Dynamic Rational Inattention and the Phillips Curve [*†]

Hassan Afrouzi [‡]
Columbia University

Choongryul Yang [§]
UT Austin

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

Abstract

We develop a tractable method for solving Dynamic Rational Inattention Problems (DRIPs) in LQG settings and propose an attention driven theory of the Phillips curve as an application of our general framework. We show that within a general equilibrium flexible price model with dynamic rational inattention, the slope of the Phillips curve is endogenous to systematic aspects of monetary policy. In our model, when the monetary authority is more committed to stabilizing nominal variables, rationally inattentive firms find it optimal to pay less attention to monetary policy shocks. Therefore, when monetary policy is more hawkish, the Phillips curve is flatter and inflation expectations are more anchored. In a quantitative exercise, we calibrate our general equilibrium model with TFP and monetary policy shocks to post-Volcker U.S. data and find that (1) our model can match the higher volatility of inflation and GDP in pre-Volcker era as non-targeted moments, and (2) our mechanism quantifies a 75% decline in the slope of the Phillips curve in the post-Volcker period.

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

[*]We are grateful to Saroj Bhattarai and Olivier Coibion for their guidance and support as well as to Eric Sims for a thoughtful discussion of an earlier draft of this paper. We would also like to thank Miguel Acosta, Mark Dean, Xavier Gabaix, Yuriy Gorodnichenko, Jennifer La'O, John Leahy, Yueran Ma, Filip Matějka, Emi Nakamura, Kris Nimark, Jón Steinsson, Luminita Stevens, Laura Veldkamp, Venky Venkateswaran, Mirko Wiederhold, Michael Woodford and seminar participants at 2018 and 2020 ASSA Meetings, Columbia University, SED Mexico City Conference, University of Texas at Austin, Federal Reserve Bank of Chicago, Cornell University, City University of New York and University of Wisconsin for helpful comments.

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

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

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

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

1 Introduction

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

While the existing New Keynesian models would suggest that such a shift in the slope of the Phillips curve is due to changes in the structural parameters of the model, we propose an attention driven theory of the Phillips curve in which the slope of the Phillips curve is endogenous to how monetary policy is conducted. In economies where the monetary authority puts a larger weight on stabilizing the nominal variables – in other words, when monetary policy is more hawkish – firms endogenously choose to pay less attention to monetary policy shocks. More specifically, our theory suggests that the change in the slope of the Phillips curve can be explained, at least partially, by the more *hawkish* monetary policy that was adopted in the beginning of the Great Moderation.[2] Since the onset of the Great Moderation, firms' nominal marginal costs are more stable relative to before that period, and accordingly they have lower incentives to track the shocks that affect this process. Therefore, prices react to nominal shocks to a lesser degree compared to the period before the Great Moderation, which in turn translates to an endogenously flatter Phillips curve.

Moreover, we show that an unexpectedly more dovish monetary policy can lead to a completely flat Phillips curve in the short-run while resulting in a steeper slope of the Phillips curve in the long-run. The intuition for this result is that in our model, a more dovish policy increases the cost of information acquisition by making nominal variables more volatile. Consequently, firms temporarily opt-out of paying attention to shocks as they find the marginal benefit of information acquisition far less than its new marginal cost. However, as time passes and firms' uncertainty accrues with the introduction of new shocks to the economy, the marginal benefit of information acquisition grows larger, and once it is as large as the new marginal cost firms restart paying attention to shocks.

[1] See, for instance, Coibion and Gorodnichenko (2015b); Blanchard (2016); Bullard (2018); Hooper et al. (2019); Del Negro et al. (2020).

[2] See Clarida et al. (2000); Coibion and Gorodnichenko (2011) for evidence on more hawkishness of monetary policy in the post-Volcker period.

However, in this new regime, due to higher volatility of the shocks, firms have to acquire information at a higher rate once they start doing so. This higher rate of information acquisition translates to a higher Kalman gain on firms' signals and increases the slope of the Phillips curve.

Furthermore, our model provides an endogenous explanation for how the conduct of monetary policy affects the anchoring of inflation expectations. Since attention is endogenous, firms' expectations of inflation is less sensitive to short-run fluctuations and co-move less with the output gap when monetary policy is more hawkish.

Our model of the Phillips curve is an application of a tractable method that we develop for solving dynamic rational inattention problems (DRIPs) with multiple shocks and actions in linear quadratic Gaussian settings. Our contribution in this area is to formulate and to characterize the full transition dynamics of DRIPs where the relevant state variable for a decision maker is his *stock of uncertainty* about the vector of shocks that they face. Specifically, we show that the transition dynamics in DRIPs is characterized by inaction regions for the decision maker's uncertainty in different dimensions of the state. Specifically, given that information is costly, a decision maker would only acquire information in a particular dimension if their uncertainty is at least as large as a reservation level. Moreover, we use our theoretical characterization of these problems to develop a computational toolbox that decreases solution times by several orders of magnitude.[3]

Our final contribution is to test the quantitative relevance of our proposed mechanism for the change in the slope of the Phillips curve. In our quantitative section, we take advantage of the efficiency of our computational toolbox and calibrate a dynamic general equilibrium version of our rational inattention model with monetary policy and TFP shocks. We use this calibrated model to perform two quantitative exercises.

The first quantitative exercise is in the spirit of Clarida et al. (2000) and is designed to check the out-of-sample fit of our calibrated model. Specifically, we replace the post-Volcker Taylor rule of monetary policy with an estimated Taylor rule for the pre-Volcker period and show that our model can quantitatively match the higher variance of inflation and GDP in the pre-Volcker era as non-targeted moments.

The second exercise directly assesses whether our proposed mechanism can explain

[3]We use our toolbox to replicate two seminal papers in the rational inattention literature and compare computational efficiency of our package to existing benchmarks. Our first replication is Maćkowiak and Wiederholt (2009a). While the original replication code takes around 280 seconds, our method allows us to find the solution in less than a second. Our second replication is the example from Sims (2010). There are no publicly available codes for this example; however, in a recent paper, Miao et al. (2020) propose a method that they report solves that example in 3 seconds. Our package allows to solve this example in 90 microseconds. All of our replication codes are publicly available at `http://github.com/afrouzi/DRIPs.jl`.

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

In regards to policy, the theory proposed in this paper offers a new point of view for the conduct of monetary policy relative to the New Keynesian models. While the slope of the Phillips curve in standard New Keynesian models is mainly pinned down by the frequency of price changes and is exogenous to how committed the monetary policy is to stabilizing the nominal variables, our model suggests a direct link between the two. Therefore, policy regimes that might seem optimal under an exogenously flat Phillips curve have completely different outlooks from the perspective of our model. For instance, from the perspective of a model where the slope of the Phillips curve is exogenous and flat, a more dovish monetary policy, or *running the economy hotter*, might seem appealing in order to reduce unemployment. Nonetheless, our model provides a different remedy: such policies would work in the short-run by pushing firms temporarily into their inaction region and paralyzing inflation. After this temporary period, however, firms would start paying more attention to monetary policy which would lead to a steeper Phillips curve and a larger sensitivity of inflation to monetary policy shocks.

Related Literature.

Dynamic rational inattention models have been applied to different settings for years.[4] Most of this literature, however, relies on simplifying assumptions – such as independence of signals – and computational methods in characterizing the solution. We provide a tractable solution method by building on a subset of this literature that has laid the ground for solving dynamic rational inattention models in LQG settings (Sims, 2003; Maćkowiak, Matějka and Wiederholt, 2018a; Fulton, 2018; Miao, Wu and Young, 2020). This literature makes two simplifying assumptions that we depart from: (1) they abstract away from transition dynamics by assuming that the cost of information is not discounted, and (2) they solve for the long-run steady-state information structure that is independent of time and state.

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

Our attention driven theory of the Phillips curve is motivated by two separate sets of empirical evidence. The first literature estimates,[5] and subsequently documents a flattening of the slope of the Phillips curve (Coibion and Gorodnichenko, 2015b; Blanchard, 2016; Bullard, 2018; Hooper, Mishkin and Sufi, 2019; Del Negro, Lenza, Primiceri and Tambalotti, 2020).[6] The second literature documents the information rigidities that economic agents exhibit in forming their expectations.[7]

Finally, we relate to the literature that considers how imperfect information affects the Phillips curve (Lucas, 1972; Mankiw and Reis, 2002; Woodford, 2003; Reis, 2006; Nimark, 2008; Angeletos and La'O, 2009; Angeletos and Huo, 2018; Angeletos and Lian, 2018). Our main departure is to derive a Phillips curve in a model with rational inattention and study how *monetary policy* shapes and alters the incentives in information acquisition of firms.

The paper is organized as follow. In Section 2, we start by setting up the dynamic rational inattention problem and then characterize the solution for the LQG case. In Section 3, we provide a simple version of our attention driven theory of the Phillips curve with analytical solutions. In Section 4, we present our quantitative model and results. Section 5 concludes. All proofs are included in the Appendix.

