The Rational Inattention Filter*

Bartosz Maćkowiak† Filip Matějka‡ Mirko Wiederholt§

April 2016

Abstract

Dynamic rational inattention problems used to be difficult to solve. This paper provides simple, analytical results for dynamic rational inattention problems. We start from the benchmark rational inattention problem. An agent tracks a variable of interest that follows a Gaussian process. The agent chooses how to pay attention to this variable. The agent aims to minimize, say, the mean squared error subject to a constraint on information flow, as in Sims (2003). We prove that if the variable of interest follows an ARMA(p,q) process, the optimal signal is about a linear combination of $\{X_t, \ldots, X_{t-p+1}\}$ and $\{\varepsilon_t, \ldots, \varepsilon_{t-q+1}\}$, where X_t denotes the variable of interest and ε_t denotes its period t innovation. The optimal signal weights can be computed from a simple extension of the Kalman filter: the usual Kalman filter equations in combination with first-order conditions for the optimal signal weights. We provide several analytical results regarding those signal weights. We also prove the equivalence of several different formulations of the information flow constraint. We conclude with general equilibrium applications from Macroeconomics.

Keywords: rational inattention, Kalman filter, macroeconomics

*An earlier February 2015 version of this paper was called "Analytical Results for Dynamic Rational Inattention Problems." We thank conference participants at EEA 2015 and SED 2015 for helpful comments. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the European Central Bank.

†European Central Bank and CEPR (bartosz.mackowiak@ecb.int)

‡CERGE-EI and CEPR (filip.matejka@cerge-ei.cz)

§Goethe University Frankfurt and CEPR (wiederholt@wiwi.uni-frankfurt.de)

1 Introduction

Economists have long studied decisions people make when they must allocate scarce resources such as wealth or factors of production. The recent literature on rational inattention investigates decision-making when human attention is scarce. Sims (1998, 2003) formalized limited attention as a constraint on information flow. He proposed to model decision-making as optimization subject to this constraint. Sims aimed to understand human behavior in a *dynamic* environment. He conjectured that rational inattention could provide a simple explanation for why a variety of economic variables, from consumption and investment to prices of goods and services, tend to display inertia in response to aggregate disturbances.

While the idea of rational inattention has received growing interest in various areas of Economics (e.g., Microeconomics, Macroeconomics, Finance, Labor Economics, International Trade), thus far most papers focus on static models or analyze economies that are independent over time. The reason is that dynamic attention choice problems are hard to solve. This difficulty has limited applicability of rational inattention, even though many economists find the idea of rational inattention plausible. The papers that do study dynamic economies that are correlated over time normally take one of three approaches to simplify the attention choice problem: (i) assume that agents act based on particular noisy signals without proving optimality of the signals (Luo, 2008, Paciello and Wiederholt, 2014), (ii) suppose that agents cannot costlessly access memory (Woodford, 2009, Stevens, 2015), or (iii) solve numerically for the actions under rational inattention sidestepping the question of what signals are being observed (Sims, 2003, Section 4, Maćkowiak and Wiederholt, 2009 and 2015).

This paper derives analytical results for dynamic rational inattention problems. We study the canonical dynamic attention choice problem proposed by Sims (2003, Section 4). An agent tracks a variable of interest that follows a Gaussian stochastic process. The agent chooses how to pay attention to this variable, i.e., the agent chooses the properties of the signals about the variable of interest, subject to the constraint on the flow of information between the signals and the variable of interest. The agent aims to minimize, say, the mean squared error between the variable of interest and the action taken based on the signals.

We prove that if the variable of interest follows an ARMA(p,q) process, any optimal signal vector with i.i.d. noise 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, where

X_t denotes the variable of interest in period t and ε_t denotes its period t innovation. Moreover, the agent can attain the optimum with a one-dimensional signal. Hence, without loss in generality, one can restrict attention to signals of the form

$$S_t = a_0 X_t + \ldots + a_{p-1} X_{t-(p-1)} + b_0 \varepsilon_t + \ldots + b_{q-1} \varepsilon_{t-(q-1)} + \psi_t,$$

where ψ_t is the i.i.d. noise and more attention implies a smaller variance of noise. For example, if the variable of interest follows an ARMA(2,1) process, the optimal signal is of the form $S_t = a_0 X_t + a_1 X_{t-1} + b_0 \varepsilon_t + \psi_t$. One only has to solve for the remaining signal weights a_0, a_1, and b_0, and for the variance of noise, σ_ψ^2. This dimensionality reduction result is important, because it is well known from Time Series Econometrics that the evolution of many aggregate economic variables can be well described by a low-order ARMA(p,q) process.

The question then becomes: What are the remaining signal weights? For example, the above statement does not rule out the possibility that $a_1 = \ldots = a_{p-1} = b_0 = \ldots = b_{q-1} = 0$. In other words, the above statement does not rule out the possibility that $S_t = X_t + \psi_t$ is an optimal signal. Our first result regarding the remaining signal weights is what we call the "dynamic attention principle." In a dynamic setting, the information choice problem is always forward-looking, because an agent cares about being well informed in the current period and entering well informed into the next period. If the variable of interest follows an AR(1) process, there is no tension between these two goals. Learning about the present and learning about the future are the same thing. For this reason, the optimal signal is $S_t = X_t + \psi_t$. Beyond an AR(1) process, there is a tension between these two goals and therefore the optimal signal is *generically not* $S_t = X_t + \psi_t$. For example, suppose that the variable of interest follows the process $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_0 \varepsilon_t$ with $\phi_1, \phi_2, \theta_0 \neq 0$. Or, suppose that the variable of interest follows the process $X_t = \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1}$ with $\phi_1, \theta_1 \neq 0$. We prove formally that in both cases the optimal signal is *never* $S_t = X_t + \psi_t$. The reason is that there is a tension between learning about the present and learning about the future, and the agent wants to enter well informed into the next period.

Our second result regarding the remaining signal weights is what we call the "rational inattention filter." The rational inattention filter is the Kalman filter with the observation equation that is optimal from the rational inattention perspective. The rational inattention filter consists of the Kalman filter equations with the information flow constraint substituted in and the $p + q - 1$ first-order conditions for the optimal signal weights. We derive those first-order conditions. Hence, anyone familiar with the Kalman filter can easily solve dynamic rational inattention problems, as in Sims (2003), without any loss in generality. Along the way, we also prove the equivalence of several different formulations of the information flow constraint that have appeared in the literature.

We illustrate these results with an application from Macroeconomics. We solve the price-setting model in Woodford (2002) with the signal that is optimal from the rational inattention perspective. The price-setting model in Woodford (2002) has become a benchmark in the literature on price-setting and in the literature on macroeconomic models with information frictions. In the original model, the information structure is exogenous. We resolve the model with the signal that is optimal from the rational inattention perspective, and we compare the resulting pricing behavior and aggregate dynamics with those in the original model.

This paper is related to another recent paper on dynamic rational inattention problems. Steiner et al. (2015) study a general dynamic model with discrete choice under rational inattention. The focus there is to show that such a model leads in general to behavior in the form of a dynamic logit model, while certain parameters specifying details of the behavior are difficult or impossible to solve for. By contrast, we focus on the class of problems studied in Sims (2003), i.e., with Gaussian stochastic processes and quadratic objectives, and we solve the problems fully.

