

Maćkowiak, Bartosz; Matějka, Filip; Wiederholt, Mirko

**Working Paper**

## The rational inattention filter

ECB Working Paper, No. 2007

**Provided in Cooperation with:**

European Central Bank (ECB)

*Suggested Citation:* Maćkowiak, Bartosz; Matějka, Filip; Wiederholt, Mirko (2017) : The rational inattention filter, ECB Working Paper, No. 2007, ISBN 978-92-899-2728-4, European Central Bank (ECB), Frankfurt a. M.,  
<https://doi.org/10.2866/339353>

This Version is available at:

<https://hdl.handle.net/10419/154440>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



EUROPEAN CENTRAL BANK

EUROSYSTEM

## Working Paper Series

Bartosz Maćkowiak, Filip Matějka,  
Mirko Wiederholt

### The rational inattention filter

No 2007 / January 2017

**Disclaimer:** This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

## 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  $\{X_t, \dots, X_{t-p+1}\}$  and  $\{\varepsilon_t, \dots, \varepsilon_{t-q+1}\}$ , where  $X_t$  denotes the variable of interest and  $\varepsilon_t$  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 (**JEL:** D83, E32)

## Non-technical summary

The recent literature on rational inattention studies decision-making when human attention is scarce. Sims (1998, 2003) formalized limited attention as a constraint on information flow and proposed to model decision-making with limited attention as optimization subject to this constraint. Sims's motivation was 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. Despite Sims's focus on intertemporal settings, most work on rational inattention thus far has solved static models or analyzed 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.

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 that the agent will receive, subject to the constraint on the flow of information between the signals and the variable of interest. The agent aims to minimize the mean squared error between the variable of interest and the action taken based on the signals.

We focus on the case when the variable of interest,  $X(t)$ , follows an ARMA( $p,q$ ) process, because it is well known from Time Series Econometrics that the evolution of many economic variables can be well described by a low-order ARMA( $p,q$ ) process. We prove that any optimal signal with i.i.d. noise is only about  $X(t)$  and the variables that appear in the best predictor of  $X(t+1)$  given full information at time  $t$ . In addition, we show that the agent can attain the optimum with a one-dimensional signal. For example, if  $X(t)$  follows an ARMA(2,1) process, one can restrict attention to signals of the form  $S(t) = a_0X(t) + a_1X(t-1) + b_0\varepsilon(t) + \psi(t)$ , where  $\varepsilon(t)$  is the innovation in  $X(t)$  in period  $t$  and  $\psi(t)$  is the i.i.d. noise. One only has to solve for the remaining signal weights ( $a_0, a_1, b_0$ ) and the variance of the noise. The remaining signal weights, the variance of the noise, and the implied actions can be computed from what we call the "rational inattention filter." The rational inattention filter is the Kalman filter with the observation equation that is optimal from the perspective of rational inattention. Hence, anyone familiar with the Kalman filter can easily solve dynamic rational inattention problems, as in Sims (2003), without any loss in generality.

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 about entering well informed into the next period. Entering well informed into the next period relaxes the agent's attention constraint. If the variable of interest follows an AR(1) process, there is no tension between these two goals. Therefore, the optimal signal is  $S(t) = X(t) + \psi(t)$ . We show that beyond an AR(1) process, there is a tension between these two goals and hence the optimal signal is *generically not*  $S(t) = X(t) + \psi(t)$ . We refer to this property as the "dynamic attention principle."

We apply the paper's analytical results in the context of two macroeconomic models, a business cycle model with news shocks about productivity and the model of price-setting proposed by Woodford (2002). Let us focus on the former application here. A popular assumption in business cycle models is that a change in productivity can be learned about before it actually occurs in production ("a news shock"). While fluctuations in expectations about future productivity 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 about productivity. 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. We point out that rational inattention on the side of firms is a force pushing labor input up after a positive news shock about productivity. The reason is that 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* on impact of the news shock. The result that rationally inattentive firms choose not to distinguish carefully between current and future increases in productivity follows directly from the dynamic attention principle.

# 1 Introduction

The recent literature on rational inattention studies decision-making when human attention is scarce. Sims (1998, 2003) formalized limited attention as a constraint on information flow and proposed to model decision-making with limited attention as optimization subject to this constraint. Sims' motivation was 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. Despite Sims's focus on intertemporal settings, most work on rational inattention thus far has solved static models or analyzed 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.

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 that the agent will receive, subject to the constraint on the flow of information between the signals and the variable of interest. The agent aims to minimize the mean squared error between the variable of interest and the action taken based on the signals.

We focus on the case when the variable of interest follows an ARMA(p,q) process, because it is well known from Time Series Econometrics that the evolution of many economic variables can be well described by a low-order ARMA(p,q) process

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

where  $X_t$  denotes the variable of interest in period  $t$  and  $\varepsilon_t$  is the innovation in  $X_t$  in period  $t$ . We prove that any optimal signal with i.i.d. noise is only about  $X_t$  and the variables that appear in the best predictor of  $X_{t+1}$  given full information at time  $t$ . In addition, we show that 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

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

where  $\psi_t$  is the i.i.d. noise. For example, if the variable of interest follows an ARMA(2,1) process, one can restrict attention to signals of the form  $S_t = a_0 X_t + a_1 X_{t-1} + b_0 \varepsilon_t + \psi_t$ . One only has to solve for the remaining signal weights  $a_0$ ,  $a_1$ , and  $b_0$ , and for the variance of noise,  $\sigma_\psi^2$ .

The question then becomes: What are the remaining signal weights and the implied actions? The optimal signal weights and the implied actions can be computed from 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 signal given in the previous paragraph has a simple state-space representation. Furthermore, one can derive first-order conditions for the optimal signal weights. The rational inattention filter consists of the usual Kalman filter equations and these first-order conditions for the optimal signal weights. Hence, anyone familiar with the Kalman filter can easily solve dynamic rational inattention problems, as in Sims (2003), without any loss in generality.

We then proceed by deriving analytical results regarding the remaining signal weights. Our first analytical result about 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 about entering well informed into the next period. Entering well informed into the next period relaxes the agent’s attention constraint. 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. Therefore, the optimal signal is  $S_t = X_t + \psi_t$ . Beyond an AR(1) process, there is a tension between these two goals and hence the optimal signal is *generically not*  $S_t = X_t + \psi_t$ .<sup>1</sup> 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 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 analytical result about the remaining signal weights is that if the information flow is very large, then it becomes optimal for the agent to process information mostly about the current optimal action  $X_t$  only.

---

<sup>1</sup>This result is not already implied by the earlier statements, because those statements do not rule out the possibility that the optimal signal weights are  $a_1 = \dots = a_{p-1} = b_0 = \dots = b_{q-1} = 0$ .

Finally, we prove the equivalence of a number of different formulations of the information flow constraint that have appeared in the literature. Namely, we prove the equivalence of a constraint on the information flow between sequences, a recursive formulation of the information flow constraint, and a constraint on a particular signal-to-noise ratio.

We apply the paper's analytical results in the context of two macroeconomic models, a business cycle model with news shocks about productivity and the model of price-setting proposed by Woodford (2002). Let us focus on the former application here. A popular assumption in business cycle models is that a change in productivity can be learned about before it actually occurs in production ("a news shock"). While fluctuations in expectations about future productivity 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 about productivity. 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. We point out that rational inattention on the side of firms is a force pushing labor input up after a positive news shock about productivity. The reason is that 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* on impact of the news shock.

In the paper we illustrate this observation in a simple model in which firms hire labor subject to rational inattention and households are hand-to-mouth consumers. Here let us explain the intuition with an example. Suppose that innovations in productivity can be learned about 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  $\phi_1, \theta_1 \neq 0$  and  $\theta_0 = 0$ . The paper's analytical results imply that a manager who makes the labor hiring decision can restrict attention to signals of the form  $S_t = a_0 z_t + b_0 \varepsilon_t + \psi_t$  with  $b_0 \neq 0$ . Hence, a manager who optimally allocates attention chooses a one-dimensional signal with a non-zero weight on current productivity,  $z_t$ , and a non-zero weight on the news shock,  $\varepsilon_t$ . The manager chooses the non-zero weight on the news shock because entering well informed into the next period relaxes the manager's attention constraint. The optimal signal has the following implications for actions. On impact of a positive news shock,  $\varepsilon_t > 0$ , the signal increases. Since the manager *chose* not to distinguish carefully between increases in current productivity and increases in future productivity,



the manager starts hiring already today.

We noted that most existing work on rational inattention solves static models or analyzes economies that are independent over time. The papers that do study dynamic economies that are correlated over time normally take one of the following three simplifying approaches: (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 by brute-force numerical optimization directly for the actions under rational inattention (Sims, 2003, Section 4, Maćkowiak and Wiederholt, 2015).<sup>2</sup>

There are two exceptions. Maćkowiak and Wiederholt (2009, Section V) find analytically an optimal signal in a special case of the dynamic attention choice problem analyzed here, the AR(1) case. Furthermore, Steiner, Stewart, and Matějka (2015) study a general dynamic model with discrete choice under rational inattention. They show that the dynamic problem can be reduced to a collection of static problems, and that the solution takes the form of a dynamic logit with endogenous biases. By contrast, this paper's solution method applies to the class of dynamic problems with Gaussian actions proposed by Sims (2003).

The following section presents the dynamic attention choice problem. Section 3 contains the main analytical results: the dimensionality reduction result for an ARMA(p,q) process, and the result that one can attain the optimum with a one-dimensional signal. Section 4 lays out the rational inattention filter and the dynamic attention principle. The application to the business cycle model with news shocks is in Section 5. The application to the price-setting model is in Section 6. Section 7 concludes.

## 2 Decision problem

In this section we present the dynamic 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 so as to minimize the mean squared error, subject to a constraint on information flow.

