Rational Inattention in the Frequency Domain*

Kyle Jurado
Duke University

September 11, 2020

Abstract

This paper solves the canonical dynamic rational inattention problem of Sims (2003) by formulating it in the frequency domain. The solution complements and extends existing results, which have been derived in the time domain. The paper provides a simple algorithm that quickly and accurately computes the solution numerically in the frequency domain, and shows how to obtain closed-form solutions when the target is an autoregressive moving average process. The key step to solve the problem is to correctly articulate the "no foresight" constraint. This constraint, often overlooked, restricts agents from processing information about future structural disturbances. The paper also analyzes the consequences of removing the no foresight constraint in a model of the real effects of nominal disturbances.

JEL classification: D83, D84, E21
Keywords: Rational inattention, spectral analysis, foresight.

---

*Department of Economics, Duke University, 419 Chapel Drive, Durham, NC 27708 (e-mail: kyle.jurado@duke.edu).

1 Introduction

Rational inattention is a theory about how purposeful agents make choices when faced with more information than they can possibly process. Since its proposal by Sims (1998, 2003), it has been increasingly adopted across many different areas of economics. Its main appeal is that it can provide an explanation for a number of "behavioral" phenomena, but in the context of a formal optimizing model in which counterfactual experiments can be conducted. One of the bottlenecks delaying even more widespread adoption of this theory is that it can be difficult to solve even some of the simplest rational inattention problems.

This paper revisits the canonical dynamic rational inattention problem initially proposed by Sims (2003). The problem is a linear-quadratic tracking problem, in which an agent chooses an action (or information structure), subject to constraints, so as to minimize the expected loss in estimating the current value of a given target process. It has received a good deal of attention in the literature, primarily because it distills many of the key issues involved in incorporating rational inattention into dynamic environments. It was originally solved using "brute-force" numerical optimization. Since then, one closed-form solution has been found, by Maćkowiak and Wiederholt (2009), in the special case that the target process is an AR(1) process. This is the only closed-form solution known in the literature, but two subsequent papers, Maćkowiak, Matějka, and Wiederholt (2018) and Miao, Wu, and Young (2020b), both present numerical algorithms that can solve this problem for any autoregressive moving average (ARMA) target processes more efficiently than brute-force optimization.

Existing approaches to solve this problem have operated in the "time domain," meaning that they have not exploited the spectral theory of random processes. The purpose of this paper is to show that it is possible to fully characterize the solution by making use of some basic insights from spectral analysis (i.e. by working in the "frequency domain"). In the frequency domain, the solution strategy becomes straightforward: form the Lagrangian, set its derivatives to zero, and then solve the resulting system of optimality conditions. It turns out that the optimality conditions reduce to one key nonlinear functional equation, which can be solved numerically in a fast and accurate way using the Fast Fourier Transform (FFT) algorithm. Unlike existing approaches, this procedure does not require the target process to have a known

ARMA form. Having a numerical procedure that does not require the input process to have a known ARMA form can be useful when the target process is itself endogenously determined, as in many equilibrium models. However, whenever the target is an ARMA process, the key equation is simple enough to admit a closed form solution in many cases. As examples, closed-form solutions for the AR(1), AR(2), MA(1), and ARMA(1,1) cases are presented. It is also proven that the solutions presented here are globally unique, something that has so far not been formally established, even for the AR(1) case.

The key step in formulating the problem in the frequency domain is properly articulating the "no foresight" constraint. What is sometimes under-appreciated is that this problem actually imposes *two informational constraints*, not one. The first restricts the rate of information flow from the target process to the agent's optimal action, and the second restricts the type of information structures the agent can entertain. More specifically, this second constraint, which will be discussed in more detail below, prevents the agent from attending to any information about future structural disturbances. Indeed, it was apparently this constraint which originally suggested to Sims that frequency domain methods would be unhelpful for solving the problem.[1] This paper shows that by appropriately articulating the no foresight constraint, frequency domain methods can prove effective.

This paper also characterizes the solution to the problem without the no foresight constraint. Since this constraint is always binding in the canonical version of the problem, its removal alters the properties of the optimal action. In particular, without the no foresight constraint it is no longer optimal for the agent to behave as if he were receiving a signal of the form: "current state vector + i.i.d. noise." The reason is that, in general, the agent can achieve a lower error variance by attending to those *frequencies* which receive the most weight according to the spectral density of the target process. Since each frequency contains information both about the past and the future, it is therefore optimal for the agent to respond to primitive disturbances in advance. In other words, it is optimal for the agent to construct an information structure with "endogenous news."

[1] "If Y [the action] could be chosen on the basis of future as well as past X [the target], it would be easy to use frequency-domain methods to solve the optimal tracking problem. For the more realistic situation where Y_t can depend only on $\{X_s, s \leq t\}$, I have not been able to find analytic formulas..." (Sims, 2003, p.672).

Results from the version of the problem both with and without the no foresight constraint are compared in the context of a general equilibrium model of the real effects of nominal disturbances. The model expresses a version of the Phelps-Lucas hypothesis, according to which rational inattention on the part of goods suppliers causes purely nominal disturbances to have real effects. A new result is that when suppliers are allowed to have foresight regarding structural disturbances, it is optimal for them to begin raising prices before the nominal disturbance occurs. This generates a temporary fall in real output, not because suppliers think that anticipated plans for future nominal stimulus "signal" a deterioration in real conditions, but because nominal expenditure is unchanged when suppliers begin raising prices. Once the disturbance occurs, real output increases due to imperfect perception of the disturbance, as it does with the no foresight constraint. This application shows how sluggish responses to nominal disturbances can coexist with anticipation effects, and how both can combine to produce endogenous waves of optimism and pessimism.

Related Literature. This paper is most closely related to Maćkowiak, Matějka, and Wiederholt (2018), who analyze the same canonical tracking problem. The main result of that paper is that, with the no foresight constraint, the optimal signal is a one-dimensional signal of the state vector plus i.i.d. noise. This paper is mostly complementary; differences are that (i) the problem is solved in the frequency domain rather than the time domain, (ii) the solution does not require the target process to have an ARMA structure, (iii) when the target does have an ARMA structure, the frequency-domain solution makes it possible to obtain closed-form solutions in many cases, (iv) all the solutions are shown to be globally unique, (v) the numerical algorithm also operates in the frequency domain, and exploits efficient numerical methods from the signal-processing literature, and (vi) this paper highlights the importance of the no foresight constraint and shows how its removal affects agents' optimal behavior.