2 Theoretical Framework

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

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

[6] While we provide an attention based theory for this phenomena, an alternative explanation is non-linearities in the slope of the Phillips curve. See, for instance, Kumar and Orrenius (2016); Babb and Detmeister (2017); Hooper et al. (2019); McLeay and Tenreyro (2020).

[7] For recent progress in this literature, see for instance, Kumar et al. (2015); Coibion and Gorodnichenko (2015a); Ryngaert (2017); Coibion et al. (2018); Roth and Wohlfart (2018); Gaglianone et al. (2019); Angeletos et al. (2020) for survey evidence, and Khaw et al. (2017); Khaw and Zorrilla (2018); Landier et al. (2019) for experimental evidence.

2.1 Environment

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

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

1. $\mathcal{S}^t$ is *rich*: for any posterior distribution on X^t, there is a set of signals $S^t \subset \mathcal{S}^t$ that generate that posterior.

2. Available signals do not expire over time: $\mathcal{S}^t \subset \mathcal{S}^{t+h}, \forall h \geq 0$.

3. Available signals at time t are not informative of future innovations to $\vec{x}_t$: $\forall S_t \in \mathcal{S}^t, \forall h \geq 1, S_t \perp \vec{x}_{t+h} | X^t$.

Information Sets and Dynamics of Beliefs. Our main assumption here is that the agent does not forget information over time, which is commonly referred to as the "no-forgetting constraint". The agent understands that any choice of information will affect their priors in the future and that information has a continuation value.[8] Formally, a sequence of information sets $\{S^t \subseteq \mathcal{S}^t\}_{t \geq 0}$ satisfy the *no-forgetting* constraint for the agent if $S^t \subseteq S^{t+\tau}, \forall t \geq 0, \tau \geq 0$.

Cost of Information and the Attention Problem. We assume cost of information is linear in Shannon's mutual information function.[9] Formally, let $\{S^t\}_{t \geq 0}$ denote a set of information sets for the agent which satisfies the no-forgetting constraint. Then, the agent's flow cost of information at time t is $\omega \mathbb{I}(X^t; S^t | S^{t-1})$, where

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

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

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

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

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

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

$$s.t.\ S^t = S^{t-1} \cup S_t, \forall t \geq 0, \qquad \text{(evolution of information set + no-forgetting)}$$

$$S^{-1} \text{ given.} \qquad \text{(initial information set)}$$

2.1.1 Two General Properties of the Solution

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

Sufficiency of Actions for Signals. An important consequence of assuming that the cost of information is linear in Shannon's mutual information function is that it implies actions are sufficient statistics for signals over time (Steiner et al., 2017; Ravid, 2019). The following lemma formalizes this result in our setting.

Lemma 2.1. *Suppose* $\{(S^t \subset \mathcal{S}^t, \vec{a}_t : S^t \to \mathbb{R}^m\}_{t=0}^{\infty} \cup S^{-1}$ *is a solution to the RI Problem.* $\forall t \geq 0$, *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.*

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 RI Problem given an initial information set* S^{-1}. *Then* $\forall t \geq 0$:

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

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

2.2 The Linear-Quadratic-Gaussian Problem

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

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

Here, $\mathbf{H} \in \mathbb{R}^{n\times m}$ has full column rank and captures the interaction of the actions with the state.[10] The assumption of $rank(\mathbf{H}) = m$ is without loss of generality; in the case that any two column of $\mathbf{H}$ are linearly dependent, we can reclassify the problem so that all colinear actions are in one class.

Moreover, we have normalized the Hessian matrix of v with respect to $\vec{a}$ to negative identity.[11]

Optimality of Gaussian Posteriors. We start by proving that optimal actions are Gaussian under quadratic payoff with a Gaussian initial prior. Maćkowiak and Wiederholt (2009b) prove a version of this result in their setup where the cost of information is given by $\lim_{T\to\infty}\frac{1}{T}\mathbb{I}(X^T; a^T)$. Our setup is slightly different as in our case the cost of information is discounted at rate β and is equal to $(1-\beta)\sum_{t=0}^{\infty}\beta^t\mathbb{I}(X^t; a^t)$, as derived in the proof of Lemma 2.1 for the derivation.

Lemma 2.3. *Suppose the initial conditional prior, $\vec{x}_0|S^{-1}$, is Gaussian. If $\{\vec{a}_t\}_{t\geq 0}$ is a solution to the RI Problem with quadratic payoff given S^{-1}, then $\forall t \geq 0$, the posterior distribution $\vec{x}_t|\{\vec{a}_\tau\}_{0\leq\tau\leq t} \cup S^{-1}$ is also Gaussian.*

The Equivalent LQG Problem. Lemma 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.

Proposition 2.1. *Suppose the initial prior $\vec{x}_0|S^{-1}$ is Gaussian and that $\{\vec{x}_t\}_{t\geq 0}$ is a Markov*

[10] 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.

[11] 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.

process with the following minimal state-space representation:

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

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

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

$$s.t. \quad \mathbf{\Sigma}_{t+1|t} = \mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{Q}\mathbf{Q}', \quad \forall t \geq 0, \qquad \text{(law of motion for priors)}$$

$$\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq 0, \quad \forall t \geq 0 \qquad \text{(no-forgetting)}$$

$$0 \prec \mathbf{\Sigma}_{0|-1} \prec \infty \quad given. \qquad \text{(initial prior)}$$

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

This characterization of the problem matches the formulation in Sims (2010) but differs from the one in Sims (2003) and Miao, Wu and Young (2020) which solve the problem by optimizing at the steady-state.[12]

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

Proposition 2.2. *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,*

[12] The implied problem under the second approach is

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

which is characterized by

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

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

$$\boldsymbol{\Sigma}_{t+1|t} = \mathbf{A}\boldsymbol{\Sigma}_{t|t}\mathbf{A}' + \mathbf{Q}\mathbf{Q}', \qquad \forall t \geq 0, \tag{law of motion for priors}$$

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

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

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

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

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

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

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

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

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

Theorem 2.1. *Let* $\boldsymbol{\Omega}_t \equiv \boldsymbol{\Omega} + \beta\mathbf{A}'(\omega\boldsymbol{\Sigma}_{t+1|t}^{-1} - \boldsymbol{\Lambda}_{t+1})\mathbf{A}$ *denote the forward-looking component of the FOC in Proposition 2.2, which represents the marginal benefit of information. Then,*

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

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

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

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

Theorem 2.2. *$\forall t \geq 0$, 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*

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

where

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

Evolution of Optimal Beliefs and Actions. While Theorems 2.1 and 2.2 provide a representation for the optimal posteriors, we are often interested in the evolution of the agents' beliefs and actions. Our next theorem characterizes how beliefs and actions evolve over time.

Proposition 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, $\hat{x}_t$ and optimal actions evolve according to:*

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

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

Transition Dynamics and the Steady State. A key property of the LQG Problem is that it is deterministic. Additionally, as it is evident from the FOC in Proposition 2.2, eigenvectors of $\mathbf{\Sigma}_{t|t}$ are jump variables except for when the no-forgetting constraint binds. Thus, on the transition path, the agent has the desire to move on to their "steady state" posterior in each orthogonalized dimension unless the no-forgetting constraint binds, in which case they have to wait until their uncertainty stabilizes, either by climbing out of the inaction region or by reaching a steady state level within that region. Using the results of Proposition 2.2 and Theorem 2.1 we can represent the steady state of the problem with three equations that characterize a triple $(\bar{\mathbf{\Sigma}}_{-1}, \bar{\mathbf{\Sigma}}, \bar{\mathbf{\Omega}})$:

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

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

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

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

2.3 Discussion and the Economics of Dynamic Rational Inattention

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

Incentives. An important property of the RI Problem is that the marginal benefit of a bit of information is increasing in the prior uncertainty of the agent, while marginal cost of a bit is assumed to be a constant ω. Accordingly, for a large enough ω, the marginal benefit of acquiring a bit of information in different dimensions (eigenspaces) of the state might fall below its marginal cost, in which case the agent will decide not to pay attention to that dimension at all. This is clear from Theorem 2.1 which shows that the optimal policy function ignores eigenvalues that are less than ω.

Underneath its technical representation, Theorem 2.1 encodes an intuitive economic result. It shows that in acquiring information, the agent first decomposes the matrix $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \mathbf{\Omega}_t \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$, which captures the marginal benefit of information, into its orthogonal eigenspaces. At the *extensive margin*, the agent ignores eigenspaces whose eigenvalues are less than ω: the marginal benefit of acquiring information in these dimensions is outweighed by its marginal cost. On the *intensive margin*, the agent acquires signals for eigenspaces whose eigenvalues are larger than ω. Moreover, Theorem 2.2 shows that the loading of each of these signals on the state $\vec{x}_t$ is given by the eigenvector associated with the signal's eigenspace.

