Dynamic Rational Inattention: Analytical Results

Bartosz Maćkowiak* Filip Matějka† Mirko Wiederholt‡

April 2018

Abstract

This paper presents analytical results for dynamic rational inattention problems as in Sims (2003). The agent tracks an optimal action that follows a Gaussian process. The agent chooses the properties of the signals that he receives, so as to minimize the mean squared difference between his action and the optimal action, subject to a constraint on information flow. We prove that the optimal signal is a one-dimensional signal about the elements of the state vector, which typically has non-zero signal weights on all elements of the state vector. The intuition for these results is that an agent with memory and limited attention wants to learn about the current optimal action and the best predictors of future optimal actions. Hence, in a dynamic economy, rational inattention creates a combination of delay in actions due to noise in the optimal signal and forward-looking actions due to forward-looking information choice. We illustrate these analytical results in a macroeconomic model of price-setting and a business cycle model with news shocks.

Keywords: rational inattention, information choice, Kalman filter, Macroeconomics

*European Central Bank and CEPR, 60640 Frankfurt am Main, Germany (bartosz.mackowiak@ecb.int).

†CERGE-EI and CEPR, P.O. Box 882, Prague, Czech Republic (filip.matejka@cerge-ei.cz).

‡Goethe University Frankfurt and CEPR, 60629 Frankfurt am Main, Germany (wiederholt@wiwi.uni-frankfurt.de); corresponding author.

1 Introduction

Most economic decision makers have to take decisions not only once but repeatedly over time. Think of the decision maker in a firm who chooses the price of a good or the scale of production. Think of a household who chooses how much to spend or which goods to consume. Think of a policy maker who chooses government spending or taxes. All economic decision makers have memory and a limited ability to process information. How should a decision maker acquire information if he or she cannot attend perfectly to all available information and information acquired today is also useful in the future because of memory and repeated decision making?

There is a growing literature that models agents' limited attention as a constraint on information flow (Sims, 2003). In this literature, agents can choose the information structure, subject to the information flow constraint ("rational inattention"). Many interesting applications of this theory are dynamic, but most papers study static models, because dynamic rational inattention problems are difficult to solve. The papers that do study dynamic models take one of the following approaches. First, the information structure and the actions are computed by brute-force numerical optimization (Sims, 2003, Maćkowiak and Wiederholt, 2015).[1] Second, the economy is assumed to be uncorrelated over time to turn a dynamic economy into a sequence of static economies (Maćkowiak and Wiederholt, 2009, Section IV). Third, a simple information structure is assumed rather than derived (Adam, 2007, Luo, 2008, Paciello and Wiederholt, 2014). Fourth, agents are assumed to have no memory or accessing memory is assumed to be as costly as paying attention to the outside world (Woodford, 2009, Stevens, 2015). In this paper, we take none of these simplifying approaches. We derive analytical results for rational inattention problems in correlated economies with memory. Our results are related to two earlier papers (Maćkowiak and Wiederholt, 2009, Section V, Steiner, Stewart, and Matějka, 2017) that are summarized in detail in the following section. The goal of this paper is to derive insights about the solution to dynamic rational inattention problems and to make it easy to solve those problems.

We start from a benchmark dynamic rational inattention problem. Time is discrete. A decision maker has to take an action in every period. The optimal action follows a Gaussian stochastic process. The decision maker chooses the properties of the signals that he receives, so as to minimize the mean squared difference between his action and the optimal action, subject to the information

[1] See also Tutino (2013) and Matějka (2016).

flow constraint. The agent has perfect memory, i.e., the signals received in the current period can also be used to improve actions in future periods.

We show that the optimal signal is a one-dimensional signal about the elements of the state vector. Furthermore, the signal weights on all elements of the state vector are typically non-zero. The optimal signal is a one-dimensional signal about the current optimal action and the best predictors of future optimal actions. Thus, in a dynamic economy, rational inattention creates a combination of *both* delay in actions due to noise in the optimal signal *and* forward-looking actions due to forward-looking information choice and the optimality of a one-dimensional signal.

We consider many extensions of the benchmark dynamic rational inattention problem: a stationary or non-stationary process for the optimal action, an optimal action that is driven by multiple shocks, alternative formulations of the information flow constraint, alternative objective functions. We show that the aforementioned results are robust findings. It seems intuitive that an agent with limited attention, memory, and repeated actions is forward-looking, because of the desire to enter well informed into future periods.

We illustrate our analytical results with two applications from Macroeconomics. We focus on applications from this field for two reasons. First, when Sims (1998) first proposed the idea of rational inattention, he was motivated by macroeconomic data. Our examples show that rational inattention indeed has important implications for macroeconomic time series. Second, solving dynamic macroeconomic models with rationally inattentive agents was perceived to be challenging. We demonstrate that, with this paper's results at hand, solving such models has become a lot easier.

We begin with the model of price-setting proposed by Woodford (2002). This is a benchmark business cycle model with exogenous imperfect information. In the model, price-setting firms observe noisy signals about nominal aggregate demand and, due to the informational friction, nominal disturbances have real effects. We resolve the model with signals that are optimal under rational inattention. Our decision makers use the available information flow more efficiently, and they choose signals that yield smaller pricing errors on impact of a shock and larger pricing errors in the longer run, compared with the Woodford economy. Hence, real effects of a nominal disturbance are weaker and at the same time more persistent than in the Woodford model.

Next, we consider a business cycle model with news shocks about productivity. By "a news shock about productivity" we mean a change in productivity that one can learn about before it actually occurs in production. Think of a success in research and development that affects productivity with a delay. While such shocks seem a plausible source of the business cycle, it has proven difficult to construct models in which the business cycle is driven by news. The key problem is that good news about future productivity makes agents wealthier and, in a neoclassical environment, this wealth effect increases both consumption and leisure, reducing labor input through a *reduction* in *labor supply*. With capital predetermined and current productivity unchanged, the decrease in labor input pushes output down. In our model, firms make a labor-hiring decision under rational inattention. Our analytical results imply that they choose not to distinguish carefully between current increases in productivity and future increases in productivity. Hence, a positive news shock causes an *increase* in *labor demand* on impact of the shock. This new effect will help models produce business cycle comovement in response to news shocks.

The following section summarizes the most closely related papers.[2] The benchmark dynamic rational inattention problem is formulated in Section 3. The paper's analytical results are in Section 4. Example solutions are in Section 5. The two applications from dynamic macroeconomics are in Sections 6 and 7, implications for expectations data are in Section 8, and Section 9 concludes.

2 Literature

In 2003, Christopher A. Sims formulated the following dynamic rational inattention problem. The agent chooses a process for his action, Y_t, to track an optimal action, X_t. Think of X_t as (the log of) a monopolist's profit-maximizing price, which follows a Gaussian stochastic process due to shocks to productivity, factor prices, or demand. The agent aims to minimize the mean squared difference between his action and the optimal action. The novel constraint is a constraint on the information flow to the agent. Formally, Sims studied the problem

$$\min_{b,c} E\left[(Y_t - X_t)^2\right] \tag{1}$$

subject to

$$X_t = \sum_{s=0}^{n} a_s \varepsilon_{t-s}, \qquad Y_t = \sum_{s=0}^{n} b_s \varepsilon_{t-s} + \sum_{s=0}^{n} c_s \nu_{t-s}, \tag{2}$$

[2] See Sims (2010), Veldkamp (2011), or Wiederholt (2010) for a review of the literature on rational inattention.

and

$$\lim_{T\to\infty}\frac{1}{T}I\left(X_1,\ldots,X_T;Y_1,\ldots,Y_T\right)\leq\kappa, \tag{3}$$

where ε_t and ν_t follow independent Gaussian white noise processes with unit variance. The optimal action, which the agent takes as given, follows a Gaussian moving average process. The actual action, which the agent chooses, can follow any Gaussian process that is a sum of two components: a component driven by the innovations to the optimal action, ε_t, and a component driven by noise, ν_t.[3] The left-hand side of the weak inequality, which is formally defined in the next section, quantifies the information flow to the agent by the amount of information contained in the (sequence of) actions about the (sequence of) optimal actions. The constraint states that the information flow to the agent cannot exceed the parameter κ.[4]

Sims solved this problem numerically. An example solution with a linearly declining sequence of weights a_s is displayed in Figure 1. He noted that the agent responds with delay to an innovation in the optimal action (compare the optimal b_s to the a_s) and that the agent's action contains noise (the optimal c_s are non-zero). Hence, Sims argued that optimal choice subject to an information flow constraint can be a foundation for both sluggishness in actions and stochasticity of actions conditional on fundamentals.

Maćkowiak and Wiederholt (2009) first generalized this problem by allowing all moving average processes to be infinite-order moving average processes and then derived an analytical solution to the problem (1)-(3) in the special case where the optimal action follows a first-order autoregressive process: $X_t=\rho X_{t-1}+a_0\varepsilon_t$, and thus $a_s=\rho^s a_0$. In this case, one can show that the information flow constraint (3) implies the following lower bound on the value of loss function (1):

$$\frac{a_0^2}{2^{2\kappa}-\rho^2}.$$

Moreover, the following coefficients yield a process for the action that attains this bound and satisfies the information flow constraint

$$b_s=\left[\rho^s-\frac{1}{2^{2\kappa}}\left(\frac{\rho}{2^{2\kappa}}\right)^s\right]a_0 \quad\text{and}\quad c_s=\sqrt{\frac{1}{2^{2\kappa}}\frac{2^{2\kappa}-1}{2^{2\kappa}-\rho^2}}\left(\frac{\rho}{2^{2\kappa}}\right)^s a_0. \tag{4}$$

[3] In the special case of $c_s=0$ for all s, there is no noise in the actual action. In the even more special case of $c_s=0$ and $b_s=a_s$ for all s, the actual action equals the optimal action at each point in time, $Y_t=X_t$. This is the very special case that is normally assumed in Economics.

[4] Sims (2003) expressed the left-hand side of the information flow constraint (3) in terms of the moving average coefficients a_s, b_s, and c_s, $s=0,1,\ldots,n$.

Hence, this process for the action solves Sims's dynamic rational inattention problem in the case of an AR(1) process for the optimal action. In addition, Maćkowiak and Wiederholt (2009) proved that the solution has an extremely simple, alternative representation

$$Y_t = E\left[X_t|\mathcal{I}_t\right], \tag{5}$$

where

$$\mathcal{I}_t = \mathcal{I}_0 \cup \{S_1, \ldots, S_t\} \quad \text{with} \quad S_t = X_t + d\nu_t. \tag{6}$$

The agent's action is identical to the action of an agent who receives a signal of the form "current optimal action plus i.i.d. noise" in every period (the coefficient d is the one that makes the information flow constraint hold with equality), remembers all past signals, and computes the conditional expectation of the current optimal action given his information set.

The following questions arise: What is the best process for the action when the optimal action does not follow an AR(1) process? In other words, what is the optimal signal when the optimal action does not follow an AR(1) process? After stating the decision problem in Section 3, we turn to the answers to these questions in the following sections.

This paper is related to another recent paper on dynamic rational inattention problems. Steiner, Stewart, and Matějka (2017) study a dynamic model of discrete choice under rational inattention. They show that the solution takes the form of a dynamic logit rule with endogenous biases.[5] They also prove that the dynamic problem can be reduced to a collection of static problems, one for each decision node. Finally, they apply these results in three examples: an example with two periods, an example where the state follows a two-state Markov chain, and an example where the state is drawn in the first period and is then constant over time. By contrast, our results apply to the class of dynamic rational inattention problems with Gaussian optimal actions proposed in Sims (2003).

3 Decision problem

In this section we formulate the dynamic rational inattention problem. The agent tracks an optimal action that follows a Gaussian process. The agent chooses the properties of the signals about the optimal action so as to minimize the mean squared difference between the action and the optimal action subject to the information flow constraint.

[5] This result extends the static logit result of Matějka and McKay (2015) to a dynamic setting.

The optimal action, denoted by X_t, can follow any stationary Gaussian ARMA(p,q) process with p,q finite. That is, the optimal action can follow an MA(q) process with $q \geq 0$

$$X_t = \theta_0 \varepsilon_t + \ldots + \theta_q \varepsilon_{t-q},$$

an AR(p) process with $p \geq 1$

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \theta_0 \varepsilon_t,$$

or an ARMA(p,q) process with $p \geq 1$ and $q \geq 1$

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \theta_0 \varepsilon_t + \ldots + \theta_q \varepsilon_{t-q}, \tag{7}$$

where ε_t follows a Gaussian white noise process with unit variance, and $\phi_1, \ldots, \phi_p$ and $\theta_0, \ldots, \theta_q$ are coefficients.[6] The MA(n) process assumed in Sims (2003) and the AR(1) process assumed in Maćkowiak and Wiederholt (2009) are obviously special cases of this assumption.

The agent's information set in any period $t \geq 1$ includes any initial information and all signals received up to this point in time

$$\mathcal{I}_t = \mathcal{I}_0 \cup \left\{ S_1^K, \ldots, S_t^K \right\}, \tag{8}$$

where $\mathcal{I}_0$ denotes the initial information set and S_t^K denotes the signal vector received in period t. The superscript $K \geq 1$ denotes the dimension of the signal vector received in period t.

The agent chooses the number of signals, K, the content of the signals, and the variance-covariance matrix of noise in the signals. Each signal in the signal vector $S_t^K = (S_{t,1}, \ldots, S_{t,K})'$ can be about a different linear combination of current and past X_t and current and past ε_t

$$S_t^K = A X_t^M + B \varepsilon_t^N + \psi_t^K, \tag{9}$$

where $X_t^M = (X_t, \ldots, X_{t-M+1})'$ denotes the vector of current and past X_t, $\varepsilon_t^N = (\varepsilon_t, \ldots, \varepsilon_{t-N+1})'$ is the vector of current and past ε_t, and $A \in \mathbb{R}^{K \times M}$ and $B \in \mathbb{R}^{K \times N}$ are matrices of coefficients. The integers $M \geq \max\{p, 1\}$ and $N \geq \max\{q, 1\}$ can be arbitrarily large. To fix ideas, think of $M = N = 10^{100}$. By choosing the matrices A and B the agent chooses the content of the signals. For example, the agent can choose to learn about the current optimal action or past optimal actions

[6] Without loss in generality, the coefficients on the largest lags are required to be non-zero, $\phi_p \neq 0$ and $\theta_q \neq 0$.

or the current innovation to the optimal action or past innovations or linear combinations of them. The vector of noise $\psi_t^K = (\psi_{t,1}, \ldots, \psi_{t,K})'$ follows a Gaussian vector white noise process with variance-covariance matrix Σ_ψ. By choosing the matrix Σ_ψ the agent chooses the precision of the K signals in S_t^K and the covariances of the noise terms in these signals.

The agent chooses the number of signals, K, the content of the signals, A and B, and the variance-covariance matrix of noise in the signals, Σ_ψ, so as to minimize the mean squared error subject to the information flow constraint. Formally, we study the problem

$$\min_{K,A,B,\Sigma_\psi} E\left[(X_t - Y_t)^2\right], \tag{10}$$

subject to (7), $Y_t = E[X_t|\mathcal{I}_t]$, (8), (9), and the information flow constraint

$$\lim_{T\to\infty} \frac{1}{T} I\left(\bar{X}_0, X_1, \ldots, X_T; S_1^K, \ldots, S_T^K\right) \leq \kappa. \tag{11}$$

The vector $\bar{X}_0$ in the information flow constraint denotes the vector of initial conditions for the process for the optimal action.[7]

A dynamic rational inattention problem is similar to a static rational inattention problem in which the optimal action depends on multiple shocks. The optimal action, X_t, is a linear combination of current and past shocks, $\varepsilon_t, \varepsilon_{t-1}, \ldots$. In a rational inattention model, the agent chooses how to learn about these shocks. The key difference between a dynamic rational inattention problem and a static rational inattention problem with multiple shocks is memory. Equation (8) implies that any information acquired in period t will also be useful in future periods.

There are four differences between Sims's dynamic rational inattention problem (1)-(3) and our dynamic rational inattention problem (7)-(11). First, we allow for a more general process for the optimal action. Second, while Sims assumes that the agent chooses directly a process for the action, we assume that the agent chooses the properties of the signals. This modelling choice is motivated by the fact that in the AR(1) case the solution to Sims's problem looked complicated in terms of actions (equation (4)) but simple in terms of signals with i.i.d. noise (equations (5)-(6)). In Section 5, we demonstrate numerically that our assumption that the vector of noise, ψ_t^K, is i.i.d. over time is not a binding restriction. Third, in the information flow constraint (11) we measure the information content of the signals rather than the information content of the actions, which has

[7] That is, $\bar{X}_0 = (X_{1-p}, \ldots, X_0)$ in the AR(p) case, $\bar{X}_0 = (\varepsilon_{1-q}, \ldots, \varepsilon_0)$ in the MA(q) case, and $\bar{X}_0 = (X_{1-p}, \ldots, X_0, \varepsilon_{1-q}, \ldots, \varepsilon_0)$ in the ARMA(p,q) case.

no effect on the solution. Fourth, in the information flow constraint (11), we measure the amount of information contained in the signals about both: (i) the optimal actions, *and* (ii) the initial conditions for the optimal actions, which has no effect on the solution and simplifies the beginning of a proof.

Quantifying information flow. The information flow constraint (11) restricts the information flow to the agent and thereby formalizes the idea that the agent has a limited amount of attention. Subject to this constraint, the agent chooses the allocation of attention (the number, content, and precision of signals). The agent decides whether to pay attention to the current optimal action, past optimal actions, innovations, and so on. The agent chooses how to deal with his own cognitive limitation, i.e., the agent is rationally attentive/inattentive.

We still need to specify how the information flow to the agent is quantified, i.e., we still need to formally define the left-hand side of the information flow constraint. Following Sims (2003), we quantify the information flow to the agent by reduction in uncertainty, where uncertainty is measured by entropy. Formally, the mutual information between two random vectors $X^T = (X_1, \ldots, X_T)$ and $S^T = (S_1, \ldots, S_T)$, denoted $I\left(X^T; S^T\right)$, equals the difference between the entropy of X^T, denoted $H\left(X^T\right)$, and the conditional entropy of X^T given knowledge of S^T, denoted $H\left(X^T | S^T\right)$,

$$I\left(X^T; S^T\right) = H\left(X^T\right) - H\left(X^T | S^T\right). \tag{12}$$

Entropy is simply a measure of uncertainty. Hence, the mutual information between X^T and S^T simply measures how much the uncertainty about the realization of $(X_1, \ldots, X_T)$ is reduced due to knowledge of $(S_1, \ldots, S_T)$.

Dividing by T on both sides of the last equation and taking the limit as $T \to \infty$ yields

$$\lim_{T\to\infty} \frac{1}{T} I\left(X^T; S^T\right) = \lim_{T\to\infty} \frac{1}{T} H\left(X^T\right) - \lim_{T\to\infty} \frac{1}{T} H\left(X^T | S^T\right). \tag{13}$$

The first term on the right-hand side measures how total uncertainty about $(X_1, \ldots, X_T)$ grows per unit of time. The second term on the right-hand side measures how total uncertainty about $(X_1, \ldots, X_T)$ grows per unit of time given knowledge of $(S_1, \ldots, S_T)$. The difference between the two terms measures the information flow to the agent.

Replacing S^T by Y^T in the last equation formally defines the left-hand side of equation (3). The fact that we condition on the signals rather than on the actions in the information flow constraint

(11) and the fact that we also measure the uncertainty reduction about the initial conditions in the information flow constraint (11) has no effect on the solution to our dynamic rational inattention problem (7)-(11).