A second complementary paper is Miao, Wu, and Young (2020b). That paper analyzes multivariate dynamic rational inattention problems, in which the target variable is a vector that is affected by multiple disturbances. This paper allows for multiple disturbances, but restricts attention only to the case that the target variable is a scalar. This case is sufficiently general to handle many economic applications of interest; in dynamic problems, the target variable will be a scalar whenever the individual agent's problem involves no more than *one endogenous state variable* (the number of control variables and exogenous state variables is not relevant). When the agent's problem involves more than one endogenous state variable, the target variable *may* still be a scalar, but this is no longer guaranteed. For an example in which the target variable is not a scalar, see the third application of Miao, Wu, and Young (2020b).

Another related paper is Chahrour and Jurado (2020). That paper also solves a model of endogenous information choice which allows for endogenous foresight. The main result of that paper is a generalization of the prediction formula of Hansen and Sargent (1980) to allow for endogenous foresight, subject to a constraint on the quantity of foresight. The difference is that that paper assumes that agents have *perfect hindsight*: they can costlessly attend to information about the current and past values of exogenous processes, and are only limited in their ability to process information about the future beyond this. In this paper, information about the past is not costless, as in most other applications of rational inattention.

This paper also participates in a tradition that has fruitfully applied frequency-domain methods to problems of macroeconomic interest. Prominent early contributions include Hansen and Sargent (1980), Futia (1981), Whiteman (1983), Taub (1989), and Kasa (2000). More recent contributions include Leeper, Walker, and Yang (2013), Kasa, Walker, and Whiteman (2014), Huo and Takayama (2018), and Miao, Wu, and Young (2020a). All of these papers analyze models in which agents' information sets are exogenously specified. The contribution of this paper is to apply frequency-domain methods to dynamic models of rational inattention, where agents' information sets are endogenous.

2 Formulation

This section formulates the canonical tracking problem of Sims (2003) in the frequency domain. It starts by stating the problem in the time domain and then translating the objective and constraints over into the frequency domain. As in many situations, much of the difficulty is in setting up the problem; solving it is purely mechanical. This section also serves to connect the frequency-domain notation and concepts to their time domain analogues, so that readers can more easily relate the two approaches.

The problem is to construct an action process $\{y_t\}$ to track a stationary Gaussian target process $\{x_t\}$ as closely as possible according to the quadratic objective[2]

$$E[(x_t - y_t)^2]. \tag{1}$$

The target is driven by a sequence of structural disturbances $\{\varepsilon_t\}$, according to the law of motion

$$x_t = \sum_{s=-\infty}^{\infty} a_s \varepsilon_{t-s}, \tag{2}$$

where $\{\varepsilon_t\}$ is an $n \times 1$ dimensional orthonormal white noise process, and $\{a_s\}$ is a sequence of $1 \times n$ dimensional coefficients. This representation generalizes the original formulation in Sims (2003) along two dimensions: first, the target can be driven by multiple structural disturbances, and second, the target can depend on future structural disturbances.

In choosing the action, the agent is subject to two informational constraints. The first is the information processing constraint, which requires that the rate of information about the target that is generated by the action be no greater than a finite amount $\kappa > 0$,

$$\lim_{T\to\infty} \frac{1}{T} I((x_{t+1}, \ldots, x_{t+T}), (y_{t+1}, \ldots, y_{t+T})) \leq \kappa, \tag{3}$$

where $I(x, y)$ denotes the mutual information between the random vectors x and y. The second constraint is the no foresight constraint, which requires the information generated by the past history of the action process about the future structural disturbances to be no greater than the information generated by the past history of the structural disturbances themselves,

$$\lim_{T\to\infty} I((\varepsilon_{t+1}, \ldots, \varepsilon_{t+T}), (y_t, y_{t-1}, \ldots)|(\varepsilon_t, \varepsilon_{t-1}, \ldots)) = 0 \tag{4}$$

where $I(x, y|z)$ denotes the conditional mutual information between the random variables x and y, conditional on z.[3] Stationarity ensures that both of these constraints are time invariant; this implies, for example, that it is possible to set $t = 0$ in both constraints without loss of generality.

---

[2]Why is the agent choosing an "action" and not an "information set"? These two ways of conceptualizing the problem are the same, because for any information set $\mathcal{S}_t$ chosen, it is always optimal in Gaussian environments to set $y_t = E[x_t|\mathcal{S}_t]$.

[3]Cover and Thomas (2006) is a standard reference on mutual and conditional mutual information.

The second of these two constraints is often articulated in a different, but equivalent way. To understand the difference, first note that it is always possible to decompose the action process by projecting it onto the space spanned by past, present, and future values of the structural disturbances,

$$y_t = \sum_{s=-\infty}^{\infty} b_s \varepsilon_{t-s} + \sum_{s=0}^{\infty} c_s v_{t-s}, \tag{5}$$

where $\{v_t\}$ is a sequence of independent random variables with zero mean and unit variance which is independent of the target process. The first term is the projection of y_t on the space spanned by $\ldots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \ldots$, and the weight on ε_{t-s} in the first term is given by the $1 \times n$ dimensional projection coefficient $b_s = E[y_t \varepsilon_{t-s}]$. The second term is the residual from this projection, expressed in terms of its Wold representation.[4] Since the action is a scalar, the residual in this projection is a scalar as well.

So far, (5) does not represent any constraint on the problem, and is purely representational. The no foresight constraint is introduced by requiring that

$$b_s = 0 \quad \text{for all } s < 0 \tag{6}$$

in (5). This means that the current action cannot be correlated with future structural disturbances. Compared to (4), this articulation of the constraint is somewhat less explicit; it is also easy to miss when it is expressed by replacing the lower summation limit in the first term of (5) with a zero. Nevertheless, (6) is interchangeable with (4). Both express the constraint that the action cannot contain more information about future structural disturbances.[5]

Now that the problem has been fully specified in the time domain, I proceed to translate it into the frequency domain. To do so, let

$$a(\lambda) = \sum_{s=-\infty}^{\infty} a_s e^{-i\lambda s}$$

denote the Fourier transformation of the coefficient sequence $\{a_s\}$, with analogous definitions for $b(\lambda)$ and $c(\lambda)$ based on the coefficient sequences in (5). These functions

[4] The concept of a Wold representation will be used repeatedly throughout this paper; for details see Sec. 5.7 of Brockwell and Davis (2006).