The variable of interest, denoted  $X_t$ , follows a stationary Gaussian process. This process can

---

<sup>2</sup>See Sims (2010), Veldkamp (2011), or Wiederholt (2010) for a review of the literature on rational inattention.

be an AR(p) process

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

an MA(q) process

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

or an ARMA(p,q) process

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \theta_0 \varepsilon_t + \dots + \theta_q \varepsilon_{t-q}, \quad (1)$$

where  $p \geq 1$  and  $q \geq 0$  are integers,  $\phi_1, \dots, \phi_p$  and  $\theta_0, \dots, \theta_q$  are coefficients, and  $\varepsilon_t$  follows a Gaussian white noise process with unit variance.<sup>3</sup>

At each time  $t \geq 1$ , the agent receives a  $K$ -dimensional signal vector,  $S_t^K = (S_{t,1}, \dots, S_{t,K})'$ , with  $K \geq 1$ , where each signal is about a potentially different linear combination of current and past  $X_t$  and current and past  $\varepsilon_t$

$$S_t^K = AX_t^M + B\varepsilon_t^N + \psi_t^K. \quad (2)$$

Here  $X_t^M = (X_t, \dots, X_{t-M+1})'$  is the vector of current and past  $X_t$ ,  $\varepsilon_t^N = (\varepsilon_t, \dots, \varepsilon_{t-N+1})'$  is the vector of current and past  $\varepsilon_t$ ,  $M \geq \max\{p, 1\}$  and  $N \geq \max\{q, 1\}$  are arbitrarily large integers, and  $A \in \mathbb{R}^{K \times M}$  and  $B \in \mathbb{R}^{K \times N}$  are matrices of coefficients. The noise vector  $\psi_t^K = (\psi_{t,1}, \dots, \psi_{t,K})'$  follows a Gaussian vector white noise process with variance-covariance matrix  $\Sigma_\psi$ . The agent chooses  $K$ ,  $A$ ,  $B$ , and  $\Sigma_\psi$ .

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

$$\mathcal{I}_t = \mathcal{I}_0 \cup \{S_1^K, \dots, S_t^K\}. \quad (3)$$

The agent chooses the number of signals,  $K$ , what the signals are about,  $A$  and  $B$ , as well as the variance-covariance matrix of noise,  $\Sigma_\psi$ . The agent aims to minimize the mean squared error, subject to a constraint on information flow. Formally, the agent solves

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

---

<sup>3</sup>Without loss in generality, the coefficients on the largest lags are required to be non-zero,  $\phi_p \neq 0$  and  $\theta_q \neq 0$ .

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

$$\lim_{T \rightarrow \infty} \frac{1}{T} I(Z_0, \varepsilon_1, \dots, \varepsilon_T; S_1^K, \dots, S_T^K) \leq \kappa. \quad (5)$$

Here  $Z_0$  is the vector of initial conditions, which consists of  $X_0, \dots, X_{1-p}$  in the AR(p) case,  $\varepsilon_0, \dots, \varepsilon_{1-q}$  in the MA(q) case, and  $X_0, \dots, X_{1-p}, \varepsilon_0, \dots, \varepsilon_{1-q}$  in the ARMA(p,q) case.

The information flow constraint (5) formalizes the idea that the agent has a limited amount of attention. The constraint restricts the information flow to the agent. Here  $\kappa$  is a parameter. In general, the mutual information between two random vectors  $X^T$  and  $S^T$ , denoted  $I(X^T; S^T)$ , equals the difference between unconditional uncertainty and conditional uncertainty

$$I(X^T; S^T) = H(X^T) - H(X^T|S^T),$$

where  $H(X^T)$  denotes the entropy of the random vector  $X^T$  and  $H(X^T|S^T)$  denotes the conditional entropy of the vector  $X^T$  given knowledge of  $S^T$ . Entropy is simply a measure of uncertainty. Hence, the term  $\lim_{T \rightarrow \infty} \frac{1}{T} H(Z_0, \varepsilon_1, \dots, \varepsilon_T)$  quantifies how total uncertainty grows with time in the absence of signals and the term  $\lim_{T \rightarrow \infty} \frac{1}{T} H(Z_0, \varepsilon_1, \dots, \varepsilon_T | S_1^K, \dots, S_T^K)$  quantifies how total uncertainty grows with time in the presence of the signals. The difference between the two terms measures the information flow to the agent.

Finally, the mean squared error in (4) is computed using the steady-state Kalman filter. A very similar dynamic rational inattention problem is formulated and studied numerically in Sims (2003, Section 4).

**Equivalent formulations of the information flow constraint.** Many formulations of the information flow constraint are equivalent. For example, each sequence on the left-hand side of the information flow constraint (5) can be replaced by any other sequence with the property that the new sequence can be computed from the original sequence and vice versa. Thus, in the AR(p) case, the information flow constraint (5) is equivalent to

$$\lim_{T \rightarrow \infty} \frac{1}{T} I(X_{1-p}, \dots, X_0, X_1, \dots, X_T; S_1^K, \dots, S_T^K) \leq \kappa. \quad (6)$$

Furthermore, dropping  $X_{1-p}, \dots, X_0$  in (6) does not affect the limit in (6).

More importantly, the information flow constraint (5) 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 to the best of our knowledge and is formally stated in the following lemma.

**Lemma 1** Let  $\xi_t$  denote the following vector

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

Let  $S^{K,t} = \{S_1^K, \dots, S_t^K\}$  denote the set of signals received up to and including time  $t$ . The information flow constraint (5) is equivalent to

$$\lim_{T \rightarrow \infty} [H(\xi_T | S^{K,T-1}) - H(\xi_T | S^{K,T})] \leq \kappa. \quad (7)$$

**Proof.** See Appendix A. ■

In words, the left-hand side of the information flow constraint (7) is simply the difference between prior uncertainty and posterior uncertainty at a given point in time about all variables that the new signal can be about. In this formulation of the constraint, the vector  $\xi_t$  can be any vector that has two properties: (i)  $X_t^M$  and  $\varepsilon_t^N$  can be computed from  $\xi_t$ , and (ii)  $\xi_t$  does not contain any redundant elements. The particular vector  $\xi_t$  defined in the lemma is only an example.<sup>4</sup> Lemma 1 establishes the equivalence of two formulations of the information flow constraint that have appeared in the literature and will be used in the following section to prove one of our main results. Furthermore, in Section 4 we show that the information flow constraint (7) is equivalent to a constraint on a particular signal-to-noise ratio once  $K = 1$ .

**Equivalent formulations of the objective.** One can also think of the agent as choosing the properties of the stochastic process for the signal vector (i.e.,  $K$ ,  $A$ ,  $B$ , and  $\Sigma_\psi$ ) in period zero so as to minimize the discounted sum of future mean squared errors

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

After the agent has chosen the properties of the signal vector in period zero, the agent receives a long sequence of signal vectors such that the conditional variance-covariance matrix of  $\xi_1$  given information in period zero equals the steady-state conditional variance-covariance matrix of  $\xi_t$  given

---

<sup>4</sup>When  $q = 0$  and  $M - p \geq N$ , the vectors  $X_t^M$  and  $\varepsilon_t^N$  can be computed from  $X_t^M$ . When  $q = 0$  and  $M - p < N$ , one needs  $N - (M - p)$  additional lags of  $X_t$  to compute the vectors  $X_t^M$  and  $\varepsilon_t^N$ . Furthermore, when  $q > 0$ , one can compute the moving-average terms  $\{\theta_0 \varepsilon_\tau + \dots + \theta_q \varepsilon_{\tau-q}\}_{\tau=t}^{t-M+p+1}$  from  $X_t^M$ . To compute the innovations  $\{\varepsilon_\tau\}_{\tau=t}^{t-N+1}$  one also needs  $\varepsilon_t, \dots, \varepsilon_{t-q+1}$  and additional lags of  $X_t$  if  $M - p + q < N$ .

information in period  $t - 1$ . The mean squared error is then the same in every period  $t \geq 1$ , and the discounted sum of future mean squared errors can be expressed as

$$\frac{\beta}{1 - \beta} E \left[ (X_t - E[X_t | \mathcal{I}_t])^2 \right].$$

Dividing by  $\beta / (1 - \beta)$ , which is simply a monotone transformation of the objective, yields objective (4).

### 3 Main analytical results

The following two propositions characterize two properties of a solution to the dynamic rational inattention problem (1)-(5).

**Proposition 1** *In the ARMA(p,q) case, any optimal signal vector is on linear combinations of  $\{X_t, \dots, X_{t-(p-1)}\}$  and  $\{\varepsilon_t, \dots, \varepsilon_{t-(q-1)}\}$  only. In the AR(p) case, any optimal signal vector is on linear combinations of  $\{X_t, \dots, X_{t-(p-1)}\}$  only. In the MA(q) case with  $q > 0$ , any optimal signal vector is on linear combinations of  $X_t$  and  $\{\varepsilon_t, \dots, \varepsilon_{t-(q-1)}\}$  only. In the white noise case, any optimal signal vector is on  $X_t$  only.*

**Proof.** See Appendix B. ■

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

**Proof.** See Appendix C. ■

These analytical results imply that one can reduce the dimensionality of the problem (1)-(5) significantly without any loss in generality. For example, in the ARMA(p,q) case, one can restrict attention to signals of the form

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

In the next section, we present a simple way of computing the remaining 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 proofs of Lemma 1 and Propositions 1-2 are well-defined. For example, let  $\Sigma_{t|t-1}$  denote the conditional variance-covariance

matrix of  $\xi_t$  given  $\mathcal{I}_{t-1}$ . Let  $\Sigma_{t|t}$  denote the conditional variance-covariance matrix of  $\xi_t$  given  $\mathcal{I}_t$ . Furthermore, let  $\Sigma_1$  and  $\Sigma_0$  denote  $\lim_{t \rightarrow \infty} \Sigma_{t|t-1}$  and  $\lim_{t \rightarrow \infty} \Sigma_{t|t}$ , respectively. Objective (4) and constraint (5) depend on  $\Sigma_1$  and  $\Sigma_0$ . The assumption that the variable  $X_t$  follows a stationary process ensures that  $\Sigma_1$  and  $\Sigma_0$  are well-defined. One can relax the stationarity assumption. Propositions 1-2 extend to the case of a non-stationary ARMA(p,q) process so long as all conditional moments appearing in the proofs of Propositions 1-2 are well-defined. This requires that the parameter  $\kappa$  is sufficiently large.<sup>5</sup>

## 4 Rational inattention filter and dynamic attention principle

In the previous section, we showed that any optimal signal vector  $S_t^K$  is only about  $X_t$  and the variables that appear in the best predictor of  $X_{t+1}$  given full information at time  $t$ . In addition, we showed that the agent can attain the optimum with a one-dimensional signal. Hence, in this section, we restrict attention to signals of this form and we focus on how to compute the remaining signal weights and the variance of noise in the optimal signal.