The rest of the paper is organized as follows. Section 2 presents the decision problem and proves the ARMA(p,q) result. Section 3 derives the rational inattention filter and proves the dynamic attention principle. Section 4 contains applications. Section 5 concludes.

2 Decision problem and dimensionality reduction

In this section we present the dynamic attention choice problem, the dimensionality reduction result, and an equivalence result about different formulations of the information flow constraint.

The dynamic attention choice problem. The agent tracks X_t which follows a stationary Gaussian ARMA(p,q) process with finite $p, q \geq 0$:

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

where $\phi_1, \ldots, \phi_p$ and $\theta_0, \ldots, \theta_q$ are coefficients and ε_t follows a Gaussian white noise process with unit variance. For ease of exposition, we require $p + q > 0$, that is, we rule out the white noise case, which we have studied before. For ease of exposition, we also require $\phi_p \neq 0$ and $\theta_q \neq 0$, which simply ensures a precise language.

At each time $t \geq 1$, the agent receives $K \geq 1$ signals, where each signal is about a potentially different linear combination of current and past X_t and current and past ε_t:

$$S_t^K = AX_t^M + B\varepsilon_t^N + \psi_t^K, \tag{2}$$

where $S_t^K = (S_{t,1}, \ldots, S_{t,K})'$, $X_t^M = (X_t, \ldots, X_{t-M+1})'$, $\varepsilon_t^N = (\varepsilon_t, \ldots, \varepsilon_{t-N+1})'$, and $\psi_t^K = (\psi_{t,1}, \ldots, \psi_{t,K})'$. Here M and N are arbitrarily large integers satisfying $M \geq p$ and $N \geq q$, $A \in \mathbb{R}^{K\times M}$ and $B \in \mathbb{R}^{K\times N}$ are matrices of coefficients, and the noise components $(\psi_{t,1}, \ldots, \psi_{t,K})$ follow independent Gaussian white noise processes with vector of variances $\Upsilon = (\sigma_1^2, \ldots, \sigma_K^2) \in \mathbb{R}_{++}^K$.

The agent's information set at any time $t \geq 1$ includes any initial information and all signals received up to and including time t:

$$\mathcal{I}_t = \mathcal{I}_0 \cup \left\{S_1^K, \ldots, S_t^K\right\}. \tag{3}$$

The agent chooses K, A, B, and Υ. The agent's decision problem reads:

$$\min_{K,A,B,\Upsilon} E\left[(X_t - E\left[X_t|\mathcal{I}_t\right])^2\right], \tag{4}$$

subject to (1), (2), (3), and the constraint on information flow

$$\lim_{T\to\infty} \frac{1}{T} I\left(X_1, \ldots, X_T; S_1^K, \ldots, S_T^K\right) \leq \kappa. \tag{5}$$

Here $I\left(X_1, \ldots, X_T; S_1^K, \ldots, S_T^K\right)$ denotes the mutual information between the sequence of variables of interest and the sequence of signals. In the case of $q > 0$, we also add the innovations $\varepsilon_T, \ldots, \varepsilon_{T-q+1}$ to the sequence $X_1, \ldots, X_T$ in the information flow constraint to ensure that the information flow constraint also measures the information contained in the signals about the innovations. In words, the agent chooses the properties of the signals so as to minimize the mean squared error subject to the information flow constraint. The expectations in (4) are computed using the steady-state Kalman filter. This dynamic attention choice problem is formulated and studied numerically in Sims (2003, Section 4).

The information flow constraint. The information flow constraint (5) is equivalent to a constraint on the difference between prior and posterior uncertainty at a given point in time.

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. *Let* ξ_t *denote the following vector*

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

Let H *denote uncertainty measured by entropy. The information flow constraint (5) is equivalent to*

$$\lim_{T\to\infty}\left[H\left(\xi_T|S^{K,T-1}\right) - H\left(\xi_T|S^{K,T}\right)\right] \leq \kappa. \tag{6}$$

Proof. See Appendix A. ■

In words, this result states that the information flow on the left-hand side of (5) is equal to the difference between prior uncertainty and posterior uncertainty about all variables that the signal (2) can be about (more formally, the difference between prior uncertainty and posterior uncertainty about all variables that are necessary to be able to compute the signal S_T^K for *any* A, B and ψ_T^K). To understand the definition of the vector ξ_t, note the following. When $q = 0$ and $N \leq M - p$, the vectors X_t^M and ε_t^N can be computed from X_t^M alone. By contrast, when $q = 0$ and $N > M - p$, one needs $N - (M - p)$ additional lags of X_t to compute the vectors X_t^M and ε_t^N. Finally, when $q > 0$, one can compute the moving-average terms $\{\theta_0\varepsilon_\tau + \ldots + \theta_q\varepsilon_{\tau-q}\}_{\tau=t}^{t-M+p+1}$ from X_t^M. To compute the innovations $\{\varepsilon_\tau\}_{\tau=t}^{t-N+1}$ one also needs $\varepsilon_t, \ldots, \varepsilon_{t-(q-1)}$ and additional lags of X_t if $M-p+q < N$. Lemma 1 shows the equivalence of two formulations of the information flow constraint that have appeared in the literature and is used to prove the following proposition.

Reduction in dimensionality without loss in generality. The following proposition states two general properties of the optimal signal weights that can be used to reduce the dimensionality of the problem (1)-(5) without any loss in generality.

Proposition 1 *If* $p > 0$ *and* $q > 0$, *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. If* $q = 0$, *any optimal signal vector is on linear combinations of* $\left\{X_t, \ldots, X_{t-(p-1)}\right\}$ *only. If* $p = 0$, *any optimal signal vector is on linear combinations of* $\left\{\varepsilon_t, \ldots, \varepsilon_{t-(q-1)}\right\}$ *only. The agent can attain the optimum with a one-dimensional signal. Hence, in the ARMA(p,q) case, for example, one can restrict attention to signals of the form*

$$S_t = a_0X_t + \ldots + a_{p-1}X_{t-(p-1)} + b_0\varepsilon_t + \ldots + b_{q-1}\varepsilon_{t-(q-1)} + \psi_t.$$

Proof. See Appendix B. ■

This result implies, for example, that if the variable of interest follows an ARMA(2,1) process, one can restrict attention to signals of the form $S_t = a_0 X_t + a_1 X_{t-1} + b_0 \varepsilon_t + \psi_t$. Furthermore, one of the coefficients $a_0, a_1, b_0, \sigma_\psi^2$ can be normalized to one, because multiplying a signal by a non-zero constant changes neither the objective nor the information flow. Figure 1 presents an example. In the next section, we show that there exists a simple and elegant way of computing the optimal signal weights a_0, a_1, and b_0, and the variance of noise, σ_ψ^2. In addition, we provide analytical results regarding those signal weights.