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

In our framework, sparsity is governed by the no-forgetting constraints. The most obvious and likely case for a binding no-forgetting constraint is when the number of actions m is strictly less than the dimension of the state n. This follows directly from Lemma 2.1 which states that the agent's actions at any given period are sufficient statistics for the underlying signals that she receives under the optimal solution.[13]

In static environments, the fact that actions are sufficient statistics for the underlying signals follows directly from optimality (Matějka and McKay, 2015). If the agent's action does not reveal the underlying signal, then he must have received information that was not used in choosing the action. Nonetheless, such a strategy is suboptimal given that information is costly. In dynamic settings, however, this is not necessarily true due to smoothing incentives. The agent might find it optimal to acquire signals about future actions before-hand in which case the history of actions at a given time is no longer sufficient for the information set of the agent. Lemma 2.1 rules this out by showing that if the chain-rule of mutual information holds, then the agent has no smoothing incentives. Thus, upon acquiring signals for every given action, the discounting of the cost induces the agent to postpone acquiring information to the period in which that action is taken.[14]

The economic consequence of this result is that rationally inattentive agents are not concerned about identification: independent of how many shocks they face, they are only interested in how those shocks affect their actions. An important reference for why this matters in an economic sense is Hellwig and Venkateswaran (2009) which shows that when firms receive signals about a sufficient statistic for their prices, they charge the right prices even though they cannot tell aggregate and idiosyncratic shocks apart.[15]

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

[14]The chain-rule of mutual information implies that for every three random variables:

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

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

[15]Hellwig and Venkateswaran (2009) do not endogenize information choice, but the exogenous signal

Information Spillovers. The intensive margin of information acquisition under dynamic rational inattention provides a microfoundation for *information spillovers* across different actions. These spillover effects are uniquely identified by the eigenvectors of $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$ with eigenvalues *larger* than ω. Therefore, information about an action can effect other actions either through a subjective correlated posterior ($\mathbf{\Sigma}_{t|t-1}$) or through complementarities or substitutabilies in actions captured by $\mathbf{\Omega}_t$.[16]

3 An Attention Driven Phillips Curve

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

3.1 Environment

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

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

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

For ease of notation, let $P_t \equiv \left[\int_0^1 P_{i,t}^{1-\theta}\right]^{\frac{1}{1-\theta}}$ denote the aggregate price index and $Q_t \equiv$

structure that they consider is optimal under our model with a particular parametrization.

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

P_tC_t be the nominal aggregate demand in this economy. The solution to the household's problem is summarized by:

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

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

$$W_t = Q_t, \qquad \forall t \geq 0. \tag{3.3}$$

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

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

where $\rho \equiv -\log(\beta)$ is the natural rate of interest, $q_t \equiv \log(P_tC_t)$ is the log of the nominal aggregate demand, and u_t is an exogenous shock to monetary policy that affects the nominal interest rates with a standard deviation of σ_u.

Lemma 3.1. *Suppose $\phi > 1$. Then, in the log-linearized version of this economy, the aggregate demand is uniquely determined by the history of monetary policy shocks, and is characterized by the following random walk process:*

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

Assuming that the monetary authority directly controls the nominal aggregate demand is a popular framework in the literature to study the effects of monetary policy on pricing.[17] We derive this as an equilibrium outcome in Lemma 3.1 in order to relate the variance of the innovations to the nominal demand to the *strength* with which the monetary authority targets its growth: a larger ϕ stabilizes the nominal demand while a larger σ_u increases its variance.

Firms. Every variety $i \in [0,1]$ is produced by a price-setting firm. Firm i hires labor $N_{i,t}$ from a competitive labor market at a subsidized wage $W_t = (1-\theta^{-1})Q_t$ where the subsidy θ^{-1} is paid per unit of worker to eliminate steady state distortions introduced by monopolistic competition. Firms produce their product with a linear technology in labor,

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

$Y_{i,t} = N_{i,t}$. Therefore, for a particular history $\{(P_t, Q_t)\}_{t\geq 0}$ and set of prices $\{P_{i,t}\}_{t\geq 0}$, the net present value of the firms' profits, discounted by the marginal utility of the household is given by

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

where the second line is a second order approximation with small letters denoting the logs of corresponding variables.[18] This approximation states that for a monopolistic competitive firms, their loss from not matching their marginal cost in pricing, which is this setting is the nominal demand, is quadratic and proportional to $\theta - 1$, with θ denoting the elasticity of demand.

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

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

where $p_i^t \equiv (p_{i,\tau})_{\tau \leq t}$ denotes the history of firm's prices over up to time t. It is important to note that $\{p_{i,t}\}_{t\geq 0}$ is a stochastic process that proxies for the underlying signals that the firm receives over time – a result that follows from Lemma 2.2.

Assuming that the distribution of q_0 conditional on p_i^{-1} is a Gaussian process, and noting that $\{q_t\}_{t\geq 0}$ is itself a Markov Gaussian process, this problem satisfies the assumptions of Proposition 2.1. Formally, let $\sigma_{i,t|t-1} \equiv \sqrt{var(q_t | p_i^{t-1})}$, $\sigma_{i,t|t} \equiv \sqrt{var(q_t | p_i^t)}$ denote the prior and posterior standard deviations of firm i belief about q_t at time t. Then, the corresponding LQG problem to the one in Proposition 2.1 is

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

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

3.2 Characterization of Solution

The solution to this problem follows from Proposition 2.2, and is characterized by the following proposition.

Proposition 3.1. *Firms only pay attention to the monetary policy shocks if their prior uncertainty is above a reservation prior uncertainty. Formally,*

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

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

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

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

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

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

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

The first part of Proposition 3.1 shows that firms pay attention to nominal demand only when they are sufficiently uncertain about it. The result follows from the fact that the marginal benefit of a bit of information is increasing in the prior uncertainty of a firm but the marginal cost is constant. Thus, for small levels of prior uncertainty where the marginal benefit of acquiring a bit of information falls below the marginal cost, the firm pays no attention to the nominal demand. However, once the prior uncertainty is at least as large as the reservation uncertainty, the firm always acquires enough information to maintain that level of uncertainty.

The second part of Proposition 2.1 shows that in the region where the firm does not pay attention to the nominal demand, their price does not respond to monetary policy shocks as the implied Kalman-gain is zero and the price is constant: $p_{i,t} = p_{i,t-1}$.

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

$$\kappa_{i,t} = \kappa \equiv \frac{\sigma_u^2}{\phi^2 \underline{\sigma}^2 + \sigma_u^2}. \qquad \text{(steady-state Kalman-gain of firms)}$$

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

Corollary 3.1. *The following hold:*

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

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:[19]

$$[\underline{\sigma}^2]_{\beta=1,\sigma_u^2 \ll \omega} \approx \frac{\sigma_u}{\phi}\sqrt{\frac{\omega}{\theta - 1}} \tag{3.10}$$

$$[\kappa]_{\beta=1,\sigma_u^2 \ll \omega} \approx \frac{\sigma_u}{\phi}\sqrt{\frac{\theta - 1}{\omega}} \tag{3.11}$$

3.3 Aggregation

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

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

pendently distributed.[20]

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

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

1. *the Phillips curve of this economy is*

$$\pi_t = \max\{0, \frac{\sigma^2_{t|t-1} - \underline{\sigma}^2}{\sigma^2_{t|t}}\} y_t \tag{3.12}$$

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

$$\pi_t = 0, \quad y_t = y_{t-1} + \frac{\sigma_u}{\phi} u_t. \tag{3.13}$$

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

$$\pi_t = (1-\kappa)\pi_{t-1} + \frac{\kappa \sigma_u}{\phi} u_t \tag{3.14}$$

$$y_t = (1-\kappa) y_{t-1} + \frac{(1-\kappa)\sigma_u}{\phi} u_t \tag{3.15}$$

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

3.4 Implications for the Slope of the Phillips Curve

Proposition 3.2 shows that this economy has a Phillips curve with a time-varying slope, *which is flat* if and when the no-forgetting constraint binds. At a time when firm's uncertainty is below the reservation uncertainty, firms pay no attention to the monetary policy and the inflation does not respond to monetary policy shocks.

Nonetheless, since nominal demand follows a random walk process and the attention problem is deterministic, Proposition 3.2 also shows that the rational inattention problem will eventually enter and remain at its steady state where firms do pay attention to

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

the nominal demand. In this section, we start by analyzing this steady state, and then consider the dynamic consequences of unanticipated disturbances (MIT shocks) to the parameters of the model.

3.4.1 The Long-run Slope of the Phillips Curve

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

$$\pi_t = \frac{\kappa}{1-\kappa} y_t \qquad \text{(long-run Phillips curve)}$$

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

Thus, in the long-run, the parameter κ is sufficient for determining the slope of the Phillips curve as well as the magnitude and persistence of the real effects of monetary policy shocks in this economy: a lower value for κ leads to a flatter Phillips curve, a more persistent process for inflation and output, and larger monetary non-neutrality. The intuition behind all of these is that a lower value for κ is equivalent to lower attention to monetary policy shocks on the part of firms. It takes longer for less attentive firms to learn about monetary policy shocks and respond to them. In the meantime, since firms are not adjusting their prices one to one with the shock, their output has to compensate. Thus, less attention, leads to a longer half-life for – and a larger degree of – monetary non-neutrality.