Equivalent formulations of the constraint. There exist many equivalent formulations of the information flow constraint. First, the mutual information in equation (12) remains unchanged when the vector S^T is replaced by any vector $\hat{S}^T$ with the property: S^T can be computed from $\hat{S}^T$ and vice versa. The reason is that the conditional entropy in equation (12) remains unchanged. The symmetry of mutual information, $I\left(X^T;S^T\right) = I\left(S^T;X^T\right) = H\left(S^T\right) - H\left(S^T|X^T\right)$, implies that the same argument applies to X^T. The mutual information in equation (12) remains unchanged when the vector X^T is replaced by any vector $\hat{X}^T$ with the property: X^T can be computed from $\hat{X}^T$ and vice versa. Furthermore, adding or dropping random variables which do not affect the limit on the left-hand side of equation (13) obviously also yields an equivalent formulation of the information flow constraint.

Second, the information flow constraint (11) is equivalent to a constraint on the difference between prior uncertainty and posterior uncertainty at a given point in time. This equivalence result is new and is therefore stated in a lemma.

Lemma 1 *Let $S^{K,t} = \left\{S_1^K, \ldots, S_t^K\right\}$ denote the set of signals received up to and including time t. The information flow constraint (11) is equivalent to*

$$\lim_{T\to\infty} \left[H\left(\varphi_T|S^{K,T-1}\right) - H\left(\varphi_T|S^{K,T}\right)\right] \leq \kappa, \tag{14}$$

where the vector φ_t can be any vector with the following two properties: (i) X_t^M and ε_t^N in equation (9) for S_t^K can be computed from φ_t, and (ii) φ_t contains no redundant elements.

Proof. See Appendix A. ■

The left-hand side of the information flow constraint (14) is the difference between prior uncertainty and posterior uncertainty about all variables that the new signal, S_T^K, can be about. The limit implies that this difference is computed once the agent has access to a long history of signals, $S^{K,T-1}$. The lemma establishes the equivalence of two formulations of the information flow constraint that have appeared in the literature and it will be used in the following section to prove our first main result. Furthermore, in Section 4 we show that the information flow constraint (14) is equivalent to a constraint on a particular signal-to-noise ratio once $K = 1$.

The objective and the initial information set. The agent aims to minimize the mean squared difference between the action, Y_t, and the optimal action, X_t. With a quadratic objective, the best action at any time t given any information set $\mathcal{I}_t$ is the conditional expectation of the optimal action, $Y_t = E[X_t|\mathcal{I}_t]$. Substituting this equation for the action into the agent's loss function (10) yields the mean squared error $E\left[(X_t - E[X_t|\mathcal{I}_t])^2\right]$.

This mean squared error can in principle vary over time, because the agent can condition on more signals as time passes. Recall that the agent's information set is specified in equation (8). To abstract from transitional dynamics in this mean squared error, we assume that – after the agent has chosen the number, content, and precision of the signals in period zero – the agent receives a long sequence of signals in period zero such that conditional second moments are independent of time.[8,9] Thus, conditional second moments can be computed using the steady-state Kalman filter.

This assumption about the information set $\mathcal{I}_0$ after the agent has made the information choice in period zero has three implications. First, it does not matter which period $t \geq 1$ one is referring to in the agent's loss function (10). Second, replacing the agent's loss function (10) by the loss function

$$E\left[\sum_{t=1}^{\infty}\beta^t (X_t - E[X_t|\mathcal{I}_t])^2\right] = \frac{\beta}{1-\beta}E\left[(X_t - E[X_t|\mathcal{I}_t])^2\right], \quad \beta \in (0,1), \tag{15}$$

is simply a monotone transformation of the objective and thus does not affect the solution to the dynamic rational inattention problem (7)-(11). Third, the conditional expectation $E[X_t|\mathcal{I}_t]$ is a time-invariant function of the signals. Hence, the action $Y_t = E[X_t|\mathcal{I}_t]$ is a time-invariant function of the shocks and one can directly compare the solution of the dynamic rational inattention problem (7)-(11) to the solution of Sims's dynamic rational inattention problem (1)-(3), where the action is *assumed to be* a time-invariant function of the shocks.

Applications. In the price-setting application of Section 6, the optimal action X_t is the (log of the) profit-maximizing price of a monopolist driven by normally distributed shocks to nominal marginal cost. In the hiring application of Section 7, the optimal action X_t is the (log of the) profit-maximizing labor input of a competitive firm driven by normally distributed shocks to productivity. After a quadratic approximation to the profit function, the loss in profit due to a suboptimal action

[8] The same assumption is made in Maćkowiak and Wiederholt (2009).

[9] Including the signals received in period zero in the information flow constraint (11) does not affect the limit on the left-hand side of the information flow constraint (11).

is proportional to the squared difference between the profit-maximizing action and the actual action. In period zero, the manager chooses the properties of the signals that he receives, so as to minimize the expected discounted sum of losses in profit due to suboptimal actions (15), subject to the information flow constraint (11). In the following periods, the manager receives the signals and sets the actual action equal to the conditional expectation of the profit-maximizing action.

4 Solution

In this section we present analytical results. We first prove that in the search for the optimal signal one can restrict attention to a one-dimensional signal about the state vector. The information flow constraint (11) then reduces to a constraint on a single signal-to-noise ratio. Next, we turn to the AR(2) case and the ARMA(1,1) case. We prove that the optimal signal is *not* on the current optimal action *only*. The optimal signal is on *both* the current optimal action *and* the best predictor of next period's optimal action. Finally, we prove that as the information flow gets very large, $\kappa \to \infty$, the optimal signal converges to a signal that is about the current optimal action only. Section 5 presents example solutions and demonstrates numerically that the solution of Sims's rational inattention problem (1)-(3) equals the solution of our rational inattention problem (7)-(11).

4.1 Two main results

In this subsection, we prove that any optimal signal vector is only about the elements of the state vector and the agent can attain the optimum with a one-dimensional signal.

Proposition 1 *In the AR(p) case with $p \geq 1$, any optimal signal vector is on linear combinations of $\left\{X_t, \ldots, X_{t-(p-1)}\right\}$ only. In the ARMA(p,q) case with $p \geq 1$ and $q \geq 1$, any optimal signal vector is on linear combinations of $\left\{X_t, \ldots, X_{t-(p-1)}\right\}$ and $\left\{\varepsilon_t, \ldots, \varepsilon_{t-(q-1)}\right\}$ only. In the white noise case, any optimal signal vector is on X_t only. In the MA(q) case with $q \geq 1$, any optimal signal vector is on linear combinations of X_t and $\left\{\varepsilon_t, \ldots, \varepsilon_{t-(q-1)}\right\}$ only.*

Proof. See Appendix B. ■

To summarize Proposition 1, note the following. The optimal action, X_t, is assumed to follow an ARMA(p,q) process. Recall from time series analysis that the optimal action, X_t, then also has a first-order vector autoregressive representation

$$\xi_{t+1} = F\xi_t + v_{t+1},$$

where the vector v_t follows a vector white noise process, the matrix F is a square matrix, and the vector ξ_t is given by

$$\xi_t = \begin{cases} \left(X_t, \ldots, X_{t-(p-1)}\right)' & \text{if } p \geq 1 \text{ and } q = 0 \\ \left(X_t, \ldots, X_{t-(p-1)}, \varepsilon_t, \ldots, \varepsilon_{t-(q-1)}\right)' & \text{if } p \geq 1 \text{ and } q \geq 1 \\ X_t & \text{if } p = 0 \text{ and } q = 0 \\ \left(X_t, \varepsilon_t, \ldots, \varepsilon_{t-(q-1)}\right)' & \text{if } p = 0 \text{ and } q \geq 1 \end{cases}. \tag{16}$$

The vector ξ_t contains all information available at time t about the current value of the optimal action and future values of the optimal action. Hence, Proposition 1 states that any optimal signal vector is only about the current optimal action and the variables that determine future values of the optimal action. The agent could in principle receive signals about variables that are not elements of the vector ξ_t, i.e., the agent could receive signals about higher lags of the optimal action and higher lags of the innovation to the optimal action, but the optimal signal weights on those higher lags are zero.

To gain intuition for this result, it is important to understand that the signal vector S_τ^K received in period τ has a dual role: it is the new information received in period τ, $\mathcal{I}_\tau = \mathcal{I}_{\tau-1} \cup \left\{S_\tau^K\right\}$, and it is part of the agent's memory $\mathcal{I}_{t-1}$ in any future period $t > \tau$. Now consider any period $t \geq 1$. The agent would like to enter the period with low *prior uncertainty* about the optimal action in the period. The prior uncertainty about the optimal action in the period is determined by signals received in the past. See the first term on the left-hand side of the information flow constraint (14). The most efficient way to reduce prior uncertainty about the optimal action in the period through signals received in the past is to have received signals about the elements of the vector ξ_τ in $\tau < t$. Since in the future the present will be the past, the agent considers receiving signals about the elements of ξ_t in period t. Furthermore, given any prior information about the optimal action in the period, the agent would like to receive *new* information about the optimal action in the period so as to minimize the mean squared error, $E\left[\left(X_t - E\left[X_t|\mathcal{I}_{t-1}, S_t^K\right]\right)^2\right]$. The most efficient way to reduce the mean squared error given any prior information, $\mathcal{I}_{t-1}$, is to receive a signal about X_t. Thus, the agent considers receiving a signal about X_t in period t. Finally, receiving signals about higher lags of the optimal action than $X_{t-(p-1)}$ or higher lags of the innovation to the optimal action than $\varepsilon_{t-(q-1)}$ is neither an efficient way to reduce prior uncertainty about the optimal action in future periods nor an efficient way to reduce the mean squared error given prior information. The optimal signal weights on those higher lags are therefore zero.

The proof of Proposition 1 is built around this intuition. First, the information flow constraint (11) is equivalent to the information flow constraint (14), which expresses the information flow to the agent as the difference between prior uncertainty and posterior uncertainty about all variables that the new signal can be about. Second, the *prior* uncertainty in the constraint (14) is a function only of the conditional variance-covariance matrix of φ_t given $\mathcal{I}_{t-1}$, denoted Σ_1. The *posterior* uncertainty in the constraint (14) is a function only of the conditional variance-covariance matrix of φ_t given $\mathcal{I}_t$, denoted Σ_0. Third, split the vector φ_t into two sub-vectors: ξ_t defined in equation (16) and a vector χ_t containing the higher lags of the optimal action and the higher lags of the innovation to the optimal action. Let d denote the number of elements in ξ_t. The information flow on the left-hand side of (14) equals the sum of the information flow about ξ_t and the information flow about χ_t given ξ_t. Furthermore, the information flow about ξ_t is a function only of the upper-left $(d \times d)$ submatrix of Σ_1 (prior) and the upper-left $(d \times d)$ submatrix of Σ_0 (posterior). Fifth, consider a signal vector that is only about the elements of ξ_t. The signal is flexible enough to go from any upper-left $(d \times d)$ submatrix of Σ_1 to any feasible upper-left $(d \times d)$ submatrix of Σ_0. Furthermore, given any upper-left $(d \times d)$ submatrix of Σ_0, the upper-left $(d \times d)$ submatrix of Σ_1 follows directly from the law of motion for X_t. Moreover, the objective (10) equals the 1,1-element of the upper-left $(d \times d)$ submatrix of Σ_0, and a signal vector that is only about the elements of ξ_t contains no information about χ_t given ξ_t. Finally, consider a signal vector that is *not only* about the elements of ξ_t. There always exists a signal vector that is *only* about the elements of ξ_t that yields the same value of the objective with a smaller information flow. The reason is simple. There always exists a signal vector that is only about ξ_t that yields the same upper-left $(d \times d)$ submatrices of Σ_0 and Σ_1 and by construction contains no information about χ_t given ξ_t.

One common theme in the rational inattention literature is that agents pay attention to the variables that are most useful to them. Proposition 1, which states that any optimal signal vector is only about the elements of the state vector, can be viewed as an example of this common theme in a dynamic setting. The following proposition is about the dimension of the signal about the state vector.

Proposition 2 *The agent can attain the optimum with a one-dimensional signal.*

Proof. See Appendix C. ■

Even though this result is not obvious, the intuition for (and proof of) this result is actually quite simple. If the signal vector S_t^K consists of multiple signals on the *same* linear combination of current and past X_t and current and past ε_t, the signal vector is obviously equivalent to a one-dimensional signal with a higher precision. If the signal vector S_t^K consists of multiple signals on *different* linear combinations of current and past X_t and current and past ε_t, the signal vector is equivalent to another signal vector in which all signals apart from one are only about the past. Furthermore, a signal vector in which some signals are only about the past cannot be optimal, because the signals that are only about the past could have been received earlier. This would reduce the value of the loss function without affecting information flow.

Propositions 1 and 2 imply that in the search for the optimal signal one can restrict attention to one-dimensional signals that have the following state-space representation

$$\xi_{t+1} = F\xi_t + v_{t+1}, \tag{17}$$

$$S_t = h'\xi_t + \psi_t, \tag{18}$$

where the state vector ξ_t is given by equation (16), the vector of signal weights h is a column vector with $\max\{1,p\}+q$ elements, and the noise ψ_t follows a Gaussian white noise process with variance $\sigma_\psi^2 > 0$.[10] The final step is to find the vector of signal weights h and the variance of noise σ_ψ^2 that minimize the loss function (10) subject to the information flow constraint (11).

[10] The matrix F is a square matrix and the length of the column vector v_t equals the length of ξ_t. The first element of v_{t+1} equals $\theta_0\varepsilon_{t+1}$ and the first row of the matrix F ensures that the first row of the state equation (17) equals the law of motion for X_t. In the AR(p) case with $p \geq 2$, the remaining elements of v_{t+1} equal zero and the remaining rows of the matrix F have a one just left of the main diagonal and zeros everywhere else. In the MA(q) case and the ARMA(p,q) case with $q \geq 1$, the vector v_{t+1} has q additional elements and the matrix F has q additional rows. The first additional element of v_{t+1} equals ε_{t+1} and the remaining additional elements of v_{t+1} equal zero. The first additional row of the matrix F contains only zeros and the remaining additional rows of the matrix F have a one just left of the main diagonal and zeros everywhere else.

4.2 The information flow constraint and the optimal signal

In this subsection, we prove that the information flow constraint (11) reduces to a constraint on a single signal-to-noise ratio in the case of the one-dimensional signal (18). Thereafter, we show how to compute the vector of signal weights h and the variance of noise σ_ψ^2 that minimize the loss function (10) subject to the information flow constraint (11).

The objective (10) and the constraint (11) depend only on conditional second moments (more on that below). Since from now on the signal has the state-space representation (17)-(18), one can use the Kalman filter to compute conditional second moments of the state vector for a given choice of h and σ_ψ^2.[11] Let $\Sigma_{t|t-1}$ denote the conditional variance-covariance matrix of ξ_t given $\mathcal{I}_{t-1}$. Let $\Sigma_{t|t}$ denote the conditional variance-covariance matrix of ξ_t given $\mathcal{I}_t$. Let Q denote the variance-covariance matrix of v_{t+1}. For a given choice of h and σ_ψ^2 and for given matrices F and Q, one can compute the variance-covariance matrices of the state vector, $\Sigma_{t|t-1}$ and $\Sigma_{t|t}$, from the usual Kalman filter equations

$$\begin{aligned} \Sigma_{t+1|t} &= F\Sigma_{t|t}F' + Q, \\ \Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}h\left(h'\Sigma_{t|t-1}h + \sigma_\psi^2\right)^{-1} h'\Sigma_{t|t-1}. \end{aligned}$$

See, for example, Hamilton (1994), Chapter 13. Furthermore, since X_t follows a stationary process, the limits $\Sigma_1 \equiv \lim_{t\to\infty} \Sigma_{t|t-1}$ and $\Sigma_0 \equiv \lim_{t\to\infty} \Sigma_{t|t}$ exist and are given by

$$\Sigma_1 = F\Sigma_0F' + Q, \tag{19}$$

$$\Sigma_0 = \Sigma_1 - \Sigma_1 h\left(h'\Sigma_1 h + \sigma_\psi^2\right)^{-1} h'\Sigma_1. \tag{20}$$

See, for example, Hamilton (1994), Propositions 13.1-13.2.

Let us simplify the information flow constraint. The information flow constraint (11) is equivalent to the information flow constraint (14) (see Lemma 1). In the case of the one-dimensional signal (18), the information flow constraint (14) in turn reduces to a constraint on a single signal-to-noise ratio.

Lemma 2 *In the case of the one-dimensional signal (18), the information flow constraint (14) reduces to*

$$\frac{1}{2}\log_2\left(\frac{h'\Sigma_1 h}{\sigma_\psi^2} + 1\right) \leq \kappa, \tag{21}$$

[11] See Hamilton (1994), Chapter 13, for an introduction to the Kalman filter.

where $h'\Sigma_1 h$ equals the variance of the informative component of the signal, conditional on $t-1$ information, and σ_ψ^2 equals the variance of the noise component of the signal.

Proof. Conditional normality implies that

$$H\left(\xi_t|S^{t-1}\right) - H\left(\xi_t|S^t\right) = \frac{1}{2}\log_2\left(\frac{\det \Sigma_{t|t-1}}{\det \Sigma_{t|t}}\right).$$

The information flow constraint (14) is thus equivalent to

$$\frac{1}{2}\log_2\left(\frac{\det \Sigma_1}{\det \Sigma_0}\right) \leq \kappa.$$

Using equation (20) to substitute for Σ_0 yields

$$\frac{\det \Sigma_1}{\det \Sigma_0} = \frac{1}{\det\left(I - \frac{1}{h'\Sigma_1 h + \sigma_\psi^2} hh'\Sigma_1\right)}.$$

Furthermore, it follows from Sylvester's determinant theorem that

$$\det\left(I - \frac{1}{h'\Sigma_1 h + \sigma_\psi^2} hh'\Sigma_1\right) = \det\left(1 - \frac{1}{h'\Sigma_1 h + \sigma_\psi^2} h'\Sigma_1 h\right) = \frac{\sigma_\psi^2}{h'\Sigma_1 h + \sigma_\psi^2}.$$

Substituting these equations into the previous weak inequality yields

$$\frac{1}{2}\log_2\left(\frac{h'\Sigma_1 h}{\sigma_\psi^2} + 1\right) \leq \kappa.$$

■

It seems intuitive that in the case of a one-dimensional Gaussian signal, information flow is a function of a signal-to-noise ratio only.[12] Furthermore, since the information flow constraint (14) restricts the difference between prior and posterior uncertainty at a given point in time, the relevant signal-to-noise ratio is the variance of the informative component of the signal divided by the variance of the noise component of the signal, *conditional on $t-1$ information*.