Non-stationarity. So far we have assumed that the variable X_t follows a stationary process. This assumption ensures that all conditional moments appearing in the proof of Lemma 1 and in the proof of Proposition 1 are well-defined. For example, 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 $\mathcal{I}_t$. 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. Objective (4) and constraint (5) depend on Σ_1 and Σ_0. The assumption that the variable X_t follows a stationary process ensures that Σ_1 and Σ_0 are well-defined. One can relax the stationarity assumption. Proposition 1 extends to the case of a non-stationary ARMA(p,q) process so long as all conditional moments appearing in the proof of Proposition 1 are well-defined. This requires that the parameter κ is sufficiently large.

3 Rational inattention filter and dynamic attention principle

In this section, we show that the remaining signal weights can be computed from an extension of the standard Kalman filter. We also provide analytical results regarding those signal weights and another equivalence result about different formulations of the information flow constraint.

The state-space representation of the signal. The variable of interest X_t follows a stationary Gaussian ARMA(p,q) process with $p > 0$:

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

According to Proposition 1, one can restrict attention to signals of the form

$$S_t = a_0 X_t + \ldots + a_{p-1} X_{t-(p-1)} + b_0 \varepsilon_t + \ldots + b_{q-1} \varepsilon_{t-(q-1)} + \psi_t. \tag{7}$$

This dynamic system has the following state-space representation (following the notation in Hamilton (1994), Chapter 13)

$$\begin{aligned} \xi_{t+1} &= F\xi_t + v_{t+1}, \\ S_t &= h'\xi_t + \psi_t. \end{aligned}$$

Here ξ_t is the following vector

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

and h is the following vector

$$h = \begin{cases} (a_0, \ldots, a_{p-1})' & \text{if } q = 0 \\ (a_0, \ldots, a_{p-1}, b_0, \ldots, b_{q-1})' & \text{if } q > 0 \end{cases}.$$

The length of the column vector v_{t+1} equals the length of the column vector ξ_{t+1}. For $q = 0$, the first element of v_{t+1} equals $\theta_0\varepsilon_{t+1}$ and its other elements equal zero. For $q > 0$, the first element of v_{t+1} equals $\theta_0\varepsilon_{t+1}$, its $p+1$ element equals ε_{t+1}, and its other elements equal zero. The matrix F is a $[(p+q) \times (p+q)]$ matrix. For $q = 0$, the first row of the matrix F equals $\left(\phi_1, \ldots, \phi_p\right)$ and the other rows of the matrix F have a one just left of the main diagonal and zeros everywhere else. For $q > 0$, the first row of the matrix F equals $\left(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q\right)$, the $p+1$ row of the matrix F contains only zeros, and the other rows of the matrix F have a one just left of the main diagonal and zeros everywhere else.

The usual Kalman filter equations for the variance-covariance matrices read

$$\begin{aligned} \Sigma_{t|t} &= \Sigma_{t|t-1} - \Sigma_{t|t-1}h\left(h'\Sigma_{t|t-1}h + \sigma_\psi^2\right)^{-1}h'\Sigma_{t|t-1} && (8) \\ \Sigma_{t+1|t} &= F\Sigma_{t|t}F' + Q. && (9) \end{aligned}$$

Here $\Sigma_{t|t}$ is the conditional variance-covariance matrix of ξ_t given $\mathcal{I}_t$, $\Sigma_{t|t-1}$ is the conditional variance-covariance matrix of ξ_t given $\mathcal{I}_{t-1}$, and Q is the variance-covariance matrix of v_{t+1}.

The information flow constraint. The information flow constraint can be expressed in terms of a single "signal-to-noise" ratio. This formulation of the constraint has a simple interpretation and has a practical advantage: one can solve the constraint for the variance of noise.

Lemma 2 *In the case of the one-dimensional signal (7), the information flow constraint (6) is equivalent to*

$$\frac{1}{2}\log_2\left[\frac{h'\Sigma_1 h}{\sigma_\psi^2}+1\right]\leq\kappa. \tag{10}$$

Here $\Sigma_1=\lim_{t\to\infty}\Sigma_{t|t-1}$. *The ratio in (10) equals the variance of the informative component of the signal, conditional on* $t-1$ *information, to the variance of the noise component of the signal.*

Proof. Conditional normality implies that

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

Hence, the information flow constraint (6) is equivalent to

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

Here $\Sigma_1=\lim_{t\to\infty}\Sigma_{t|t-1}$ and $\Sigma_0=\lim_{t\to\infty}\Sigma_{t|t}$. Equation (8) in terms of Σ_1 and Σ_0 reads

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

Using the last equation to substitute for Σ_0 in the information flow constraint yields

$$\frac{1}{2}\log_2\left[\frac{1}{\det\left(I-\frac{1}{h'\Sigma_1 h+\sigma_\psi^2}hh'\Sigma_1\right)}\right]\leq\kappa.$$

It follows from Sylvester's determinant theorem that

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

The lemma follows. ■

The decision problem. Collecting results we arrive at the following statement. Without loss in generality, the decision problem (1)-(5) can be stated as:

$$\min_{(a_0,\ldots,a_{p-1},b_0,\ldots,b_{q-1})\in\mathbb{R}^{p+q},\sigma_\psi^2\in\mathbb{R}_{++}}\begin{pmatrix}1&0&\cdots&0\end{pmatrix}\left[\Sigma-\Sigma h\left(h'\Sigma h+\sigma_\psi^2\right)^{-1}h'\Sigma\right]\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix},$$

subject to

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

Here Σ is the matrix that we called Σ_1 before and it is given by

$$\Sigma=F\left[\Sigma-\Sigma h\left(h'\Sigma h+\sigma_\psi^2\right)^{-1}h'\Sigma\right]F'+Q.$$

The objective is simply the (1,1)-element of the matrix Σ_0. The information flow constraint is simply a constraint on a particular signal-to-noise ratio, as explained above. Finally, the matrix Σ_1 is given by the usual Kalman filter equation for the conditional variance-covariance matrix of the state vector in period t given information in period $t-1$.

The information flow constraint is always binding and thus, $\forall\kappa>0$,

$$\sigma_\psi^2=\frac{h'\Sigma h}{2^{2\kappa}-1}. \tag{11}$$

Using the information flow constraint to substitute for the variance of noise yields the following statement of the decision problem, $\forall\kappa>0$:

$$\min_{(a_0,\ldots,a_{p-1},b_0,\ldots,b_{q-1})\in\mathbb{R}^{p+q}}\begin{pmatrix}1 & 0 & \cdots & 0\end{pmatrix}\left[\Sigma-\frac{\left(1-2^{-2\kappa}\right)}{h'\Sigma h}\Sigma hh'\Sigma\right]\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}. \tag{12}$$

Here Σ is given by

$$\Sigma=F\left[\Sigma-\frac{\left(1-2^{-2\kappa}\right)}{h'\Sigma h}\Sigma hh'\Sigma\right]F'+Q, \tag{13}$$

where F, h, and Q are defined at the beginning of the section.[1]

It is worth pointing out that, for $p>1$, the objective equals the (2,2)-element of the matrix Σ. The reason is simple. For $p>1$, the second element of the state vector ξ_t is X_{t-1}. Furthermore, the matrix Σ is the conditional variance-covariance matrix of the state vector in period t given information *in period* $t-1$. Hence, the (2,2)-element of the matrix Σ equals $Var\left(X_{t-1}|\mathcal{I}_{t-1}\right)$, which equals $Var\left(X_t|\mathcal{I}_t\right)$. However, when $p=1$, this is not the case. To cover the case of $p=1$,

[1] One can also endogenize κ by augmenting the vector of choice variables by κ and by adding a cost function for κ to the objective. In the rational inattention filter presented below, this will simply add another first-order condition.

we have to keep the more complicated looking objective, which equals the (1,1)-element of the conditional variance-covariance matrix of the state vector in period t given information *in period* t.