[5] In the existing literature, this constraint is imposed in Equation (13) of Sims (2003) and Equation (11) of Maćkowiak, Matějka, and Wiederholt (2018).

are complex-valued and are defined on the interval $[-\pi, \pi]$. In terms of these functions, the objective in (1) equals

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} a(\lambda)a(\lambda)^* + b(\lambda)b(\lambda)^* + |c(\lambda)|^2 - 2a(\lambda)b(\lambda)^* d\lambda,$$

where the asterisk denotes complex conjugate transposition. The integral of the first term is the variance of x_t, the integral of the second two terms is the variance of y_t, and the integral of the third term is twice the covariance of x_t and y_t.

For the information processing constraint, an important result due to Pinsker (1954) is that

$$\lim_{T\to\infty}\frac{1}{T}I((x_1,\ldots,x_T),(y_1,\ldots,y_T)) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\ln\left(1-\frac{|f_{xy}(\lambda)|^2}{f_x(\lambda)f_y(\lambda)}\right)d\lambda,$$

where the integrand is considered equal to zero whenever $f_x(\lambda)f_y(\lambda) = 0$. In this expression, $f_x(\lambda)$ is the spectral density of the target, $f_y(\lambda)$ is the spectral density of the action, and $f_{xy}(\lambda)$ is their cross spectral density. By writing these densities in terms of the functions $a(\lambda)$, $b(\lambda)$, and $c(\lambda)$, the information processing constraint takes the form[6]

$$-\frac{1}{4\pi}\int_{-\pi}^{\pi}\ln\left(1-\frac{b(\lambda)b(\lambda)^*}{b(\lambda)b(\lambda)^* + |c(\lambda)|^2}\right)d\lambda \leq \kappa.$$

Lastly, the no foresight constraint requires that $b_s = 0$ for all $s < 0$. Since $\{b_s\}$ are the Fourier coefficients of $b(\lambda)$, b_s can be obtained from $b(\lambda)$ using the inverse Fourier transform,

$$b_s = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda s}b(\lambda)d\lambda.$$

This means that the no foresight constraint can be expressed in the frequency domain as the sequence of restrictions that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda s}b(\lambda)d\lambda = 0 \quad \text{for all } s < 0.$$

To summarize, the canonical tracking problem can be formulated in the frequency domain as follows.

[6] Replacing natural logarithms with base 2 logarithms changes the interpretation of the units of κ from "nats" to "bits."

Problem 1 (Canonical tracking problem)**.** *Choose the* $1 \times n$ *dimensional function* $b(\lambda)$ *with elements in* $L^2[-\pi, \pi]$ *and the scalar function* $c(\lambda)$ *in* $L^2[-\pi, \pi]$ *to minimize*

$$\frac{1}{4\pi}\int_{-\pi}^{\pi} a(\lambda)a(\lambda)^* + b(\lambda)b(\lambda)^* + |c(\lambda)|^2 - 2a(\lambda)b(\lambda)^* d\lambda$$

subject to

$$-\frac{1}{4\pi}\int_{-\pi}^{\pi} \ln\left(1 - \frac{b(\lambda)b(\lambda)^*}{b(\lambda)b(\lambda)^* + |c(\lambda)|^2}\right) d\lambda \leq \kappa \qquad \text{(finite processing)}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\lambda s} b(\lambda) d\lambda = 0 \quad \textit{for all } s < 0. \qquad \text{(no foresight)}$$

Note that in stating this problem, the objective function has been re-scaled by 1/2, which helps simplify some expressions later on. Also note that the choice of $c(\lambda)$ is indeterminate; only $|c(\lambda)|^2$ appears in the objective and constraints. Following the normalization in (5), the coefficients $\{c_s\}$ can be defined as the coefficients from the Wold representation of the process with spectral density $\frac{1}{2\pi}|c(\lambda)|^2$.

3 Solution

This section uses standard methods in constrained optimization to characterize the solution to Problem (1). The first step is to set up the Lagrangian by attaching multipliers to each of the constraints and then adding them to the objective. In this case, the Lagrangian is

$$L = \frac{1}{4\pi}\int_{-\pi}^{\pi} \Bigg(|a(\lambda)|^2 + |b(\lambda)|^2 + |c(\lambda)|^2 - 2a(\lambda)b(\lambda)^* \\ - \theta\left[\ln\left(1 - \frac{|b(\lambda)|^2}{|b(\lambda)|^2 + |c(\lambda)|^2}\right) + 2\kappa\right] + 2\sum_{s=-\infty}^{-1} \psi_s e^{-i\lambda s} b(\lambda)^* \Bigg) d\lambda,$$

where $\theta \geq 0$ is the multiplier on the information processing constraint, and $\{\psi_s\}$ is the sequence of $1 \times n$ dimensional multipliers corresponding to the sequence of restrictions that make up the no foresight constraint.[7]

[7] Here I have used the fact that $\int_{-\pi}^{\pi} e^{i\lambda s} b(\lambda) d\lambda = \int_{-\pi}^{\pi} e^{-i\lambda s} b(\lambda)^* d\lambda$.

To find the optimality conditions, (Gateaux) differentiate the Lagrangian with respect to $b(\lambda)$ and $c(\lambda)$ and set the derivatives equal to zero. This yields the two first order conditions

$$b(\lambda) = a(\lambda) - \psi(\lambda) - \theta \frac{b(\lambda)}{b(\lambda)b(\lambda)^* + |c(\lambda)|^2} \tag{7}$$

$$|c(\lambda)|^2 = \theta \frac{b(\lambda)b(\lambda)^*}{b(\lambda)b(\lambda)^* + |c(\lambda)|^2}, \tag{8}$$

where the sequence of multipliers associated with the no foresight constraint have been collected into the single function

$$\psi(\lambda) \equiv \sum_{s=-\infty}^{-1} \psi_s e^{-i\lambda s}.$$

Combining the two first order conditions, I obtain expressions for $b(\lambda)$ and $|c(\lambda)|^2$ only in terms of the function $a(\lambda)$ and the Lagrange multipliers,

$$b(\lambda) = (a(\lambda) - \psi(\lambda)) \left(1 - \frac{\theta}{(a(\lambda) - \psi(\lambda))(a(\lambda) - \psi(\lambda))^*} \right) \tag{9}$$

$$|c(\lambda)|^2 = \theta \left(1 - \frac{\theta}{(a(\lambda) - \psi(\lambda))(a(\lambda) - \psi(\lambda))^*} \right). \tag{10}$$