**The state-space representation of the signal.** Propositions 1 and 2 imply that one can restrict attention to signals that have the following state-space representation

$$\xi_{t+1} = F\xi_t + v_{t+1}, \quad (8)$$

$$S_t = h'\xi_t + \psi_t, \quad (9)$$

where the state vector  $\xi_t$  is given by

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

The matrix  $F$  is a square matrix and the length of the column vector  $v_t$  equals the length of  $\xi_t$ . 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

---

<sup>5</sup>Moreover, Propositions 1-2 also hold for more general objectives. Proposition 1 holds for any objective that is a function only of the elements of  $\Sigma_0$ , while Proposition 2 holds for any objective that is a function only of the elements of  $\Sigma_0$  and has the property that pure delay in the arrival of signals makes the agent worse off.

row of the state equation (8) equals the law of motion for  $X_t$ . In the AR( $p$ ) case with  $p > 1$ , 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  $q > 0$ , 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  $\varepsilon_{t+1}$  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 vector of signal weights  $h$  is a column vector with  $\max\{1, p\} + q$  elements and the noise term  $\psi_t$  follows a Gaussian white noise process with variance  $\sigma_\psi^2 > 0$ . The vector of signal weights  $h$  and the variance of noise  $\sigma_\psi^2$  are the objects of interest of this section.

Let  $\Sigma_{t|t}$  denote the conditional variance-covariance matrix of  $\xi_t$  given  $\mathcal{I}_t$ , let  $\Sigma_{t|t-1}$  denote the conditional variance-covariance matrix of  $\xi_t$  given  $\mathcal{I}_{t-1}$ , and let  $Q$  denote the variance-covariance matrix of  $v_{t+1}$ . For a given matrix  $F$ , a given matrix  $Q$ , a given vector  $h$ , and a given scalar  $\sigma_\psi^2$ , one can compute the variance-covariance matrices of the state vector,  $\Sigma_{t|t}$  and  $\Sigma_{t|t-1}$ , from the usual Kalman filter equations

$$\Sigma_{t+1|t} = F\Sigma_{t|t}F' + Q, \quad (10)$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}h(h'\Sigma_{t|t-1}h + \sigma_\psi^2)^{-1}h'\Sigma_{t|t-1}. \quad (11)$$

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

$$\Sigma_1 = F\Sigma_0F' + Q, \quad (12)$$

$$\Sigma_0 = \Sigma_1 - \Sigma_1h(h'\Sigma_1h + \sigma_\psi^2)^{-1}h'\Sigma_1. \quad (13)$$

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

**The rational inattention filter.** In the following, we call the Kalman filter with the observation equation that is optimal from a rational inattention perspective the “rational inattention filter.” We have already proved that without loss in generality one can restrict attention to signals of the form (9). We now show how one can compute the optimal vector of signal weights,  $h^*$ , and the optimal variance of noise,  $\sigma_\psi^{2*}$ .

In the case of a one-dimensional signal, the information flow constraint reduces to a constraint on a particular “signal-to-noise” ratio.

**Lemma 2** *In the case of the one-dimensional signal (9), the information flow constraint (7) reduces to a constraint on a particular “signal-to-noise” ratio*

$$\frac{1}{2} \log_2 \left( \frac{h' \Sigma_1 h}{\sigma_\psi^2} + 1 \right) \leq \kappa, \quad (14)$$

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

**Proof.** Conditional normality implies that

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

The information flow constraint (7) is thus equivalent to

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

Using equation (13) to substitute for  $\Sigma_0$  yields

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

Furthermore, it follows from Sylvester’s determinant theorem that

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

Substituting these equations into the previous weak inequality yields

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

■

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

$$\min_{h \in \mathbb{R}^{\max\{1,p\}+q}, \sigma_\psi^2 > 0} \left( 1 \ 0 \ \dots \ 0 \right) \Sigma_0 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad (15)$$



subject to

$$\frac{1}{2} \log_2 \left( \frac{h' \Sigma_1 h}{\sigma_\psi^2} + 1 \right) \leq \kappa, \quad (16)$$

where the conditional variance-covariance matrices of the state vector,  $\Sigma_0$  and  $\Sigma_1$ , are given by the usual Kalman filter equations (12)-(13).

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 (16) 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

$$\sigma_\psi^2 = \frac{h' \Sigma_1 h}{2^{2\kappa} - 1}. \quad (17)$$

Using the binding information flow constraint (17) to substitute for the variance of noise in the Kalman filter equation (13) yields the following statement of the decision problem, for all  $\kappa > 0$ ,

$$\min_{h \in \mathbb{R}^{\max\{1,p\}+q}} \left( \begin{array}{cccc} 1 & 0 & \cdots & 0 \end{array} \right) \Sigma_0 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad (18)$$

where the matrices  $\Sigma_0$  and  $\Sigma_1$  are given by

$$\Sigma_1 = F \Sigma_0 F' + Q, \quad (19)$$

$$\Sigma_0 = \Sigma_1 - \frac{(1 - 2^{-2\kappa})}{h' \Sigma_1 h} \Sigma_1 h h' \Sigma_1. \quad (20)$$

Expression (18) is the objective. Equations (19)-(20) are the usual Kalman filter equations, but where the information flow constraint (17) has been used to substitute for the variance of noise in equation (13). Equations (19)-(20) give the prior variance-covariance matrix of the state vector,  $\Sigma_1$ , and the posterior variance-covariance matrix of the state vector,  $\Sigma_0$ , as implicit functions of the matrices  $F$  and  $Q$ , the vector of signal weights  $h$ , and the information flow parameter  $\kappa$ .<sup>6</sup> Solving

---

<sup>6</sup>One can also endogenize  $\kappa$  by augmenting the vector of choice variables by  $\kappa$  and by adding a cost function for  $\kappa$  to the objective. In the set of first-order conditions presented below, this will simply add a first-order condition. Note also that any cost function for  $\kappa$  can be expressed as a cost function for the signal-to-noise ratio by substituting in the binding information flow constraint (16).

this problem yields the optimal vector of signal weights,  $h^*$ . Substituting this vector into equation (17) yields the optimal variance of noise,  $\sigma_\psi^{2*}$ .

**First-order conditions.** The problem (18)-(20) can be solved in many different ways, for example, by brute-force numerical optimization or by stating and solving first-order conditions. We found that solving the first-order conditions works well and therefore we state these first-order conditions in the next paragraphs. We also use these first-order conditions below to derive more analytical results regarding the optimal signal weights.

Since the optimal signal weight on  $X_t$  is different from zero and multiplying a signal by a non-zero constant does not change the matrices  $\Sigma_0$  and  $\Sigma_1$ , one can normalize the weight on  $X_t$  to one without loss in generality (i.e.,  $h_1 = 1$ ). If  $\max\{1, p\} + q > 1$ , there are  $\max\{1, p\} + q - 1$  remaining optimal signal weights that one needs to solve for. The first-order conditions for these signal weights can be derived as follows.

Substituting equation (20) into equation (19) and rearranging yields

$$\Sigma_1 - F \left[ \Sigma_1 - \frac{(1 - 2^{-2\kappa})}{h' \Sigma_1 h} \Sigma_1 h h' \Sigma_1 \right] F' - Q = 0. \quad (21)$$

This equation gives  $\Sigma_1$  as an implicit function of  $F$ ,  $Q$ ,  $h$ , and  $\kappa$ . Changing a single element of the vector of signal weights  $h$  potentially affects all elements of  $\Sigma_1$ . Let  $\Sigma_{1,ij}$  denote the  $i, j$ -element of  $\Sigma_1$  and let  $d = \max\{1, p\} + q$ . The derivatives  $(d\Sigma_{1,ij}/dh_l)$  for  $i, j = 1, \dots, d$  and  $l = 2, \dots, d$  are given by

$$\forall l = 2, \dots, d : \sum_{i=1}^d \sum_{j=1}^d Z^{ij} \frac{d\Sigma_{1,ij}}{dh_l} + Z^l = 0. \quad (22)$$

Here the  $(d \times d)$  matrix  $Z$  denotes the left-hand side of equation (21), the  $(d \times d)$  matrix  $Z^{ij}$  denotes the derivative of  $Z$  with respect to  $\Sigma_{1,ij}$ , i.e.,

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

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

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

Appendix D gives closed form expressions for  $Z^{ij}$  and  $Z^l$ , which depend on  $\Sigma_1$ . Finally, when  $p > 1$ , the objective (18) equals  $\Sigma_{1,22}$ , because the second element of the state vector  $\xi_t$  equals  $X_{t-1}$  and  $\Sigma_1$  is the variance-covariance matrix of the state vector given information at time  $t - 1$ . In this case, the first-order conditions for the optimal signal weights are simply<sup>7</sup>

$$\forall l = 2, \dots, d : \frac{d\Sigma_{1,22}}{dh_l} = 0. \quad (23)$$

Solving the system of equations (21)-(23) for the  $(d \times d)$  symmetric matrix  $\Sigma_1$ , the  $d^2(d - 1)$  derivatives  $(d\Sigma_{1,ij}/dh_l)$ , and the  $(d - 1)$  signal weights  $h_2, \dots, h_d$  yields signal weights that satisfy the first-order conditions.

**Proposition 3** *The optimal signal weights have to satisfy equations (21)-(23). Given the optimal signal weights, the conditional expectations of the state vector can be computed from the usual Kalman filter equations.*

**The dynamic attention principle.** We now ask whether the optimal signal can have the simple form  $S_t = X_t + \psi_t$  beyond an AR(1) process. In other words, we ask whether the optimal signal vector can have the property that the weights on all variables apart from  $X_t$  equal zero. Proposition 1 does not rule out this possibility. The answer is: Beyond an AR(1) process, the optimal signal is generically not  $S_t = X_t + \psi_t$ . We call this the “dynamic attention principle.”

The intuition is that in a dynamic setting an agent cares about being well informed in the current period and about 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. Thus, 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  $S_t = X_t + \psi_t$ .

To begin, consider an AR(1) process,  $X_t = \phi_1 X_{t-1} + \theta_0 \varepsilon_t$ . Propositions 1 and 2 imply that one can restrict attention to signals of the form  $S_t = a_0 X_t + \psi_t$ . Substituting  $F = \phi_1$ ,  $Q = \theta_0^2$ , and  $h = a_0$  into equations (19)-(20) yields  $\Sigma_1 = \phi_1^2 \Sigma_0 + \theta_0^2$  and  $\Sigma_0 = 2^{-2\kappa} \Sigma_1$ . Substituting these two

<sup>7</sup>In the case of  $p \leq 1$  and  $\max\{1, p\} + q > 1$  (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 6.

equations for the conditional variances of  $X_t$  given information at time  $t - 1$  and  $t$  into equation (17) yields the following proposition.