Collecting results we arrive at the following statement of the dynamic rational inattention problem (7)-(11): Propositions 1 and 2, Lemmas 1 and 2, and equations (19)-(20) imply that the dynamic rational inattention problem (7)-(11) reduces to

$$\min_{h \in \mathbb{R}^{\max\{1,p\}+q}, \sigma_\psi^2 > 0} \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} \Sigma_0 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{22}$$

[12] Adam (2007), Appendix A.3, derives a similar result for a static economy.

subject to

$$\frac{1}{2}\log_2\left(\frac{h'\Sigma_1 h}{\sigma_\psi^2}+1\right)\leq\kappa, \tag{23}$$

where the conditional variance-covariance matrices of the state vector, Σ_1 and Σ_0, are given by

$$\Sigma_1 = F\Sigma_0 F' + Q, \tag{24}$$

$$\Sigma_0 = \Sigma_1 - \Sigma_1 h\left(h'\Sigma_1 h+\sigma_\psi^2\right)^{-1}h'\Sigma_1. \tag{25}$$

The statement of the problem can be simplified further, because the information flow constraint is always binding and the information flow constraint can be solved explicitly for the variance of noise. The fact that the information flow constraint (23) is always binding implies

$$\frac{h'\Sigma_1 h}{\sigma_\psi^2}=2^{2\kappa}-1.$$

For all $\kappa > 0$, this constraint can also be expressed as

$$\sigma_\psi^2=\frac{h'\Sigma_1 h}{2^{2\kappa}-1}. \tag{26}$$

Using the binding information flow constraint (26) to substitute for the variance of noise in the Kalman filter equation (20) yields the following statement of the decision problem, for all $\kappa > 0$,

$$\min_{h\in\mathbb{R}^{\max\{1,p\}+q}}\begin{pmatrix}1 & 0 & \cdots & 0\end{pmatrix}\Sigma_0\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \tag{27}$$

where the matrices Σ_1 and Σ_0 are given by

$$\Sigma_1 = F\Sigma_0 F' + Q, \tag{28}$$

$$\Sigma_0 = \Sigma_1 - \frac{\left(1-2^{-2\kappa}\right)}{h'\Sigma_1 h}\Sigma_1 hh'\Sigma_1. \tag{29}$$

The loss function (27) is simply another way of writing the loss function (10). Equations (28)-(29) are obtained by using the information flow constraint (23) to substitute for the variance of noise in the Kalman filter equation (25). The equations (28)-(29) give Σ_1 and Σ_0 as implicit functions of the vector of signal weights h, the matrices F and Q, and the parameter κ. Solving the problem (27)-(29) yields the vector of optimal signal weights, h.[13,14]

The problem (27)-(29) can be solved in many different ways, for example, by numerical optimization or by stating and solving first-order conditions. Solving the problem (27)-(29) (or the problem (22)-(25)) by numerical optimization should be easy for anyone familiar with the Kalman filter. Moreover, we found that solving the first-order conditions also works well and we therefore state these first-order conditions in Appendix D. We also use these first-order conditions below to derive analytical results regarding the optimal vector of signal weights, h.

Since multiplying the signal (18) by a non-zero constant does not change the matrices Σ_1 and Σ_0 (see equation (20)), it is helpful to normalize either an element of h or σ_ψ^2 before solving the problem (22)-(25) or the problem (27)-(29).

4.3 Analytical results about the optimal vector of signal weights, h

We now present analytical results regarding the optimal signal weights in the optimal observation equation (18). The main insight in this subsection is that the optimal signal is on the current optimal action *and* the best predictors of future optimal actions.

Let us begin with the case of an AR(1) process for the optimal action. In this case, there is a single state variable, $\xi_t = X_t$, and one can therefore obtain a closed-form solution to the dynamic rational inattention problem (7)-(11).

Proposition 3 *If the optimal action follows an AR(1) process, $X_t = \phi_1 X_{t-1} + \theta_0 \varepsilon_t$, one can restrict attention to signals of the form $S_t = X_t + \psi_t$. For all $\kappa > 0$, the variance of noise in the optimal signal equals*

$$\sigma_\psi^2 = \frac{2^{2\kappa}}{2^{2\kappa} - 1} \frac{\theta_0^2}{2^{2\kappa} - \phi_1^2},$$

and the value of the objective at the solution equals

$$\Sigma_0 = \frac{\theta_0^2}{2^{2\kappa} - \phi_1^2}.$$

[13]Substituting the vector of optimal signal weights, h^*, into equation (26) yields the optimal variance of noise, σ_ψ^{2*}.

[14]One can also endogenize κ by adding a cost function for κ to the objective and by augmenting the set of choice variables by κ. In the set of first-order conditions presented below, this will simply add a first-order condition. Finally, note that any cost function for κ can be expressed as an equivalent cost function for the signal-to-noise ratio $\left(h'\Sigma_1 h/\sigma_\psi^2\right)$ by substituting the binding information flow constraint (23) into the original cost function for κ.

Proof. Propositions 1 and 2 imply that one can restrict attention to signals of the form $S_t = hX_t + \psi_t$. Furthermore, without loss in generality, one can normalize the scalar h to one. Substituting $F = \phi_1$, $Q = \theta_0^2$, and $h = 1$ into equations (28)-(29) yields $\Sigma_1 = \phi_1^2 \Sigma_0 + \theta_0^2$ and $\Sigma_0 = 2^{-2\kappa}\Sigma_1$. These two equations can be solved for Σ_1 and Σ_0. Substituting the solution for Σ_1 into equation (26) yields the solution for σ_ψ^2. ■

In the AR(1) case, the best predictor of any future optimal action is a linear function of the current optimal action only (e.g., the best predictor of X_{t+1} given full information in period t equals $\phi_1 X_t$, the best predictor of X_{t+2} given full information in period t equals $\phi_1^2 X_t$, and so on). Thus, in the AR(1) case, receiving a signal on the best predictor of any future optimal action is equivalent to receiving a signal on the current optimal action. For this reason, the optimal signal is of the form $S_t = hX_t + \psi_t$. The signal weight h can be normalized to one (or to ϕ_1 or to ϕ_1^2 or to any other non-zero real number) and the optimal variance of noise, σ_ψ^{2*}, is the smallest variance of noise that satisfies the information flow constraint.[15]

Next, consider an AR(2) process for the optimal action. Propositions 1 and 2 imply that one can restrict attention to signals of the form $S_t = h_1 X_t + h_2 X_{t-1} + \psi_t$. The next proposition states that the optimal signal weight on X_{t-1} is *generically different* from zero (the word generically refers to the condition $\phi_1, \phi_2 \neq 0$ in Proposition 4). Thus the optimal signal is not on the current optimal action only; the optimal signal is on *both* the current optimal action *and* the best predictor of next period's optimal action.

Proposition 4 *If the optimal action follows an AR(2) process, $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_0 \varepsilon_t$, one can restrict attention to signals of the form $S_t = h_1 X_t + h_2 X_{t-1} + \psi_t$. Furthermore, $\phi_1, \phi_2 \neq 0$ implies $h_2 \neq 0$.*

Proof. See Appendix E. ■

To see the link to a signal on the best predictor of next period's optimal action, multiply the signal $S_t = h_1 X_t + h_2 X_{t-1} + \psi_t$ by a non-zero constant ϱ to arrive at a signal about a weighted average of the current optimal action and the best predictor of next period's optimal action. Formally,

[15] Maćkowiak and Wiederholt (2009) already derived the solution in the AR(1) case but without providing any intuition for why a signal of the form $S_t = X_t + \psi_t$ is optimal in the AR(1) case. See their Propositions 3-4 or Section 2. Luo and Young (2014) *assume* the signal $S_t = X_t + \psi_t$ and then also derive the variance of noise.

$S_t' = \varrho S_t = \omega X_t + (1-\omega)(\phi_1 X_t + \phi_2 X_{t-1}) + \varrho \psi_t$.[16] The term $(\phi_1 X_t + \phi_2 X_{t-1})$ is the best predictor of X_{t+1} given full information in period t. This best predictor no longer depends on X_t only. The statement $h_2 \neq 0$ is equivalent to the statement $(1-\omega) \neq 0$. Hence, Proposition 4 states that the optimal signal is generically on *both* the current optimal action *and* the best predictor of next period's optimal action.

The key difference between a dynamic rational inattention problem and a static rational inattention problem with multiple shocks is memory. Due to memory the signal received in the current period has a dual role: it is the new information received in the current period and it affects the agent's prior beliefs in future periods. Since the agent is aware of his memory, the agent's information choice is forward-looking. The result stated in Proposition 4 that the optimal signal is generically on *both* the current optimal action *and* the best predictor of next period's optimal action simply means that the optimal information choice is affected by *both* the desire to take a good action in the current period *and* the desire to enter well informed into the next period. In the case of $\phi_2 = 0$, the optimal action follows an AR(1) process and these two goals coincide; a signal on the current optimal action is equivalent to a signal on the best predictor of any future optimal action. Finally, in the case of $\phi_1 = 0$, even time periods and odd time periods are completely separated in the sense that shocks that affect the optimal actions in even time periods do not affect the optimal actions in odd time periods. In this case, the agent chooses to learn separately about the optimal actions in even periods and the optimal actions in odd periods. Furthermore, a signal about the current optimal action in an even period is equivalent to a signal about the best predictor of any future optimal action in an even period. Hence, once the agent has separated learning about even and odd periods, the goal to learn about the present once again coincides with the goal to learn about the future.

Figure 2 displays example solutions holding constant $\phi_1 + \phi_2$, θ_0, and κ and varying ϕ_2. The optimal signal weight on last period's optimal action, h_2, is non-zero for all $\phi_1, \phi_2 \neq 0$ and non-monotonic in ϕ_2 for the reason described at the end of the last paragraph. Figure 3 shows example solutions holding constant ϕ_1, ϕ_2, and θ_0 and varying κ. The optimal signal weight on the best predictor of next period's optimal action, $(1-\omega)$, is monotonically decreasing in κ. Intuition for this monotonicity result is provided at the end of this subsection.

[16] The coefficients ϱ and ω solve $\varrho h_1 = \omega + (1-\omega)\phi_1$ and $\varrho h_2 = (1-\omega)\phi_2$. Hence, $(1-\omega) = \frac{\frac{h_2}{h_1}}{\phi_2 + (1-\phi_1)\frac{h_2}{h_1}}$.

The next proposition characterizes the optimal signal when the optimal action follows an ARMA(1,1) process; more specifically, it characterizes the optimal signal in the case of news shocks. In Macroeconomics, it is often assumed that nature draws shocks that affect fundamentals with a delay. For example, a success in research and development affects productivity with a delay or a policy change is announced but implemented with a delay. The delay implies that future changes in, say, productivity, interest rates, and taxes can be learned about to some extent in advance. In this case, the productivity shock or the policy shock is called a "news shock" because first the news arrives and then the change occurs. An abstract example is $X_t = \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1}$ with $\phi_1, \theta_1 \neq 0$. The shock is realized and can be learned about in period $t-1$ but only affects the optimal action in period t. Formally, the optimal action follows an ARMA(1,1) process with $\theta_0 = 0$. Propositions 1 and 2 imply that one can restrict attention to signals of the form $S_t = h_1 X_t + h_2 \varepsilon_t + \psi_t$. The next proposition states that the optimal signal weight on ε_t is different from zero. Thus the optimal signal is not on the current optimal action only. The optimal signal is on both the current optimal action and the best predictor of next period's optimal action.

Proposition 5 *If the optimal action follows an ARMA(1,1) process, $X_t = \phi_1 X_{t-1} + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1}$, one can restrict attention to signals of the form $S_t = h_1 X_t + h_2 \varepsilon_t + \psi_t$. Furthermore, $\phi_1, \theta_1 \neq 0$ and $\theta_0 = 0$ implies $h_2 \neq 0$.*

Proof. See Appendix F. ■

To see the link to a signal on the best predictor of next period's optimal action, multiply the signal $S_t = h_1 X_t + h_2 \varepsilon_t + \psi_t$ by a non-zero constant ϱ to arrive at a signal about a weighted average of the current optimal action and the best predictor of next period's optimal action. Formally, $S_t' = \varrho S_t = \omega X_t + (1-\omega)(\phi_1 X_t + \theta_1 \varepsilon_t) + \varrho \psi_t$.[17] The term $(\phi_1 X_t + \theta_1 \varepsilon_t)$ is the best predictor of X_{t+1} given full information in period t. The statement $h_2 \neq 0$ is equivalent to the statement $(1-\omega) \neq 0$. Hence, when one changes the process for the optimal action from an AR(1) process, $X_t = \phi_1 X_{t-1} + \theta_0 \varepsilon_t$, to a process where the innovation can be learned about one period in advance, $X_t = \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1}$, a signal on the current optimal action is no longer equivalent to a signal on the best predictor of next period's optimal action. As a result, one can see that the optimal signal is on *both* the current optimal action *and* the best predictor of next period's optimal action.

[17] The coefficients ϱ and ω solve $\varrho h_1 = \omega + (1-\omega)\phi_1$ and $\varrho h_2 = (1-\omega)\theta_1$. Thus, $(1-\omega) = \frac{\frac{h_2}{h_1}}{\theta_1 + (1-\phi_1)\frac{h_2}{h_1}}$.

Finally, we study the $\kappa \to \infty$ limit.

Proposition 6 *As $\kappa \to \infty$, the information capacity devoted to other components of uncertainty than to X_t approaches 0, i.e., $h' \to (1, 0, 0, \ldots)$.*

Proof. See Appendix G. ■

If the information flow is very large, it is optimal for the agent to process information mostly about the current optimal action X_t only. The reason is that for large information capacity, in each period t the prior uncertainty about X_t is much larger than uncertainty about any of the past states X_{t-s} for $s > 0$. Gains from processing information in models with rational inattention increase with the level of uncertainty, and thus paying attention to the more uncertain X_t in period t is more valuable. This holds even if information about X_{t-s} were useful in the future.

For large κ, posterior uncertainty is always almost zero, which translates into low prior uncertainty about past states in the next period. While the agent already acquired lots of information about the past states, he has not acquired any about the current shock ε_t yet, and thus uncertainty about X_t is much larger.

4.4 Extensions

We already pointed out in Section 3 that many different formulations of the information flow constraint are in fact equivalent and hence yield the same solution. In this subsection, we discuss other modifications of the decision problem formulated in Section 3 that can be relevant in economic applications: the optimal action is non-stationary, the optimal action is driven by multiple shocks, and the objective depends on multiple elements of Σ_0.

Non-stationary optimal action. In applications, the optimal action may follow a non-stationary process. So far we assumed that the optimal action, X_t, follows a stationary process. The stationarity assumption ensures that all conditional moments appearing in the proofs of Lemmas 1 and 2 and Propositions 1-6 are well-defined. Let $\Sigma_{t|t-1}$ denote the conditional variance-covariance matrix of φ_t given $\mathcal{I}_{t-1}$. Let $\Sigma_{t|t}$ denote the conditional variance-covariance matrix of φ_t given $\mathcal{I}_t$. Let Σ_1 and Σ_0 denote $\lim_{t\to\infty} \Sigma_{t|t-1}$ and $\lim_{t\to\infty} \Sigma_{t|t}$, respectively. Objective (10) and constraint (11) depend on Σ_1 and Σ_0. The assumption that the optimal action follows a stationary process ensures that Σ_1 and Σ_0 are well-defined. One can relax the stationarity assumption. Lemmas

1 and 2 and Propositions 1-5 extend to the case of a non-stationary ARMA(p,q) process for the optimal action so long as all moments appearing in the proofs of the lemmas and propositions are well-defined. This requires that the agent is given some initial information and that the parameter κ is sufficiently large.[18]

Multiple shocks. In reality, the optimal action may be driven by multiple shocks. To fix ideas, assume that the optimal action equals the sum of two components, $X_t = X_{1,t} + X_{2,t}$, where $X_{1,t}$ and $X_{2,t}$ follow independent Gaussian processes, $X_{1,t}$ follows an ARMA(p_1, q_1) process, and $X_{2,t}$ follows an ARMA(p_2, q_2) process, with p_1, p_2, q_1, q_2 finite. In the price-setting application presented at the end of Section 3, $X_{1,t}$ can be the log of the desired markup driven by shocks to the elasticity of substitution between goods and $X_{2,t}$ can be the log of marginal cost driven by productivity shocks. The agent chooses the number of signals, the content of the signals, and the variance-covariance matrix of noise in the signals. The agent's information set is given by equation (8). Each signal in the signal vector $S_t^K = (S_{t,1}, \ldots, S_{t,K})'$ can be about a different linear combination of current and past $X_{1,t}$, current and past $X_{2,t}$, and current and past innovations

$$S_t^K = A_1 X_{1,t}^M + A_2 X_{2,t}^M + B_1 \varepsilon_{1,t}^N + B_2 \varepsilon_{2,t}^N + \psi_t^K,$$

where $X_{j,t}^M = (X_{j,t}, \ldots, X_{j,t-M+1})'$ and $\varepsilon_{j,t}^N = (\varepsilon_{j,t}, \ldots, \varepsilon_{j,t-N+1})'$ for $j = 1, 2$. The agent aims to minimize the mean squared error (10), subject to the information flow constraint. In the information flow constraint, we now measure the information flow about both components of the optimal action

$$\lim_{T \to \infty} \frac{1}{T} I\left(\bar{X}_{1,0}, \bar{X}_{2,0}, X_{1,1}, X_{2,1}, \ldots, X_{1,T}, X_{2,T}; S_1^K, \ldots, S_T^K\right) \leq \kappa. \quad (30)$$

What are the properties of the optimal signal vector? The analytical results of Sections 3 and 4.1 extend in a straightforward way to the case of multiple shocks. The version of Lemma 1 for multiple shocks reads: The information flow constraint (30) is equivalent to the information flow constraint (14), where the vector φ_t is any vector with the properties: (i) $X_{1,t}^M, X_{2,t}^M, \varepsilon_{1,t}^N, \varepsilon_{2,t}^N$ can be computed from φ_t, and (ii) φ_t contains no redundant elements. The version of Proposition 1 for multiple shocks reads: Any optimal signal vector is on linear combinations of the elements of $\xi_{1,t}$ and $\xi_{2,t}$ only, where $\xi_{1,t}$ and $\xi_{2,t}$ are defined as in equation (16) for $X_{1,t}$ and $X_{2,t}$ separately. The

[18] In the AR(1) case and in the ARMA(1,q) case, the precise condition is $2^{2\kappa} > \phi_1^2$. For the precise condition in the AR(2) case, see the proof of Proposition 4.

version of Proposition 2 for multiple shocks reads: The agent can attain the optimum with $K \leq 2$. That is, the agent can restrict attention to one- and two-dimensional signals.

A new question arises: Is the optimal signal vector one-dimensional, even though the optimal action is driven by multiple shocks? Consider a simple case. Suppose that each component of the optimal action follows an AR(1) process, $X_{1,t} = \phi_{1,1} X_{1,t-1} + \theta_{1,0} \varepsilon_{1,t}$ and $X_{2,t} = \phi_{2,1} X_{2,t-1} + \theta_{2,0} \varepsilon_{2,t}$. The results stated in the previous paragraph imply that any optimal signal vector is on linear combinations of $X_{1,t}$ and $X_{2,t}$ only and that the agent can restrict attention to signals with $K \leq 2$. If $\phi_{1,1} = \phi_{2,1} \equiv \phi_1$, the optimal action follows an AR(1) process with innovation $(\theta_{1,0}\varepsilon_{1,t} + \theta_{2,0}\varepsilon_{2,t})$:

$$X_t = \phi_1 X_{t-1} + (\theta_{1,0}\varepsilon_{1,t} + \theta_{2,0}\varepsilon_{2,t}) \, .$$

In this case, it is straightforward to show that the optimal signal is of the form $S_t = hX_t + \psi_t$. That is, the optimal signal is one-dimensional and it is directly on the optimal action, $X_t = X_{1,t} + X_{2,t}$. However, if $\phi_{1,1} \neq \phi_{2,1}$, the optimal action in period t and the optimal action in period $t+1$ depend on different linear combinations of $X_{1,t}$ and $X_{2,t}$, because $X_t = X_{1,t} + X_{2,t}$ while $X_{t+1} = \phi_{1,1} X_{1,t} + \phi_{2,1} X_{2,t} + (\theta_{1,0}\varepsilon_{1,t+1} + \theta_{2,0}\varepsilon_{2,t+1})$. We studied this case numerically. We find that the optimal signal is one-dimensional, the optimal signal is on a linear combination of $X_{1,t}$ and $X_{2,t}$, and it converges to a signal on $X_t = X_{1,t} + X_{2,t}$ as the information flow parameter κ becomes large. We also studied numerically the general case where $X_{1,t}$ follows an ARMA(p_1, q_1) process and $X_{2,t}$ follows an ARMA(p_2, q_2) process. The optimal signal is one-dimensional, it is on a linear combination of the elements of $\xi_{1,t}$ and $\xi_{2,t}$ (with non-zero weights on all elements of $\xi_{1,t}$ and $\xi_{2,t}$), and the optimal signal converges to a signal on $X_t = X_{1,t} + X_{2,t}$ as the information flow parameter κ becomes large.