To simplify the solution further, I introduce the operator

$$[h(\lambda)]_- \equiv \sum_{s=-\infty}^{-1} h_s e^{-i\lambda s},$$

which sets all the non-negative Fourier coefficients of a function $h(\lambda)$ with elements in $L^2[-\pi, \pi]$ to zero. Applying this operator to both sides of (9), using the fact that the no foresight constraint implies that $[b(\lambda)]_- = 0$, and solving for $\psi(\lambda)$ delivers the relation[8]

$$\psi(\lambda) = \left[a(\lambda) - \theta \frac{a(\lambda) - \psi(\lambda)}{(a(\lambda) - \psi(\lambda))(a(\lambda) - \psi(\lambda))^*} \right]_- . \tag{11}$$

The multiplier θ is chosen so that the information processing constraint holds with equality. Equations (8) and (10) imply that

$$\frac{b(\lambda)b(\lambda)^*}{b(\lambda)b(\lambda)^* + |c(\lambda)|^2} = 1 - \frac{\theta}{(a(\lambda) - \psi(\lambda))(a(\lambda) - \psi(\lambda))^*},$$

[8]Whenever $a_s = 0$ for all $s < 0$, $[a(\lambda)]_- = 0$, so the first term inside the square brackets vanishes.

which can be substituted into the information processing constraint to obtain an expression for the multiplier θ only in terms of $a(\lambda)$, $\psi(\lambda)$, and κ,

$$\theta = \exp\left(-2\kappa + \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[(a(\lambda) - \psi(\lambda))(a(\lambda) - \psi(\lambda))^{*}]d\lambda\right). \tag{12}$$

Substitution of (12) into (11) delivers the key nonlinear equation of the model, which implicitly determines $\psi(\lambda)$ as a function of the primitive objects $a(\lambda)$ and κ,

$$\boxed{\begin{aligned}\psi(\lambda) = \Bigg[a(\lambda) - \exp\left(-2\kappa + \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[(a(\lambda) - \psi(\lambda))(a(\lambda) - \psi(\lambda))^{*}]d\lambda\right) \\ \times \frac{a(\lambda) - \psi(\lambda)}{(a(\lambda) - \psi(\lambda))(a(\lambda) - \psi(\lambda))^{*}}\Bigg]_{-}\end{aligned}} \tag{13}$$

Given a solution to this equation, θ can be computed from (12), and then $b(\lambda)$ and $|c(\lambda)|^2$ can be computed from (9) and (10). Therefore, solving the canonical tracking problem reduces to solving (13). Fortunately, it turns out that solving it is not too difficult; the next section describes two methods of doing so. The first assumes that $\{x_t\}$ has ARMA dynamics, which makes it possible to obtain a closed-form solution. The second does not assume that $\{x_t\}$ has ARMA dynamics, and uses a simple numerical algorithm to compute the solution.

Finally, it is important to note that the constraint set in Problem (1) is convex, and the objective function is strictly convex. Therefore, these optimality conditions are both necessary and sufficient for a global optimum. Once a solution to (13) has been obtained, it is possible to conclude that that solution is *globally unique*.

Lemma 1. *Any solution to* (9), (10), (12), *and* (13) *is globally unique.*

4 Computation

The previous section has shown that the canonical tracking problem can be reduced to solving a fixed-point problem in the space of square-integrable functions whose non-negative Fourier coefficients are all zero (Equation 13). This section first describes how to solve that fixed-point problem in closed form whenever the target process has ARMA dynamics. It then presents a simple algorithm that can solve the problem numerically for arbitrary target dynamics, by using standard numerical procedures from the signal processing literature.

4.1 Closed-form solutions

To illustrate how to find a closed form solution to Equation (13), I consider the case that the target process has ARMA(1,1) dynamics. This case will be worked out in detail, making it easier to see the main steps involved. The discussion also illustrates how this procedure can be generalized to handle any ARMA process, just at the cost of additional notation and algebraic calculations. Analogous results for the AR(1), AR(2), and MA(1) cases are presented in Appendix (A).

Suppose that the target process has ARMA(1,1) dynamics,

$$x_t = \sigma \frac{(1-\alpha L)}{(1-\rho L)} \varepsilon_t \equiv A(L)\varepsilon_t,$$

where $\sigma > 0$, $|\alpha| < 1$, $|\rho| < 1$, and L is the lag operator. Viewing the linear filter $A(L)$ as a function of the complex variable $z = e^{-i\lambda}$, the function $a(\lambda)$ is given by

$$a(\lambda) = A(e^{-i\lambda}).$$

Throughout this section, I will follow the convention that upper-case letters denote functions of complex variables, and lower-case letters denote the value of their upper-case counterpart evaluated at $e^{-i\lambda}$ and viewed as a function of the real number λ.

Given the function $a(\lambda)$ and the parameter κ, the solution procedure is a simple application of the method of undetermined coefficients:

(i) Conjecture a functional form for $\psi(\lambda)$ with unknown coefficients, which is rational in $e^{-i\lambda}$ and with all non-negative Fourier coefficients equal to zero.

(ii) Substitute this conjecture into (13) to obtain a system of nonlinear equations in the unknown coefficients, and solve them.

Regarding step (i), the difficulty is coming up with a good conjecture. Some practice with manipulation of Equation (13) suggests that whenever the target process is ARMA(p,q), a good conjecture is that $\psi(\lambda) = \Psi(e^{-i\lambda})$, where $L\Psi(L^{-1})$ is the linear filter associated with an ARMA$(\max(p-1,q), \max(p-1,q-1))$ process. For example, when the target is an ARMA(1,1), this corresponds to

$$\Psi(z) = \sigma\omega \frac{z^{-1}}{1-\phi z^{-1}} \tag{14}$$

where $|\phi|< 1$ and ω are unknown coefficients (σ appears only as a convenient rescaling). Conjecturing more ARMA terms than this will still deliver the correct solution, since Equation (13) will simply require these additional terms to drop out. Conjecturing fewer terms than this will lead to a contradiction in Equation (13), making it easy to recognize the mistake. Regarding step (ii), the difficulties are to compute the integral inside the exponential and evaluate the operator $[\quad]_-$. Fortunately, some basic results from the theory of analytic functions, discussed below, greatly simplify these tasks.