The rational inattention filter. The rational inattention filter is the Kalman filter with the observation equation that is optimal from a rational inattention perspective. Mathematically, its main building blocks are the Kalman filter equation for the conditional variance-covariance matrix with the information flow constraint substituted in (equation (13)), and a set of first-order conditions for the optimal signal weights. In the following, we normalize the weight on X_t to one without loss in generality (i.e., $a_0 = 1$). There are $p+q-1$ first-order conditions for the remaining signal weights.

One can derive these first-order conditions as follows. Equation (13) gives Σ as an implicit function of $(a_0, \ldots, a_{p-1}, b_0, \ldots, b_{q-1}) \in \mathbb{R}^{p+q}$ and can be written as

$$\Sigma - F\Sigma F' - Q + \frac{\left(1 - 2^{-2\kappa}\right)}{h'\Sigma h} F\Sigma h h'\Sigma F' = 0. \tag{14}$$

Let the $[(p+q) \times (p+q)]$ matrix $Z = \Sigma - F\Sigma F' - Q + \frac{\left(1-2^{-2\kappa}\right)}{h'\Sigma h} F\Sigma h h'\Sigma F'$ denote the left-hand side of the equation (14). Furthermore, let the $[(p+q) \times (p+q)]$ matrix Z^{ij} denote the derivative of Z with respect to Σ_{ij}, i.e.,

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

Let the $[(p+q) \times (p+q)]$ matrix Z^l denote the derivative of Z with respect to h_l, i.e.,

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

Let 1^{ij} denote a $[(p+q) \times (p+q)]$ matrix whose i,j-element equals one and whose other elements equal zero. Let 1^l denote a $[(p+q) \times 1]$ vector whose lth element equals one and whose other elements equal zero. It is straightforward to show that

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

and

$$Z^l = -\frac{\left(1-2^{-2\kappa}\right)}{\left(h'\Sigma h\right)^2}\left[\left(1^l\right)'\Sigma h + h'\Sigma\left(1^l\right)\right]F\Sigma h h'\Sigma F' + \frac{\left(1-2^{-2\kappa}\right)}{h'\Sigma h}F\left[\Sigma\left(1^l\right)h'\Sigma + \Sigma h\left(1^l\right)'\Sigma\right]F'.$$

Next, the derivatives $(d\Sigma_{ij}/dh_l)$ for $i=1,\ldots,p+q$, $j=1,\ldots,p+q$, and $l=2,\ldots,p+q$ are given by

$$\forall l = 2,\ldots,p+q : \sum_{i=1}^{p+q}\sum_{j=1}^{p+q} Z^{ij}\frac{d\Sigma_{ij}}{dh_l} + Z^l = 0. \tag{15}$$

Finally, when $p>1$, the objective (12) reduces to the (2,2)-element of the matrix Σ and the first-order conditions for the optimal signal weights are simply[2]

$$\forall l = 2,\ldots,p+q : \frac{d\Sigma_{22}}{dh_l} = 0. \tag{16}$$

Solving the system of equations (14)-(16) for the $[(p+q)\times(p+q)]$ symmetric matrix Σ, the $(p+q)^2(p+q-1)$ derivatives $(d\Sigma_{ij}/dh_l)$, and the $(p+q-1)$ signal weights h_l yields the optimal signal weights.

Proposition 2 *The optimal signal weights have to satisfy equations (14)-(16). Given the optimal signal weights, the actions can be computed from the usual Kalman filter equations.*

The dynamic attention principle. In a dynamic setting, an agent cares about being well informed in the current period and entering well informed into the next period. If the variable of interest follows an AR(1) process, there is no tension between the goals of being well informed today and entering well informed into the next period. Learning about the present and learning about the future are the exact same thing. For this reason, the optimal signal is $S_t = X_t + \psi_t$. Beyond an AR(1) process, there is a tension between these two goals and therefore the optimal signal is generically not $S_t = X_t + \psi_t$. We call this the "dynamic attention principle."

The following proposition formally states the dynamic attention principle in the AR(2) case. Suppose the variable of interest is given by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_0\varepsilon_t.$$

[2] We add the version of equation (16) for the case of $p=1$ later. For an example, see the proof of Proposition 4.

It follows from Proposition 1 that an optimal one-dimensional signal has the form

$$S_t = a_0 X_t + a_1 X_{t-1} + \psi_t.$$

However, this statement does not rule out the possibility that the optimal signal weight a_1 equals zero (i.e., the statement does not rule out the possibility that $S_t = X_t + \psi_t$ is an optimal signal.) Proposition 3 states that if $\phi_1, \phi_2 \neq 0$ then $a_1 \neq 0$. In words, if learning about the present and learning about the future are not the exact same thing ($\phi_2 \neq 0$) and the process cannot be written as overlapping AR(1)'s in a lower frequency ($\phi_1 \neq 0$), the optimal signal weight on X_{t-1} is non-zero.

Proposition 3 *Suppose that the variable of interest follows an AR(2) process. If the process is non-stationary, suppose further that κ is large enough so that*

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

Then

$$\phi_1, \phi_2 \neq 0 \Rightarrow a_1 \neq 0.$$

Proof. See Appendix C. ■

The two inequalities ensure that all conditional moments are well-defined in the case of a non-stationary AR(2) process.

The next proposition formally states the dynamic attention principle in the case of news shocks. In Macroeconomics, it is often assumed that a change in fiscal policy is announced before the change in government spending or taxes actually occurs or that an increase in productivity can be learned about before it actually occurs in production. In this case, the shock is called a "news shock." A standard example is

$$X_t = \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1} \quad \text{with} \quad \phi_1, \theta_1 \neq 0.$$

The shock can be learned about in period $t-1$ but only affects the variable of interest (e.g., some optimal action) in period t. In our language, the variable of interest follows an ARMA(1,1) process with $\theta_0 = 0$. It follows from Proposition 1 that an optimal one-dimensional signal has the form

$$S_t = a_0 X_t + b_0 \varepsilon_t + \psi_t.$$

This statement does not rule out the possibility that the optimal signal weight b_0 equals zero (i.e., the statement does not rule out the possibility that $S_t = X_t + \psi_t$ is an optimal signal.) Proposition

4 states that the optimal signal weight b_0 is *always* different from zero, even though the variable of interest in period t does not depend on ε_t. The reason is that the agent would like to enter well informed into the next period.