More general objective. Propositions 1-2 also hold for more general objectives. Proposition 1 holds for any objective that is a function only of the elements of Σ_0, while Proposition 2 holds for any objective that is a function only of the elements of Σ_0 and has the property that pure delay in the arrival of signals makes the agent worse off.

Restrictions on signal vector. Two restrictions on the signal vector (9) may in principle be binding. First, M and N are assumed to be finite. Second, the vector of signal noise, ψ_t^K, is assumed to follow a vector white noise process. However, the nature of the solution (i.e., the fact that any optimal signal vector is on linear combinations of the elements of ξ_t only) suggests that the first restriction is not binding. Furthermore, in the next section we demonstrate numerically that the solution of Sims's rational inattention problem (1)-(3) equals the solution of the rational inattention problem (7)-(11). This numerical finding suggests that the second restriction is also not binding.

5 Examples

This section presents example solutions and demonstrates numerically that the solution of Sims's rational inattention problem (1)-(3) equals the solution of our rational inattention problem (7)-(11).

Figure 1 displays an example solution of Sims's problem. Recall that in Sims's problem the agent chooses directly a process for the action, Y_t. Let us solve our problem (7)-(11) assuming the *same* finite-order MA process for the optimal action, X_t, as in Figure 1. That is, let us compute the optimal signal about X_t and the implied process for the action, $Y_t = E[X_t \mid \mathcal{I}_t]$. Figure 4 shows the implied process for the action. Compare Figure 4 with Figure 1. The two figures are identical. The solution of Sims's problem equals the solution of our problem.

Figure 5 presents two other examples. In the left column, we suppose that X_t follows an AR(3) process whose characteristic polynomial has complex roots. The top panel displays the solution for Y_t by "brute-force" optimization (Sims's problem). To obtain this solution, we compute the (infinite-order) MA representation of X_t and truncate it at a large but finite lag n. We then solve numerically for the coefficients b_s and c_s in the MA representation of Y_t (see equation (2)). The bottom panel shows the corresponding solution of our problem. Here we compute the optimal signal about X_t and the implied process for the action, $Y_t = E[X_t \mid \mathcal{I}_t]$. The two solutions are identical. The solution of Sims's problem equals the solution of our problem.

In the right column of Figure 5 we suppose that X_t follows an ARMA(2,1) process. Compare the solution of Sims's problem (top panel) with the solution of our problem (bottom panel). Again, the two solutions are identical. Note also that in each column of Figure 5, as in Figures 1 and 4, the agent responds with delay to an innovation in the optimal action and the agent's action contains noise.

We have solved many other examples. The solution of Sims's problem always equals the solution of our problem.

We think that solving a dynamic rational inattention problem via our approach has advantages.

We provide an interpretation of actions under rational inattention in a dynamic setting as actions based on optimal signals. Moreover, the optimal signals have a simple form and there is clear intuition, given in Section 4, for why the optimal signals have this particular simple form.

There are also computational advantages. With “brute-force” optimization, a truncation of the MA representation of X_t is necessary in the AR(p) case and in the ARMA(p,q) case. Moreover, the number of unknowns can be very large (in principle, $2(n+1)$ where n is a large number) implying that numerical optimization can be time-consuming. By contrast, when the solution is based on an optimal signal there is no need to approximate X_t or Y_t. The number of unknowns is equal to the number of weights in the optimal signal, and thus in many cases the number of unknowns is small. In each example in Figure 5 there are only three unknowns implying a very short computational time (less than one fifth of a second on a laptop).

6 Optimal signals in a model of price-setting

Solving for signals that are optimal under rational inattention yields new insights regarding dynamic economic behavior. Let us illustrate this point using two macroeconomic models.

In this section, we consider the model of price-setting proposed by Woodford (2002). Woodford supposes that monopolistically competitive firms set prices based on noisy signals about nominal aggregate demand. He shows that in this environment a nominal disturbance can have large and persistent real effects. His model has become a benchmark in the literature on price-setting and in the literature on business cycle models with information frictions. Woodford assumes that firms set prices based on signals of the form “nominal aggregate demand plus i.i.d. noise.” We resolve the model with signals that are optimal under rational inattention.

Model. Woodford’s model features an economy with a continuum of firms indexed by $i \in [0,1]$. Firm i sells good i. In every period, firm i sets the price of good i to maximize the present discounted value of profits. Since the firm can reset the price next period, this is equivalent to setting the price to maximize current profit. After a log-quadratic approximation to the profit function, the loss in profit in the case of a deviation of the actual price, p_{it}, from the profit-maximizing price, p_{it}^*, is proportional to $(p_{it}^* - p_{it})^2$. Therefore, a firm’s optimal price given any information set $\mathcal{I}_{it}$ is

$p_{it} = E\left[p_{it}^* \mid \mathcal{I}_{it}\right]$. The profit-maximizing price can be written as

$$p_{it}^* = \xi q_t + (1 - \xi) p_t, \tag{31}$$

where q_t is nominal aggregate demand, p_t is the aggregate price level, and $\xi \in (0, 1]$ is a parameter reflecting the degree of strategic complementarity in price setting.

Woodford assumes that in every period the decision-maker in firm i observes a signal about nominal aggregate demand given by

$$S_{it} = q_t + v_{it}, \tag{32}$$

where v_{it} is a Gaussian white noise error term, distributed independently both of the history of fundamental disturbances and of the observation errors of all other firms. The information set of the decision-maker who is setting the price in period t includes any initial information and all signals received up to this point in time, $\mathcal{I}_{it} = \mathcal{I}_{i,0} \cup \{S_{i1}, \ldots, S_{it}\}$. Nominal aggregate demand follows an exogenous stochastic process given by

$$q_t = (1 + \rho) q_{t-1} - \rho q_{t-2} + \sigma \varepsilon_t, \tag{33}$$

where $\rho \in [0, 1)$ and $\sigma > 0$ are parameters and ε_t follows a Gaussian white noise process with unit variance. The price level is given by the integral over all the individual prices

$$p_t = \int_0^1 p_{it} di, \tag{34}$$

and output y_t can be computed from the relation $y_t = q_t - p_t$.

The dynamic rational inattention problem. In the rational inattention version of the model, firms choose the signal properties to maximize expected profit subject to the information flow constraint. In period zero, the decision-maker in firm i solves

$$\min_{K,A,B,\Sigma_\psi} E\left[\sum_{t=1}^{\infty} \beta^t \left(p_{it}^* - E\left[p_{it}^* \mid \mathcal{I}_{it}\right]\right)^2\right],$$

where $\beta \in (0, 1)$ is a parameter, subject to the information flow constraint (11) with $X_t = p_{it}^*$,

$$\mathcal{I}_{it} = \mathcal{I}_{i,0} \cup \left\{S_{i1}^K, \ldots, S_{it}^K\right\},$$

and

$$S_{it}^K = A \begin{pmatrix} p_{it}^* \\ \vdots \\ p_{i,t-M+1}^* \end{pmatrix} + B \begin{pmatrix} \varepsilon_t \\ \vdots \\ \varepsilon_{t-N+1} \end{pmatrix} + \psi_{it}^K,$$

where ψ_{it}^K follows a Gaussian vector white noise process with variance-covariance matrix Σ_ψ. The decision-maker minimizes the expected discounted sum of profit losses due to suboptimal pricing. She understands that in every period $t \geq 1$ she will set the price equal to the conditional expectation of the profit-maximizing price and she will remember past signals. The signal vector in period t can be K-dimensional and can be about any linear combination of current and past values of the profit-maximizing price and current and past values of the nominal shock.

After the decision-maker has chosen the signal properties in period zero, she receives a long sequence of signals in period zero such that conditional second moments are independent of time (and hence can be evaluated using the steady-state Kalman filter). In particular, the mean squared error, $E\left[\left(p_{it}^* - E\left[p_{it}^* \mid \mathcal{I}_{it}\right]\right)^2\right]$, is constant for all $t \geq 1$. Therefore, the firms' objective simplifies to $\beta/\left(1-\beta\right)$ times $E\left[\left(p_{it}^* - E\left[p_{it}^* \mid \mathcal{I}_{it}\right]\right)^2\right]$ and is simply a monotone transformation of the objective in Section 2 (equation (10)).[19]

Price-setting behavior and real effects of nominal shocks. To see how rational inattention affects price-setting behavior, let us focus on the case when $\xi = 1$ (no strategic complementarity in price setting). The profit-maximizing price is then equal to nominal aggregate demand, $p_{it}^* = q_t$ (equation (31)). Since nominal aggregate demand follows an AR(2) process (equation (33)), the profit-maximizing price simply follows an AR(2) process. The analytical results in Section 4 imply that in this case one can restrict attention to signals of the form

$$S_{it} = h_1 q_t + h_2 q_{t-1} + \psi_{it}, \tag{35}$$

where ψ_{it} follows a Gaussian white noise process.[20] Notice that Woodford's assumption (32) amounts to a simple restriction $h_2 = 0$. In effect, Woodford assumes that the firm learns only about the current profit-maximizing price. However, Proposition 4 implies that if $\rho \neq 0$ then $h_2 \neq 0$. The optimal signal is on both the current profit-maximizing price and the best predictor of next period's profit-maximizing price. The best predictor of next period's profit-maximizing price equals $\left(1+\rho\right) q_t - \rho q_{t-1}$. Hence, the optimal signal weight on q_{t-1} is non-zero. Since the price is set based on a signal on both the current profit-maximizing price and the best predictor of next period's profit-maximizing price, price-setting behavior is forward-looking.

To investigate to what extent Woodford's restriction on the signal matters, we assume his

[19] Woodford (2002) also uses the steady-state Kalman filter to compute conditional second moments.

[20] We follow Woodford (2002) in assuming that the noise term ψ_{it} is independent across firms.

parameter values.[21] Furthermore, we suppose that the information flow in the model with optimal signals (the model with S_{it} given by equation (35)) is equal to the information flow in the Woodford model (the model with S_{it} given by equation (32)). Thus decision-makers process the same amount of information in both models. The only difference is that in one model decision-makers use optimal signals, whereas in the other model decision-makers use restricted signals.[22]

The top panel in Figure 6 compares the equilibrium impulse responses of output to a nominal disturbance in the two models. Woodford's restriction on the signal matters a lot. The real effects of a nominal disturbance are much weaker in the model with optimal signals than in the Woodford model. With optimal signals the variance of output falls by 60 percent. A decision-maker in the model with optimal signals uses a given amount of information as efficiently as possible. Consequently, the tracking of the profit-maximizing price is more accurate and a nominal shock has weaker real effects than in the Woodford model. Moreover, the difference between the two models can be sizable.

While the real effects are weaker in the model with optimal signals, they are *more persistent* than in the Woodford model. In the top panel of Figure 6, the impulse response of output to a nominal shock decays more slowly in the model with optimal signals than in the Woodford model. The shape of the impulse response function changes due to forward-looking price-setting behavior. Let us explain. First, as in Section 4.3 one can write the optimal signal (equation (35)) as a signal about a weighted average of current nominal aggregate demand and the best predictor of next period's nominal aggregate demand, $S_{it}' = \omega q_t + (1-\omega)\left((1+\rho)\,q_t - \rho q_{t-1}\right) + \varrho\psi_{it}$, where ϱ is a constant. By contrast, Woodford's signal (equation (32)) has a weight of *one* on current nominal aggregate demand, a weight of *zero* on the best predictor of next period's nominal aggregate demand, and a greater precision (the variance of the noise term in Woodford's signal is smaller than the variance of the noise term in S_{it}').[23] Second, the best predictor of next period's nominal aggregate demand assumes different values on impact of a shock and in the long run. On impact the best predictor

[21]The exception is that for the moment we set $\xi = 1$ whereas Woodford focuses on the case when $\xi = 0.15$.

[22]The information flow in the Woodford model, κ_W, can be computed from the formula $\kappa_W = (1/2)\log_2\left(\sigma_{q,1}^2/\sigma_{q,0}^2\right)$, where $\sigma_{q,1}^2$ is the prior variance of nominal aggregate demand and $\sigma_{q,0}^2$ is the posterior variance of nominal aggregate demand, in steady state. We solve for h_2 and σ_ψ^2 in the model with optimal signals by applying the analytical results from Section 4 and setting $\kappa = \kappa_W$.

[23]Recall that the optimal signal and Woodford's signal are parameterized to yield the same information flow.

of q_{t+1} equals $(1+\rho)\,q_t$. The optimal signal then has a weight on q_t of $\omega + (1-\omega)(1+\rho)$, or close to 2 in our example, larger than the weight of one in Woodford's signal.[24] With a larger weight on q_t in the signal, prices react more strongly in the rational inattention model implying that on impact the real effects are smaller than in the Woodford model. Third, as time goes to infinity the best predictor of q_{t+1} approaches q_t. Hence in the long run the optimal signal has a weight of one on q_t, the same weight as Woodford's signal. Since in addition Woodford's signal has a greater precision, eventually prices react faster in the Woodford model implying that the real effects are more persistent in the model with optimal signals.

Next, consider the case of $\xi = 0.15$ as in Woodford (2002). With $\xi \neq 1$ the profit-maximizing price depends on an endogenous variable, the price level. We guess that in equilibrium the profit-maximizing price follows an ARMA(p,q) process where $p \geq 1$ and $q \geq 0$ are integers. Given the guess, we apply the analytical results from Section 4 to establish the form of an optimal signal. For example, if the profit-maximizing price follows an ARMA(2,2) process, one can restrict attention to signals of the form

$$S_{it} = p_{it}^{*} + h_2 p_{i,t-1}^{*} + h_3 \varepsilon_t + h_4 \varepsilon_{t-1} + \psi_{it}. \tag{36}$$

We let the decision-maker in firm i choose the optimal signal weights (h_2, h_3, and h_4 in the ARMA(2,2) example) and the variance of noise σ_ψ^2 to maximize profits, subject to the information flow constraint. We then obtain the price level from the relation $p_t = \int_0^1 p_{it} di$, and we compute the profit-maximizing price using equation (31). As before, the information flow in the model with optimal signals is set equal to the information flow in the Woodford model.[25]

The bottom panel in Figure 6 compares the equilibrium impulse responses of output to a nominal disturbance in the two models in the case of $\xi = 0.15$. The real effects of a nominal disturbance are still weaker in the model with optimal signals than in the Woodford model. The reason is that the same amount of information is always used more efficiently in the former model than in the latter. However, the difference between the two models is now small. With strategic complementarity in price setting ($\xi < 1$), the response of the profit-maximizing price to a nominal shock weakens implying that the firms' tracking problem becomes easier. The Lagrange multiplier

[24] In our example $\rho = 0.9$ and $1 - \omega = 0.987$.

[25] We verify that we cannot reduce the difference between the guessed profit-maximizing price and the actual profit-maximizing price by adding parameters to the law of motion for p_{it}^{*}.

on the information flow constraint falls. With a small Lagrange multiplier on the information flow constraint, it is unimportant that the firm is receiving a suboptimal signal in the Woodford model.[26]

An implication for microeconomic price data. Carlsson and Skans (2012) test models of price-setting using firm-level data on prices and unit labor costs from Sweden. The authors find that prices depend on current *and* expected future marginal cost. They conclude that this finding is consistent with the well-known Calvo model of price-setting. Our analysis makes it clear that this finding is also consistent with rational inattention. Think of the profit-maximizing price, p_{it}^*, as proportional to marginal cost. Our analysis implies that, in general, firms subject to rational inattention are learning about both current and expected future marginal cost. Therefore, prices set under rational inattention depend on both current and expected future marginal cost.

Conclusion. Price-setting behavior under rational inattention is forward-looking. Therefore, a model of price-setting under rational inattention makes different predictions, in particular about the real effects of nominal shocks, than a model of price-setting with exogenous imperfect information such as the model of Woodford (2002).

7 Optimal signals in a business cycle model with news shocks

As another application, in this section we consider a simple business cycle model with news shocks. Here by a "news shock" we mean a change in productivity that one can learn about before it actually occurs in production.[27] While news shocks seem a plausible source of the business cycle, it has proven difficult to construct models in which the business cycle is driven by news shocks. The key problem is that good news about future productivity makes agents wealthier and, in a neoclassical environment, this wealth effect increases both consumption *and* leisure, reducing labor input through a reduction in labor supply. With capital predetermined and current productivity unchanged, the decrease in labor input pushes output *down*.[28]

[26] One can introduce strategic substitutability in price-setting ($\xi > 1$). The difference between the two models becomes larger, rather than smaller, compared with the case of $\xi = 1$.

[27] For simplicity, in this section we write "news shocks" instead of "news shocks about productivity."

[28] The most popular model generating a business cycle expansion in response to a positive news shock is Jaimovich and Rebelo (2009). The model has three key elements: preferences that allow the modeler to parameterize the strength of short-run wealth effects on the labor supply, variable capital utilization, and adjustment costs to investment. See Lorenzoni (2011) for a review of the literature on news shocks.

Rational inattention is a force pushing labor input *up* after a positive news shock, because rationally inattentive firms *choose* not to distinguish carefully between current and future increases in productivity and thus a news shock causes an increase in labor demand. We illustrate this point in a simple model in which firms make a labor-hiring decision subject to rational inattention and households are hand-to-mouth consumers. In the model, rational inattention causes labor input and output to rise following a positive news shock. To the best of our knowledge, no one has proposed this explanation before.

Model. There is a continuum of firms indexed by $i \in [0,1]$. All firms produce the same good using an identical technology represented by the production function

$$Y_{it} = e^{z_t} L^{\alpha} N_{it}^{1-\alpha},$$

where Y_{it} is output of firm i in period t, N_{it} is labor input, and $\alpha \in (0,1)$ is a parameter. The owner of the firm provides an entrepreneurial input L and chooses labor input in every period. Productivity follows the process

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

with $\rho \in (0,1)$, $\sigma > 0$, $\varepsilon_t \sim iid\ N(0,1)$, and $k > 0$. The fact that the productivity shock has a subscript $t-k$ means that a productivity shock drawn in period $t-k$ affects actual productivity with a k period delay. As a result, one can learn about productivity changes k periods in advance. The productivity shock is also called a news shock.

There is a representative household. In every period, the household chooses labor supply so as to maximize period utility

$$\frac{C_t^{1-\gamma} - 1}{1-\gamma} - \frac{N_t^{1+\psi}}{1+\psi},$$

subject to the budget constraint

$$C_t = W_t N_t,$$

where C_t is consumption, W_t is the real wage, and N_t is labor supply. The preference parameters satisfy $\gamma > 0$ and $\psi \geq 0$. For simplicity, we assume that the household cannot save. This assumption can be relaxed. See the discussion below.