Comparative statics of κ with respect to the underlying parameters of the model are derived in Corollary 3.1. In particular, we would like to focus on how the rule of monetary policy affects the slope of the Phillips curve and consequently the persistence and the magnitude of the real effect so of monetary policy shocks.

Corollary 3.1 shows that κ is increasing with $\frac{\sigma_u}{\phi}$. We interpret this ratio as a measure for how dovish the monetary policy is in this economy since a larger $\frac{\sigma_u}{\phi}$ corresponds to a lower relative weight on stabilizing inflation. It follows that in the long-run, the Phillips curve is steeper in more dovish economies. If the monetary authority opts for a lower weight on the stabilization of the nominal variables, the firms face a more volatile process for their marginal cost and optimally choose to pay more attention to monetary policy shocks in the steady state of their attention problem. As a result, such firms are more responsive to monetary policy shocks and are quicker in adjusting their prices.

3.4.2 The Aftermath of An Unexpectedly More Hawkish Monetary Policy

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

Corollary 3.2. *Suppose the economy is in the steady state of its attention problem, and consider an unexpected decrease in $\frac{\sigma_u}{\phi}$. Then, the economy immediately jumps to a new steady state of the attention problem, in which:*

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

The comparative statics follow directly from Corollary 3.1 and are straight forward; however, the reason that the economy jumps to its new steady state needs some intuition. The reason for this jump is that a more hawkish economy has a less volatile nominal demand process and firms have lower reservation uncertainties in less volatile environments. Therefore, once the monetary policy rule becomes more hawkish, firms find themselves with a prior uncertainty that is higher than their new reservation uncertainty. Consequently, they acquire enough information to immediately reduce their uncertainty to the new reservation level. The key observation is that once they reach this new lower level of uncertainty they need a lower rate of information acquisition to maintain that level of uncertainty. Hence, while the reservation uncertainty decreases with a more hawkish rule, the steady state Kalman-gain also decreases and leads to flatter Phillips curve and a higher persistence in responses of output and inflation.

Conceptually, our results speak to, and are consistent with, the post-Volcker era in the U.S. monetary policy. A large strand of the literature has documented that the slope of the Phillips curve has become flatter in the last few decades.[21] Our theory provides a new perspective on this issue. Firms do not need to be attentive to monetary policy in an environment where the policy makers follow a hawkish rule.

3.4.3 The Aftermath of An Unexpectedly More Dovish Monetary Policy

The model is non-symmetric in response to changes in the rule of monetary policy. While the economy jumps to the new steady state of the attention problem after a decreases in

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

$\frac{\sigma_u}{\phi}$, as shown in Corollary 3.2, the reverse is not true. An unexpected increase in $\frac{\sigma_u}{\phi}$ has different short-run implications due to its effect on reservation uncertainty.

Corollary 3.3. *Suppose the economy is in the steady state of its attention problem, and consider an unexpected increase in $\frac{\sigma_u}{\phi}$. Then,*

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

The intuition follows from Corollary 3.1. An increase in $\frac{\sigma_u}{\phi}$ makes the nominal demand more volatile and raises the reservation uncertainty of firms. Hence, immediately after such a shock, firms find themselves with an uncertainty that is below this reservation level; the no-forgetting constraint binds and they temporarily stop paying attention to the monetary policy shocks until their uncertainty grows to its new reservation level. In the meantime, the Phillips curve is flat and inflation is non-responsive to monetary policy shocks.

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

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

3.5 Implications for Anchoring of Inflation Expectations

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

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

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

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

$$\hat{\pi}_t = (1-\kappa)\hat{\pi}_{t-1} + \frac{\kappa^2}{(2-\kappa)(1-\kappa)} y_t \tag{3.16}$$

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

$$\hat{\pi}_t = 2(1-\kappa)\hat{\pi}_{t-1} - (1-\kappa)^2\hat{\pi}_{t-2} + \frac{\kappa^2}{2-\kappa}\frac{\sigma_u}{\phi}u_t \tag{3.17}$$

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

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

The AR(2) nature of these expectations indicate the inherent inertia that expectations inherit from firms' imperfect information – the counterfactual being full-information rational expectations, in which case both inflation and inflation expectations are i.i.d. over time.[22]

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

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

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

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

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

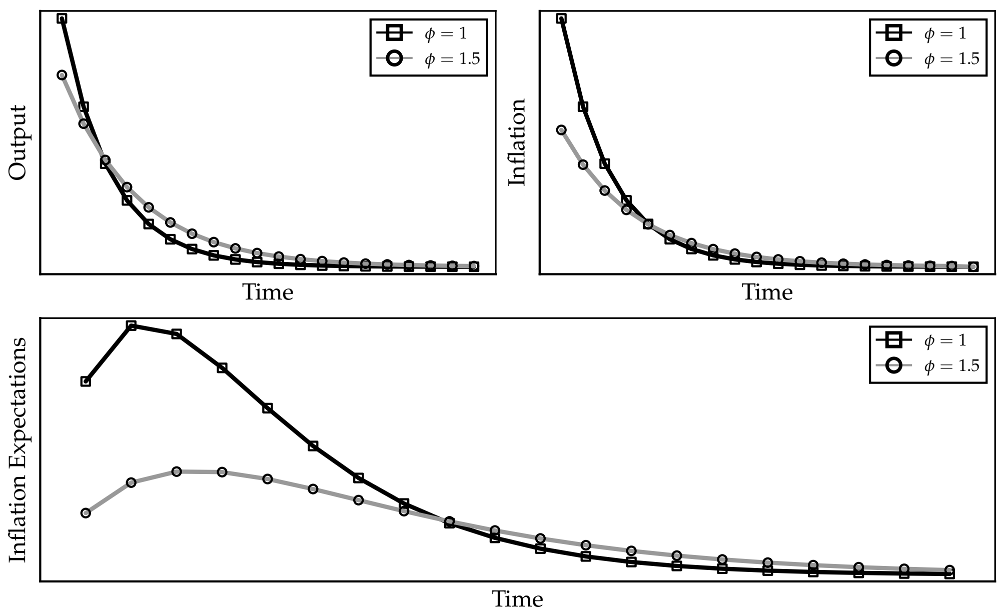

Figure 1: Impulse Responses to a 1 Std. Dev. Expansionary Monetary Policy Shock

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

Figure 1 illustrates these results. The top two panels in the Figure show the impulse responses of output and inflation to a one standard deviation expansionary monetary policy shock under two different values for $\phi \in \{1, 1.5\}$. Moreover, the bottom panel of Figure 1 shows the impulse responses of firms average inflation expectations under these two parameters: with more Hawkish monetary policy, expectations are less sensitive to monetary policy shocks, but their responses are more persistent.[23]

[23]While in our setup higher anchoring of the expectation are generated by a combination of higher order

4 Quantitative Analysis

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

Our exercise in this section is very much in the spirit of the literature that interprets the Great Moderation, at least partially, through the lens of a shift in monetary policy in the post-Volcker era (Clarida et al., 2000; Coibion and Gorodnichenko, 2011). In particular, we are interested in the following question: can the shift in the rule of monetary policy in the post-Volcker era explain the decline in the slope of the Phillips curve, and if so by how much?

To answer this question we calibrate a quantitative version of our model with TFP and monetary policy shocks to the U.S. inflation and output data in the post-Volcker era and examine (1) how well the model fits the pre-Volcker data and (2) if the model can match both the level and the change in the slope of the Phillips curve estimates in the empirical literature by repeating the same estimation methods in the simulated data from our calibrated model.

4.1 Setup of the Quantitative Model

Household. A representative household is fully rational and maximize her life-time utilities:

$$\begin{aligned} \max_{\{C_t,\{C_{i,t},L_{i,t}\}_{i\in[0,1]},B_t\}_{t\geq 0}} \quad & \mathbb{E}_t^f\left[\sum_{t=0}^{\infty}\beta^t\left(\frac{C_t^{1-\sigma}}{1-\sigma}-\frac{\int_0^1 L_{i,t}^{1+\psi}di}{1+\psi}\right)\right] \\ \text{s.t.} \quad & \int_0^1 P_{i,t}C_{i,t}di + B_t \leq R_{t-1}B_{t-1} + \int_0^1 W_{i,t}L_{i,t}di + \Pi_t, \quad \text{for all } t \end{aligned} \tag{4.1}$$

where

$$C_t = \left(\int C_{i,t}^{\frac{\theta-1}{\theta}}di\right)^{\frac{\theta}{\theta-1}}.$$

Here $\mathbb{E}_t^f[\cdot]$ is the full information rational expectation operator at time t. Since the main purpose of this paper is to study the effects of rational inattention among firms, we as-

beliefs and lower information acquisition on the part of firms, it is also important to note that higher persistence and anchoring can be generated in a context that takes the role of strategic interactions into account (see, e.g., Angeletos and Huo, 2018).

sume that the household is fully informed about all prices and wages. B_t is the demand for nominal bond and R_{t-1} is the nominal interest rate. $L_{i,t}$ is firm-specific labor supply of the household, $W_{i,t}$ is the firm-specific nominal wage, and Π_t is the aggregate profit from the firms. C_t is the aggregator over the consumption for goods produced by firms. θ is the constant elasticity of substitution across different firms.