Proposition 4 *Suppose that the variable of interest follows an ARMA(1,1) process. If the process is non-stationary, suppose further that κ is large enough so that*

$$2^{2\kappa} > \phi_1^2.$$

Then

$$\phi_1, \theta_1 \neq 0, \theta_0 = 0 \Rightarrow b_0 \neq 0.$$

Proof. See Appendix D. ■

The inequality ensures that all conditional moments are well-defined in the case of a non-stationary ARMA(1,1) process with $\theta_0 = 0$.

4 Application to a model of price-setting

This section illustrates the usefulness of the paper's analytical results in the context of an economic model. We consider the model of price-setting under imperfect information proposed by Woodford (2002). Woodford supposes that firms set prices based on noisy signals about aggregate nominal 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. From our perspective, it is interesting that the signals assumed by Woodford are not optimal in his model economy. Let us resolve his model letting firms set prices based on optimal signals.

Woodford's model features an economy with a continuum of firms indexed by $i \in [0,1]$. Firm i sells good i. Every period $t = 1, 2, ...$, firm i sets the price of good i to maximize the present discounted value of profits. After a log-quadratic approximation, the objective function of the firm can be expressed as

$$E\left[\sum_{t=0}^{\infty} \beta^t \left(p_{it}^* - p_{it}\right)^2\right], \tag{17}$$

where $\beta \in (0,1)$ is a parameter, p_{it}^* is the profit-maximizing price of firm i in period t, $p_{it} = E\left[p_{it}^* \mid \mathcal{I}_{it}\right]$ is the actual price set by the firm, and $\mathcal{I}_{it} = \mathcal{I}_{i,0} \cup \{s_{i1}, \ldots, s_{it}\}$ is the information set of the decision-maker who is setting the price.[3] $\mathcal{I}_{i,0}$ denotes the information set of the decision-maker in period zero. In period zero the decision-maker receives information such that the conditional variance of p_{it}^* given information in period t, $E\left[(p_{it}^* - p_{it})^2 \mid \mathcal{I}_{it}\right]$, is constant for all $t \geq 0$. Hence objective (17) simplifies to $1/(1-\beta)$ times

$$E\left[(p_{it}^* - p_{it})^2\right], \tag{18}$$

where the expectations can be computed using the steady-state Kalman filter. Expression (18) is identical to the objective function in Section 2.

The profit-maximizing price in the Woodford model can be written as

$$p_{it}^* = \xi q_t + (1-\xi) p_t, \tag{19}$$

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. Nominal aggregate demand follows an exogenous stochastic process given by

$$q_t = (1+\rho) q_{t-1} - \rho q_{t-2} + u_t, \tag{20}$$

where $\rho \in [0,1)$ is a parameter and u_t is a zero-mean Gaussian white noise process. Woodford assumes that in every period the decision-maker in firm i observes a signal about nominal aggregate demand given by

$$s_{it} = q_t + v_{it}, \tag{21}$$

where v_{it} is a zero-mean Gaussian white noise error term, distributed independently both of the history of fundamental disturbances and of the observation errors of all other firms.

It is straightforward to solve for the equilibrium of the Woodford model numerically. An object of particular interest is the impulse response of output, $y_t = q_t - p_t$, to an innovation in nominal aggregate demand. Below we study how this impulse response changes when we relax Woodford's restriction that signals must be of the form "nominal aggregate demand plus i.i.d. noise."

4.1 The case without strategic complementarity in price-setting

To develop intuition, it is helpful to start with the case when $\xi = 1$ (no strategic complementarity in price-setting). The profit-maximizing price is then equal to nominal aggregate demand, $p_{it}^* = q_t$ (see

[3] For simplicity, we drop a constant multiplying the objective function after the log-quadratic approximation.

equation (19)). Since nominal aggregate demand follows a Gaussian AR(2) process (see equation (20)), we can apply the results in Section 2 to establish that the optimal signal in this case has the form

$$s_{it}^{*} = q_t + a_1 q_{t-1} + \psi_{it}, \tag{22}$$

where ψ_{it} is a zero-mean Gaussian white noise error term distributed independently of the history of fundamental disturbances.[4] To maximize profits (i.e., to minimize (18)) the decision-maker in firm i chooses a_1 and the variance of noise ψ_{it}, subject to the information flow constraint, and afterwards in every period the decision-maker computes p_{it} from the relation $p_{it} = E\left[p_{it}^{*} \mid \mathcal{I}_{it}\right]$ where $\mathcal{I}_{it} = \mathcal{I}_{i,0} \cup \{s_{i1}^{*}, \ldots, s_{it}^{*}\}$. Notice that assumption (21) in this case amounts to a simple restriction $a_1 = 0$.

To investigate to what extent Woodford's restriction on the signal matters we assume his parameter values.[5] Furthermore, we suppose that the information flow in the model with optimal signals (the model with s_{it}^{*} given by equation (22)) is equal to the information flow in the Woodford model (the model with s_{it} given by equation (21)).[6] 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 Woodford's restricted signals.[7]

The top panel in Figure 1 compares the equilibrium impulse responses of output to a nominal disturbance in the two models. Woodford's restriction matters a lot. The real effects of a nominal disturbances are much stronger in his model than in the model with optimal signals. With Woodford's restriction the variance of output rises by a factor of 2.5. At the same time, the marginal value of information and the profit loss from imperfect information each increase by about 20 percent. A decision-maker in the model with optimal signals uses a given amount of information as efficiently

[4] We follow Woodford in assuming that ψ_{it} is also independent of the observation errors of all other firms.

[5] The exception is that for the moment we set $\xi = 1$, whereas Woodford focuses on the case when $\xi = 0.15$. We study the case of $\xi = 0.15$ below.

[6] 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.

[7] To solve the model with optimal signals, we write the model in state-space form with (20) as the state equation and (22) as the observation equation. The information flow constraint is $(1/2)\log_2\left(\det \Sigma_1 / \det \Sigma_0\right) = \kappa_W$, where Σ_1 is the prior variance-covariance matrix of the state vector and Σ_0 is the posterior variance-covariance matrix of the state vector, in steady state.

as possible. Consequently, in the Woodford model the tracking of the profit-maximizing price is less accurate, an extra unit of information is more valuable, and a nominal shock has stronger real effects compared with in the model with optimal signals. Furthermore, the differences between the two models can be sizable.