**Proposition 4** *Suppose the variable of interest follows an AR(1) process,  $X_t = \phi_1 X_{t-1} + \theta_0 \varepsilon_t$ . Then, one can restrict attention to signals of the form  $S_t = a_0 X_t + \psi_t$ , the value of the objective at the solution equals*

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

and for all  $\kappa > 0$ , the noise-to-signal ratio in an optimal signal equals

$$\frac{\sigma_\psi^2}{a_0^2} = \frac{2^{2\kappa}}{2^{2\kappa} - 1} \frac{\theta_0^2}{2^{2\kappa} - \phi_1^2}.$$

Next, consider an AR(2) process,  $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_0 \varepsilon_t$ . Propositions 1 and 2 imply that one can restrict attention to signals of the form  $S_t = a_0 X_t + a_1 X_{t-1} + \psi_t$ . The following proposition formally states the dynamic attention principle in the AR(2) case.

**Proposition 5** *Suppose the variable of interest follows an AR(2) process,  $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \theta_0 \varepsilon_t$ . Then, one can restrict attention to signals of the form  $S_t = a_0 X_t + a_1 X_{t-1} + \psi_t$ , and  $\phi_1, \phi_2 \neq 0$  implies  $a_1 \neq 0$ .*

**Proof.** See Appendix E. ■

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 ( $\phi_1 \neq 0$ ), then the optimal signal weight on  $X_{t-1}$  is non-zero ( $a_1 \neq 0$ ). The proposition is proved by showing that the first-order condition for the optimal signal weight on  $X_{t-1}$  is satisfied at  $a_1 = 0$  if and only if  $\phi_1 \phi_2 = 0$ .<sup>8</sup>

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 productivity can be learned about before it actually occurs in production or that a change in fiscal policy is announced before the change in government spending or taxes actually occurs. In this case, the shock is called a “news shock.” A standard example is  $X_t = \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1}$  with  $\phi_1, \theta_1 \neq 0$ . The shock is realized and can be

<sup>8</sup>The proof extends to the case of a non-stationary AR(2) process if  $\kappa$  is sufficiently large so that all conditional moments appearing in the proof are finite and well-defined. The following two inequalities ensure that this is the case:  $2^{2\kappa} > \phi_2$  and  $(2^{2\kappa} - \phi_2^2) [2^{4\kappa} + \phi_2^2 - 2^{2\kappa} (\phi_1^2 + 2\phi_2)] > 0$ .

learned about in period  $t - 1$  but only affects the variable of interest in period  $t$ . Formally, the variable of interest 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  $S_t = a_0 X_t + b_0 \varepsilon_t + \psi_t$ . The next proposition states that the optimal signal weight on  $\varepsilon_t$  is always different from zero, even though  $X_t$  does not depend on  $\varepsilon_t$ . The reason is that the agent would like to enter well informed into the next period.

**Proposition 6** *Suppose the variable of interest follows an ARMA(1,1) process,  $X_t = \phi_1 X_{t-1} + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1}$ . Then, one can restrict attention to signals of the form  $S_t = a_0 X_t + b_0 \varepsilon_t + \psi_t$ , and  $\phi_1 \neq 0$ ,  $\theta_0 = 0$ , and  $\theta_1 \neq 0$  implies  $b_0 \neq 0$ .*

**Proof.** See Appendix F. ■

The proposition is again proved by showing that the first-order condition for the optimal signal weight on  $\varepsilon_t$  is violated at  $b_0 = 0$  if  $\phi_1 \neq 0$ ,  $\theta_0 = 0$ , and  $\theta_1 \neq 0$ .<sup>9</sup>

In sum, in a dynamic setting an agent cares about being well informed in the current period and about entering well informed into the next period because this relaxes the agent's attention constraint. Beyond an AR(1) process, there is a tension between these two goals and therefore the optimal signal is generically not  $S_t = X_t + \psi_t$ .

**Examples.** Figure 1 shows two solved examples of the dynamic attention choice problem. In the top panel  $X_t$  follows an ARMA(2,1) process. In the bottom panel  $X_t$  follows an AR(2) process whose characteristic polynomial has complex roots.

**Information flow very large.** To understand the dynamic attention principle, it is also useful to consider the  $\kappa \rightarrow \infty$  limit.

**Proposition 7** *As  $\kappa \rightarrow \infty$ , the information capacity devoted to other components of uncertainty than to  $X_t$  approaches 0, i.e.,  $h' \rightarrow (1, 0, 0, \dots)$ .*

**Proof.** See Appendix G. ■

If the information flow is very large, then 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,

---

<sup>9</sup>The proof extends to the case of a non-stationary ARMA(1,1) process if  $\kappa$  is sufficiently large so that all conditional moments appearing in the proof are finite and well-defined. The following condition ensures that this is the case:  $2^{2\kappa} > \phi_1^2$ .

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  $\kappa$ , 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  $\varepsilon_t$  yet, and thus uncertainty about  $X_t$  is much larger.

## 5 Application to a business cycle model with news shocks

We apply the paper’s analytical results in the context of two macroeconomic models. 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 can be learned about before it actually occurs in production.<sup>10</sup> 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*.<sup>11</sup>

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

---

<sup>10</sup>For simplicity, in this section we write “news shocks” instead of “news shocks about productivity.”

<sup>11</sup>The most popular model generating a boom in response to a positive news shock is Jaimovich and Rebelo (2009). The model has three key elements: preferences that allow the modeller 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.

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  $\rho \in (0, 1)$ ,  $\sigma > 0$ , and  $\varepsilon_t \sim i.i.d.N(0, 1)$ . 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, productivity changes can be learned about  $k$  periods in advance. When  $k > 0$ , 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

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

**Statement of the attention choice 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

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

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  $n_{it} = E[n_{it}^* | \mathcal{I}_{it}]$ .

The attention choice problem of the entrepreneur is a special case of the problem presented in Section 2. The entrepreneur tracks a variable of interest, here  $n_{it}^*$ , that follows a Gaussian process. The entrepreneur chooses how to pay attention to the variable. The signals in period  $t$  can be about any linear combination of current and past  $n_{it}^*$  and current and past  $\varepsilon_t$ . The entrepreneur remembers all signals and aims to minimize the mean squared error, here  $E[(n_{it}^* - E[n_{it}^* | \mathcal{I}_{it}])^2]$ .<sup>12</sup> When the variable of interest follows a Gaussian ARMA(p,q) process, one can directly apply the results of Sections 3-4.

We now show that a positive news shock creates a boom 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  $\frac{1}{\alpha} \ln [(1-\alpha) L^\alpha] = 0$  and consider the log-linearized labor market clearing condition<sup>13</sup>

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

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

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

---

<sup>12</sup>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 2.

<sup>13</sup>The original labor market clearing condition yields the same equilibrium under perfect information.



Thus, the market clearing wage equals

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

and the equilibrium labor input equals

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

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.

**A boom 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  $w_t = \varphi z_t$ , and thus the profit-maximizing labor input equals  $n_{it}^* = \frac{1}{\alpha} (1 - \varphi) z_t$ . Furthermore, suppose that innovations in productivity can be learned about 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 imply that an entrepreneur subject to rational inattention can restrict attention to signals of the form

$$S_{it} = a_0 z_t + b_0 \varepsilon_t + \psi_{it},$$

where  $\psi_{it}$  is a noise term that follows a Gaussian white noise process. Proposition 6 implies that  $b_0 \neq 0$ . Hence, an entrepreneur who optimally allocates attention chooses a one-dimensional signal with a non-zero weight on current productivity,  $z_t$ , and a non-zero weight on the news shock,  $\varepsilon_t$ . Recall that the entrepreneur chooses the non-zero weight on the news shock because entering well informed into the next period relaxes the entrepreneur's attention constraint. The optimal signal has the following implications for actions. On impact of a positive news shock,  $\varepsilon_t > 0$ , the signal increases. Since the entrepreneur chose not to distinguish carefully between increases in current productivity and increases in future productivity, the entrepreneur starts hiring already today.

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  $w_t = \varphi z_t$  and thus the profit-maximizing labor input that entrepreneurs are tracking is no longer simply equal

to  $n_{it}^* = \frac{1}{\alpha}(1 - \varphi)z_t$ . 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,  $w_t = \frac{+\gamma}{1-\gamma}n_t$ , 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

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

Actions of different firms are strategic substitutes, because 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 results in Section 3 to establish the form of an optimal signal and the results in Section 4 to compute the optimal signal weights and the implied actions. We then compute 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 2 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  $k = 8$ .<sup>14</sup> Labor input rises coincident with a positive innovation in  $\varepsilon_t$ , and a boom develops before productivity actually rises. Positive news shocks produce a boom, 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 2 also shows the equilibrium labor input in the 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

---

<sup>14</sup>We use a value of the information-processing parameter  $\kappa$  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.

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.

## 6 Application to a model of price-setting

As another application, 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 macroeconomic 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 from a rational inattention perspective.

**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, and since the firm can reset the price in the next period, this is equivalent to setting the price to maximize current profits. The price maximizing current profits in the Woodford model can be written as

$$p_{it}^* = \xi q_t + (1 - \xi) p_t, \quad (24)$$

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. 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  $(p_{it}^* - p_{it})^2$  and a firm’s optimal price given any information set  $\mathcal{I}_{it}$  is therefore  $p_{it} = E[p_{it}^* | \mathcal{I}_{it}]$ .

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

$$S_{it} = q_t + v_{it}, \quad (25)$$

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 includes any initial information and all past signals,  $\mathcal{I}_{it} = \mathcal{I}_{i,0} \cup \{S_{i1}, \dots, S_{it}\}$ . Nominal aggregate demand follows an exogenous stochastic process given

by

$$q_t = (1 + \rho) q_{t-1} - \rho q_{t-2} + \sigma \varepsilon_t, \quad (26)$$

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

$$p_t = \int_0^1 p_{it} di. \quad (27)$$

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."

**Statement of the attention choice 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

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

where  $\beta \in (0, 1)$  is a parameter, subject to the information flow constraint (5) and

$$p_{it} = E[p_{it}^* | \mathcal{I}_{it}],$$

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

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  $\psi_{it}^K$  follows a Gaussian vector white noise process with variance-covariance matrix  $\Sigma_\psi$ . In words, the decision-maker aims to minimize the expected discounted sum of profit losses due to suboptimal pricing. He understands that in every period  $t \geq 1$  he will set the price equal to the conditional expectation of the profit-maximizing price and he 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.<sup>15</sup>

After the decision-maker has chosen the signal properties in period zero, he receives information such that the conditional variance-covariance matrix of the state vector in period one given information in period zero equals the steady-state conditional variance-covariance matrix of the state vector in period  $t$  given information in period  $t - 1$ . The mean squared error,  $E \left[ (p_{it}^* - E[p_{it}^* | \mathcal{I}_{it}])^2 \right]$ , is then constant for all  $t \geq 1$ . Hence, the objective simplifies to  $\beta / (1 - \beta)$  times

$$E \left[ (p_{it}^* - E[p_{it}^* | \mathcal{I}_{it}])^2 \right],$$

and this mean squared error can be evaluated using the steady-state Kalman filter. In other words, the firms' objective equals the objective in Section 2.