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

$$Y_{i,t} = A_t L_{i,t}^d,$$

where $L_{i,t}^d$ is firm i's demand for labor. I assume that shocks to A_t are independently and identically distributed and the log of the productivity shock, $a_t \equiv \log(A_t)$, follows a AR(1) process:

$$a_t = \rho_a a_{t-1} + \varepsilon_{a,t}, \;\; \varepsilon_{a,t} \sim N(0, \sigma_a^2). \tag{4.2}$$

Then, firm i's nominal profit from sales of all goods at prices $P_{i,t}$ is given by

$$\Pi_{i,t}(P_{i,t}, A_t, W_{i,t}, P_t, Y_t) = \left(P_{i,t} - \frac{W_{i,t}}{A_t}\right)\left(\frac{P_{i,t}}{P_t}\right)^{-\theta} Y_t, \tag{4.3}$$

where Y_t is the nominal aggregate demand.

At each period, firms optimally decide their prices and signals subject to costs of processing information. Firms are rationally inattentive in a sense that they choose their optimal information set by taking into account the cost of obtaining and processing information. At the beginning of period t, firm i wakes up with its initial information set, S_i^{t-1}. Then it chooses optimal signals, $s_{i,t}$, from a set of available signals, $\mathcal{S}_{i,t}$, subject to the cost of information which is linear in Shannon's mutual information function. Denote ω as the marginal cost of information processing. Firm i forms a new information set, $S_i^t = S_i^{t-1} \cup s_{i,t}$, and sets its new prices, $P_{i,t}$, based on that.

The firm i chooses a set of signals to observe over time $(s_{i,t} \in \mathcal{S}_{i,t})_{t=0}^{\infty}$ and a pricing strategy that maps the set of its prices at $t-1$ and its information set at t to its optimal price at any given period, $P_{i,t} : (S_i^t) \to \mathbb{R}$ where $S_i^t = S_i^{t-1} \cup s_{i,t} = S_i^{-1} \cup \{s_{i,\tau}\}_{\tau=0}^{t}$ is the firm's information set at time t. Then, the firm i's problem is to maximize the net present value of its life time profits given an initial information set:

$$\max_{\{s_{i,t}\in\mathcal{S}_{i,t},P_{i,t}(S_i^t)\}_{t\geq 0}} \quad \mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t\Lambda_t\left\{\Pi_{i,t}(P_{i,t},A_t,W_{i,t},P_t,Y_t)\right.\right. \tag{4.4}$$

$$\left.\left. -\omega\mathbb{I}(S_i^t;(A_\tau,W_{i,\tau},P_\tau,Y_\tau)_{\tau\leq t}|S_i^{t-1})\right\}\middle|S_i^{-1}\right]$$

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

where Λ_t is the stochastic discount factor and $\mathbb{I}(S_i^t;(A_\tau,W_{i,\tau},P_\tau,Y_\tau)_{\tau\leq t}|S_i^{t-1})$ is the Shannon's mutual information function.

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

$$\frac{R_t}{\bar{R}} = \left(\frac{R_{t-1}}{\bar{R}}\right)^{\rho}\left(\left(\frac{P_t}{P_{t-1}}\right)^{\phi_\pi}\left(\frac{Y_t}{Y_t^n}\right)^{\phi_x}\left(\frac{Y_t}{Y_{t-1}}\right)^{\phi_{\Delta y}}\right)^{1-\rho}\exp(u_t) \tag{4.5}$$

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

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

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

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

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

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

1. given the set of prices and $\{\Omega_i^F\}_{i\in[0,1]}$, the household's allocation solves its problem as specified in Equation (4.1) ;
2. given the set of prices and Ω^H, and the implied labor supply and output demand, firms' allocation solve their problem as specified in Equation (4.4) ;

3. given the set of prices, Ω^H, and $\left\{\Omega_i^F\right\}_{i\in[0,1]}$, the monetary policy satisfies the specified rule in Equation (4.5) ;

4. and, all markets clear: for all $t \geq 0$,

$$\begin{aligned} Y_{i,t} =& C_{i,t}, \qquad \text{for all } i \in [0,1], \\ L_{i,t} =& L_{i,t}^d, \qquad \text{for all } i \in [0,1], \\ B_t =& 0, Y_t = C_t. \end{aligned}$$

4.2 Computing the Equilibrium

We use our theoretical framework for solving dynamic rational inattention problems to solve for the equilibrium by taking a second order approximation to firms' profit function in Equation (4.3).[24] This problem is given by a similar expression in our simple model, but now includes a role for strategic complementarities:

$$\min_{\{p_{i,t}\}_{t\geq 0}} \sum_{t=0}^{\infty} \beta^t \mathbb{E}\left[(\theta-1)(p_{i,t}-p_t-\alpha x_t)^2 + \omega \mathbb{I}(p_{i,t}, \{p_{t-j}+\alpha x_{t-j}\}_{j=0}^{\infty} | p_i^{t-1}) | p_i^{-1}\right] \tag{4.6}$$

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

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

$$x_t = \mathbb{E}_t^f\left[x_{t+1} - \frac{1}{\sigma}(i_t - \pi_{t+1})\right] + \mathbb{E}_t^f\left[y_{t+1}^n\right] - y_t^n \tag{4.7}$$

$$p_{i,t} = \mathbb{E}_{i,t}\left[p_t + \alpha x_t\right] \tag{4.8}$$

$$i_t = \rho i_{t-1} + (1-\rho)\left(\phi_\pi \pi_t + \phi_x x_t + \phi_{\Delta y}\Delta y_t\right) + u_t \tag{4.9}$$

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

[24] This is a common approach to turn firms' problems to quadratic objectives – see e.g. Maćkowiak and Wiederholt 2009a, Afrouzi 2016, and Yang 2019).

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

4.3 Calibration

Our benchmark model is calibrated at quarterly frequency with a time discount factor of $\beta = 0.99$ to the post-Volcker U.S. data ending at the onset of the Great Recession (1983–2007). A summary of calibrated values of the parameters is presented in Table 1. In the remainder of this section we go over the details of our calibration strategy.

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

Monetary policy rule(s). We set the standard deviation of monetary policy shocks (σ_u) in our benchmark model to match the size of the identified monetary policy shocks constructed by Romer and Romer (2004) for the period 1983–2007.[25]

Furthermore, for the parameters describing the monetary policy rule (ρ, ϕ_π, $\phi_{\Delta y}$, ϕ_x), we estimate the Taylor rule in Equation (4.9) using real-time U.S. data. Specifically, following Coibion and Gorodnichenko (2011), we use the Greenbook forecasts of inflation and the real GDP growth. The measure of the output gap is also based on Greenbook forecasts. We consider two time samples: pre-Volcker period(1969–1978) and post-Volcker period(1983–2007).[26] The point estimates are reported in Panel B of Table 1, and more detailed results including standard errors are reported in Appendix Table A.1.

[25]Original monetary policy shocks data in Romer and Romer (2004) are available until 1996, while we use extended data, which are available until 2007, from Coibion et al. (2017).

[26]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.

Table 1: Calibrated and Assigned Parameters

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

Notes: The table presents the baseline parameters for the general equilibrium model. Panel A shows the calibrated parameters which match the three key moments shown in Table 2. Panel B shows values and the source of the assigned model parameters. Panel C shows the parameters for the counterfactual analysis in Section 4.5. See Section 4.3 for details.

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).[27]

Finally, for our counterfactual analysis in later sections, we do a similar estimation of

[27] 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, since 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 were monthly, whereas the sample from 1983 through 2007 were six-weekly. Thus, we convert the estimated AR(1) parameters from monthly or six-weekly frequency to quarterly and use the converted parameters for our model simulations.

these parameters for the pre-Volcker era (1969–1978). The point estimates are reported in Panel C of Table 1, and more detailed results including standard errors are reported in Appendix Table A.1.

Calibrated Parameters. We calibrate the three remaining parameters of the model – marginal costs of information processing (ω) as well as the persistence (ρ_a) and the size (σ_a) of productivity shocks – jointly by targeting the covariance matrix of inflation and real GDP in post-Volcker U.S. data (1983–2007). The covariance matrix is measured after we detrend the CPI core inflation and real GDP data using log-quadratic trends. The three moments (variances of inflation and GDP along with their covariance) exactly identify the three model parameters, as reported in Table 1.

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

Moreover, the calibrated cost of information processing, $\omega\mathbb{I}(.,.)$, is 0.1 percent of firms' steady-state real revenue.[28] This small calibrated cost implies that imperfect information models do not require large information costs to match the data. The cost is negligible compared to firms' revenue. One relevant measure that one could use to relate the degree of information acquisition to that of professional forecasters is the firms' Kalman gain on their signals under the optimal information structure. The implied Kalman gain for firms in the model is 0.8 which implies a large degree of information acquisition relative to professional forecaster – Coibion and Gorodnichenko (2015b) estimate professional forecaster's Kalman gain to be around 0.5.