4.2 The case with strategic complementarity in price-setting

Now consider the case of $\xi = 0.15$, as assumed in Woodford (2002), implying a significant degree of strategic complementarity in price-setting. We guess that in equilibrium the profit-maximizing price follows an ARMA(p,q) process for some finite $p \geq 1$ and $q \geq 0$. Given a guess we apply the results in Section 2 to establish the form of the optimal signal. For example, if the profit-maximizing price follows an ARMA(2,2) process, the optimal signal has the form

$$s_{it}^{*} = p_{it}^{*} + a_1 p_{i,t-1}^{*} + b_0 u_t + b_1 u_{t-1} + \psi_{it}. \tag{23}$$

We let the decision-maker in firm i choose the weights in the optimal signal (a_1, b_0, and b_1 in the ARMA(2,2) example) and the variance of noise ψ_{it} to maximize profits, subject to the information flow constraint. We then compute the aggregate price level using the relation $p_t = \int_0^1 p_{it} di$ and the profit-maximizing price using equation (19), and we iterate until convergence to a fixed point. As before, we suppose that the information flow in the model with optimal signals is equal to the information flow in the Woodford model.[8]

The bottom panel in Figure 1 compares the equilibrium impulse responses of output to a nominal disturbance in the two models in the case of $\xi = 0.15$. Woodford's restriction matters much less than with $\xi = 1$. With $\xi = 0.15$ the real effects of a nominal disturbance are only somewhat stronger in the Woodford model than in the model with optimal signals.[9] The Woodford model predicts stronger real effects than the model with optimal signals for any value of ξ, because information is always used less efficiently in the former model than in the latter model. By the same token, the marginal value of information is always greater in the Woodford model than in

[8] Again, we write the model in state-space form. The state equation is the law of motion for the profit-maximizing price, and the observation equation has the form of equation (23). We verify that we cannot improve convergence to a fixed point if we add parameters to the law of motion for the profit-maximizing price.

[9] With Woodford's restriction the variance of output, the marginal value of information, and the profit loss from imperfect information each increase by about 5 percent.

the model with optimal signals. At the same time, the difference between the size of real effects in the two models decreases as ξ falls from one to zero (as the degree of strategic complementarity in price-setting rises). To understand why, note that the firms' tracking problem becomes simpler as ξ falls from one to zero. The reason is that the response of the profit-maximizing price to a nominal shock weakens as the degree of strategic complementarity in price-setting rises. Thus the gap between the marginal value of information in the two models decreases as ξ falls from one to zero (Woodford's restriction on the signal becomes less harmful to the decision-maker), implying that the difference between the size of real effects in the models diminishes.

5 Conclusions

Solving dynamic rational inattention problems has become straightforward and intuitive. In the canonical dynamic attention choice problem an optimal signal has a simple form. One can solve for the optimal signal using the rational inattention filter, which is just the Kalman filter with the observation equation that is optimal from the perspective of rational inattention. The resulting behavior satisfies the dynamic attention principle implying that the agent is learning optimally both about the present and about the future.

A Proof of Lemma 1

First, we prove the lemma for $q = 0$. Let $X^t = \{X_1, \ldots, X_t\}$ denote the history of X's at time t and let $S^{K,t} = \left\{S_1^K, \ldots, S_t^K\right\}$ denote the history of signals at time t. The mutual information between two random vectors is symmetric and equals the difference between entropy and conditional entropy

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

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

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

Equation (2) for the signal implies that S_t^K depends only on X_t^M, ε_t^N, and ψ_t^K. In the following, let $\tau = \max\{M + p, N + p\}$. For $t \in \{\tau, \ldots, T\}$, one can compute X_t^M and ε_t^N from X^T. Furthermore, one can compute X_t^M and ε_t^N from the vector ξ_t defined in Lemma 1. Hence, for $t \in \{\tau, \ldots, T\}$,

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

Collecting results yields that, $\forall T \geq \tau$, the mutual information between X^T and $S^{K,T}$ equals

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

It follows from the symmetry of mutual information that

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

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

$$I\left(X^T; S^{K,T}\right) = H\left(S^{K,\tau-1}\right) - H\left(S^{K,\tau-1}|X^T\right) + \sum_{t=\tau}^{T} \left[H\left(\xi_t|S^{K,t-1}\right) - H\left(\xi_t|S^{K,t}\right)\right]. \tag{24}$$

We show in Appendix B that the following limit exists

$$\lim_{T\to\infty} \left[H\left(\xi_T|S^{K,T-1}\right) - H\left(\xi_T|S^{K,T}\right)\right]. \tag{25}$$

Equation (24) and Cesaro mean then imply

$$\lim_{T\to\infty}\frac{1}{T}I\left(X^T;S^{K,T}\right)=\lim_{T\to\infty}\left[H\left(\xi_T|S^{K,T-1}\right)-H\left(\xi_T|S^{K,T}\right)\right]. \tag{26}$$

Second, we prove the lemma for $q>0$. Let $\varepsilon_t^q=\{\varepsilon_t,\ldots,\varepsilon_{t-q+1}\}$. The mutual information between X^T and ε_T^q, on the one hand, and $S^{K,T}$, on the other hand, equals

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

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

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

Equation (2) for the signal implies that S_t^K depends only on X_t^M, ε_t^N, and ψ_t^K. In the following, let $\tau=\max\{M+p,N+p\}$. For $t\in\{\tau,\ldots,T\}$, one can compute X_t^M and ε_t^N from X^T and ε_T^q. Furthermore, one can compute X_t^M and ε_t^N from the vector ξ_t defined in Lemma 1. Hence, for $t\in\{\tau,\ldots,T\}$,

$$H\left(S_t^K|S^{K,t-1},X^T,\varepsilon_T^q\right)=H\left(S_t^K|S^{K,t-1},\xi_t\right)=H\left(\psi_t^K\right).$$

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

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

It follows from the symmetry of mutual information that

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

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

$$I\left(X^T,\varepsilon_T^q;S^{K,T}\right)=H\left(S^{K,\tau-1}\right)-H\left(S^{K,\tau-1}|X^T,\varepsilon_T^q\right)+\sum_{t=\tau}^{T}\left[H\left(\xi_t|S^{K,t-1}\right)-H\left(\xi_t|S^{K,t}\right)\right]. \tag{27}$$

We show in Appendix B that the following limit exists

$$\lim_{T\to\infty}\left[H\left(\xi_T|S^{K,T-1}\right)-H\left(\xi_T|S^{K,T}\right)\right]. \tag{28}$$

Equation (27) and Cesaro mean then imply

$$\lim_{T\to\infty}\frac{1}{T}I\left(X^T,\varepsilon_T^q;S^{K,T}\right)=\lim_{T\to\infty}\left[H\left(\xi_T|S^{K,T-1}\right)-H\left(\xi_T|S^{K,T}\right)\right]. \tag{29}$$

B Proof of Proposition 1

First, the signal (2) has the following state-space representation (following the notation in Hamilton (1994), Chapter 13)

$$\xi_{t+1} = F\xi_t + v_{t+1}, \tag{30}$$

$$S_t^K = G'\xi_t + \psi_t^K. \tag{31}$$

Here ξ_t is the vector defined in Lemma 1. The matrix F is a square matrix and the length of the column vector v_{t+1} equals the length of the column vector ξ_{t+1}. In the case of $q = 0$, the first element of v_{t+1} equals $\theta_0\varepsilon_{t+1}$ and the other elements of v_{t+1} equal zero. Furthermore, for $q = 0$, the first row of the matrix F ensures that the first row of equation (30) equals equation (1) and the other rows of the matrix F have a one just left of the main diagonal and zeros everywhere else. In the case of $q > 0$, the first element of v_{t+1} equals ε_{t+1}, the $q+1$ element of v_{t+1} equals $\theta_0\varepsilon_{t+1}$, and the other elements of v_{t+1} equal zero. Furthermore, for $q > 0$, the first row of the matrix F contains only zeros, the $q+1$ row of F ensures that the $q+1$ row of equation (30) equals equation (1), and the other 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 (31) equals equation (2). Such a matrix exists because X_t^M and ε_t^N can be computed from ξ_t.

