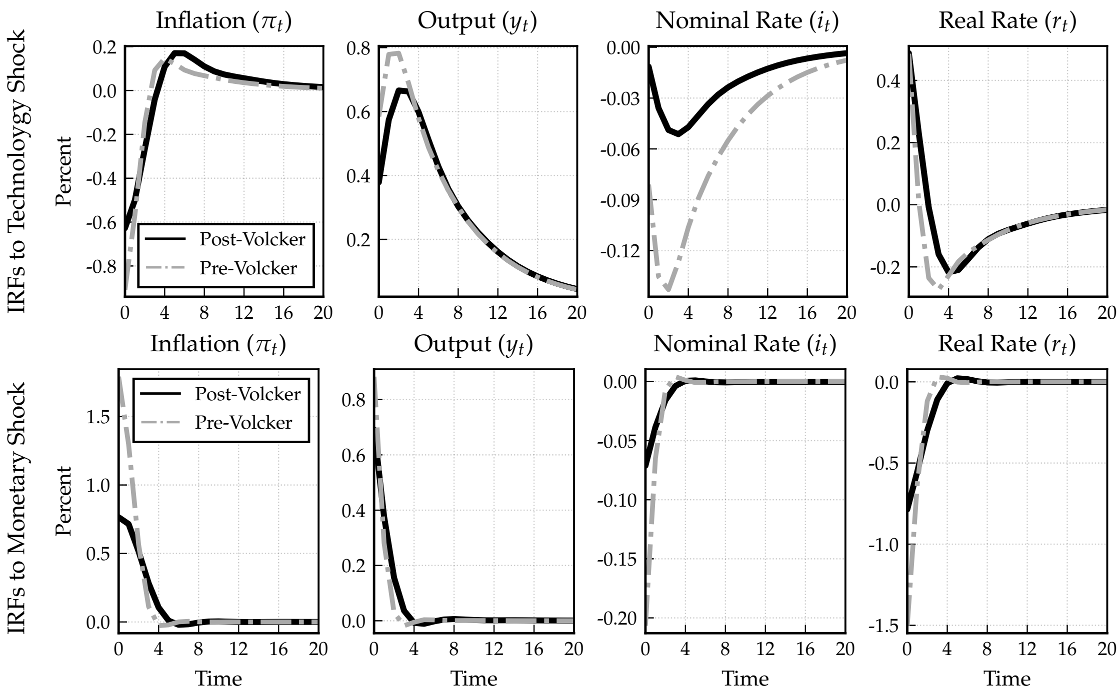

Figure G.1: Impulse Responses to Technology and Monetary Shocks

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.

Table G.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.

Table G.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 NKPC using simulated data from the baseline model presented in Section 4.2. Column (1) and (2) show the estimates of the NKPC 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.