4.4 Model Fit

Targeted Moments. Table 2 reports our targeted moments both in the data and as implied by the model. All three targeted moments, variances of GDP and inflation and their covariance, are matched by the model.

Non-targeted moments. To examine the model's ability in capturing the out of sample behavior of the GDP and inflation, 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.

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

Table 2: Targeted Moments

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

Notes: The table presents moments of the data and simulated series from the model parameterized at the baseline values in Table 1. See Section 4.3 for details.

To do so, we first replace the parameters related to monetary policy with the pre-Volcker era estimates. Specifically, we replace the estimates of Taylor rule for post-Volcker period with our estimates for the pre-Volcker period. Furthermore, we re-estimate the standard deviation of monetary policy shocks (σ_u) using the pre-Volcker period monetary policy shock series from Romer and Romer (2004). As shown in Panel C of Table 1, monetary policy is less responsive to inflation and output growth in the pre-Volcker period than in the post-Volcker period. Also, the monetary shock is more volatile in the pre-Volcker period than in the post-Volcker period.

Table 3: Non-targeted Moments

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

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

We then simulate the model under the calibrated values for the cost of attention and the process for the TFP shocks and calculate the variance-covariance matrix for GDP and inflation. Table 3 report the model generated moments and their analogs in the data. While we only target the volatility of inflation and GDP for the post-Volcker period, our model is able to match the high volatility of inflation and GDP in the pre-Volcker period as a consequence of more dovish monetary policy during that period.

Furthermore, for both pre- and post-Volcker parameterization of monetary policy, Figure 2 shows the impulse response functions of the variables of the model with respect to one standard-deviation TFP and monetary policy shocks. The main takeaway from these

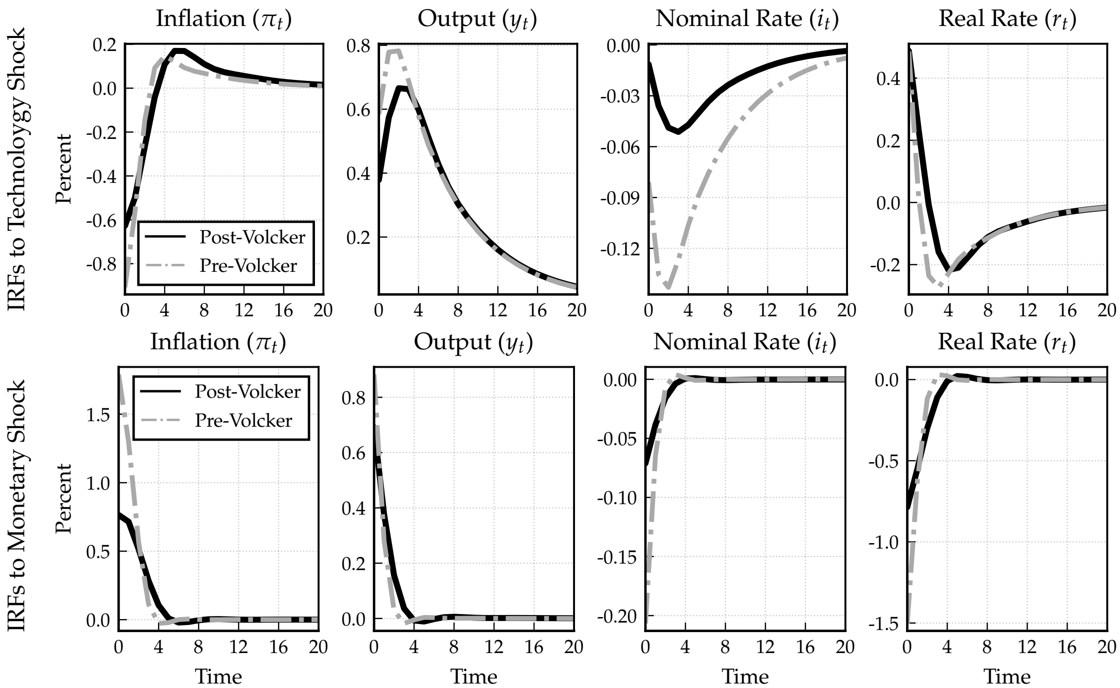

Figure 2: Impulse Responses to Technology and Monetary Shocks

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.

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

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

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

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

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

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

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

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

$$\pi_t = constant + \gamma \mathbb{E}_t[\pi_{t+1}] + (1-\gamma)\pi_{t-1} + \kappa x_t + \varepsilon_t. \tag{4.10}$$

We use four lags of both inflation and output gap as instruments. Table 4 shows the parameter estimates of the New Keynesian Phillips curve using the simulated data from our model.

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

Table 4: Estimates of the New Keynesian Phillip Curve Using Simulated Data

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

Notes: This table shows the estimation results of the New Keynesian Phillips curve using simulated data from the baseline model presented in Section 4.2. Column (1) and (2) show the estimates of the New Keynesian Phillips curve (4.10) using the simulated output gap and output data, respectively. Column (3) shows the estimates using the simulated output gap data, which are adjusted by subtracting moving averages of natural level of output from actual output. Four lags of inflation and output gap (or output) are used as instruments for the GMM estimation. A constant term is included in the regressions but not reported. Newey-West standard errors are reported in parentheses. ***, **, * denotes statistical significance at 1%, 5%, and 10% levels respectively.

to which we control for the supply shocks.

Column (1) in Table 4 shows the estimates of the New Keynesian Phillips curve when we use the true output gap (fully controlling for the supply shocks). In this scenario, the model predicts that the slope of the Phillips curve declined from 1.16 in the pre-Volcker era to 0.30 in the post-Volcker period – a 75% decline. The benefit of this specification is that we are controlling for the true output gap, which eliminates the concerns regarding the downward bias induced by omitted variables bias. Therefore, all the decline in the slope that we observe in this specification is due to the change in the information acquisition incentives of firms.

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

To illustrate this point further, Column (2) in Table 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 pre- to post-Volcker period.

Finally, we consider an interim case in Column (3) where we partially control for the supply shocks by subtracting a moving average of the natural level of output from realized output in the model to construct the output gap. Again, the model predicts a decline in the slope of the Phillips curve from pre- to post-Volcker period.[29]

5 Concluding Remarks

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

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

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

References

Afrouzi, Hassan, "Strategic Inattention, Inflation Dynamics and the Non-Neutrality of Money," 2016. Manuscript.

Angeletos, George-Marios and Chen Lian, "Incomplete information in macroeconomics: Accommodating frictions in coordination," in "Handbook of macroeconomics," Vol. 2, Elsevier, 2016, pp. 1065–1240.

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

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

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

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

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

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

Babb, Nathan and Alan K Detmeister, "Nonlinearities in the Phillips Curve for the United States: Evidence Using Metropolitan Data," FEDS Working Paper 2017-070 2017.

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

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

Bullard, James, "The Case of the Disappearing Phillips Curve," in "The 2018 ECB Forum on Central Banking, Sintra, Portugal, June," Vol. 19 2018.

Caplin, Andrew, Mark Dean, and John Leahy, "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy," Working Paper 23652, National Bureau of Economic Research 2017.

Caplin, Andrew S and Daniel F Spulber, "Menu Costs and the Neutrality of Money," *The Quarterly Journal of Economics*, 1987, *102* (4), 703–725.

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

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

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

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

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

_ , _ , and Tiziano Ropele, "Inflation Expectations and Firm Decisions: New Causal Evidence," Working Paper 25412, National Bureau of Economic Research 2018.

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

Cover, Thomas M and Joy A Thomas, *Elements of Information Theory*, John Wiley & Sons, 2012.

da Silveira, Rava Azeredo, Yeji Sung, and Michael Woodford, "Noisy Memory and Over-Reaction to News," 2019. Manuscript.

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

Denti, Tommaso, "Unrestricted Information Acquisition," 2015. Manuscript.

Fulton, Chad, "Mechanics of Static Quadratic Gaussian Rational Inattention Tracking Problems," Working Paper, Board of Governors of the Federal Reserve System 2018.

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

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

Gali, Jordi and Mark Gertler, "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, 1999, *44* (2), 195–222.

Gertler, Mark and John Leahy, "A Phillips Curve with an Ss Foundation," *Journal of Political Economy*, 2008, *116* (3), 533–572.

Golosov, Mikhail and Robert E Lucas Jr, "Menu Costs and Phillips curves," *Journal of Political Economy*, 2007, *115* (2), 171–199.

Hébert, Benjamin and Jennifer La'O, "Information Acquisition, Efficiency, and Non-Fundamental Volatility," 2019. Manuscript.

_ and Michael Woodford, "Information Costs and Sequential Information Sampling," Working Paper 25316, National Bureau of Economic Research 2018.

Hellwig, Christian and Venky Venkateswaran, "Setting the Right Prices for the Wrong Reasons," *Journal of Monetary Economics*, 2009, *56*, S57–S77.