Second, let $\Sigma_{t|t-1}$ denote the conditional variance-covariance matrix of ξ_t given $S^{K,t-1}$. Let $\Sigma_{t|t}$ denote the conditional variance-covariance matrix of ξ_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

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

$$\Sigma_0 = \Sigma_1 - \Sigma_1 G\left(G'\Sigma_1 G + R\right)^{-1} G'\Sigma_1,$$

where Q denotes the variance-covariance matrix of the innovation in the state equation (30) and R denotes the variance-covariance matrix of the innovation in the observation equation (31).

Third, one can express the information flow constraint (5) in terms of the matrices Σ_1 and Σ_0. According to Lemma 1 the information flow constraint (5) is equivalent to

$$\lim_{T\to\infty}\left[H\left(\xi_T|S^{K,T-1}\right) - H\left(\xi_T|S^{K,T}\right)\right] \leq \kappa. \tag{32}$$

Conditional normality implies that

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

Hence, the information flow constraint (5) is equivalent to

$$\frac{1}{2}\log_2\left(\frac{\det\Sigma_1}{\det\Sigma_0}\right)\leq\kappa. \tag{33}$$

Fourth, if one compares two signal vectors with $M>p$ and $N>q$ that yield the same upper-left $[(p+q)\times(p+q)]$ sub-matrix of Σ_0 and one signal vector has the property

$$\begin{array}{l}\forall j>p:A_{ij}=0\\ \forall j>q:B_{ij}=0\end{array}, \tag{34}$$

while the other signal vector does not have this property, then the first signal vector yields a lower value of the left-hand side of the information flow constraint (32) although both signal vectors imply the same value of the objective (4). Both signal vectors imply the same value of the objective (4), because the objective equals an element of the upper-left $[(p+q)\times(p+q)]$ sub-matrix of Σ_0. To see that the first signal vector yields a lower value of the left-hand side of the information flow constraint (32), let us define the vector ξ_t^{up}, which is a sub-vector of ξ_t,

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

Let ξ_t^{low} be the vector that contains the remaining elements of ξ_t. The information flow constraint (32) can be written as

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

Furthermore, two signal vectors that yield the same upper-left $[(p+q)\times(p+q)]$ sub-matrix of Σ_0 generate the same limit

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

because two signal vectors that yield the same upper-left $[(p+q)\times(p+q)]$ sub-matrix of Σ_0 also generate the same upper-left $[(p+q)\times(p+q)]$ sub-matrix of Σ_1, because X_t follows an ARMA(p,q) process. Finally, the difference

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

is non-negative, since conditioning weakly reduces entropy, and it equals zero if and only if condition (34) is satisfied.

Fifth, we show that for any signal vector violating condition (34) there exists a signal vector satisfying condition (34) that yields the same upper-left $[(p+q) \times (p+q)]$ 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 $\left\{K, \hat{A}, \hat{B}, \hat{\Upsilon}\right\}$ violating condition (34) 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 $[(p+q) \times (p+q)]$ sub-matrix of $\Sigma_1 - \Sigma_0$ is positive semi-definite, too. Using the statement above, this implies that there exists a signal $\{K, A, B, \Upsilon\}$ satisfying condition (34) that generates the upper-left $[(p+q) \times (p+q)]$ sub-matrix of Σ_0 from the upper-left $[(p+q) \times (p+q)]$ sub-matrix of Σ_1.

Sixth, the agent can attain the optimum with a one-dimensional signal. Suppose the agent chooses a signal with $K > 1$. Then, the agent can combine the different elements of $S_t^K = (S_{t,1}, \ldots, S_{t,K})'$ into a single signal by using the Kalman weights of the steady-state Kalman filter. The new signal has $K = 1$ and yields the same matrices Σ_1 and Σ_0, implying that the new signal yields the same value of the left-hand side of information flow constraint (33) and of objective (4).

C Proof of Proposition 3

Equation (15) at $p = 2$ and $a_1 = 0$ reads

$$Z^{11}\frac{d\Sigma_{11}}{da_1} + Z^{12}\frac{d\Sigma_{12}}{da_1} + Z^{21}\frac{d\Sigma_{21}}{da_1} + Z^{22}\frac{d\Sigma_{22}}{da_1} + Z^2 = 0, \tag{35}$$

with

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

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

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

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

and

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

Equation (14) at $p = 2$ and $a_1 = 0$ implies

$$\Sigma_{22} = 2^{-2\kappa}\Sigma_{11}, \tag{36}$$

$$\Sigma_{12} = \Sigma_{21} = \frac{\phi_1}{2^{2\kappa} - \phi_2}\Sigma_{11}, \tag{37}$$

$$\Sigma_{11} = \frac{\theta_0^2}{1 - 2^{-2\kappa}\left(\phi_1^2 + \phi_2^2\right) - 2^{-2\kappa}\frac{2\phi_1^2\phi_2}{2^{2\kappa}-\phi_2} + \frac{(1-2^{-2\kappa})\phi_1^2\phi_2^2}{(2^{2\kappa}-\phi_2)^2}}. \tag{38}$$

In the case of a stationary AR(2) process, the denominators in equations (37) and (38) 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 (35) is a system of four linear equations in $\frac{d\Sigma_{11}}{da_1}$, $\frac{d\Sigma_{12}}{da_1}$, $\frac{d\Sigma_{21}}{da_1}$, and $\frac{d\Sigma_{22}}{da_1}$. Solving this system for $\frac{d\Sigma_{22}}{da_1}$ and using equations (36)-(38) yields

$$\frac{d\Sigma_{22}}{da_1} = -2\phi_1\phi_2\underbrace{\frac{\left(1 - 2^{-2\kappa}\right)\Sigma_{11}}{\left(2^{2\kappa} - \phi_2\right)^2}}_{>0}.$$

Hence, in the AR(2) case, the first-order condition (16) is satisfied at $a_1 = 0$ if and only if $\phi_1\phi_2 = 0$.

D Proof of Proposition 4

Equation (15) at $p = 1$, $q = 1$, and $b_0 = 0$ reads

$$Z^{11}\frac{d\Sigma_{11}}{db_0} + Z^{12}\frac{d\Sigma_{12}}{db_0} + Z^{21}\frac{d\Sigma_{21}}{db_0} + Z^{22}\frac{d\Sigma_{22}}{db_0} + Z^2 = 0,$$

with

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

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

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

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

and

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

These equations imply

$$\frac{d\Sigma_{12}}{db_0} = \frac{d\Sigma_{21}}{db_0} = \frac{d\Sigma_{22}}{db_0} = 0,$$

and

$$\left[1 - \phi_1^2 2^{-2\kappa} - \theta_1^2\left(1-2^{-2\kappa}\right)\frac{\Sigma_{12}\Sigma_{21}}{\Sigma_{11}^2}\right]\frac{d\Sigma_{11}}{db_0} + \left(1-2^{-2\kappa}\right)\frac{\Sigma_{11}\Sigma_{22} - \Sigma_{12}\Sigma_{21}}{\Sigma_{11}}\left[2\phi_1\theta_1 + \theta_1^2\frac{\Sigma_{12}+\Sigma_{21}}{\Sigma_{11}}\right] = 0.$$