To solve the problem, first note that the fact that $a(\lambda)$ is a scalar function satisfying $[a(\lambda)]_-$ implies that (13) reduces to

$$\psi(\lambda) = -\exp\left(-2\kappa + \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln |a(\lambda) - \psi(\lambda)|^2 \, d\lambda\right) \left[\frac{1}{a(\lambda)^* - \psi(\lambda)^*}\right]_- . \tag{15}$$

To solve this equation, start with the conjecture for $\Psi(z)$ in (14) and write

$$\begin{aligned} A(z) - \Psi(z) &= \sigma \left[\frac{(1-\alpha z)(1-\phi z^{-1}) - \omega z^{-1}(1-\rho z)}{(1-\rho z)(1-\phi z^{-1})}\right] \\ &= \sigma \alpha r_2 \frac{(1 - r_1 z^{-1})(1 - r_2^{-1} z)}{(1-\rho z)(1-\phi z^{-1})}, \end{aligned} \tag{16}$$

where r_1 and r_2 are the two roots of the polynomial

$$\mathcal{P}(r) = \alpha r^2 - (1 + \alpha\phi + \omega\rho) r + (\phi + \omega).$$

Let $|r_1|< 1$ and $|r_2|> 1$, which will be ensured later on. For now, it is enough to note that r_1 and r_2 satisfy

$$r_1 + r_2 = \frac{1 + \alpha\phi + \omega\rho}{\alpha} \tag{17}$$

$$r_1 r_2 = \frac{\phi + \omega}{\alpha} \tag{18}$$

Using the expression in (16), it is possible to compute the integral in (15) using a well-known formula due to Kolmogorov (1941), which says that

$$\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln f(\lambda) d\lambda\right) = \frac{1}{2\pi}|\Gamma(0)|^2,$$

where $\Gamma(L)$ is the linear filter from the Wold representation of a process with spectral density $f(\lambda)$.[9] Setting $f(\lambda) \equiv \frac{1}{2\pi}|a(\lambda) - \psi(\lambda)|^2$, this formula becomes

$$\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|a(\lambda) - \psi(\lambda)|^2 \, d\lambda\right) = |\Gamma(0)|^2, \tag{19}$$

which means that computing the integral in (15) reduces to finding the Wold representation associated with the spectral density $f(\lambda)$. Finding the Wold representation in this case amounts to representing $f(\lambda)$ in the form

$$f(\lambda) = \frac{1}{2\pi}\frac{|P(e^{-i\lambda})|^2}{|Q(e^{-i\lambda})|^2},$$

where the polynomials $P(z)$ and $Q(z)$ do not have any zeros inside the unit circle.[10] From (16), it is easy to see that setting $P(z) = \sigma\alpha r_2(1 - r_1 z)(1 - r_2^{-1}z)$ and $Q(z) = (1 - \rho z)(1 - \phi z)$ delivers the appropriate representation. Therefore,

$$\Gamma(z) = \sigma\alpha r_2 \left[\frac{(1 - r_1 z)(1 - r_2^{-1}z)}{(1 - \rho z)(1 - \phi z)}\right].$$

By Kolmogorov's formula, it follows that the integral in (15) is

$$\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln|a(\lambda) - \psi(\lambda)|^2 \, d\lambda\right) = (\sigma\alpha r_2)^2. \tag{20}$$

Next, it is necessary to evaluate $[1/f(\lambda)^*]_-$. To do so, first note that (16) implies

$$\frac{1}{A(z^{-1}) - \Psi(z^{-1})} = \frac{1}{\sigma\alpha r_2}\frac{(z - \rho)(1 - \phi z)}{(1 - r_1 z)(z - r_2^{-1})}$$

This function is analytic in the unit circle except at the point $z = r_2^{-1}$, and has a convergent Laurent series expansion in the annulus $r_2^{-1} < |z| < r_1^{-1}$. By definition, evaluation of the operator $[\quad]_-$ amounts to finding the singular part of this expansion (the part with only negative powers of z) about r_2^{-1}. Letting $S(z)$ denote this singular part, it follows that

$$\left[\frac{1}{a(\lambda)^* - \psi(\lambda)^*}\right]_- = S(e^{-i\lambda}). \tag{21}$$

[9] See, e.g. Brockwell and Davis (2006) Thm. 5.8.1.

[10] See Rozanov (1967) Ch. 2, Thm. 5.3.

To find the principal part $S(z)$, first compute the residue of the Laurent expansion at r_2^{-1}. This is given by

$$\begin{aligned}\text{Res}\left(\frac{1}{A(z^{-1})-\Psi(z^{-1})}, r_2^{-1}\right) &= \lim_{z\to 1/r_2}(z-r_2^{-1})\frac{1}{A(z^{-1})-\Psi(z^{-1})}\\ &= \frac{1}{\sigma\alpha r_2}\frac{(r_2^{-1}-\rho)(1-\phi/r_2)}{(1-r_1/r_2)}\end{aligned}$$

Since r_2^{-1} is a simple pole of order one, the residue can be used to compute the singular part as follows,[11]

$$S(z) = \frac{(r_2^{-1}-\rho)(1-\phi/r_2)}{\sigma\alpha r_2(1-r_1/r_2)}\frac{1}{(z-r_2^{-1})}.$$

and therefore by (21),

$$\left[\frac{1}{a(\lambda)^*-\psi(\lambda)^*}\right]_- = \frac{(r_2^{-1}-\rho)(1-\phi/r_2)}{\sigma\alpha r_2(1-r_1/r_2)}\frac{1}{(e^{i\lambda}-r_2^{-1})}. \tag{22}$$

Plugging the initial conjecture for $\psi(\lambda)$ into the left side of (15), and plugging (20) and (22) into the right side, I obtain

$$\sigma\omega\frac{1}{(e^{i\lambda}-\phi)} = \frac{(r_2^{-1}-\rho)(1-\phi/r_2)}{\sigma\alpha r_2(1-r_1/r_2)}\frac{1}{(e^{i\lambda}-r_2^{-1})}.$$

Matching coefficients, the initial conjecture is correct if

$$\phi = r_2^{-1} \tag{23}$$

$$\omega = -e^{-2\kappa}\alpha r_2\frac{(r_2^{-1}-\rho)(1-\phi/r_2)}{(1-r_1/r_2)}. \tag{24}$$

The system of four equations, (17), (18), (23), and (24), then jointly determine the value of the four coefficients (ω, ϕ, r_1, r_2). This solution for $\psi(\lambda)$ can be plugged into (12) to obtain the solution for θ. Then (9) and (10) can be used to determine $b(\lambda)$ and $c(\lambda)$. The results of these additional algebraic manipulations are summarized in the following proposition.