**No strategic complementarity in price-setting.** To develop intuition for the implications of rational inattention, 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 equation (24)), and nominal aggregate demand follows a Gaussian AR(2) process (see equation (26)). The results in Section 3 imply that in this case one can restrict attention to signals of the form

$$S_{it}^* = q_t + a_1 q_{t-1} + \psi_{it}, \tag{28}$$

where  $\psi_{it}$  follows a Gaussian white noise process.<sup>16</sup> Notice that assumption (25) in this case amounts to a simple restriction  $a_1 = 0$ . However, Proposition 5 implies that if  $\rho \neq 0$  then  $a_1 \neq 0$ . Recall that the decision-maker chooses the non-zero weight on  $q_{t-1}$  in the period  $t$  signal because entering well informed into the next period relaxes the decision-makers' attention constraint.

To investigate to what extent Woodford's restriction on the signals matters, we assume his parameter values.<sup>17</sup> Furthermore, we suppose that the information flow in the model with optimal signals (the model with  $S_{it}^*$  given by equation (28)) is equal to the information flow in the Woodford model (the model with  $S_{it}$  given by equation (25)).<sup>18</sup> Thus decision-makers process the same

---

<sup>15</sup>This decision problem is a simplified version of the decision problem in Maćkowiak and Wiederholt (2009). There are no idiosyncratic shocks and the decision-maker does not choose the amount of attention allocated to aggregate conditions.

<sup>16</sup>We follow Woodford in assuming that the noise term  $\psi_{it}$  is also independent across firms.

<sup>17</sup>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.

<sup>18</sup>The information flow in the Woodford model,  $\kappa_W$ , can be computed from the formula  $\kappa_W = (1/2) \log_2 (\sigma_{q,1}^2 / \sigma_{q,0}^2)$ ,

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.

The top panel in Figure 3 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 disturbance 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 as possible. Consequently, in the model with optimal signals the tracking of the profit-maximizing price is more accurate, an extra unit of information is less valuable, and a nominal shock has weaker real effects compared with the Woodford model. Furthermore, the differences between the two models can be sizable.

**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 where  $p \geq 1$  and  $q \geq 0$  are integers. Given the guess, we apply the results in Section 3 to establish the form of an optimal signal. For example, if the profit-maximizing price follows an ARMA(2,2) process, we restrict attention to signals of the form

$$S_{it}^* = p_{it}^* + a_1 p_{i,t-1}^* + b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \psi_{it}. \quad (29)$$

We let the decision-maker in firm  $i$  choose the optimal signal weights ( $a_1$ ,  $b_0$ , and  $b_1$  in the ARMA(2,2) example) and the variance of noise  $\sigma_\psi^2$  to maximize profits, subject to the information flow constraint. We then obtain the aggregate price level from the relation  $p_t = \int_0^1 p_{it} di$ , and we compute the profit-maximizing price using equation (24). As before, the information flow in the model with optimal signals is held equal to the information flow in the Woodford model.<sup>19</sup>

The bottom panel in Figure 3 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 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 the values of  $a_1$  and  $\sigma_\psi^2$  in the model with optimal signals by applying the results in Section 4 and setting  $\kappa = \kappa_W$ .

<sup>19</sup>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  $p_{it}^*$ .

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.<sup>20</sup> The Woodford model predicts stronger real effects than the model with optimal signals for any value of  $\xi$ , 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 the model with optimal signals. At the same time, the difference between the size of real effects in the two models decreases as  $\xi$  falls from one to zero (as the degree of strategic complementarity in price-setting rises). 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. Hence, the firms' tracking of the equilibrium profit-maximizing price improves, the marginal value of information falls, and Woodford's restriction on the signals becomes less harmful to the decision-makers, implying that the difference between the size of real effects in the models diminishes.

In sum, Woodford's restriction on the signals increases the size of real effects predicted by the model, strongly in the case without strategic complementarity in price setting and weakly in the case with a significant degree of strategic complementarity in price setting.

## 7 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.

---

<sup>20</sup>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.

## A Proof of Lemma 1

Let  $S^{K,t} = \{S_1^K, \dots, S_t^K\}$  denote the history of signals at time  $t$  and let  $\varepsilon^t = \{\varepsilon_1, \dots, \varepsilon_t\}$  denote the history of innovations to the variable of interest at time  $t$ . The left-hand side of the information flow constraint (5) can be written as

$$\lim_{T \rightarrow \infty} \frac{1}{T} I(Z_0, \varepsilon_1, \dots, \varepsilon_T; S_1^K, \dots, S_T^K) = \lim_{T \rightarrow \infty} \frac{1}{T} I(Z_0, \varepsilon^T; S^{K,T}).$$

We now show that

$$\lim_{T \rightarrow \infty} \frac{1}{T} I(Z_0, \varepsilon^T; S^{K,T}) = \lim_{T \rightarrow \infty} [H(\xi_T | S^{K,T-1}) - H(\xi_T | S^{K,T})].$$

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

$$I(Z_0, \varepsilon^T; S^{K,T}) = I(S^{K,T}; Z_0, \varepsilon^T) = H(S^{K,T}) - H(S^{K,T} | Z_0, \varepsilon^T).$$

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

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

and

$$H(S^{K,T} | Z_0, \varepsilon^T) = H(S^{K,\tau-1} | Z_0, \varepsilon^T) + \sum_{t=\tau}^T H(S_t^K | S^{K,t-1}, Z_0, \varepsilon^T).$$

The signal  $S_t^K$  depends only on  $X_t^M$ ,  $\varepsilon_t^N$ , and  $\psi_t^K$  for given  $A$  and  $B$ . In the following, let  $\tau = \max\{M, N, 2\}$ . For  $t \in \{\tau, \dots, T\}$ , one can compute  $X_t^M$  and  $\varepsilon_t^N$  from  $Z_0, \varepsilon^T$ , and one can compute  $X_t^M$  and  $\varepsilon_t^N$  from the vector  $\xi_t$  defined in Lemma 1. Hence, for  $t \in \{\tau, \dots, T\}$ ,

$$H(S_t^K | S^{K,t-1}, Z_0, \varepsilon^T) = H(\psi_t^K) = H(S_t^K | S^{K,t-1}, \xi_t).$$

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

$$I(Z_0, \varepsilon^T; S^{K,T}) = H(S^{K,\tau-1}) - H(S^{K,\tau-1} | Z_0, \varepsilon^T) + \sum_{t=\tau}^T [H(S_t^K | S^{K,t-1}) - H(S_t^K | S^{K,t-1}, \xi_t)].$$

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

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



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

$$I(Z_0, \varepsilon^T; S^{K,T}) = H(S^{K,\tau-1}) - H(S^{K,\tau-1}|Z_0, \varepsilon^T) + \sum_{t=\tau}^T [H(\xi_t|S^{K,t-1}) - H(\xi_t|S^{K,t})]. \quad (30)$$

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

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

Equation (30) and Cesaro mean then imply

$$\lim_{T \rightarrow \infty} \frac{1}{T} I(Z_0, \varepsilon^T; S^{K,T}) = \lim_{T \rightarrow \infty} [H(\xi_T|S^{K,T-1}) - H(\xi_T|S^{K,T})]. \quad (31)$$

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

$$\xi_{t+1} = F\xi_t + v_{t+1}, \quad (32)$$

$$S_t^K = G'\xi_t + \psi_t^K. \quad (33)$$

Here the vector  $\xi_t$  contains the same elements as the vector  $\xi_t$  defined in Lemma 1 but with a slightly different ordering for the case of  $q > 0$

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

The different ordering simplifies the exposition below. The matrix  $F$  is a square matrix and the length of the column vector  $v_t$  equals the length of the column vector  $\xi_t$ . 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 (32) equals the law of motion for  $X_t$ . The next  $\max\{M, N+p-q\} - 1$  elements of  $v_{t+1}$  equal zero and the next  $\max\{M, N+p-q\} - 1$  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  $q > 0$  (i.e.,  $q > 0$ ), 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  $\varepsilon_{t+1}$  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 (33) equals equation (2). Such a matrix exists because  $X_t^M$  and  $\varepsilon_t^N$  can be computed from  $\xi_t$ .

Second, let  $\Sigma_{t|t-1}$  denote the conditional variance-covariance matrix of  $\xi_t$  given  $S^{K,t-1}$  and let  $\Sigma_{t|t}$  denote the conditional variance-covariance matrix of  $\xi_t$  given  $S^{K,t}$ . Furthermore, let  $\Sigma_1$  and  $\Sigma_0$  denote  $\lim_{t \rightarrow \infty} \Sigma_{t|t-1}$  and  $\lim_{t \rightarrow \infty} \Sigma_{t|t}$ , respectively. Recall that  $X_t$  follows a stationary process. It follows from Propositions 13.1-13.2 in Hamilton (1994) that  $\lim_{t \rightarrow \infty} \Sigma_{t|t-1}$  and  $\lim_{t \rightarrow \infty} \Sigma_{t|t}$  exist

and are given by

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

where  $Q$  denotes the variance-covariance matrix of the innovation in the state equation (32) and  $R$  denotes the variance-covariance matrix of the innovation in the observation equation (33).