Equation (14) at $p = 1$, $q = 1$, and $b_0 = 0$ reads

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

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

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

and the previous equation can be written as

$$\frac{d\Sigma_{11}}{db_0} = -\frac{\left(1-2^{-2\kappa}\right)}{\left(1-\phi_1^2 2^{-2\kappa}\right)} 2\phi_1\theta_1. \tag{39}$$

Furthermore, using equation (11) to substitute for the variance of noise in equation (8) yields

$$\Sigma_0 = \Sigma - \frac{\left(1-2^{-2\kappa}\right)}{h'\Sigma h}\Sigma h h' \Sigma.$$

In the case of $p = 1$ and $q = 1$, the last equation reads

$$\begin{bmatrix} \Sigma_{11,0} & \Sigma_{12,0} \\ \Sigma_{21,0} & \Sigma_{22,0} \end{bmatrix} = \begin{bmatrix} \Sigma_{11} - \left(1-2^{-2\kappa}\right)\frac{(\Sigma_{11}+b_0\Sigma_{12})(\Sigma_{11}+b_0\Sigma_{21})}{\Sigma_{11}+b_0\Sigma_{12}+b_0\Sigma_{21}+b_0^2\Sigma_{22}} & \Sigma_{12} - \left(1-2^{-2\kappa}\right)\frac{(\Sigma_{11}+b_0\Sigma_{12})(\Sigma_{12}+b_0\Sigma_{22})}{\Sigma_{11}+b_0\Sigma_{12}+b_0\Sigma_{21}+b_0^2\Sigma_{22}} \\ \Sigma_{21} - \left(1-2^{-2\kappa}\right)\frac{(\Sigma_{21}+b_0\Sigma_{22})(\Sigma_{11}+b_0\Sigma_{21})}{\Sigma_{11}+b_0\Sigma_{12}+b_0\Sigma_{21}+b_0^2\Sigma_{22}} & \Sigma_{22} - \left(1-2^{-2\kappa}\right)\frac{(\Sigma_{21}+b_0\Sigma_{22})(\Sigma_{12}+b_0\Sigma_{22})}{\Sigma_{11}+b_0\Sigma_{12}+b_0\Sigma_{21}+b_0^2\Sigma_{22}} \end{bmatrix}.$$

The upper-left equation implies that the derivative of $\Sigma_{11,0}$ with respect to b_0 at the point $b_0 = 0$ equals

$$\frac{d\Sigma_{11,0}}{db_0} = 2^{-2\kappa}\frac{d\Sigma_{11}}{db_0}. \tag{40}$$

It follows from equations (39) and (40) 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.

References

[1] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (2013). *Nonlinear programming: theory and algorithms*. John Wiley & Sons.

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

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

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

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

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

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

[8] Steiner, Jakub, Colin Stewart, and Filip Matějka (2015). "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," Discussion paper, CERGE-EI, University of Edinburgh, and University of Toronto.

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

[10] 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.

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

Figure 1: An ARMA(2,1) example

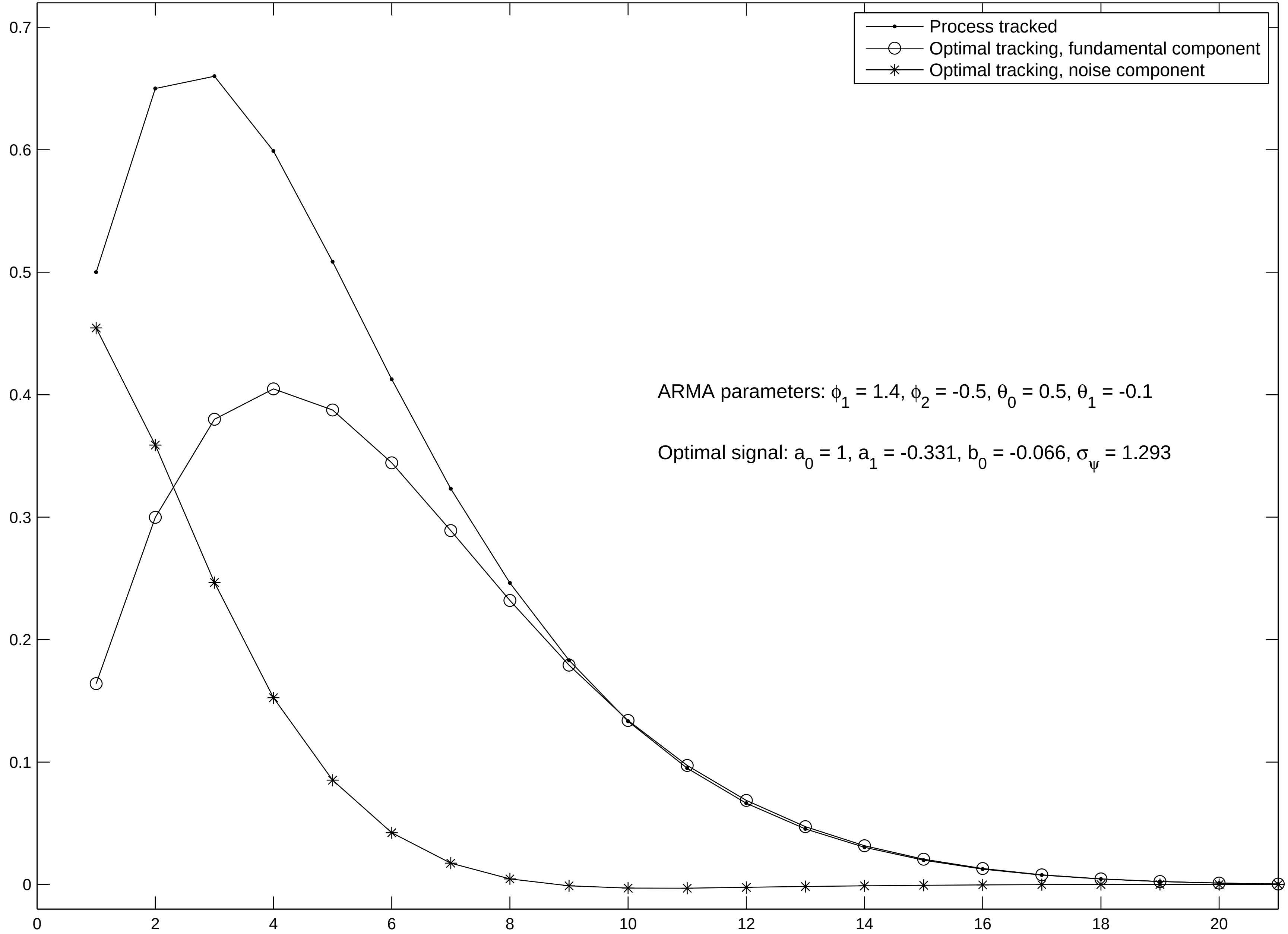

Figure 2: Impulse responses of output to a nominal shock, Woodford model and model with optimal signals