Proposition 1. *When the target is an ARMA(1,1) process, so*

$$a(\lambda) = \sigma\frac{(1-\alpha e^{-i\lambda})}{(1-\rho e^{-i\lambda})}$$

[11] For details, see Ch. 5 of Conway (1973). By way of comparison, this is the same step used by Hansen and Sargent (1980) in deriving their prediction formula.

with $\sigma > 0$, $|\alpha| < 1$, $|\rho| < 1$, and $\alpha \neq \rho$, the unique solution to Problem (1) *is*

$$b(\lambda) = \sigma_b \frac{(1-\phi e^{-i\lambda})(1-\phi e^{-i\lambda})}{(1-\rho e^{-i\lambda})(1-r_1 e^{-i\lambda})}$$

$$c(\lambda) = \sigma_c \frac{(1-\phi e^{-i\lambda})}{(1-r_1 e^{-i\lambda})},$$

where the parameters in these expressions are given by

$$\sigma_b = \sigma \frac{\alpha(r_1 - \rho e^{-2\kappa})}{\phi^2} \quad \sigma_c = \sigma \sqrt{e^{-2\kappa} \frac{\alpha^2(r_1 - \rho e^{-2\kappa})}{\phi^3}} \quad r_1 = \frac{\alpha(1-\phi^2) - \phi(1-\rho\phi)}{\alpha(\rho - \phi)},$$

and ϕ is the root of the quartic equation

$$\mathcal{P}(\phi) = (\alpha - \rho)\phi^4 + (1 + \alpha\rho - \alpha^2(1 + e^{-2\kappa}))\phi^3 - 3\alpha(1 - \alpha\rho e^{-2\kappa})\phi^2$$

$$+ (\alpha\rho + 2\alpha^2 - 3\alpha^2\rho^2 e^{-2\kappa})\phi - \alpha^2\rho(1 - \rho^2 e^{-2\kappa})$$

which satisfies $|\phi| < 1$ and ensures that $|r_1| < 1$.

This proposition directly characterizes the dynamics of the optimal action $\{y_t\}$. For some purposes, it may be interesting to describe the optimal information structure not in terms of the optimal action, but in terms of a *subjective signal* that agents use to construct their optimal action. The precise nature of this signal is not uniquely determined by the problem (only the action is). However, consistent with Proposition 1 of Maćkowiak, Matějka, and Wiederholt (2018), the solution presented here implies that the optimal action can be generated by a signal of the special form "current state + i.i.d. noise."

Corollary 1. *The optimal action in Proposition* (1) *is equal to $y_t = E[x_t|s^t]$ when the subjective signal s_t is given by*

$$s_t = \varsigma_x x_t + \varsigma_\varepsilon \varepsilon_t + \varsigma_v v_t,$$

where $\{v_t\}$ is a sequence of independent normal random variables with zero mean and unit variance that is independent of $\{\varepsilon_t\}$,

$$\varsigma_x = \frac{\phi - \rho}{\alpha - \rho} \qquad \varsigma_\varepsilon = \sigma \left(\frac{\alpha - \phi}{\alpha - \rho} \right) \qquad \varsigma_v = \frac{\sigma \sigma_c}{\sigma_b},$$

and where ϕ, σ_b, and σ_c are defined as in Proposition (1).

Proof. Starting from the solution in Proposition (1), define the signal

$$s_t \equiv \frac{\sigma}{\sigma_b}\frac{(1-r_1L)}{(1-\phi L)}y_t = \sigma\frac{1-\phi L}{1-\rho L}\varepsilon_t + \frac{\sigma\sigma_c}{\sigma_b}v_t.$$

Since $|\phi|< 1$ and $|r_1|< 1$, this transformation does not change the agent's optimal estimate of the target variable; i.e. $y_t = E[x_t|y^t] = E[x_t|s^t]$. Now it only remains to rewrite the first term as a linear combination of x_t and ε_t. To do this set

$$\varsigma_x x_t + \varsigma_\varepsilon \varepsilon_t = \sigma\frac{1-\phi L}{1-\rho L}\varepsilon_t.$$

Substituting in the law of motion for $\{x_t\}$ into the left side and matching coefficients in L delivers the expressions for ς_x and ς_ε stated in the corollary. □

4.2 Numerical solution

The procedure described in the previous subsection can, in principle, be used to obtain the closed-form solution in the case of arbitrary ARMA dynamics. This section describes an algorithm that can quickly and efficiently compute the numerical solution, regardless of whether the target process has ARMA dynamics.

The idea is simple: construct a grid on the interval $[-\pi, \pi]$, make an initial guess for the values of $\psi(\lambda)$ on this grid, and then iterate on Equation (13) until convergence. As in the closed-form procedure, the only difficulties arise in computing the integral in (13) and evaluating the $[\quad]_-$ operator.

Given a function $a(\lambda)$ and an iterate of $\psi(\lambda)$, a direct way to compute the integral in (13) is to approximate the integral on the discretely sampled values of these functions using, for example, trapezoidal numerical integration. This will work, but it is apparently less efficient than another alternative, in the sense of requiring a finer grid to achieve the same level of accuracy. This alternative is to appeal once again to Kolmogorov's formula. Equation (19) says that integration can be substituted with finding a Wold representation. Fortunately, finding a Wold representation only requires numerical computation of Fourier coefficients, which is fast and accurate thanks to the Fast Fourier Transform (FFT) algorithm.[12]

To use (19) it is necessary to factor the spectral density

$$f(\lambda) \equiv \frac{1}{2\pi}(a(\lambda) - \psi(\lambda))(a(\lambda) - \psi(\lambda))^*$$

[12]For details on this algorithm, see Brockwell and Davis (2006), Sec. 10.7.