Third, one can express the information flow constraint (5) in terms of the matrices  $\Sigma_1$  and  $\Sigma_0$ . According to Lemma 1 the information flow constraint (5) is equivalent to

$$\lim_{T \rightarrow \infty} [H(\xi_T | S^{K, T-1}) - H(\xi_T | S^{K, T})] \leq \kappa. \quad (34)$$

Conditional normality implies that

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

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

$$\frac{1}{2} \log_2 \left( \frac{\det \Sigma_1}{\det \Sigma_0} \right) \leq \kappa. \quad (35)$$

Fourth, let us split the vector  $\xi_t$  into two sub-vectors, denoted  $\xi_t^{up}$  and  $\xi_t^{low}$ . The vector  $\xi_t^{up}$  is defined as

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

The vector  $\xi_t^{low}$  contains the remaining elements of  $\xi_t$ . Note that the conditional variance-covariance matrix of  $\xi_t^{up}$  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  $\xi_t^{up}$ . Furthermore, note that the objective (4) is simply an element of the upper-left  $(d \times d)$  sub-matrix of  $\Sigma_0$ . Moreover, the information flow constraint (34) can be written as

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

Next, compare two signal vectors with  $M > p$  and  $N > q$  that yield the same upper-left  $(d \times d)$  sub-matrix of  $\Sigma_0$ . One signal vector has the property

$$\begin{aligned} \forall j > p : A_{ij} &= 0 \\ \forall j > q : B_{ij} &= 0 \end{aligned}, \quad (36)$$

while the other signal vector does not have this property. Both signal vectors imply the same value of the objective (4), because the objective is an element of the upper-left  $(d \times d)$  sub-matrix of  $\Sigma_0$ . Furthermore, both signal vectors generate the same limit

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

because both signal vectors yield the same upper-left  $(d \times d)$  sub-matrix of  $\Sigma_0$  by assumption and they also generate the same upper-left  $(d \times d)$  sub-matrix of  $\Sigma_1$ , because  $X_t$  follows an ARMA(p,q) process. Finally, the difference

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

is non-negative, since conditioning weakly reduces entropy, and it equals zero if and only if condition (36) 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 (36) there exists a signal vector satisfying condition (36) that yields the same upper-left  $(d \times d)$  sub-matrix of  $\Sigma_0$ . 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  $\{K, \hat{A}, \hat{B}, \hat{\Sigma}\}$  violating condition (36) that yields the variance-covariance matrices  $\Sigma_1$  and  $\Sigma_0$ . Since  $\Sigma_0$  is generated from  $\Sigma_1$  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 (36) that generates the upper-left  $(d \times d)$  sub-matrix of  $\Sigma_0$  from the upper-left  $(d \times d)$  sub-matrix of  $\Sigma_1$ .

## C Proof of Proposition 2

The proof consists of two steps. Let  $\xi_t$  denote a column vector containing the variables given in Proposition 1. For example,  $\xi_t = (X_t, \dots, X_{t-(p-1)})'$  in the AR(p) case and  $\xi_t = (X_t, \varepsilon_t, \dots, \varepsilon_{t-(q-1)})'$  in the MA(q) case. 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  $\Sigma_\psi^{-1}$ . 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  $\Sigma_1$  is the prior variance-covariance matrix of  $\xi_t$  and  $\Sigma_0$  is the posterior variance-covariance matrix of  $\xi_t$ . Objective (4) and constraint (5) depend only on  $\Sigma_1$  and  $\Sigma_0$ . The objective is an element of  $\Sigma_0$ , and the information flow is given by the ratio of determinants of  $\Sigma_1$  and  $\Sigma_0$ . Therefore, it suffices to show that for any  $\Sigma_1$  and  $\Sigma_0$  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  $\Sigma_1$  and  $\Sigma_0$ .

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  $\Sigma_\psi^{-1}$  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  $\Sigma_{\psi}^{-1}$  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  $\Sigma_{\psi}^{-1}$ . 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  $p > 1$ . In this case, we have  $\xi_t = (X_t, \dots, X_{t-(p-1)})'$  and the first signal is on a linear combination of  $X_t, \dots, X_{t-(p-1)}$ , the second signal is on a linear combination of  $X_{t-1}, \dots, X_{t-(p-1)}$ , the third signal is on a linear combination of  $X_{t-2}, \dots, X_{t-(p-1)}$ , and the  $p$ th 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  $X_{t-1}, \dots, X_{t-(p-1)}$  is replaced by a signal on  $X_t, \dots, X_{t-(p-2)}$  in every period. The only change in the history of signals  $S^{K,T} = \{S_1^K, \dots, S_T^K\}$  is that the signal on  $X_T, \dots, X_{T-(p-2)}$  is added and the signal on  $X_0, \dots, X_{-p+2}$  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  $T \rightarrow \infty$  of the conditional variance of  $X_T$  given  $S^{K,T}$ ,  $\lim_{T \rightarrow \infty} Var(X_T | S^{K,T})$ . The loss of the signal on  $X_0, \dots, X_{-p+2}$  does not affect the value of this loss function, while the addition of the signal on  $X_T, \dots, X_{T-(p-2)}$  reduces the value of this loss function. The information flow equals

$$\lim_{T \rightarrow \infty} [H(X_T, \dots, X_{T-(p-1)} | S^{K,T-1}) - H(X_T, \dots, X_{T-(p-1)} | S^{K,T})].$$

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

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

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

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

Recall that the only change in the history of signals  $S^{K,T} = \{S_1^K, \dots, S_T^K\}$  is that the signal on  $X_T, \dots, X_{T-(p-2)}$  is added and the signal on  $X_0, \dots, X_{-p+2}$  is lost. The addition of the signal on  $X_T, \dots, X_{T-(p-2)}$  to  $S^{K,T}$  does not change the second term. Similarly, the addition of the signal on  $X_{T-1}, \dots, X_{T-(p-1)}$  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  $p$ th 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  $p$ th row of the matrix  $G'$  contain only zeros.

When the variable of interest follows an MA( $q$ ) process with  $q > 0$ , the proof needs to be modified slightly, because the vector  $\xi_t = (X_t, \varepsilon_t, \dots, \varepsilon_{t-(q-1)})'$  contains more than one element that depends on the current shock  $\varepsilon_t$ : the variable of interest,  $X_t$ , and the shock itself,  $\varepsilon_t$ . In this case, one can go through the same two steps with the state vector  $\hat{\xi}_t = (\varepsilon_t, \dots, \varepsilon_{t-(q-1)}, \varepsilon_{t-q})'$  instead of the state vector  $\xi_t = (X_t, \varepsilon_t, \dots, \varepsilon_{t-(q-1)})'$ : (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  $\Sigma_\psi^{-1}$ , 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  $\xi_t$ .

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

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

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  $\xi_t$ .

## D The matrices $Z^{ij}$ and $Z^l$

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  $l$ th element equals one and whose other elements equal zero.

It is straightforward to show that

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

and

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



## E Proof of Proposition 5

Equation (22) at  $p = 2$  and  $a_1 = 0$  reads

$$Z^{11} \frac{d\Sigma_{1,11}}{da_1} + Z^{12} \frac{d\Sigma_{1,12}}{da_1} + Z^{21} \frac{d\Sigma_{1,21}}{da_1} + Z^{22} \frac{d\Sigma_{1,22}}{da_1} + Z^2 = 0, \quad (37)$$

with

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

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

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

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

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

Equation (21) at  $p = 2$  and  $a_1 = 0$  implies

$$\Sigma_{1,22} = 2^{-2\kappa} \Sigma_{1,11}, \quad (38)$$

$$\Sigma_{1,12} = \Sigma_{1,21} = \frac{\phi_1}{2^{2\kappa} - \phi_2} \Sigma_{1,11}, \quad (39)$$

$$\Sigma_{1,11} = \frac{\theta_0^2}{1 - 2^{-2\kappa} (\phi_1^2 + \phi_2^2) - 2^{-2\kappa} \frac{2\phi_1^2\phi_2}{2^{2\kappa} - \phi_2} + \frac{(1 - 2^{-2\kappa})\phi_1^2\phi_2^2}{(2^{2\kappa} - \phi_2)^2}}. \quad (40)$$

In the case of a stationary AR(2) process, the denominators in equations (39) and (40) are positive.

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

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

Equation (37) is a system of four linear equations in  $\frac{d\Sigma_{1,11}}{da_1}$ ,  $\frac{d\Sigma_{1,12}}{da_1}$ ,  $\frac{d\Sigma_{1,21}}{da_1}$ , and  $\frac{d\Sigma_{1,22}}{da_1}$ . Solving this system for  $\frac{d\Sigma_{1,22}}{da_1}$  and using equations (38)-(40) yields

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

Hence, in the AR(2) case, the first-order condition (23) is satisfied at  $a_1 = 0$  if and only if  $\phi_1\phi_2 = 0$ .

## F Proof of Proposition 6

Equation (22) at  $p = 1$ ,  $q = 1$ , and  $b_0 = 0$  reads

$$Z^{11} \frac{d\Sigma_{1,11}}{db_0} + Z^{12} \frac{d\Sigma_{1,12}}{db_0} + Z^{21} \frac{d\Sigma_{1,21}}{db_0} + Z^{22} \frac{d\Sigma_{1,22}}{db_0} + Z^2 = 0,$$

with

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

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

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

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

and

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

These equations imply

$$\frac{d\Sigma_{1,12}}{db_0} = \frac{d\Sigma_{1,21}}{db_0} = \frac{d\Sigma_{1,22}}{db_0} = 0,$$

and

$$0 = \begin{bmatrix} 1 - \phi_1^2 2^{-2\kappa} - \theta_1^2 (1 - 2^{-2\kappa}) \frac{\Sigma_{1,12}\Sigma_{1,21}}{\Sigma_{1,11}^2} \end{bmatrix} \frac{d\Sigma_{1,11}}{db_0} + (1 - 2^{-2\kappa}) \frac{\Sigma_{1,11}\Sigma_{1,22} - \Sigma_{1,12}\Sigma_{1,21}}{\Sigma_{1,11}} \begin{bmatrix} 2\phi_1 \theta_1 + \theta_1^2 \frac{\Sigma_{1,12} + \Sigma_{1,21}}{\Sigma_{1,11}} \end{bmatrix}.$$

Equation (21) at  $p = 1$ ,  $q = 1$ , and  $b_0 = 0$  reads

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

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

$$\frac{d\Sigma_{1,11}}{db_0} = -\frac{(1-2^{-2\kappa})}{(1-\phi_1^2 2^{-2\kappa})} 2\phi_1 \theta_1. \quad (41)$$

Furthermore, in the case of  $p = 1$  and  $q = 1$ , equation (20) reads

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

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

$$\frac{d\Sigma_{0,11}}{db_0} = 2^{-2\kappa} \frac{d\Sigma_{1,11}}{db_0}. \quad (42)$$

It follows from equations (41) and (42) 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  $\varepsilon_t$  is always non-zero.

## G Proof of Proposition 7

Let  $h$  be a vector of signal weights and  $\Sigma_0, \Sigma_1$  denote posterior and prior variance-covariance matrices that are solutions to Kalman filter equations (19)-(20) for  $h$ . Let  $\hat{h} = (1, 0, 0, \dots)$ , 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 (19)-(20). Note that if  $\kappa$  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  $\kappa$  is devoted to  $X_t$ , i.e.,  $\kappa' = 0$ .