Hooper, Peter, Frederic S Mishkin, and Amir Sufi, "Prospects for Inflation in a High Pressure Economy: Is the Phillips Curve Dead or is It Just Hibernating?," Working Paper 25792, National Bureau of Economic Research 2019.

Ilut, Cosmin and Rosen Valchev, "Economic Agents As Imperfect Problem Solvers," 2017. Manuscript.

Kamdar, Rupal, "The Inattentive Consumer: Sentiment and Expectations," 2018. Manuscript.

Khaw, Mel Win and Oskar Zorrilla, "Deeper Habits," 2018. Manuscript.

_ , Luminita Stevens, and Michael Woodford, "Discrete Adjustment to a Changing Environment: Experimental Evidence," *Journal of Monetary Economics*, 2017, *91*, 88–103.

Kőszegi, Botond and Filip Matějka, "An Attention-Based Theory of Mental Accounting," 2019. Manuscript.

Kumar, Anil and Pia M Orrenius, "A Closer Look at the Phillips Curve Using State-Level Data," *Journal of Macroeconomics*, 2016, *47*, 84–102.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko, "Inflation Targeting Does Not Anchor Inflation Expectations: Evidence from Firms in New Zealand," *Brookings Papers on Economic Activity*, 2015, pp. 187–226.

Landier, Augustin, Yueran Ma, and David Thesmar, "Biases in Expectations: Experimental Evidence," 2019. Manuscript.

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

Lucas, Robert E, "Expectations and the Neutrality of Money," *Journal of Economic Theory*, 1972, *4* (2), 103–124.

Luo, Yulei, "Consumption Dynamics under Information Processing Constraints," *Review of Economic Dynamics*, 2008, *11* (2), 366–385.

_ , Jun Nie, and Eric R Young, "Robustness, Information Processing Constraints, and the Current Account in Small Open Economies," *Journal of International Economics*, 2012, *88* (1), 104–120.

Maćkowiak, Bartosz and Mirko Wiederholt, "Optimal Sticky Prices under Rational Inattention," *The American Economic Review*, 2009, *99* (3), 769–803.

_ and _ , "Optimal Sticky Prices under Rational Inattention," Working Paper Series 1009, European Central Bank February 2009.

Maćkowiak, Bartosz and Mirko Wiederholt, "Business Cycle Dynamics under Rational Inattention," *The Review of Economic Studies*, 2015, *82* (4), 1502–1532.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Dynamic Rational Inattention: Analytical Results," *Journal of Economic Theory*, 2018, *176*, 650 – 692.

_ , Filip Matějka, and Mirko Wiederholt, "Survey: Rational inattention, a disciplined behavioral model," 2018.

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

Matějka, Filip, "Rationally Inattentive Seller: Sales and Discrete Pricing," *The Review of Economic Studies*, 2015, *83* (3), 1125–1155.

Matějka, Filip and Alisdair McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, January 2015, *105* (1), 272–98.

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

Melosi, Leonardo, "Estimating Models with Dispersed Information," *American Economic Journal: Macroeconomics*, 2014, *6* (1), 1–31.

Miao, Jianjun, Jieran Wu, and Eric Young, "Multivariate Rational Inattention," January 2020. Boston University - Department of Economics.

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

Mondria, Jordi and Thomas Wu, "The Puzzling Evolution of the Home Bias, Information Processing and Financial Openness," *Journal of Economic Dynamics and Control*, 2010, *34* (5), 875–896.

Nagel, Stefan and Zhengyang Xu, "Asset Pricing with Fading Memory," Working Paper 26255, National Bureau of Economic Research 2019.

Nakamura, Emi and Jon Steinsson, "Monetary Non-Neutrality in a Multisector Menu Cost Model," *The Quarterly journal of economics*, 2010, *125* (3), 961–1013.

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

Nimark, Kristoffer, "Dynamic Pricing and Imperfect Common Knowledge," *Journal of Monetary Economics*, 2008, *55* (2), 365–382.

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

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

Pasten, Ernesto and Raphael Schoenle, "Rational Inattention, Multi-Product Firms and the Neutrality of Money," *Journal of Monetary Economics*, 2016, *80*, 1–16.

Peng, Lin and Wei Xiong, "Investor Attention, Overconfidence and Category Learning," *Journal of Financial Economics*, 2006, *80* (3), 563–602.

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

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

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

Roberts, John M, “New Keynesian Economics and the Phillips Curve,” *Journal of Money, Credit and Banking*, 1995, *27* (4), 975–984.

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

Roth, Christopher and Johannes Wohlfart, “How Do Expectations About the Macroeconomy Affect Personal Expectations and Behavior?,” 2018.

Rudd, Jeremy and Karl Whelan, “New Tests of the New-Keynesian Phillips Curve,” *Journal of Monetary Economics*, 2005, *52* (6), 1167 – 1181.

Ryngaert, Jane, “What do (and Don’t) Forecasters Know About US Inflation,” 2017. Manuscript.

Sims, Christopher A, “Implications of Rational Inattention,” *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

_ , “Rational Inattention and Monetary Economics,” in “Handbook of Monetary Economics,” Vol. 3, Elsevier, 2010, pp. 155–181.

Steiner, Jakub, Colin Stewart, and Filip Matějka, “Rational Inattention Dynamics: Inertia and Delay in Decision-Making,” *Econometrica*, 2017, *85* (2), 521–553.

Stevens, Luminita, “Coarse Pricing Policies,” *The Review of Economic Studies*, 07 2019.

Tutino, Antonella, “Rationally Inattentive Consumption Choices,” *Review of Economic Dynamics*, 2013, *16* (3), 421–439.

Woodford, Michael, “Imperfect Common Knowledge and the Effects of Monetary Policy,” *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, 2003.

_ , “Information-Constrained State-Dependent Pricing,” *Journal of Monetary Economics*, 2009, *56*, S100–S124.

Yang, Choongryul, “Rational Inattention, Menu Costs, and Multi-Product Firms: Micro Evidence and Aggregate Implications,” 2019. Manuscript.

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

APPENDIX

A Proofs

Proof of Lemma 2.1. First, note that observing $\{a^t\}_{t=0}^{\infty}$ induces the same action payoffs over time as $\{S^t\}_{t=0}^{\infty}$ because at any time t and for every possible realization of S^t, the agent gets $a(S^t)$ – the optimal action induced by that realization – as a direct signal. Suppose now that a^t is not a sufficient statistic for S^t relative to X^t. Then, we can show that $\{a^t\}_{t=0}^{\infty}$ costs less in terms of information than $\{S^t\}_{t=0}^{\infty}$. To see this, note that for any $t \geq 1$ and S^t, consecutive applications of the chain-rule of mutual information imply

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

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

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

Thus,

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

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

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

Finally, note that $X^t \to S^t \to a^t$ 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

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

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

***Proof of Lemma* 2.2.** The chain-rule implies $\mathbb{I}(X^t;a^t|a^{t-1})=\mathbb{I}(X^t;a_t,a^{t-1}|a^{t-1})=\mathbb{I}(X^t;a_t|a^{t-1})$. Moreover, it also implies

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

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

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

■

***Proof of Lemma* 2.3.** We prove this Proposition by showing that for any sequence of actions, we can construct a Gaussian process that costs less in terms of information costs, but generates the exact same payoff sequence. To see this, take an action sequence $\{\vec{a}_t\}_{t\geq 0}$, and let $a^t\equiv\{\vec{a}_\tau:0\leq\tau\leq t\}\cup S^{-1}$ denote the information set implied by this action sequence. Now define a sequence of Gaussian variables $\{\hat{a}_t\}_{t\geq 0}$ such that for $t\geq 0$,

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

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.[30] ■

Proof of Proposition 2.1. 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 $\vec{a}_t=\mathbf{H}'\vec{x}_{t|t}$ which is allowed as $\mathbf{H}$ has full column rank.

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

Additionally, observe that given a^t:

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

Thus, the RI Problem becomes:

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

$$s.t. \quad \mathbf{\Sigma}_{t+1|t} = \mathbf{A}\mathbf{\Sigma}_{t|t}\mathbf{A}' + \mathbf{Q}\mathbf{Q}', \quad \forall t \geq 0, \tag{law of motion for priors}$$

$$\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} \succeq 0, \quad \forall t \geq 0 \tag{no-forgetting}$$

$$0 \prec \mathbf{\Sigma}_{0|-1} = var(\vec{x}_0|S^{-1}) \prec \infty \quad \text{given.} \tag{initial prior}$$

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. ■

***Proof of Proposition* 2.2.** We start by writing the Lagrangian. Let $\mathbf{\Gamma}_t$ be a symmetric matrix whose k'th row is the vector of shadow costs on the k'th column of the evolution of prior at time t. Moreover, let λ_t be the vector of shadow costs on the no-forgetting constraint which can be written as $\text{eig}(\mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t}) \geq 0$ where eig(.) denotes the vector of eigenvalues of a matrix.