in the form

$$f(\lambda) = \frac{1}{2\pi}|\Gamma(e^{-i\lambda})|^2$$

where all the negative Fourier coefficients of the scalar function $\Gamma(e^{-i\lambda})$ vanish. It is not difficult to see that the solution to this problem is given by

$$\Gamma(e^{-i\lambda}) = \sqrt{2\pi}\exp\left(\frac{1}{2}\vartheta_0 + \sum_{s=1}^{\infty}\vartheta_s e^{-i\lambda s}\right), \tag{25}$$

where $\{\vartheta_s\}$ is the sequence of coefficients from the Fourier series expansion of $\ln f(\lambda)$,

$$\ln f(\lambda) = \sum_{s=-\infty}^{\infty}\vartheta_s e^{-i\lambda s}.$$

Given discrete values of the function $\ln f(\lambda)$, these coefficients can be efficiently computed using the inverse FFT; for example, using the function `ifft` in Matlab. By Kolmogorov's formula,[13]

$$\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left[(a(\lambda)-\psi(\lambda)(a(\lambda)-\psi(\lambda))^*\right]d\lambda\right) = 2\pi e^{\vartheta_0}.$$

Evaluation of the operator $[\quad]_-$ is similarly easy. This just requires computing the Fourier expansion

$$a(\lambda) - \theta\frac{a(\lambda)-\psi(\lambda)}{(a(\lambda)-\psi(\lambda))(a(\lambda)-\psi(\lambda))^*} = \sum_{s=-\infty}^{\infty}\varrho_s e^{-i\lambda s}$$

and then setting

$$\left[a(\lambda) - \theta\frac{a(\lambda)-\psi(\lambda)}{(a(\lambda)-\psi(\lambda))(a(\lambda)-\psi(\lambda))^*}\right]_- = \sum_{s=-\infty}^{-1}\varrho_s e^{-i\lambda s}.$$

The first step can be performed in Matlab using `ifft`, and then the second can be performed, after changing all the negative Fourier coefficients to zero, using `fft`.

Matlab functions that perform these computations are presented in Appendix (B). The function `wold` computes the Wold representation given a vector of values of the spectral density evaluated on a frequency grid, and the function `singular` computes

[13] Since ϑ_0 is the only coefficient from the expansion of $\ln f(\lambda)$ that is needed, it may be possible to increase speed by avoiding computing the unnecessary coefficients. However, because the algorithm is already so fast, I have not pursued this further.

the singular part of a function. Finally, the function `track` computes the optimal values of $b(\lambda)$ and $c(\lambda)$ on the grid, given the values of $a(\lambda)$ on the same grid, and a value for the parameter κ. To obtain the time-domain coefficients $\{b_s\}$ and $\{c_s\}$ associated with these functions, one can execute the commands `bs = ifft(b)` and `cs = ifft(c)`.

5 Foresight

So far the paper has provided a characterization of the solution to the canonical tracking problem with both information constraints: the processing constraint and the no foresight constraint. This section characterizes the solution to the same problem when the no foresight constraint is removed. This exercise reveals that in general it is optimal for a rationally inattentive agent to pay attention to more information about the target process than what is contained in the past history of structural disturbances.

Without the no foresight constraint, the Lagrange multiplier $\psi(\lambda)$ disappears from (7), which becomes

$$b(\lambda) = a(\lambda) - \theta \frac{b(\lambda)}{b(\lambda)b(\lambda)^* + |c(\lambda)|^2}. \tag{26}$$

Taking the complex conjugate transpose of both sides of (26), left-multiplying by $b(\lambda)$, and using the expression for $|c(\lambda)|^2$ in (8), it follows that

$$b(\lambda)b(\lambda)^* + |c(\lambda)|^2 = b(\lambda)a(\lambda)^*. \tag{27}$$

The main difference is that now it is important to explicitly account for the possibility that $b(\lambda)$ may be zero on a set of positive measure (previously, the no foresight constraint prevented this from happening). Consider each of the two possible cases. First, for all λ such that $b(\lambda) \neq 0$, (27) can be substituted into (26) to obtain

$$b(\lambda) = a(\lambda) \left(1 - \frac{\theta}{a(\lambda)a(\lambda)^*}\right),$$

which in turn implies, from (27), that

$$b(\lambda)b(\lambda)^* + |c(\lambda)|^2 = a(\lambda)a(\lambda)^* - \theta.$$

Second, for all λ such that $b(\lambda) = 0$, (27) implies that $b(\lambda)b(\lambda)^* + |c(\lambda)|^2 = 0$ as well.

Combining the results from each of these two cases, one can conclude that

$$b(\lambda) = \begin{cases} a(\lambda)\left(1 - \dfrac{\theta}{a(\lambda)a(\lambda)^*}\right) & a(\lambda)a(\lambda)^* > \theta \\ 0 & a(\lambda)a(\lambda)^* \leq \theta. \end{cases} \tag{28}$$

Finally, the multiplier θ is chosen to ensure that the information processing constraint holds with equality. Combination of (27) and (28) implies that

$$\frac{b(\lambda)b(\lambda)^*}{b(\lambda)b(\lambda)^* + |c(\lambda)|^2} = \max\left\{1 - \frac{\theta}{a(\lambda)a(\lambda)^*}, 0\right\}. \tag{29}$$

By substituting this into the processing constraint, it follows that θ solves the non-linear equation

$$\boxed{\frac{1}{4\pi}\int_{-\pi}^{\pi} \max\left\{\ln\left(\frac{a(\lambda)a(\lambda)^*}{\theta}\right), 0\right\} d\lambda = \kappa.} \tag{30}$$

Given the solution to this equation, $b(\lambda)$ can be computed from (28), and, using (29) and the second optimality condition (8), $|c(\lambda)|^2$ can be computed from

$$|c(\lambda)|^2 = \max\left\{\theta\left(1 - \frac{\theta}{a(\lambda)a(\lambda)^*}\right), 0\right\}. \tag{31}$$

Therefore, solving the tracking problem without the no foresight constraint reduces to solving Equation (30) for the unknown parameter θ. This can be done numerically using a nonlinear equation solver; e.g. `fzero` in Matlab.[14]

The intuition behind this solution is transparent. The scalar function $a(\lambda)a(\lambda)^*$ assigns a value to each frequency on the interval $[-\pi, \pi]$ based on the dynamics of the target process. The parameter θ is the shadow price of one unit of information. The agent optimally chooses to ignore frequencies whose value is below the shadow price of information.[15]

Figure (1) illustrates the relationship between $a(\lambda)a(\lambda)^*$ and the shadow price θ. Given the finite processing capacity κ, the agent optimally allocates attention only

[14]Because $a(\lambda)a(\lambda)^*$ is symmetric on $[-\pi, \pi]$, the integral only needs to be evaluated over $[0, \pi]$.