In every period, each entrepreneur makes the hiring decision under rational inattention. The representative household makes the labor supply decision under perfect information. The labor market is perfectly competitive, i.e., entrepreneurs and the representative household take the real wage as given. The real wage adjusts so as to equate labor supply and labor demand

$$N_t = \int_0^1 N_{it} di.$$

The dynamic rational inattention problem. Profits of firm i in period t equal

$$e^{z_t} L^\alpha N_{it}^{1-\alpha} - W_t N_{it}.$$

The profit-maximizing labor input of firm i in period t is given by

$$N_{it}^* = \left[\frac{W_t}{(1-\alpha) e^{z_t} L^\alpha}\right]^{-\frac{1}{\alpha}}.$$

Taking logs and letting small letters denote logs of capital letters yields

$$n_{it}^* = \frac{1}{\alpha} \ln\left[(1-\alpha) L^\alpha\right] + \frac{1}{\alpha}(z_t - w_t).$$

After a log-quadratic approximation to the profit function, the profit loss in the case of a deviation of the actual labor input, n_{it}, from the profit-maximizing labor input, n_{it}^*, is proportional to $(n_{it}^* - n_{it})^2$ and a firm's optimal hiring decision given any information set $\mathcal{I}_{it}$ is $n_{it} = E[n_{it}^* \mid \mathcal{I}_{it}]$.

The dynamic rational inattention problem of the entrepreneur is a special case of the problem from Section 3. The entrepreneur tracks an optimal action, n_{it}^*, that follows a serially correlated Gaussian process. The entrepreneur chooses how to pay attention to n_{it}^*. The signals in period t can be about any linear combination of current and past n_{it}^* and current and past ε_t. The entrepreneur remembers all signals and aims to minimize the mean squared error, $E\left[(n_{it}^* - E[n_{it}^* \mid \mathcal{I}_{it}])^2\right]$.[29]

We now show that a positive news shock raises output in the initial periods under rational inattention, even though it has no effect on output in the initial periods under perfect information.

No output response on impact under perfect information. As a benchmark, we first present the equilibrium under perfect information. For ease of exposition, assume $\frac{1}{\alpha}\ln\left[(1-\alpha) L^\alpha\right] = 0$ and consider the log-linearized labor market clearing condition[30]

$$n_t = \int_0^1 n_{it} di.$$

[29] Recall that one can think of the agent as choosing the properties of the signal in an initial period so as to minimize the discounted sum of future mean squared errors. After the agent has chosen the properties of the signal, the agent receives a long sequence of signals such that the mean squared error is the same in every period and thus the discounted sum of mean squared errors is proportional to the mean squared error in any period. See Section 3.

[30] The original labor market clearing condition yields the same equilibrium under perfect information.

Under perfect information, the household chooses the utility-maximizing labor supply, all entrepreneurs choose the profit-maximizing labor input, and the labor market clearing condition reads

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

Thus, the market clearing wage equals

$$w_t = \frac{\frac{1}{\alpha}}{\frac{1-\gamma}{\psi+\gamma}+\frac{1}{\alpha}} z_t \equiv \zeta z_t,$$

and the equilibrium labor input equals

$$n_t = \frac{1}{\alpha}(1-\zeta) z_t.$$

A positive news shock has no effect on labor input and output until productivity actually increases, because there is no reason for firms to hire more labor before productivity actually increases. A news shock increases output with a k period delay.

An expansion on impact under rational inattention. To develop intuition for the implications of rational inattention, imagine for the moment that a measure zero of firms are subject to rational inattention and all other firms have perfect information. Since all other firms have perfect information, the market clearing wage still equals $w_t = \zeta z_t$, and thus the profit-maximizing labor input equals $n_{it}^* = \frac{1}{\alpha}(1-\zeta) z_t$. Furthermore, suppose that one can learn about innovations in productivity one period in advance, and thus productivity follows an ARMA(1,1) process $z_t = \phi_1 z_{t-1} + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1}$ with $\phi_1 = \rho$, $\theta_0 = 0$, and $\theta_1 = \sigma$. Propositions 1 and 2 (and the fact that n_{it}^* is proportional to z_t) imply that an entrepreneur subject to rational inattention can restrict attention to signals of the form

$$S_{it} = h_1 z_t + h_2 \varepsilon_t + \psi_{it},$$

where ψ_{it} is a noise term that follows a Gaussian white noise process. Proposition 5 implies that $h_2 \neq 0$. The optimal signal is on both current productivity and the best predictor of next period's productivity. The best predictor of next period's productivity equals $\rho z_t + \theta_1 \varepsilon_t$. Hence, the optimal signal weight on the news shock, ε_t, is non-zero. On impact of a positive news shock, $\varepsilon_t > 0$, the signal increases. (Think of the signal as analogous to a stock price that moves in response to innovations in both current productivity and expected future productivity.) Since the entrepreneur has chosen not to distinguish carefully between changes in current productivity and changes in expected future productivity, the entrepreneur starts hiring already today.

The optimal signal has very different implications for actions than a signal of the form "current productivity plus i.i.d. noise" or a signal of the form "the news shock plus i.i.d. noise." With either of those alternative signals, the entrepreneur does not hire more labor on impact of a positive news shock. The entrepreneur hires more labor only when productivity actually increases (in the following period in this example).

Solving the model in the case when all firms are subject to rational inattention is slightly more complicated, because the market clearing real wage is no longer simply equal to $w_t = \zeta z_t$ and thus the profit-maximizing labor input that entrepreneurs are tracking is no longer simply equal to $n_{it}^* = \frac{1}{\alpha}(1-\zeta) z_t$. In general, the profit-maximizing labor input depends on the real wage, which depends on the behavior of other firms. Formally, substituting the household's optimality condition, $w_t = \frac{\psi+\gamma}{1-\gamma} n_t$, and the labor market clearing condition, $n_t = \int_0^1 n_{it} di$, into the equation for the profit-maximizing labor input, $n_{it}^* = \frac{1}{\alpha}(z_t - w_t)$, yields

$$n_{it}^* = \frac{1}{\alpha} z_t - \frac{1}{\alpha} \frac{\psi+\gamma}{1-\gamma} \int_0^1 n_{it} di.$$

With $\gamma < 1$ actions of different firms are strategic substitutes. The profit-maximizing labor input is a decreasing function of the real wage, which is an increasing function of the aggregate labor input. To solve for the equilibrium law of motion for the profit-maximizing labor input, we employ a guess and verify method. We guess that the profit-maximizing labor input follows an ARMA(p,q) process. Given the guess, we apply the analytical results from Section 4 to establish the form of an optimal signal and we compute the optimal signal weights and the implied actions. We then calculate the actual law of motion for the profit-maximizing labor input from the last equation. If the actual law of motion for the profit-maximizing labor input differs from our guess, we update the guess until a fixed point is reached.

Figure 7 plots the equilibrium impulse response of aggregate labor input to a news shock assuming $\gamma = 1/3$, $\psi = 0$, $\alpha = 3/4$, $\rho = 0.9$, $\sigma = 0.01$, and $k = 8$.[31] Labor input rises coincident with a positive innovation in ε_t, and a boom develops before productivity actually rises. Positive news shocks produce an expansion, because rationally inattentive firms choose not to distinguish carefully between current and future increases in productivity and thus a positive news shock causes an increase in labor demand. For comparison, Figure 7 also shows the equilibrium labor input in the

[31] We use a value of the information-processing parameter κ such that the equilibrium per period profit loss from rational inattention, expressed as a fraction of the steady-state wage bill, is equal to 0.0001.

case when entrepreneurs in all firms have perfect information. In that case, a productivity shock drawn in period one affects labor input only in period nine.

Discussion. One can relax the model's simplifying assumptions. For example, if one supposed that the representative household can save and has preferences as in Greenwood, Hercowitz, and Huffman (1988), the solution of the model would be identical for particular parameter values. Furthermore, with standard preferences and variable capital, a news shock would have additional effects (the reduction in labor supply due to the wealth effect after a positive news shock, and the fall in investment to finance the rise in consumption). It is a quantitative question whether the effect of rational inattention identified here would be strong enough to produce an increase in hours worked on impact of a positive news shock in that standard neoclassical setup.

Conclusion. With firms subject to rational inattention, a positive news shock about productivity increases labor demand on impact. This new effect will help models produce business cycle comovement in response to news shocks.

8 Implications for expectations data

In this section, we discuss implications of the optimal rational inattention signal for expectations data. The main takeaway is that the optimal rational inattention signal generates a combination of backward-looking expectations and forward-looking expectations.

The optimal rational inattention signal generates backward-looking expectations, because the noise in the signal implies that the agent puts weight on the prior. Consider the AR(1) case as an example. In this case, the variable being tracked, X_t, follows an AR(1) process, the optimal rational inattention signal is $S_t = X_t + \psi_t$ (Proposition 3), and the standard Kalman filter equations imply that the conditional expectation of X_{t+1} given information in period t equals

$$E\left[X_{t+1}|\mathcal{I}_t\right] = \phi_1 E\left[X_t|\mathcal{I}_t\right] = \phi_1\left[E\left[X_t|\mathcal{I}_{t-1}\right] + K\left(X_t + \psi_t - E\left[X_t|\mathcal{I}_{t-1}\right]\right)\right],$$

where K is the weight on the period t signal. Note that the period t expectation of the variable in period $t+1$ is a function of the period t signal and the period $t-1$ expectation of the variable in period t. The noise in the signal implies that $K < 1$ and thus the agent puts weight on the prior.

Muth (1960) made a similar observation by pointing out that an optimal forecast can look like "adaptive expectations." In his model, an agent forecasts next period's value of a variable of interest,

Z_{t+1}, based on observations of current and past values of the variable, Z_t, Z_{t-1}, Z_{t-2}, ... The variable of interest, Z_t, equals the sum of two components – a permanent component, X_t, following a random walk and a transitory component, ψ_t, following a white noise process. The optimal forecast of the permanent component is given by the previous equation with $\phi_1 = 1$. Noting that the forecast of the variable equals the forecast of the permanent component, $E[Z_{t+1}|\mathcal{I}_t] = E[X_{t+1}|\mathcal{I}_t]$, and using that the variable equals the sum of the permanent and the transitory component, $Z_t = X_t + \psi_t$, yields

$$E[Z_{t+1}|\mathcal{I}_t] = E[Z_t|\mathcal{I}_{t-1}] + K(Z_t - E[Z_t|\mathcal{I}_{t-1}]).$$

Muth (1960) pointed out that the last equation is identical to the "adaptive expectations" equation assumed by Cagan (1956) and many others

$$Z^e_{t+1} = Z^e_t + \beta(Z_t - Z^e_t),$$

where Z^e_t is the time $t-1$ expectation of Z_t and β is an exogenous coefficient. Of course, Muth (1960) also emphasized that the K in the optimal forecast is endogenous, whereas the β in the "adaptive expectations" equation is exogenous.

In the rational inattention model, the variable ψ_t is the noise in the optimal signal rather than the transitory component of the variable of interest. Moreover, beyond the AR(1) case, the optimal rational inattention signal is on the current optimal action and the best predictors of future optimal actions (see Section 4). Hence, beyond the AR(1) case, the rational inattention model generates *a combination of backward-looking expectations* (due to the noise in the signal and the resulting weight on the prior, see the previous paragraphs) and *forward-looking expectations* (due to the fact that the optimal signal is also on the best predictors of future optimal actions).

9 Conclusions

We have characterized analytically the optimal signals for dynamic rational inattention problems as in Sims (2003). The optimal signals have a simple, intuitive form. The analytical results make it clear that information choice and actions under rational inattention in a dynamic setting are forward-looking.

With this paper's results at hand, it is straightforward to solve for optimal signals in economic applications. With firms subject to rational inattention, a positive news shock increases labor demand on impact, which helps a model produce business cycle comovement in response to news shocks. Prices depend on current and expected future marginal cost, and the real effects of nominal disturbances are weaker but more persistent than in a benchmark model with exogenous imperfect information.

In this paper, we have focused on dynamic rational inattention problems in which the objective is quadratic, the optimal action follows a Gaussian process, and the agent makes a date-zero information choice. It would be interesting to investigate whether this paper's results also hold in settings with a non-quadratic objective, non-Gaussian shocks, and sequential information choice. We do find the setup with date-zero information choice quite realistic, because it captures the idea that agents think infrequently about the optimal allocation of attention and then allocate their attention accordingly.

Acknowledgments

We thank Laura Veldkamp, two anonymous referees, as well as seminar and conference participants at CEU Budapest, Cornell University, EEA 2015, Federal Reserve Bank of New York, SED 2015, and the 2016 conference on Rational Inattention and Related Theories in Prague for helpful comments. An earlier February 2015 version of this paper was called "Analytical Results for Dynamic Rational Inattention Problems" and an earlier September 2016 version of this paper was called "The Rational Inattention Filter." The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the European Central Bank. This research was funded by GA CR grant P402-14-30724S.

A Proof of Lemma 1

On the left-hand side of the information flow constraint (11), the sequence of optimal actions can be replaced by the sequence of innovations to the optimal action, because $\left(\bar{X}_0, X_1, \ldots, X_T\right)$ can be computed from $\left(\bar{X}_0, \varepsilon_1, \ldots, \varepsilon_T\right)$ and vice versa

$$\lim_{T\to\infty}\frac{1}{T}I\left(\bar{X}_0, X_1, \ldots, X_T; S_1^K, \ldots, S_T^K\right) = \lim_{T\to\infty}\frac{1}{T}I\left(\bar{X}_0, \varepsilon_1, \ldots, \varepsilon_T; S_1^K, \ldots, S_T^K\right).$$

Let $\varepsilon^t = \{\varepsilon_1, \ldots, \varepsilon_t\}$ denote the history of innovations to the optimal action at time t. Let $S^{K,t} = \left\{S_1^K, \ldots, S_t^K\right\}$ denote the history of signals at time t. We now show that

$$\lim_{T\to\infty}\frac{1}{T}I\left(\bar{X}_0, \varepsilon^T; S^{K,T}\right) = \lim_{T\to\infty}\left[H\left(\varphi_T|S^{K,T-1}\right) - H\left(\varphi_T|S^{K,T}\right)\right],$$

where the vector φ_T is defined in Lemma 1.

The mutual information between two random vectors is symmetric and equals the difference between entropy and conditional entropy. Thus

$$I\left(\bar{X}_0, \varepsilon^T; S^{K,T}\right) = I\left(S^{K,T}; \bar{X}_0, \varepsilon^T\right) = H\left(S^{K,T}\right) - H\left(S^{K,T}|\bar{X}_0, \varepsilon^T\right).$$

The chain rule for entropy implies that, $\forall \tau \geq 2$ and $\forall T \geq \tau$,

$$H\left(S^{K,T}\right) = H\left(S^{K,\tau-1}\right) + \sum_{t=\tau}^{T} H\left(S_t^K|S^{K,t-1}\right),$$

and

$$H\left(S^{K,T}|\bar{X}_0, \varepsilon^T\right) = H\left(S^{K,\tau-1}|\bar{X}_0, \varepsilon^T\right) + \sum_{t=\tau}^{T} H\left(S_t^K|S^{K,t-1}, \bar{X}_0, \varepsilon^T\right).$$

We now rewrite the last term on the right-hand side of the last equation. The signal S_t^K depends only on X_t^M, ε_t^N, and ψ_t^K for given A and B. One can compute X_t^M and ε_t^N from the vector φ_t defined in Lemma 1, implying $H\left(S_t^K|S^{K,t-1}, \varphi_t\right) = H\left(\psi_t^K\right)$. In the following, let $\tau = \max\{M, N, 2\}$. For $t \in \{\tau, \ldots, T\}$, one can also compute X_t^M and ε_t^N from $\bar{X}_0$ and ε^T. Hence, for $t \in \{\tau, \ldots, T\}$,

$$H\left(S_t^K|S^{K,t-1}, \bar{X}_0, \varepsilon^T\right) = H\left(\psi_t^K\right) = H\left(S_t^K|S^{K,t-1}, \varphi_t\right).$$

Collecting results yields that, $\forall T \geq \tau$,

$$I\left(\bar{X}_0, \varepsilon^T; S^{K,T}\right) = H\left(S^{K,\tau-1}\right) - H\left(S^{K,\tau-1}|\bar{X}_0, \varepsilon^T\right) + \sum_{t=\tau}^{T}\left[H\left(S_t^K|S^{K,t-1}\right) - H\left(S_t^K|S^{K,t-1}, \varphi_t\right)\right].$$

Next, it follows from the definition and the symmetry of mutual information that

$$H\left(S_t^K|S^{K,t-1}\right)-H\left(S_t^K|S^{K,t-1},\varphi_t\right)=I\left(S_t^K;\varphi_t|S^{K,t-1}\right)=H\left(\varphi_t|S^{K,t-1}\right)-H\left(\varphi_t|S^{K,t-1},S_t^K\right).$$

Collecting results yields that, $\forall T\geq\tau$,

$$I\left(\bar{X}_0,\varepsilon^T;S^{K,T}\right)=H\left(S^{K,\tau-1}\right)-H\left(S^{K,\tau-1}|\bar{X}_0,\varepsilon^T\right)+\sum_{t=\tau}^{T}\left[H\left(\varphi_t|S^{K,t-1}\right)-H\left(\varphi_t|S^{K,t}\right)\right]. \quad (37)$$

Finally, we show in Appendix B that the following limit exists

$$\lim_{T\to\infty}\left[H\left(\varphi_T|S^{K,T-1}\right)-H\left(\varphi_T|S^{K,T}\right)\right].$$

Equation (37) and Cesaro mean then imply

$$\lim_{T\to\infty}\frac{1}{T}I\left(\bar{X}_0,\varepsilon^T;S^{K,T}\right)=\lim_{T\to\infty}\left[H\left(\varphi_T|S^{K,T-1}\right)-H\left(\varphi_T|S^{K,T}\right)\right]. \quad (38)$$

B Proof of Proposition 1

We now show that if a signal vector does not have the property stated in Proposition 1, then there exists another signal vector that yields the same value of the objective with strictly less information flow.

First, the signal (9) has the following state-space representation

$$\varphi_{t+1} = F\varphi_t+v_{t+1}, \quad (39)$$

$$S_t^K = G'\varphi_t+\psi_t^K, \quad (40)$$

where the vector φ_t is defined as

$$\varphi_t=\begin{cases}\left(X_t,\ldots,X_{t-\max\{M,N+p\}+1}\right)' & \text{if } q=0\\ \left(\varepsilon_t,\ldots,\varepsilon_{t-q+1},X_t,\ldots,X_{t-\max\{M,N+p-q\}+1}\right)' & \text{if } q\geq 1\end{cases}.$$

The matrix F is a square matrix and the length of the column vector v_{t+1} equals the length of the column vector φ_{t+1}. In the AR(p) case and the white noise case (i.e., $q=0$), the first element of v_{t+1} equals $\theta_0\varepsilon_{t+1}$ and the first row of the matrix F ensures that the first row of equation (39) equals the law of motion for X_t. The next $\max\{M,N+p\}-1$ elements of v_{t+1} equal zero and the next $\max\{M,N+p\}-1$ rows of the matrix F have a one just left of the main diagonal and zeros everywhere else. In the ARMA(p,q) case and the MA(q) case with $q \geq 1$ (i.e., $q \geq 1$), the vector v_{t+1} has q additional elements on top and the matrix F has q additional rows on top. The first additional element of v_{t+1} equals ε_{t+1} and the remaining additional elements of v_{t+1} equal zero. The first additional row of the matrix F contains only zeros and the remaining additional rows of the matrix F have a one just left of the main diagonal and zeros everywhere else. Finally, the matrix G is the matrix for which equation (40) equals equation (9). Such a matrix exists because X_t^M and ε_t^N can be computed from φ_t.