The Kalman filter equation (20) for the vector of weights  $h$  implies:

$$\Sigma_0^{1,1} = 2^{-2(\kappa - \kappa')} \left( \theta_0^2 + (F\Sigma_0 F')^{1,1} \right), \quad (43)$$

where  $M^{1,1}$  denotes the 1-1 element of a matrix  $M$ . In (43), we used the fact that devoting an amount  $C$  of information capacity to tracking a normally distributed random variable of variance  $\sigma^2$  implies posterior variance of  $\sigma^2 2^{-2C}$ . Similarly for the weights  $\hat{h}$ , we get:

$$\hat{\Sigma}_0^{1,1} = 2^{-2\kappa} \left( \theta_0^2 + (F\hat{\Sigma}_0 F')^{1,1} \right). \quad (44)$$

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

$$\Sigma_0^{1,1} - \hat{\Sigma}_0^{1,1} = 2^{-2\kappa} \left[ (2^{2\kappa'} - 1)\theta_0^2 + \left( 2^{2\kappa'} (F\Sigma_0 F')^{1,1} - (F\hat{\Sigma}_0 F')^{1,1} \right) \right]. \quad (45)$$

For any  $\kappa$ , if  $h$  is optimal, then the RHS of (45) must always be less or equal to zero.

We will show by contradiction that as  $\kappa \rightarrow \infty$ , in a sequence of optimal  $h$  for such levels of information flows,  $\kappa'$  approaches zero. Let us thus assume that  $\kappa'$  does not approach zero. In this case, there exists a lower bound  $m$ , such that there always exists an arbitrarily large  $\kappa$  for which  $(2^{2\kappa'} - 1)\theta_0^2 > m$ . Moreover, we will show that  $\left( 2^{2\kappa'} (F\Sigma_0 F')^{1,1} - (F\hat{\Sigma}_0 F')^{1,1} \right)$ , the second term in the bracket on the RHS of (45), is positive or approaches zero. Therefore, the RHS of (45) is for any such  $\kappa$  positive, which implies that  $h$  is suboptimal to  $h^*$ , which is a contradiction.

First, the term  $2^{2\kappa'} (F\Sigma_0 F')^{1,1}$  is positive. Second,  $(F\hat{\Sigma}_0 F')^{1,1}$  approaches zero as  $\kappa \rightarrow \infty$ .  $(F\hat{\Sigma}_0 F')^{1,1}$  is the part of prior uncertainty about  $X_t$  that is driven by all past shocks, excluding the current shock  $\varepsilon_t$ . If the agent did not pay any attention to any shock, this variance would be

equal to some  $\sigma_{tot}^2$ , which is finite, because  $X_t$  follows a stationary process. Since the agent devotes under  $\hat{h}$  information capacity of exactly  $\kappa$  to each shock, then

$$\left(F\hat{\Sigma}_0F'\right)^{1,1} = 2^{-2\kappa}\sigma_{tot}^2,$$

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

Putting this together implies that in a sequence of optimal strategies,  $\kappa'$  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  $\kappa$  for which the RHS of (45) would be positive, and thus  $h$  would not be an optimal strategy, which is a contradiction.

## References

- [1] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. (2013). *Nonlinear programming: theory and algorithms*. John Wiley & Sons.
- [2] Greenwood, Jeremy, Zvi Hercowitz, and Gregory Huffman (1988). “Investment, Capacity Utilization, and the Real Business Cycle,” *American Economic Review*, 78(3), 402-417.
- [3] Jaimovich, Nir, and Sergio Rebelo (2009). “Can News about the Future Drive the Business Cycle?” *American Economic Review*, 99(4), 1097-1118.
- [4] Lorenzoni, Guido (2011). “News and Aggregate Demand Shocks,” *Annual Review of Economics*, 3(1), 537-557.
- [5] Luo, Yulei (2008). “Consumption Dynamics under Information Processing Constraints,” *Review of Economic Dynamics*, 11(2), 366-385.
- [6] Maćkowiak, Bartosz, and Mirko Wiederholt (2009). “Optimal Sticky Prices under Rational Inattention,” *American Economic Review*, 99(3), 769-803.
- [7] Maćkowiak, Bartosz, and Mirko Wiederholt (2015). “Business Cycle Dynamics under Rational Inattention,” *Review of Economic Studies*, 82(4), 1502-1532.
- [8] Paciello, Luigi, and Mirko Wiederholt (2014). “Exogenous Information, Endogenous Information, and Optimal Monetary Policy,” *Review of Economic Studies*, 81(1), 356-388.
- [9] Sims, Christopher A. (1998). “Stickiness,” *Carnegie-Rochester Conference Series on Public Policy*, 49(1), 317-356.
- [10] Sims, Christopher A. (2003). “Implications of Rational Inattention,” *Journal of Monetary Economics*, 50(3), 665-690.
- [11] 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.

- [12] 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.
- [13] Stevens, Luminita (2015). “Coarse Pricing Policies,” Discussion paper, Federal Reserve Bank of Minneapolis and University of Maryland.
- [14] Veldkamp, Laura L. (2011). *Information Choice in Macroeconomics and Finance*. Princeton University Press.
- [15] 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.
- [16] 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.
- [17] Woodford, Michael (2009). “Information-Constrained State-Dependent Pricing,” *Journal of Monetary Economics*, 56(S), 100-124.

Figure 1: Solved examples of the dynamic attention choice problem

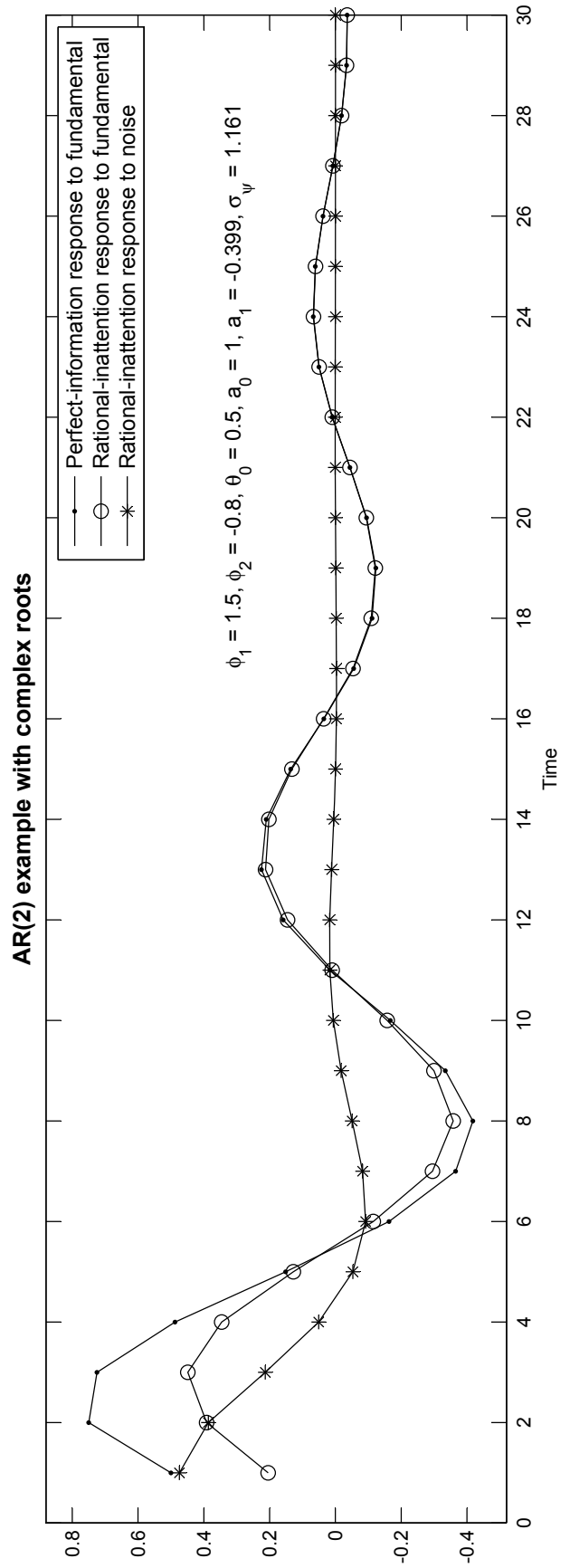
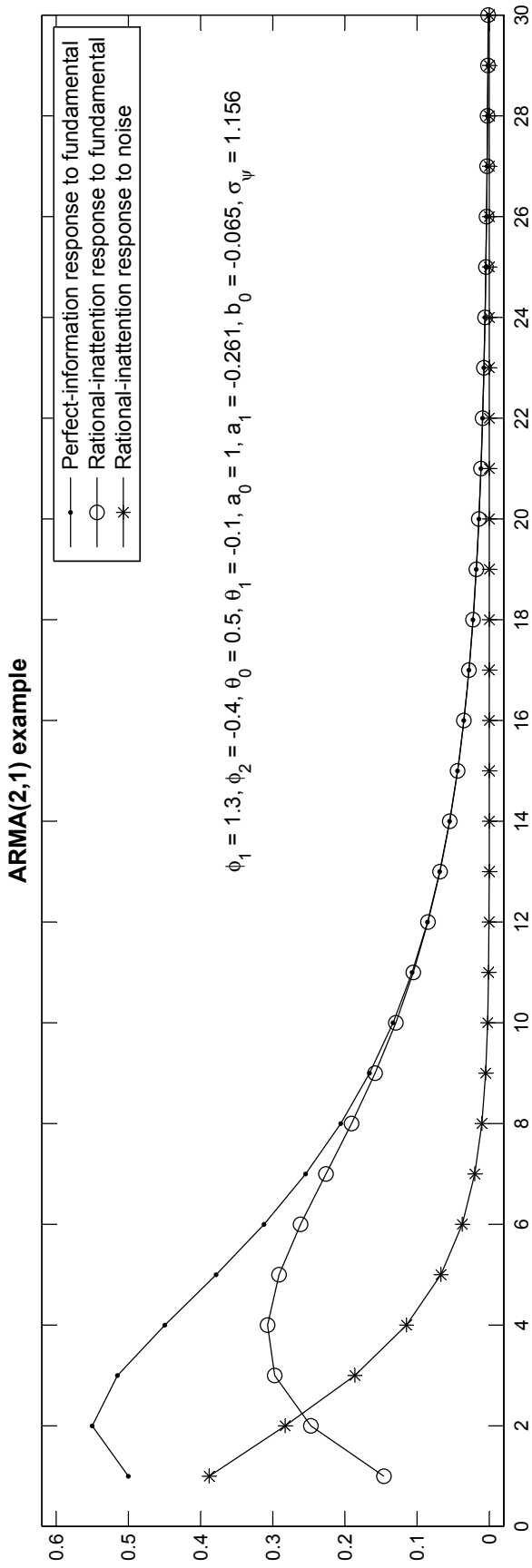