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

But notice that

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

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

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

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

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

Moreover, note that complementary slackness for this constraint requires:

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

re-writing the Lagrangian we get:

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

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

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

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

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

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

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

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

Thus,

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

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

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

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

$$(\mathbf{AA}'\mathbf{w} = 0) \wedge (\mathbf{QQ}'\mathbf{w} = 0) \Rightarrow (\mathbf{AA}' + \mathbf{QQ}')\mathbf{w} = 0 \tag{A.4}$$

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.

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

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

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:

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

∎

***Proof of Theorem* 2.1.** From the FOC in Proposition 2.2 observe that

$$\omega \mathbf{\Sigma}_{t|t}^{-1} = \mathbf{\Omega}_t + \mathbf{\Lambda}_t \Rightarrow \mathbf{\Sigma}_{t|t-1} - \mathbf{\Sigma}_{t|t} = \mathbf{\Sigma}_{t|t-1} - \omega(\mathbf{\Omega}_t + \mathbf{\Lambda}_t)^{-1}. \tag{A.5}$$

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

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

where we have imposed the complementarity slackness $\mathbf{X}_t\mathbf{\Lambda}_t = 0$. Finally, multiply this equation by $\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}$ from right and $\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}$ from left.[31] We have

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

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

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

Now, for ease of notation let

$$\hat{\mathbf{X}}_t \equiv \mathbf{U}_t' \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \mathbf{X}_t \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}} \mathbf{U}_t \tag{A.6}$$

$$\hat{\mathbf{\Lambda}}_t \equiv \mathbf{U}_t' \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \mathbf{\Lambda}_t \mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}} \mathbf{U}_t \tag{A.7}$$

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

$$\hat{\mathbf{X}}_t \mathbf{D}_t - \hat{\mathbf{\Lambda}}_t = \mathbf{D}_t - \omega \mathbf{I} \tag{A.8}$$

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

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

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

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

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

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

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

Now, using Equation A.7 we get:

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

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

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

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

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

■

***Proof of Theorem* 2.2.** 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

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

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

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

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

$$\mathbf{\Sigma}_{t|t} = \mathbf{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{Y}_t'\mathbf{\Sigma}_{t|t-1} \Rightarrow \mathbf{K}_t\mathbf{Y}_t' = \mathbf{I} - \mathbf{\Sigma}_{t|t}\mathbf{\Sigma}_{t|t-1}^{-1} \tag{A.16}$$

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

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

where $d_{i,t}$ is the i'th eigenvalue in $\mathbf{D}_t$ and $\mathbf{y}_{i,t}$ is the i'th column of the matrix $\mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}\mathbf{U}_t$. 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 $\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}$:

$$\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}\mathbf{y}_{i,t} = \mathbf{\Sigma}_{t|t-1}^{-\frac{1}{2}}(\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}}\mathbf{\Omega}_t\mathbf{\Sigma}_{t|t-1}^{\frac{1}{2}})\mathbf{u}_{i,t} = d_{i,t}\mathbf{\Sigma}_{i,t}^{-\frac{1}{2}}\mathbf{u}_{i,t} = d_{i,t}\mathbf{y}_{i,t} \tag{A.18}$$

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

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

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

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

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.16

we have:

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

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

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

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:

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

■

Proof of Proposition 2.3. From the proof of the last Theorem, recall that the Kalman gain for predicting the state is given by

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

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

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

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

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

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

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

■

Proof of Lemma 3.1. The log-linearized Euler equation from the household side is

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

Combining this with the monetary policy rule, we have

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

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. ■

Proof of Proposition 3.1. **Part 1.** For ease of notation we drop the firm index i in the proof. The FOC in Proposition 2.2 in this case reduces to

$$\lambda_t = 1 - \theta + \frac{\omega}{\sigma^2_{t|t}} - \frac{\beta\omega}{\sigma^2_{t+1|t}} + \beta\lambda_{t+1} \tag{A.30}$$

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

$$\sigma^4_{t|t} + \left[\frac{\sigma^2_u}{\phi^2} - (1-\beta)\frac{\omega}{\theta-1}\right]\sigma^2_{t|t} - \frac{\omega}{\theta-1}\frac{\sigma^2_u}{\phi^2} = 0 \tag{A.31}$$

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 $\underline{\sigma}^2 \le \sigma^2_{t|t-1}$. Thus,

$$\sigma^2_{t|t} = \min\{\sigma^2_{t|t-1}, \underline{\sigma}^2\}. \tag{A.32}$$

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

$$\sigma^2_{i,t|t} = (1-\kappa_{i,t})\sigma^2_{i,t|t-1} \Rightarrow \kappa_{i,t} = 1 - \min\{1, \frac{\underline{\sigma}^2}{\sigma^2_{i,t|t-1}}\} = \max\{0, 1 - \frac{\underline{\sigma}^2}{\sigma^2_{i,t|t-1}}\}. \tag{A.33}$$

■

Proof of Corollary 3.1. Follows from the characterization of $\underline{\sigma}^2$ in Proposition 3.1. ∎

Proof of Proposition 3.2. **Part 1.** Recall from the proof of Proposition 3.1 that

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

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

$$\pi_t = \frac{\kappa_t}{1-\kappa_t} y_t. \tag{A.35}$$

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

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

Part 3. If $\sigma^2_{T|T-1} \geq \underline{\sigma}^2$, 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. ∎

Proof of Corollary 3.2. 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

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

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

Proof of Corollary 3.3. The transition to the new steady state follows from the fact that reservation uncertainty increases with a positive shock to $\underline{\sigma}^2$. The policy function of the firm in Proposition 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. ∎

Proof of Proposition 3.3. Note that in the steady state of the attention problem, inflation and nominal demand, $\vec{s}_t \equiv \begin{bmatrix} q_t \\ \pi_t \end{bmatrix}$, jointly evolve according to

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

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

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

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

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

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

$$\mathbf{\Sigma}_s \equiv \begin{bmatrix} \frac{1}{\kappa} & \frac{1}{2-\kappa} \\ \frac{1}{2-\kappa} & \frac{(3-2\kappa)\kappa}{(2-\kappa)^3} \end{bmatrix} \frac{\sigma_u^2}{\phi^2} \tag{A.37}$$

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

$$\mathbf{K}_s = \kappa\frac{\mathbf{\Sigma}_s\mathbf{e}_1\mathbf{e}_1'}{\mathbf{e}_1'\mathbf{\Sigma}_s\mathbf{e}_1}$$
$$= \begin{bmatrix} \kappa \\ \frac{\kappa^2}{2-\kappa} \end{bmatrix} \mathbf{e}_1 \tag{A.38}$$

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

$$\hat{\pi}_t = (1-\kappa)\hat{\pi}_{t-1} + \frac{\kappa^2}{2-\kappa}(q_t - p_{t-1}) = (1-\kappa)\hat{\pi}_{t-1} + \frac{\kappa^2}{(2-\kappa)(1-\kappa)}y_t \tag{A.39}$$

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

$$\hat{\pi}_t - (1-\kappa)\hat{\pi}_{t-1} = (1-\kappa)\hat{\pi}_{t-1} - (1-\kappa)^2\hat{\pi}_{t-2} + \frac{\kappa^2}{(2-\kappa)(1-\kappa)}(y_t - (1-\kappa)y_{t-1})$$
$$= (1-\kappa)\hat{\pi}_{t-1} - (1-\kappa)^2\hat{\pi}_{t-2} + \frac{\kappa^2}{2-\kappa}\frac{\sigma_u}{\phi}u_t. \tag{A.40}$$

■

Proof of Corollary 3.4. 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). ■

B Computing the Equilibrium

B.1 Matrix Representation and Solution Algorithm

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

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

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

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

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

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

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

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

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

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

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

Moreover, note that

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

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

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

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

Now, note that

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

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

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

Also, the Taylor rule gives:

$$
\begin{aligned}
& i_t = \rho_1 i_{t-1} + \rho_2 i_{t-2} + (1-\rho_1-\rho_2)\left(\phi_\pi \pi_t + \phi_x x_t + \phi_{\Delta y}\left(y_t - y_{t-1}\right)\right) + u_t \\
\implies & \left(\mathbf{I} - \rho_1 \mathbf{M} - \rho_2 \mathbf{M}^2\right)\mathbf{H}_i = (1-\rho_1-\rho_2)\,\phi_\pi \mathbf{H}_\pi + (1-\rho_1-\rho_2)\,\phi_x \mathbf{H}_x \\
& + (1-\rho_1-\rho_2)\,\phi_{\Delta y}\,(\mathbf{I}-\mathbf{M})\left(\mathbf{H}_x + \frac{1+\psi}{\sigma+\psi}\mathbf{H}_a\right) + \mathbf{H}_u
\end{aligned}
$$

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

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

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

C Appendix Tables

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

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

Notes: This table reports least squares estimates of the Taylor rule. We use the Greenbook forecasts of current and future macroeconomic variables. The interest rate is the target federal funds rate set at each meeting from the Fed. The measure of the output gap is based on Greenbook forecasts. We consider two time samples: 1969–1978 and 1983–2002. Newey-West standard errors are reported in parentheses. ***, **, * denotes statistical significance at 1%, 5%, and 10% levels respectively.

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

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

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