Second, let $\Sigma_{t|t-1}$ denote the conditional variance-covariance matrix of φ_t given $S^{K,t-1}$ and let $\Sigma_{t|t}$ denote the conditional variance-covariance matrix of φ_t given $S^{K,t}$. Furthermore, let Σ_1 and Σ_0 denote $\lim_{t\to\infty}\Sigma_{t|t-1}$ and $\lim_{t\to\infty}\Sigma_{t|t}$, respectively. Recall that X_t follows a stationary process. It follows from Propositions 13.1-13.2 in Hamilton (1994) that $\lim_{t\to\infty}\Sigma_{t|t-1}$ and $\lim_{t\to\infty}\Sigma_{t|t}$ exist and are given by

$$\begin{aligned}\Sigma_1 &= F\left[\Sigma_1 - \Sigma_1 G\left(G'\Sigma_1 G + R\right)^{-1} G'\Sigma_1\right]F' + Q,\\ \Sigma_0 &= \Sigma_1 - \Sigma_1 G\left(G'\Sigma_1 G + R\right)^{-1} G'\Sigma_1,\end{aligned}$$

where Q denotes the variance-covariance matrix of the innovation in the state equation (39) and R denotes the variance-covariance matrix of the innovation in the observation equation (40).

Third, one can express the information flow constraint (11) in terms of the matrices Σ_1 and Σ_0. According to Lemma 1 the information flow constraint (11) is equivalent to

$$\lim_{T\to\infty}\left[H\left(\varphi_T|S^{K,T-1}\right) - H\left(\varphi_T|S^{K,T}\right)\right] \leq \kappa. \tag{41}$$

Conditional normality implies that

$$H\left(\varphi_T|S^{K,T-1}\right) - H\left(\varphi_T|S^{K,T}\right) = \frac{1}{2}\log_2\left(\frac{\det\Sigma_{T|T-1}}{\det\Sigma_{T|T}}\right).$$

Hence, the information flow constraint (11) is equivalent to

$$\frac{1}{2}\log_2\left(\frac{\det\Sigma_1}{\det\Sigma_0}\right) \leq \kappa. \tag{42}$$

Fourth, let us split the vector φ_t into two sub-vectors, denoted ξ_t and χ_t. The vector ξ_t is defined as

$$\xi_t = \begin{cases} X_t & \text{if } p = 0 \text{ and } q = 0 \\ \left(X_t, \ldots, X_{t-(p-1)}\right)' & \text{if } p \geq 1 \text{ and } q = 0 \\ \left(\varepsilon_t, \ldots, \varepsilon_{t-(q-1)}, X_t\right)' & \text{if } p = 0 \text{ and } q \geq 1 \\ \left(\varepsilon_t, \ldots, \varepsilon_{t-(q-1)}, X_t, \ldots, X_{t-(p-1)}\right)' & \text{if } p \geq 1 \text{ and } q \geq 1 \end{cases}.$$

The vector χ_t contains the remaining elements of φ_t. Note that the conditional variance-covariance matrix of ξ_t given $S^{K,t}$ is the upper-left $(d \times d)$ sub-matrix of $\Sigma_{t|t}$, where d is the number of elements of the vector ξ_t. Furthermore, note that the objective (10) is simply an element of the upper-left $(d \times d)$ sub-matrix of Σ_0. Moreover, the information flow constraint (41) can be written as

$$\lim_{T\to\infty} \left[H\left(\xi_T|S^{K,T-1}\right) - H\left(\xi_T|S^{K,T}\right) + H\left(\chi_T|S^{K,T-1}, \xi_T\right) - H\left(\chi_T|S^{K,T}, \xi_T\right)\right] \leq \kappa.$$

Next, compare two signal vectors with $M > p$ and $N > q$ that yield the same upper-left $(d \times d)$ sub-matrix of Σ_0. One signal vector has the property

$$\begin{matrix} \forall j > p : A_{ij} = 0 \\ \forall j > q : B_{ij} = 0 \end{matrix}, \tag{43}$$

while the other signal vector does not have this property. Both signal vectors imply the same value of the objective (10), because the objective is an element of the upper-left $(d \times d)$ sub-matrix of Σ_0. Furthermore, both signal vectors generate the same limit

$$\lim_{T\to\infty} \left[H\left(\xi_T|S^{K,T-1}\right) - H\left(\xi_T|S^{K,T}\right)\right],$$

because both signal vectors yield the same upper-left $(d \times d)$ sub-matrix of Σ_0 by assumption and they also generate the same upper-left $(d \times d)$ sub-matrix of Σ_1, because X_t follows an ARMA(p,q) process. Finally, the difference

$$H\left(\chi_T|S^{K,T-1}, \xi_T\right) - H\left(\chi_T|S^{K,T}, \xi_T\right)$$

is non-negative, since conditioning weakly reduces entropy, and it equals zero if and only if condition (43) is satisfied. Hence, both signal vectors imply the same value of the objective, but the second signal vector is associated with strictly more information flow.

Fifth, we show that for any signal vector violating condition (43) there exists a signal vector satisfying condition (43) that yields the same upper-left $(d \times d)$ sub-matrix of Σ_0. For any variance-covariance matrices $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_0$, there exists a signal generating the posterior variance-covariance matrix $\tilde{\Sigma}_0$ from the prior variance-covariance matrix $\tilde{\Sigma}_1$ if and only if $\tilde{\Sigma}_1 - \tilde{\Sigma}_0$ is positive semi-definite. Consider a signal $\left\{K, \hat{A}, \hat{B}, \hat{\Sigma}_\psi\right\}$ violating condition (43) that yields the variance-covariance matrices Σ_1 and Σ_0. Since Σ_0 is generated from Σ_1 by the signal, then $\Sigma_1 - \Sigma_0$ must be positive semi-definite. By Sylvester's criterion (Bazaraa et al., 2013), the upper-left $(d \times d)$ sub-matrix of $\Sigma_1 - \Sigma_0$ is positive semi-definite, too. Using the statement above, this implies that there exists a signal $\{K, A, B, \Sigma_\psi\}$ satisfying condition (43) that generates the upper-left $(d \times d)$ sub-matrix of Σ_0 from the upper-left $(d \times d)$ sub-matrix of Σ_1.

C Proof of Proposition 2

The proof consists of two steps. Recall that the vector ξ_t is defined in equation (16). In the first step, we show that without loss in generality one can restrict attention to signal vectors $S_t^K = G'\xi_t + \psi_t^K$ with lower triangular matrix G and diagonal, positive definite precision matrix Σ_ψ^{-1}. In the second step, we show that any optimal signal vector of this form has the property that all rows of G' apart from the first one contain only zeros.

First, without loss in generality one can restrict attention to signal vectors $\tilde{S}_t^K = \xi_t + \tilde{\psi}_t^K$ with positive semi-definite precision matrix $\Sigma_{\tilde{\psi}}^{-1}$. Bayesian updating implies

$$\Sigma_0^{-1} = \Sigma_{\tilde{\psi}}^{-1} + \Sigma_1^{-1},$$

where Σ_1 is the prior variance-covariance matrix of ξ_t and Σ_0 is the posterior variance-covariance matrix of ξ_t. Objective (10) and constraint (11) depend only on Σ_1 and Σ_0. The objective is an element of Σ_0, and the information flow is given by the ratio of determinants of Σ_1 and Σ_0. Therefore, it suffices to show that for any Σ_1 and Σ_0 such that $\Sigma_1 - \Sigma_0$ is positive semi-definite, there exists a positive semi-definite precision matrix $\Sigma_{\tilde{\psi}}^{-1}$ such that

$$\Sigma_{\tilde{\psi}}^{-1} = \Sigma_0^{-1} - \Sigma_1^{-1}.$$

Note that if $\Sigma_1 - \Sigma_0$ is positive semi-definite, then so is $\Sigma_0^{-1} - \Sigma_1^{-1}$. Signals of the given form thus suffice to reproduce any feasible Σ_1 and Σ_0.

Next, for any signal vector $\tilde{S}_t^K = \xi_t + \tilde{\psi}_t^K$ with positive semi-definite precision matrix $\Sigma_{\tilde{\psi}}^{-1}$, there exists a signal vector $S_t^K = G'\xi_t + \psi_t^K$ with lower triangular matrix G and diagonal, positive definite precision matrix Σ_{ψ}^{-1} that contains the same information. In the case of a positive definite precision matrix $\Sigma_{\tilde{\psi}}^{-1}$, the triangular factorization of $\Sigma_{\tilde{\psi}}^{-1}$ implies that

$$\Sigma_{\tilde{\psi}}^{-1} = LDL',$$

where L is a lower triangular matrix with ones along the principal diagonal and D is a diagonal matrix with $D_{ii} > 0$ for all i. The matrix L' is invertible. Multiplying the original signal vector $\tilde{S}_t^K$ by L' yields the new signal vector $S_t^K = L'\xi_t + \psi_t^K$ with diagonal precision matrix D, and multiplying the new signal vector S_t^K by L'^{-1} recovers the original signal vector $\tilde{S}_t^K$. Hence, the two signal vectors contain the same information. When $\Sigma_{\tilde{\psi}}^{-1}$ is not positive definite, some signals have zero precision. In this case, one can define a new signal vector, which contains only the signals with positive precision, construct a new signal vector as before, and note that the new signal vector is again of the form $S_t^K = G'\xi_t + \psi_t^K$, where now some rows of G contain only zeros.

Second, thus far we have shown that without loss in generality one can restrict attention to signal vectors $S_t^K = G'\xi_t + \psi_t^K$ with lower triangular matrix G and diagonal, positive definite precision matrix Σ_{ψ}^{-1}. We now show that any optimal signal vector of this form has the property that all rows of G' apart from the first row contain only zeros.

In the case when X_t follows an AR(1) process or a white noise process, we have $\xi_t = X_t$ and $S_t^K = G'\xi_t + \psi_t^K$ is a one-dimensional signal. Next, consider the case when X_t follows an AR(p) process with $p \geq 1$. In this case, we have $\xi_t = \left(X_t, \ldots, X_{t-(p-1)}\right)'$ and the first signal is on a linear combination of $X_t, \ldots, X_{t-(p-1)}$, the second signal is on a linear combination of $X_{t-1}, \ldots, X_{t-(p-1)}$, the third signal is on a linear combination of $X_{t-2}, \ldots, X_{t-(p-1)}$, and the pth signal is on $X_{t-(p-1)}$. Note that all signals apart from the first signal are only about the past. We now show that any optimal signal vector of the form $S_t^K = G'\xi_t + \psi_t^K$ must have the property that all rows of G' apart from the first row contain only zeros. Suppose that the second row of G' contained a non-zero element. Generate a new matrix $\tilde{G}'$ by shifting the elements of the second row of the original matrix G' to the left. In words, the signal on $X_{t-1}, \ldots, X_{t-(p-1)}$ is replaced by a signal on $X_t, \ldots, X_{t-(p-2)}$ in every period. The only change in the history of signals $S^{K,T} = \left\{S_1^K, \ldots, S_T^K\right\}$ is that the signal on $X_T, \ldots, X_{T-(p-2)}$ is added and the signal on $X_0, \ldots, X_{-p+2}$ is lost. This change in the matrix G' reduces the value of the loss function without affecting information flow. The loss function is the limit as $T \to \infty$ of the conditional variance of X_T given $S^{K,T}$, $\lim_{T\to\infty} Var\left(X_T | S^{K,T}\right)$. The loss of the signal on $X_0, \ldots, X_{-p+2}$ does not affect the value of this loss function, while the addition of the signal on $X_T, \ldots, X_{T-(p-2)}$ reduces the value of this loss function. The information flow equals

$$\lim_{T\to\infty}\left[H\left(X_T,\ldots,X_{T-(p-1)}|S^{K,T-1}\right)-H\left(X_T,\ldots,X_{T-(p-1)}|S^{K,T}\right)\right].$$

Using the chain rule for entropy the information flow can be written as

$$\lim_{T\to\infty}\left[\begin{array}{c} H\left(X_{T-1},\ldots,X_{T-(p-1)}|S^{K,T-1}\right)+H\left(X_T|X_{T-1},\ldots,X_{T-(p-1)},S^{K,T-1}\right) \\ -H\left(X_T,\ldots,X_{T-(p-2)}|S^{K,T}\right)-H\left(X_{T-(p-1)}|X_T,\ldots,X_{T-(p-2)},S^{K,T}\right)\end{array}\right].$$

The limit of the first term equals the limit of the third term, and thus information flow equals

$$\lim_{T\to\infty}\left[H\left(X_T|X_{T-1},\ldots,X_{T-(p-1)},S^{K,T-1}\right)-H\left(X_{T-(p-1)}|X_T,\ldots,X_{T-(p-2)},S^{K,T}\right)\right].$$

Recall that the only change in the history of signals $S^{K,T}=\left\{S_1^K,\ldots,S_T^K\right\}$ is that the signal on $X_T,\ldots,X_{T-(p-2)}$ is added and the signal on $X_0,\ldots,X_{-p+2}$ is lost. The addition of the signal on $X_T,\ldots,X_{T-(p-2)}$ to $S^{K,T}$ does not change the second term. Similarly, the addition of the signal on $X_{T-1},\ldots,X_{T-(p-1)}$ to $S^{K,T-1}$ does not change the first term. Hence, information flow remains unchanged. The same argument can be repeated for the third row to the pth row of the matrix G'. It follows that any optimal signal vector of the form $S_t^K=G'\xi_t+\psi_t^K$ must have the property that the second row to the pth row of the matrix G' contain only zeros.

When the optimal action follows an MA(q) process with $q\geq 1$, the proof needs to be modified slightly, because the vector $\xi_t=\left(X_t,\varepsilon_t,\ldots,\varepsilon_{t-(q-1)}\right)'$ contains more than one element that depends on the current shock ε_t: the optimal action, X_t, and the shock itself, ε_t. In this case, one can go through the same two steps with the state vector $\hat{\xi}_t=\left(\varepsilon_t,\ldots,\varepsilon_{t-(q-1)},\varepsilon_{t-q}\right)'$ instead of the state vector $\xi_t=\left(X_t,\varepsilon_t,\ldots,\varepsilon_{t-(q-1)}\right)'$: (i) without loss in generality one can restrict attention to signal vectors $S_t^K=G'\hat{\xi}_t+\psi_t^K$ with lower triangular matrix G and diagonal, positive definite precision matrix Σ_ψ^{-1}, and (ii) any optimal signal vector of this form has the property that all rows of G' apart from the first row contain only zeros. To complete the proof, one only needs to note that any one-dimensional signal on $\hat{\xi}_t$ can also be written as a one-dimensional signal on ξ_t.

Similarly, when the optimal action follows an ARMA(p,q) process, the proof needs to be modified slightly, because the vector $\xi_t=\left(X_t,\ldots,X_{t-(p-1)},\varepsilon_t,\ldots,\varepsilon_{t-(q-1)}\right)'$ contains more than one element that depends on the current shock ε_t: the optimal action, X_t, and the shock itself, ε_t. In this case, one can go through the same two steps with the state vector

$$\hat{\xi}_t = \begin{cases} \left(X_t, \ldots, X_{t-(p-1)}, X_t - \theta_0 \varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-(q-1)}\right)' & \text{if } \theta_0 \neq 0 \\ \left(X_{t+1}, X_t, \ldots, X_{t-(p-1)}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-(q-1)}\right)' & \text{if } \theta_0 = 0 \end{cases}.$$

The state vector $\hat{\xi}_t$ has the property that only one element depends on the current shock: X_t if $\theta_0 \neq 0$ and X_{t+1} if $\theta_0 = 0$. After going through the same two steps as before with the state vector $\hat{\xi}_t$, one only needs to note that any one-dimensional signal on $\hat{\xi}_t$ can also be written as a one-dimensional signal on ξ_t.

D First-order conditions

In this appendix, we first show how one can compute the partial derivatives of each element of Σ_1 with respect to the elements of the vector of signal weights, h. We then state first-order conditions for the optimal signal weights. These first-order conditions are necessary conditions.

Substituting equation (29) into equation (28) and rearranging yields

$$\Sigma_1 - F\left[\Sigma_1 - \frac{\left(1 - 2^{-2\kappa}\right)}{h'\Sigma_1 h}\Sigma_1 h h' \Sigma_1\right]F' - Q = 0. \tag{44}$$

This equation gives Σ_1 as an implicit function of F, Q, h, and κ. Changing a single element of the vector of signal weights h potentially affects all elements of Σ_1. Let $\Sigma_{1,ij}$ denote the i,j-element of Σ_1 and let $d = \max\{1, p\} + q$. The derivatives $(d\Sigma_{1,ij}/dh_l)$ for $i, j = 1, \ldots, d$ and $l = 2, \ldots, d$ are given by

$$\forall l = 2, \ldots, d : \sum_{i=1}^{d}\sum_{j=1}^{d} Z^{ij}\frac{d\Sigma_{1,ij}}{dh_l} + Z^l = 0. \tag{45}$$

Here the $(d \times d)$ matrix Z denotes the left-hand side of equation (44), the $(d \times d)$ matrix Z^{ij} denotes the derivative of Z with respect to $\Sigma_{1,ij}$, i.e.,

$$Z^{ij} = \begin{bmatrix} \frac{\partial Z_{11}}{\partial \Sigma_{1,ij}} & \cdots & \frac{\partial Z_{1,d}}{\partial \Sigma_{1,ij}} \\ \vdots & \ddots & \vdots \\ \frac{\partial Z_{d,1}}{\partial \Sigma_{1,ij}} & \cdots & \frac{\partial Z_{d,d}}{\partial \Sigma_{1,ij}} \end{bmatrix},$$

and the $(d \times d)$ matrix Z^l denotes the derivative of Z with respect to h_l, i.e.,

$$Z^l = \begin{bmatrix} \frac{\partial Z_{11}}{\partial h_l} & \cdots & \frac{\partial Z_{1,d}}{\partial h_l} \\ \vdots & \ddots & \vdots \\ \frac{\partial Z_{d,1}}{\partial h_l} & \cdots & \frac{\partial Z_{d,d}}{\partial h_l} \end{bmatrix}.$$

One can give closed form expressions for Z^{ij} and Z^l, which depend on Σ_1. Let 1^{ij} denote a $(d \times d)$ matrix whose i,j-element equals one and whose other elements equal zero. Let 1^l denote a $(d \times 1)$ vector whose lth element equals one and whose other elements equal zero. It is straightforward to show that

$$\begin{aligned} Z^{ij} &= 1^{ij} - F1^{ij}F' - \frac{\left(1-2^{-2\kappa}\right)}{\left(h'\Sigma_1 h\right)^2} h'1^{ij}hF\Sigma_1 hh'\Sigma_1 F' \\ &\quad + \frac{\left(1-2^{-2\kappa}\right)}{h'\Sigma_1 h} F\left[1^{ij}hh'\Sigma_1 + \Sigma_1 hh'1^{ij}\right]F', \end{aligned}$$

and

$$\begin{aligned} Z^l &= -\frac{\left(1-2^{-2\kappa}\right)}{\left(h'\Sigma_1 h\right)^2}\left[\left(1^l\right)'\Sigma_1 h + h'\Sigma_1\left(1^l\right)\right]F\Sigma_1 hh'\Sigma_1 F' \\ &\quad + \frac{\left(1-2^{-2\kappa}\right)}{h'\Sigma_1 h}F\left[\Sigma_1\left(1^l\right)h'\Sigma_1 + \Sigma_1 h\left(1^l\right)'\Sigma_1\right]F'. \end{aligned}$$

Finally, when $p \geq 2$, the objective (27) equals $\Sigma_{1,22}$, because the second element of the state vector ξ_t equals X_{t-1} and Σ_1 is the variance-covariance matrix of the state vector given information at time $t-1$. In this case, normalizing the first signal weight to one ($h_1 = 1$), the first-order conditions for the optimal signal weights are simply

$$\forall l = 2, \ldots, d : \frac{d\Sigma_{1,22}}{dh_l} = 0. \tag{46}$$

In the case of $p \leq 1$ and $\max\{1, p\} + q \geq 2$ (i.e., in the case of an ARMA(1,q) process or an MA(q) process), the first-order conditions are only marginally more complicated and are omitted here to save space. For an example, see the proof of Proposition 5.