Figure 2: The impulse response of labor input to a productivity shock

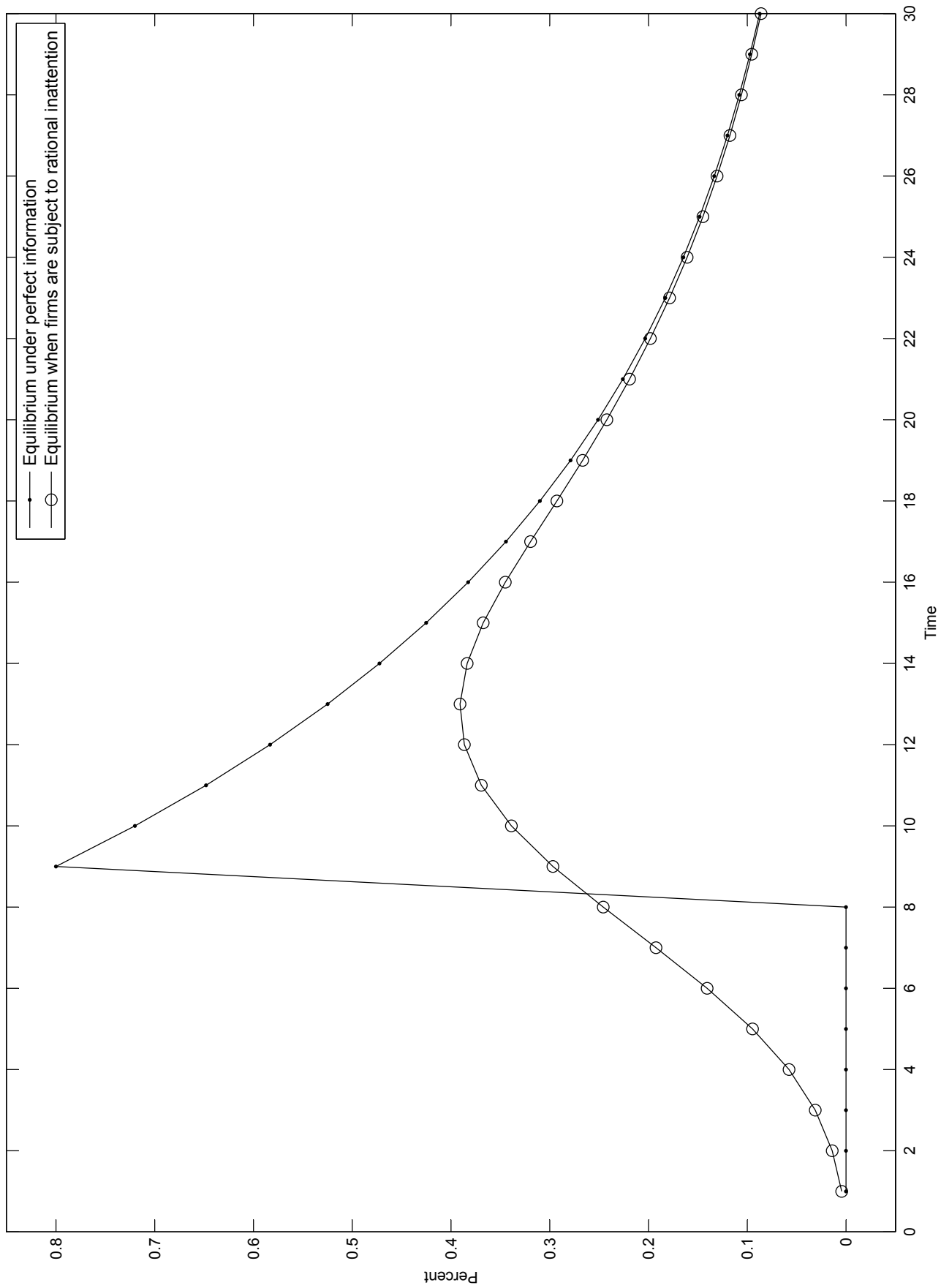
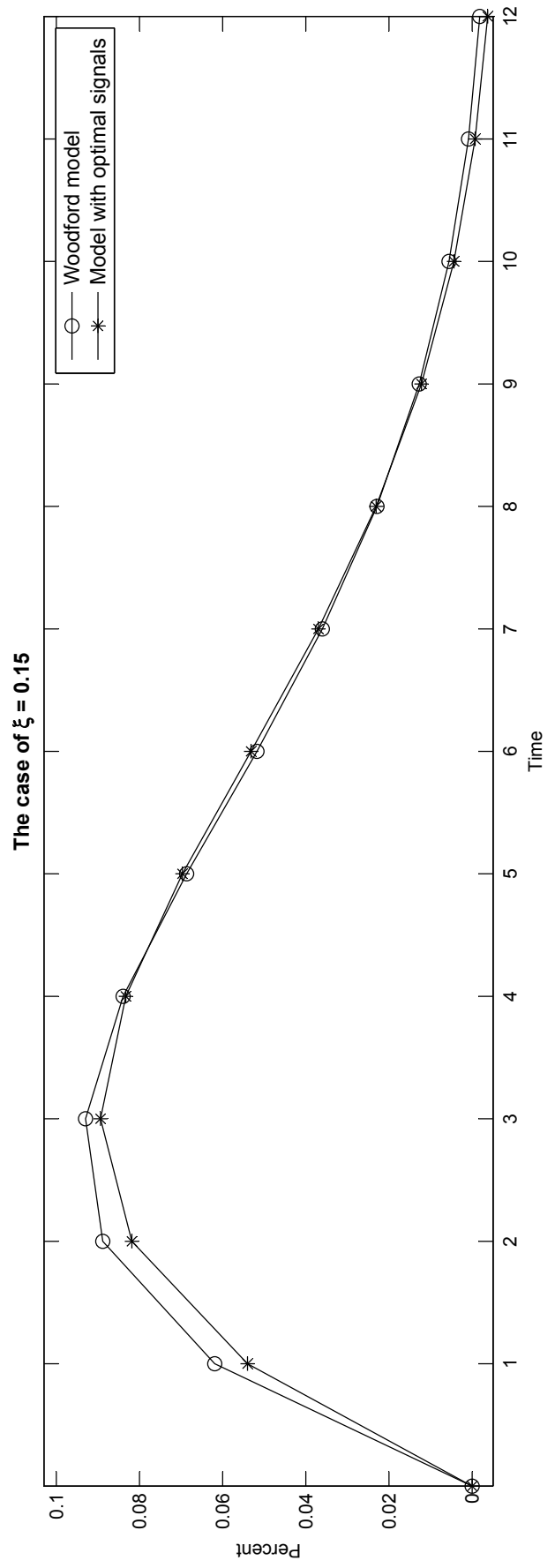
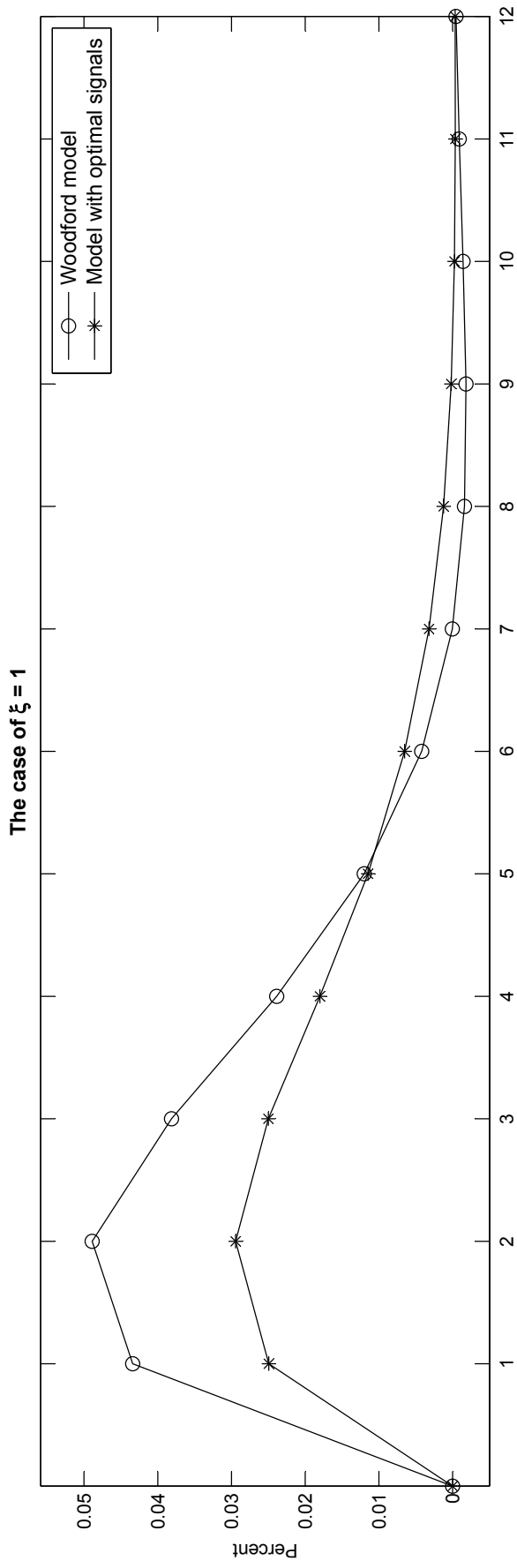


Figure 3: Impulse response of output to a nominal shock, Woodford model and model with optimal signals



## Acknowledgements

We thank seminar and conference participants at CEU Budapest, EEA 2015, SED 2015, and the 2016 conference on Rational Inattention and Related Theories in Prague 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.

### Bartosz Maćkowiak

European Central Bank and CEPR; email: [bartosz.mackowiak@ecb.int](mailto:bartosz.mackowiak@ecb.int)

### Filip Matějka

CERGE-EI and CEPR; email: [filip.matejka@cerge-ei.cz](mailto:filip.matejka@cerge-ei.cz)

### Mirko Wiederholt

Goethe University Frankfurt and CEPR; email: [wiederholt@wiwi.uni-frankfurt.de](mailto:wiederholt@wiwi.uni-frankfurt.de)

## © European Central Bank, 2017

Postal address 60640 Frankfurt am Main, Germany  
Telephone +49 69 1344 0  
Website [www.ecb.europa.eu](http://www.ecb.europa.eu)

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from [www.ecb.europa.eu](http://www.ecb.europa.eu), from the [Social Science Research Network electronic library](#) or from [RePEc: Research Papers in Economics](#). Information on all of the papers published in the ECB Working Paper Series can be found on the [ECB's website](#).

ISSN	1725-2806 (pdf)	DOI	10.2866/339353 (pdf)
ISBN	978-92-899-2728-4 (pdf)	EU catalogue No	QB-AR-17-018-EN-N (pdf)