Solving the system of equations (44)-(46) for the $(d \times d)$ symmetric matrix Σ_1, the $d^2(d-1)$ derivatives $(d\Sigma_{1,ij}/dh_l)$, and the $(d-1)$ signal weights $h_2, \ldots, h_d$ yields signal weights that satisfy the first-order conditions.

E Proof of Proposition 4

Equation (45) at $p = 2$ and $h_2 = 0$ reads

$$Z^{11}\frac{d\Sigma_{1,11}}{dh_2} + Z^{12}\frac{d\Sigma_{1,12}}{dh_2} + Z^{21}\frac{d\Sigma_{1,21}}{dh_2} + Z^{22}\frac{d\Sigma_{1,22}}{dh_2} + Z^2 = 0, \tag{47}$$

with

$$Z^{11} = \begin{bmatrix} 1 - \phi_1^2 2^{-2\kappa} - \phi_2^2\left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,12}\Sigma_{1,21}}{\Sigma_{1,11}^2} & -\phi_1 2^{-2\kappa} \\ -\phi_1 2^{-2\kappa} & -2^{-2\kappa} \end{bmatrix},$$

$$Z^{12} = \begin{bmatrix} -\phi_1\phi_2 2^{-2\kappa} + \phi_2^2\left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,21}}{\Sigma_{1,11}} & 1 \\ -\phi_2 2^{-2\kappa} & 0 \end{bmatrix},$$

$$Z^{21} = \begin{bmatrix} -\phi_1\phi_2 2^{-2\kappa} + \phi_2^2\left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,12}}{\Sigma_{1,11}} & -\phi_2 2^{-2\kappa} \\ 1 & 0 \end{bmatrix},$$

$$Z^{22} = \begin{bmatrix} -\phi_2^2 & 0 \\ 0 & 1 \end{bmatrix},$$

$$Z^2 = \left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,11}\Sigma_{1,22} - \Sigma_{1,12}\Sigma_{1,21}}{\Sigma_{1,11}}\begin{bmatrix} 2\phi_1\phi_2 + \phi_2^2\frac{\Sigma_{1,12}+\Sigma_{1,21}}{\Sigma_{1,11}} & \phi_2 \\ \phi_2 & 0 \end{bmatrix}.$$

Equation (44) at $p = 2$ and $h_2 = 0$ implies

$$\Sigma_{1,22} = 2^{-2\kappa}\Sigma_{1,11}, \tag{48}$$

$$\Sigma_{1,12} = \Sigma_{1,21} = \frac{\phi_1}{2^{2\kappa} - \phi_2}\Sigma_{1,11}, \tag{49}$$

$$\Sigma_{1,11} = \frac{\theta_0^2}{1 - 2^{-2\kappa}\left(\phi_1^2 + \phi_2^2\right) - 2^{-2\kappa}\frac{2\phi_1^2\phi_2}{2^{2\kappa}-\phi_2} + \frac{(1-2^{-2\kappa})\phi_1^2\phi_2^2}{(2^{2\kappa}-\phi_2)^2}}. \tag{50}$$

In the case of a stationary AR(2) process, the denominators in equations (49) and (50) are positive. In the case of a non-stationary AR(2) process, these denominators are positive if and only if

$$2^{2\kappa} - \phi_2 > 0 \text{ and } \left(2^{2\kappa} - \phi_2^2\right)\left[2^{4\kappa} + \phi_2^2 - 2^{2\kappa}\left(\phi_1^2 + 2\phi_2\right)\right] > 0.$$

Equation (47) is a system of four linear equations in $\frac{d\Sigma_{1,11}}{dh_2}$, $\frac{d\Sigma_{1,12}}{dh_2}$, $\frac{d\Sigma_{1,21}}{dh_2}$, and $\frac{d\Sigma_{1,22}}{dh_2}$. Solving this system for $\frac{d\Sigma_{1,22}}{dh_2}$ and using equations (48)-(50) yields

$$\frac{d\Sigma_{1,22}}{dh_2} = -2\phi_1\phi_2\underbrace{\frac{\left(1 - 2^{-2\kappa}\right)\Sigma_{1,11}}{\left(2^{2\kappa} - \phi_2\right)^2}}_{>0}.$$

Hence, in the AR(2) case, the first-order condition (46) is satisfied at $h_2 = 0$ if and only if $\phi_1\phi_2 = 0$.

F Proof of Proposition 5

Equation (45) at $p = 1$, $q = 1$, and $h_2 = 0$ reads

$$Z^{11}\frac{d\Sigma_{1,11}}{dh_2} + Z^{12}\frac{d\Sigma_{1,12}}{dh_2} + Z^{21}\frac{d\Sigma_{1,21}}{dh_2} + Z^{22}\frac{d\Sigma_{1,22}}{dh_2} + Z^2 = 0,$$

with

$$Z^{11} = \begin{bmatrix} 1 - \phi_1^2 2^{-2\kappa} - \theta_1^2\left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,12}\Sigma_{1,21}}{\Sigma_{1,11}^2} & 0 \\ 0 & 0 \end{bmatrix},$$

$$Z^{12} = \begin{bmatrix} -\phi_1\theta_1 2^{-2\kappa} + \theta_1^2\left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,21}}{\Sigma_{1,11}} & 1 \\ 0 & 0 \end{bmatrix},$$

$$Z^{21} = \begin{bmatrix} -\phi_1\theta_1 2^{-2\kappa} + \theta_1^2\left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,12}}{\Sigma_{1,11}} & 0 \\ 1 & 0 \end{bmatrix},$$

$$Z^{22} = \begin{bmatrix} -\theta_1^2 & 0 \\ 0 & 1 \end{bmatrix},$$

and

$$Z^2 = \left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,11}\Sigma_{1,22} - \Sigma_{1,12}\Sigma_{1,21}}{\Sigma_{1,11}}\begin{bmatrix} 2\phi_1\theta_1 + \theta_1^2\frac{\Sigma_{1,12}+\Sigma_{1,21}}{\Sigma_{1,11}} & 0 \\ 0 & 0 \end{bmatrix}.$$

These equations imply

$$\frac{d\Sigma_{1,12}}{dh_2} = \frac{d\Sigma_{1,21}}{dh_2} = \frac{d\Sigma_{1,22}}{dh_2} = 0,$$

and

$$\begin{aligned} 0 \;=\; & \left[1 - \phi_1^2 2^{-2\kappa} - \theta_1^2\left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,12}\Sigma_{1,21}}{\Sigma_{1,11}^2}\right]\frac{d\Sigma_{1,11}}{dh_2} \\ & + \left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,11}\Sigma_{1,22} - \Sigma_{1,12}\Sigma_{1,21}}{\Sigma_{1,11}}\left[2\phi_1\theta_1 + \theta_1^2\frac{\Sigma_{1,12} + \Sigma_{1,21}}{\Sigma_{1,11}}\right]. \end{aligned}$$

Equation (44) at $p = 1$, $q = 1$, and $h_2 = 0$ reads

$$\Sigma_1 = \begin{bmatrix} \phi_1^2 2^{-2\kappa}\Sigma_{1,11} + \phi_1\theta_1 2^{-2\kappa}\left(\Sigma_{1,12} + \Sigma_{1,21}\right) + \theta_1^2\left[\Sigma_{1,22} - \left(1 - 2^{-2\kappa}\right)\frac{\Sigma_{1,21}\Sigma_{1,12}}{\Sigma_{1,11}}\right] & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \theta_0^2 & \theta_0 \\ \theta_0 & 1 \end{bmatrix}.$$

When $\theta_0 = 0$, the last equation implies

$$\Sigma_{1,22} = 1, \Sigma_{1,12} = \Sigma_{1,21} = 0, \Sigma_{1,11} = \frac{\theta_1^2}{1 - \phi_1^2 2^{-2\kappa}},$$

and the previous equation can be written as

$$\frac{d\Sigma_{1,11}}{dh_2} = -\frac{\left(1-2^{-2\kappa}\right)}{\left(1-\phi_1^2 2^{-2\kappa}\right)} 2\phi_1\theta_1. \tag{51}$$

Furthermore, in the case of $p=1$ and $q=1$, equation (29) reads

$$\begin{bmatrix} \Sigma_{0,11} & \Sigma_{0,12} \\ \Sigma_{0,21} & \Sigma_{0,22} \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{1,11} - \left(1-2^{-2\kappa}\right)\frac{(\Sigma_{1,11}+h_2\Sigma_{1,12})(\Sigma_{1,11}+h_2\Sigma_{1,21})}{\Sigma_{1,11}+h_2\Sigma_{1,12}+h_2\Sigma_{1,21}+h_2^2\Sigma_{1,22}} & \Sigma_{1,12} - \left(1-2^{-2\kappa}\right)\frac{(\Sigma_{1,11}+h_2\Sigma_{1,12})(\Sigma_{1,12}+h_2\Sigma_{1,22})}{\Sigma_{1,11}+h_2\Sigma_{1,12}+h_2\Sigma_{1,21}+h_2^2\Sigma_{1,22}} \\ \Sigma_{1,21} - \left(1-2^{-2\kappa}\right)\frac{(\Sigma_{1,21}+h_2\Sigma_{1,22})(\Sigma_{1,11}+h_2\Sigma_{1,21})}{\Sigma_{1,11}+h_2\Sigma_{1,12}+h_2\Sigma_{1,21}+h_2^2\Sigma_{1,22}} & \Sigma_{1,22} - \left(1-2^{-2\kappa}\right)\frac{(\Sigma_{1,21}+h_2\Sigma_{1,22})(\Sigma_{1,12}+h_2\Sigma_{1,22})}{\Sigma_{1,11}+h_2\Sigma_{1,12}+h_2\Sigma_{1,21}+h_2^2\Sigma_{1,22}} \end{bmatrix}.$$

The upper-left equation implies that the derivative of $\Sigma_{0,11}$ with respect to h_2 at the point $h_2=0$ equals

$$\frac{d\Sigma_{0,11}}{dh_2} = 2^{-2\kappa}\frac{d\Sigma_{1,11}}{dh_2}. \tag{52}$$

It follows from equations (51) and (52) that in the ARMA(1,1) case with $\phi_1 \neq 0$, $\theta_1 \neq 0$, and $\theta_0 = 0$ the optimal signal weight on ε_t is always non-zero.

G Proof of Proposition 6

Let h be a vector of signal weights and Σ_0, Σ_1 denote posterior and prior variance-covariance matrices that are solutions to Kalman filter equations (28)-(29) for h. Let $\hat{h} = (1,0,0,...)$, i.e., it represents a strategy that puts zero weights on past states; and let $\hat{\Sigma}_0, \hat{\Sigma}_1$ denote the corresponding solutions to (28)-(29). Note that if κ approaches infinity, then uncertainty and thus losses under $\hat{h}$ approach zero, and thus so must also losses from any optimal strategy.

Under any h, the agent allocates information capacity $(\kappa - \kappa')$ to the current optimal action X_t, while $\kappa' \geq 0$ is devoted to components of the uncertainty that are orthogonal to uncertainty about X_t. Under $\hat{h}$, all of the capacity κ is devoted to X_t, i.e., $\kappa' = 0$.

The Kalman filter equation (29) for the vector of weights h implies:

$$\Sigma_0^{1,1} = 2^{-2(\kappa-\kappa')}\left(\theta_0^2 + \left(F\Sigma_0 F'\right)^{1,1}\right), \tag{53}$$

where $M^{1,1}$ denotes the 1-1 element of a matrix M. In (53), we used the fact that devoting an amount C of information capacity to tracking a normally distributed random variable of variance

σ^2 implies posterior variance of $\sigma^2 2^{-2C}$. Similarly for the weights $\hat{h}$, we get:

$$\hat{\Sigma}_0^{1,1} = 2^{-2\kappa}\left(\theta_0^2 + \left(F\hat{\Sigma}_0 F'\right)^{1,1}\right). \tag{54}$$

Now, we can express the difference between expected losses from an optimal h and $\hat{h}$.

$$\Sigma_0^{1,1} - \hat{\Sigma}_0^{1,1} = 2^{-2\kappa}\left[(2^{2\kappa'}-1)\theta_0^2 + \left(2^{2\kappa'}\left(F\Sigma_0 F'\right)^{1,1} - \left(F\hat{\Sigma}_0 F'\right)^{1,1}\right)\right]. \tag{55}$$

For any κ, if h is optimal, then the RHS of (55) must always be less or equal to zero.

We will show by contradiction that as $\kappa \to \infty$, in a sequence of optimal h for such levels of information flows, κ' approaches zero. Let us thus assume that κ' does not approach zero. In this case, there exists a lower bound m, such that there always exists an arbitrarily large κ for which $(2^{2\kappa'}-1)\theta_0^2 > m$. Moreover, we will show that $\left(2^{2\kappa'}\left(F\Sigma_0 F'\right)^{1,1} - \left(F\hat{\Sigma}_0 F'\right)^{1,1}\right)$, the second term in the bracket on the RHS of (55), is positive or approaches zero. Therefore, the RHS of (55) is for any such κ positive, which implies that h is suboptimal to h^*, which is a contradiction.

First, the term $2^{2\kappa'}\left(F\Sigma_0 F'\right)^{1,1}$ is positive. Second, $\left(F\hat{\Sigma}_0 F'\right)^{1,1}$ approaches zero as $\kappa \to \infty$. $\left(F\hat{\Sigma}_0 F'\right)^{1,1}$ is the part of prior uncertainty about X_t that is driven by all past shocks, excluding the current shock ε_t. If the agent did not pay any attention to any shock, this variance would be equal to some σ_{tot}^2, which is finite, because X_t follows a stationary process. Since the agent devotes under $\hat{h}$ information capacity of exactly κ to each shock, then

$$\left(F\hat{\Sigma}_0 F'\right)^{1,1} = 2^{-2\kappa}\sigma_{tot}^2,$$

which approaches zero as $\kappa \to \infty$.

Putting this together implies that in a sequence of optimal strategies, κ' must approach zero, i.e., the agent does not resolve any uncertainty beyond that about X_t. Otherwise there would always exist an arbitrarily large κ for which the RHS of (55) would be positive, and thus h would not be an optimal strategy, which is a contradiction.

References

[1] Adam, Klaus (2007). "Optimal Monetary Policy with Imperfect Common Knowledge." *Journal of Monetary Economics*, 54(2), 267-301.

[2] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (2013). *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons.

[3] Cagan, Phillip (1956). "The Monetary Dynamics of Hyperinflation." in Milton Friedman, ed., Studies in the Quantity Theory of Money, Chicago: University of Chicago Press.

[4] Carlsson, Mikael, and Oskar Nordström Skans (2012). "Evaluating Microfoundations for Aggregate Price Rigidities: Evidence from Matched Firm-Level Data on Product Prices and Unit Labor Cost." *American Economic Review*, 102(4), 1571-1595.

[5] Greenwood, Jeremy, Zvi Hercowitz, and Gregory Huffman (1988). "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review,* 78(3), 402-417.

[6] Jaimovich, Nir, and Sergio Rebelo (2009). "Can News about the Future Drive the Business Cycle?" *American Economic Review*, 99(4), 1097-1118.

[7] Lorenzoni, Guido (2011). "News and Aggregate Demand Shocks." *Annual Review of Economics*, 3(1), 537-557.

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

[9] Luo, Yulei, and Eric Young (2014). "Signal Extraction and Rational Inattention." *Economic Inquiry*, 52(2), 811-829.

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

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

[12] Matějka, Filip (2016). "Rationally Inattentive Seller: Sales and Discrete Pricing." *Review of Economic Studies*, 83(3), 1125-1155.

[13] Matějka, Filip and Alisdair McKay (2015). "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review*, 105(1), 272-298.

[14] Muth, John F. (1960). "Optimal Properties of Exponentially Weighted Forecasts." *Journal of the American Statistical Association*, 55(290), 299-306.

[15] Paciello, Luigi, and Mirko Wiederholt (2014). "Exogenous Information, Endogenous Information, and Optimal Monetary Policy." *Review of Economic Studies*, 81(1), 356-388.

[16] Sims, Christopher A. (1998). "Stickiness." *Carnegie-Rochester Conference Series on Public Policy*, 49(1), 317-356.

[17] Sims, Christopher A. (2003). "Implications of Rational Inattention." *Journal of Monetary Economics*, 50(3), 665-690.

[18] Sims, Christopher A. (2010). "Rational Inattention and Monetary Economics." in *Handbook of Monetary Economics, Volume 3*, edited by Benjamin M. Friedman and Michael Woodford, 155-181. Elsevier.

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

[20] Stevens, Luminita (2015). "Coarse Pricing Policies." Discussion paper, Federal Reserve Bank of Minneapolis and University of Maryland.

[21] Tutino, Antonella (2013). "Rationally Inattentive Consumption Choices." *Review of Economic Dynamics*, 16(3), 421-439.

[22] Veldkamp, Laura L. (2011). *Information Choice in Macroeconomics and Finance*. Princeton University Press.

[23] Wiederholt, Mirko (2010). "Rational Inattention." in *The New Palgrave Dictionary of Economics, online edition*, edited by Steven N. Durlauf and Lawrence E. Blume, Palgrave Macmillan.

[24] Woodford, Michael (2002). "Imperfect Common Knowledge and the Effects of Monetary Policy." in P. Aghion et al., eds., Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps, Princeton and Oxford: Princeton University Press.

[25] Woodford, Michael (2009). "Information-Constrained State-Dependent Pricing." *Journal of Monetary Economics*, 56(S), 100-124.

Figure 1: Dynamic rational inattention problem in Sims (2003), example

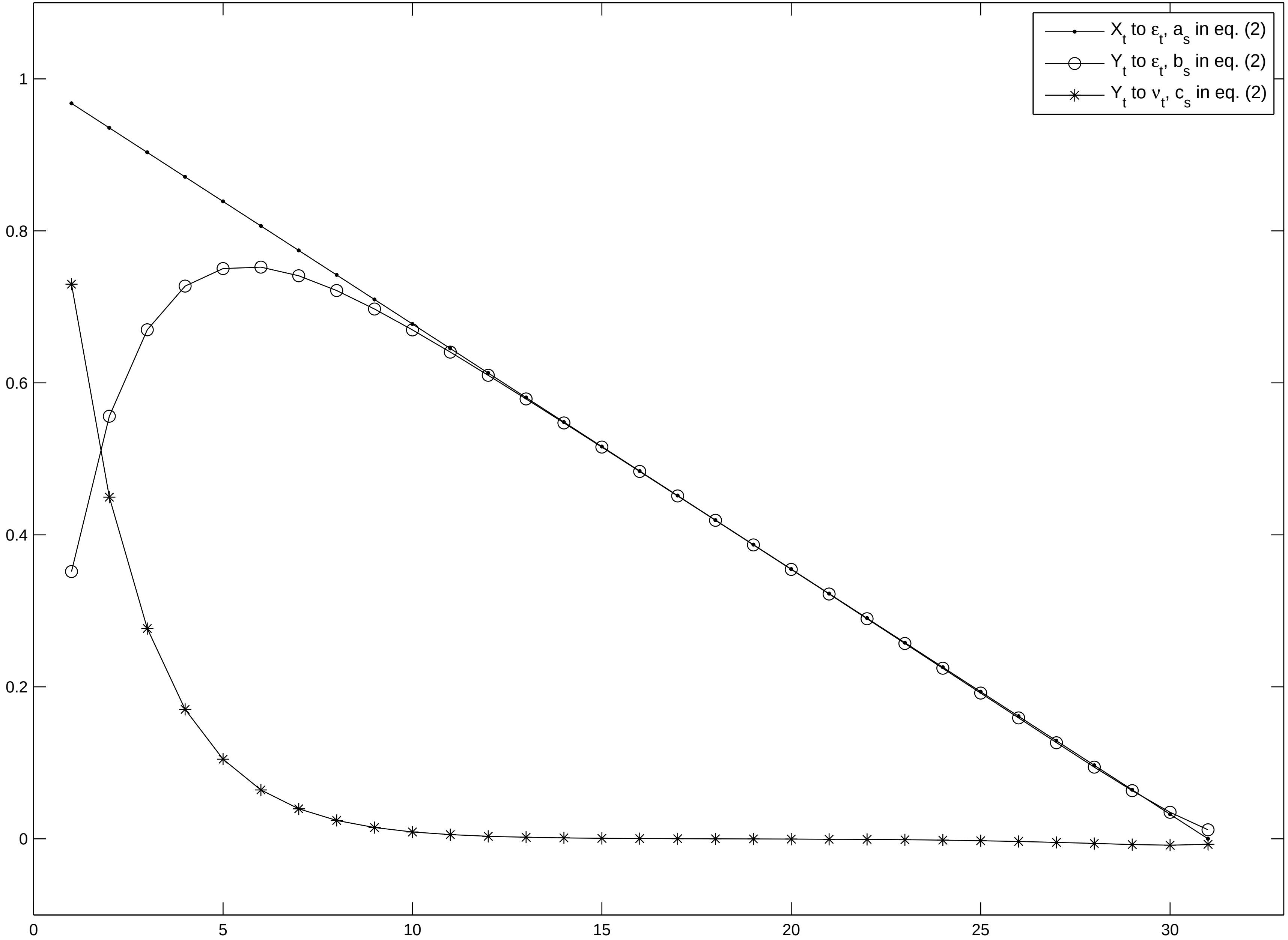


Figure 2: Optimal signal, AR(2) processes with $\phi_1 + \phi_2 = 0.9$, κ constant

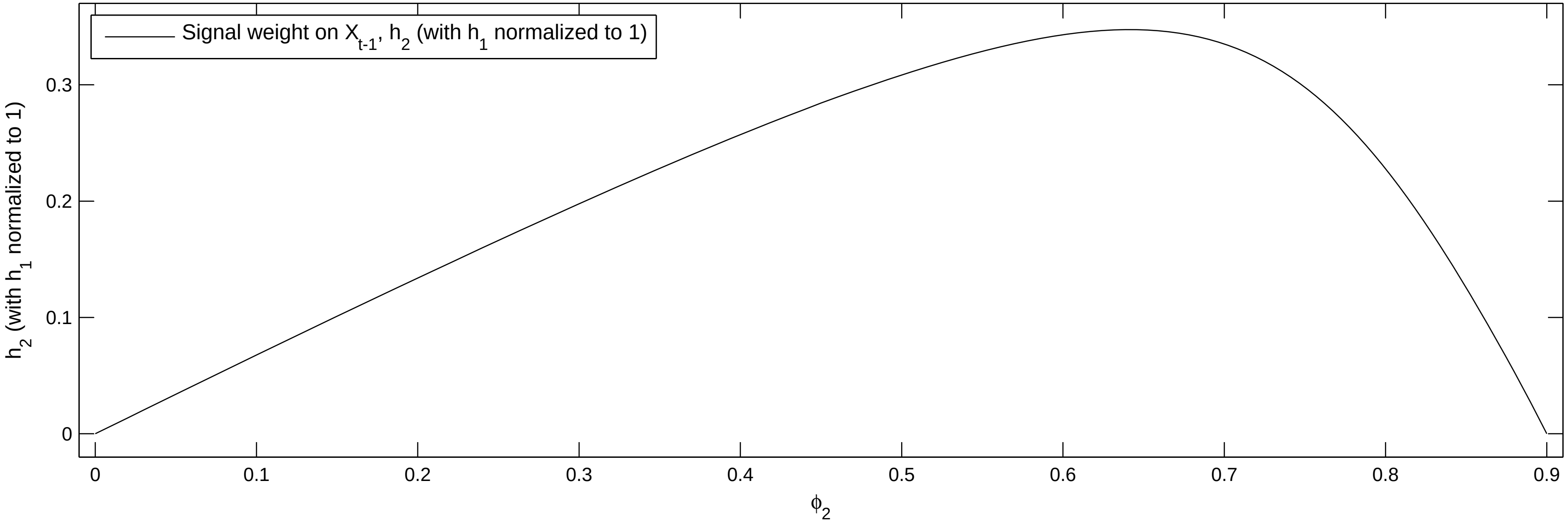


Figure 3: Optimal signal as function of κ, AR(2) example

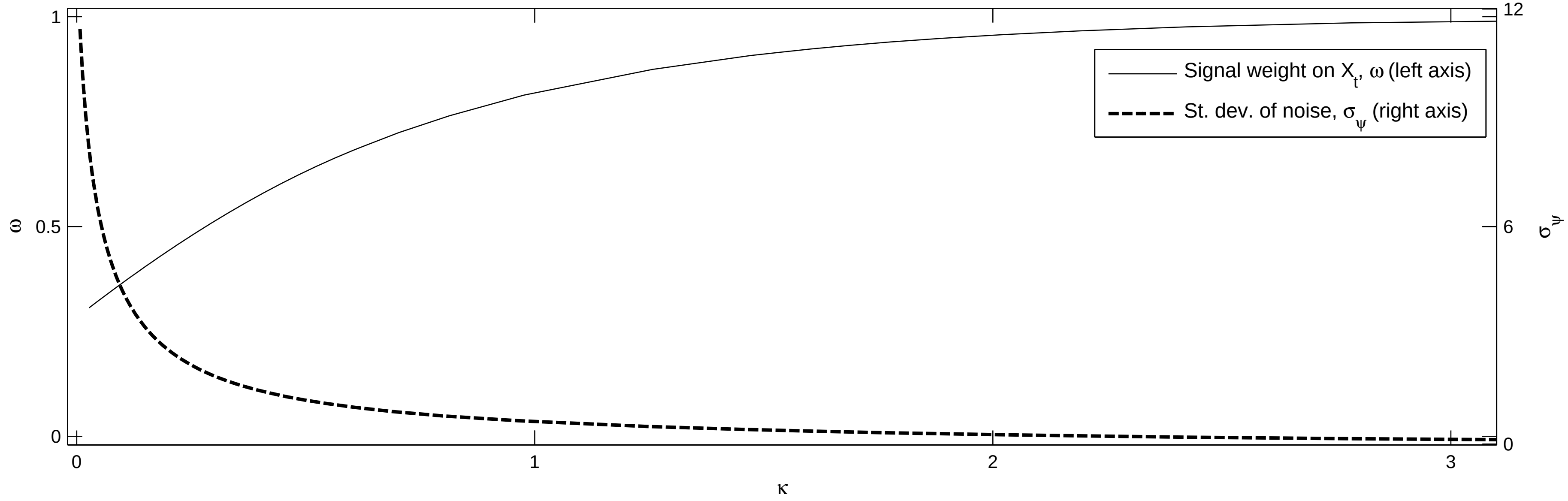


Figure 4: Dynamic rational inattention problem in this paper, X_t assumed identical to Figure 1

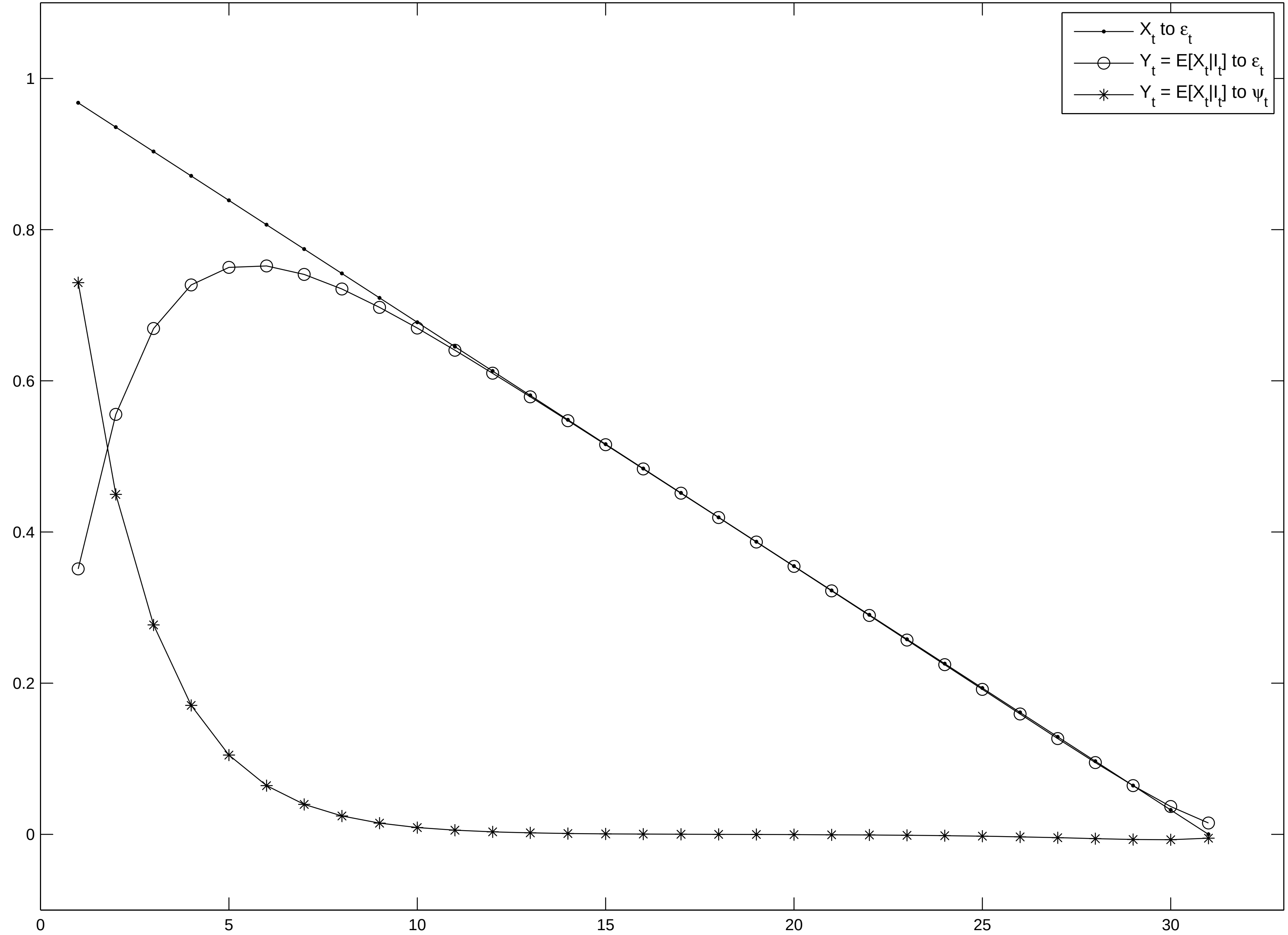
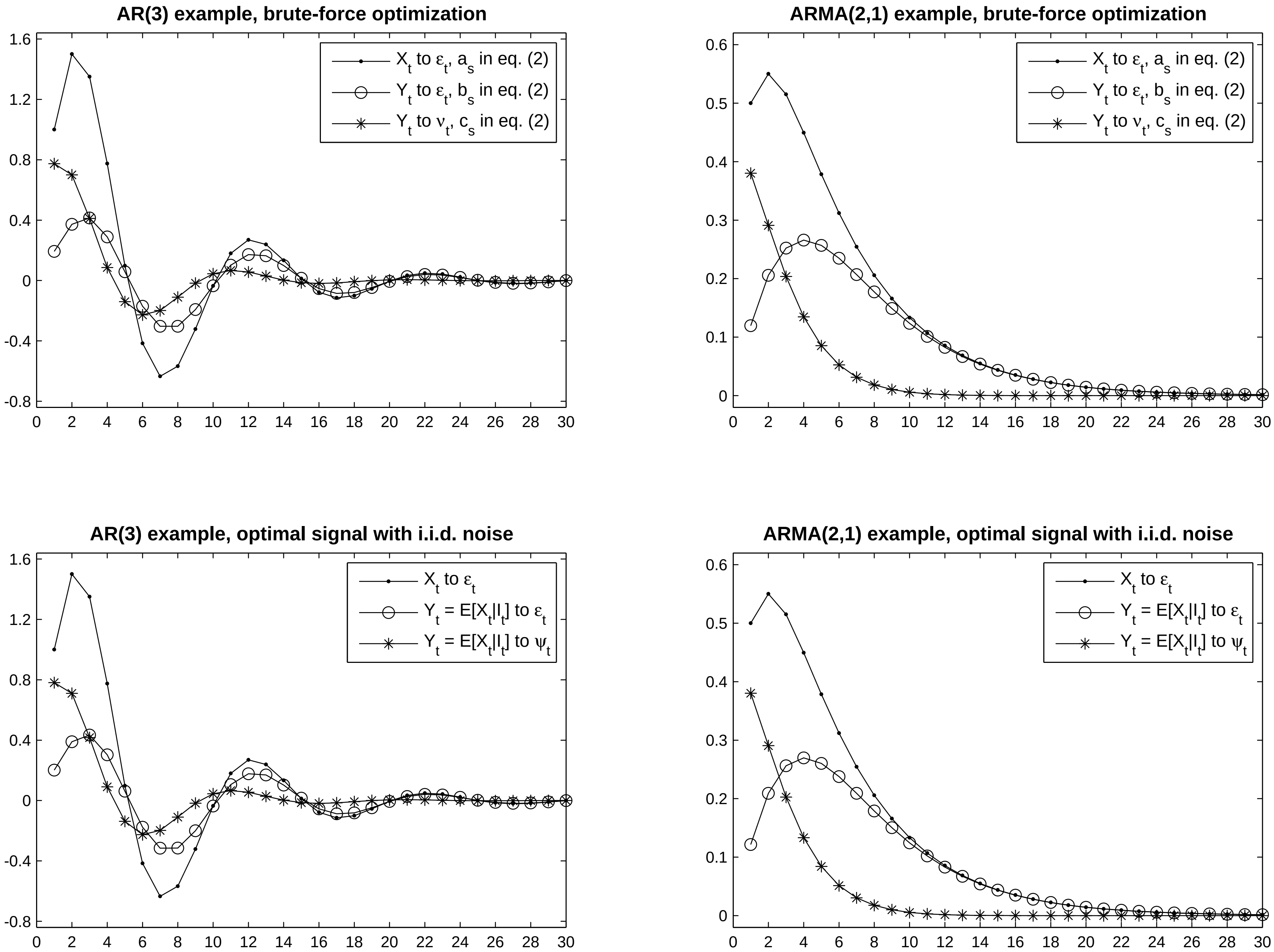


Figure 5: Examples of the dynamic rational inattention problem



Left column: $\phi_1 = 1.5$, $\phi_2 = -0.9$, $\phi_3 = -0.1$, $\theta_0 = 1$, $h_1 = 1$, $h_2 = -0.475$, $h_3 = 0.053$, $\sigma_\psi = 3.879$.

Right column: $\phi_1 = 1.3$, $\phi_2 = -0.4$, $\theta_0 = 0.5$, $\theta_1 = -0.1$, $h_1 = 1$, $h_2 = -0.275$, $h_3 = -0.069$, $\sigma_\psi = 1.349$.

Figure 6: Impulse response of output to a nominal shock, Woodford model and model with optimal signals

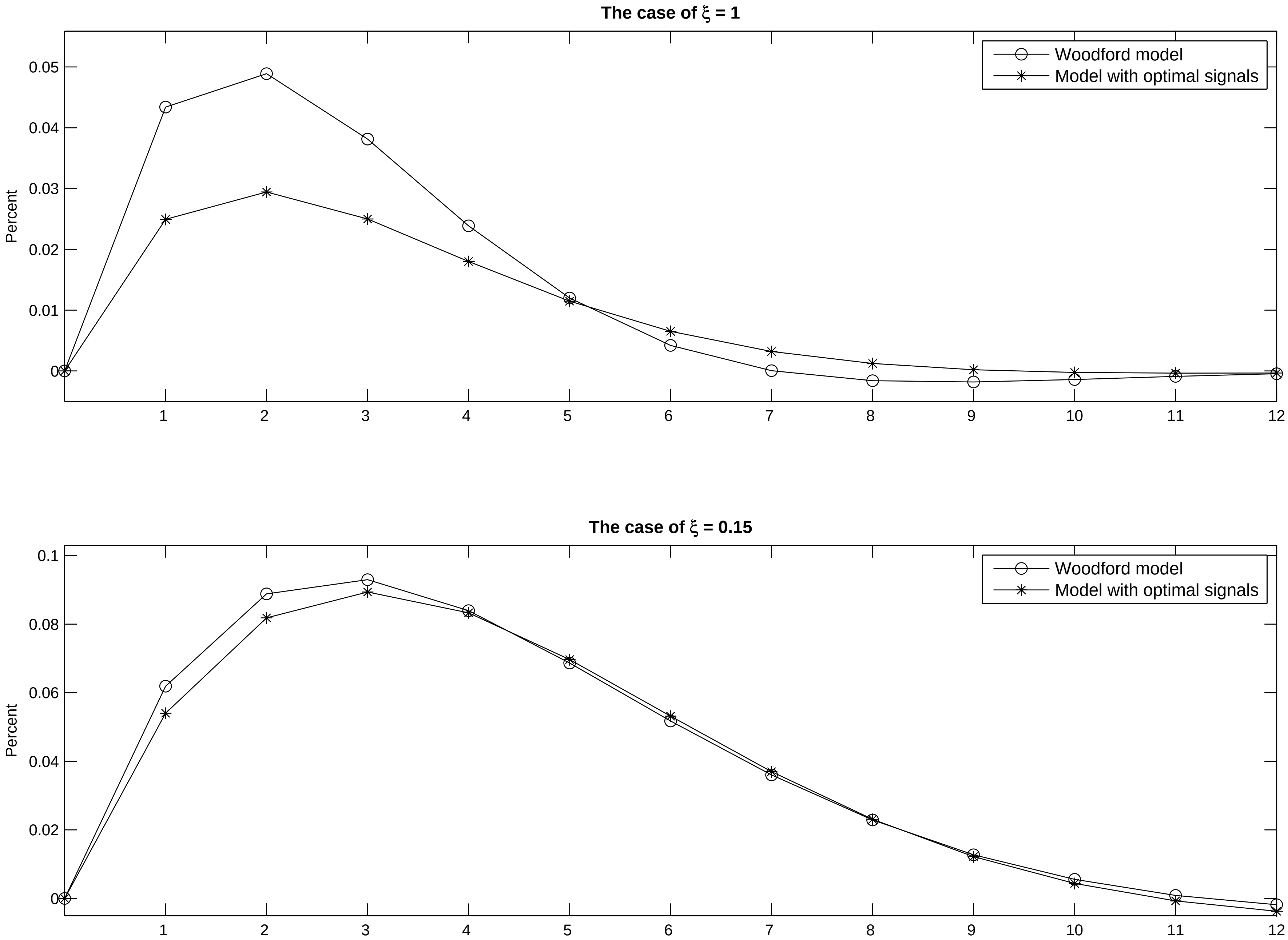


Figure 7: The impulse response of labor input to a productivity shock