[15]For readers familiar with the engineering literature on rate distortion theory, the fact that this solution has a familiar "reverse water-filling" form is not a coincidence; the rational inattention problem is the *dual* of the rate distortion problem. The solution to the rate distortion problem with parallel Gaussian sources is presented in McDonald and Schultheiss (1964); see also the discussion in Ch. 10 of Cover and Thomas (2006). The relationship between rate distortion theory and rational inattention is also highlighted by Miao, Wu, and Young (2020b).

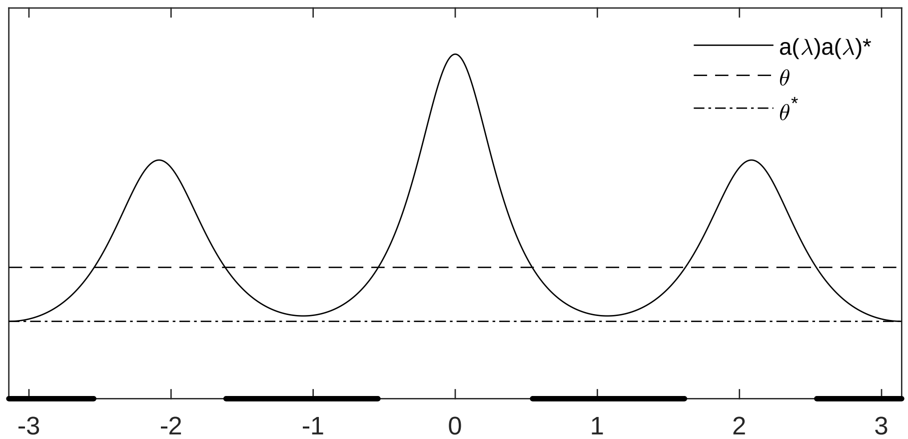

Figure 1: Optimal band-pass filter.

to the most important frequencies. In the figure, the most important frequencies are those for which the value $a(\lambda)a(\lambda)^*$ exceeds the horizontal cutoff level (the dashed line). The remaining frequencies, which correspond to the dark regions along the horizontal axis, are ignored. Depending on the dynamic properties of the target, the set of frequencies that are ignored does not need to be connected (as in the figure). The figure illustrates how, given the primitives $a(\lambda)$ and κ, the agent effectively constructs an *endogenous band-pass filter*, which removes frequencies that least contribute to the variance of the target.

Figure (1) also illustrates how in some cases, the agent may choose not to ignore any frequency. The shadow price θ is a decreasing function of κ; the more information processing capacity available, the more frequencies can be attended to. If κ is sufficiently large, the horizontal line in Figure (1) may fall completely below the graph of $a(\lambda)a(\lambda)^*$. The critical value of κ that determines whether any frequencies need to be ignored is the one that makes $\theta = \theta^*$, where

$$\theta^* \equiv \min_{\lambda \in [0,\pi]} a(\lambda)a(\lambda)^*.$$

Plugging this into (30), it follows that the critical value of κ is given by

$$\kappa^* = \frac{1}{4\pi} \int_{-\pi}^{\pi} \ln\left(\frac{a(\lambda)a(\lambda)^*}{\theta^*}\right) d\lambda.$$

(See the dash-dotted line in Figure 1.) For any $\kappa \geq \kappa^*$, (30) admits the closed-form solution

$$\theta = \exp\left(-2\kappa + \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[a(\lambda)a(\lambda)^*] d\lambda\right).$$

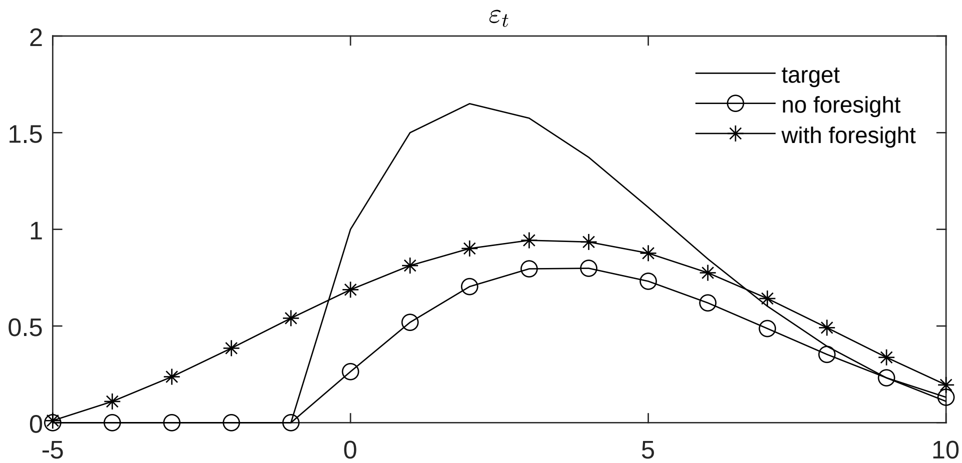

Figure 2: Rational inattention with and without foresight. This figure plots impulse responses with respect to the disturbance ε_t. The target process is an AR(2) of the form $x_t = 1.5x_{t-1} - 0.6x_{t-2} + \varepsilon_t$, and $\kappa = 0.1$.

6 Comparison

This section compares the solutions of the rational inattention tracking problem with and without the no foresight constraint. To have a concrete example, suppose that the target is an AR(2) process of the form

$$x_t = 1.5x_{t-1} - 0.6x_{t-2} + \varepsilon_t, \quad \varepsilon_t \overset{\text{iid}}{\sim} N(0,1),$$

and processing capacity is constrained to be no greater than $\kappa = 0.1$ units of information. Figure (2) plots the impulse response coefficients of the target and the optimal action to the structural disturbance ε_t, both with and without foresight. The horizontal axis indicates periods since the disturbance has occurred, so negative values indicate periods of *anticipation*. The target exhibits hump-shaped dynamics, in a manner suggestive of how an aggregate macroeconomic time series might respond to a structural disturbance of interest.

With the no foresight constraint, the agent is not permitted to respond to the structural disturbance in advance. This can be seen in the fact that the coefficients marked with circles are zero during the period of anticipation. Once the disturbance occurs, the agent begins to respond, but with a delay, in the sense that the peak response of the action occurs after the peak response of the target. The effect of the disturbance on the optimal action is also more persistent than the effect on the target.

The optimal action with rational inattention exhibits "stickiness," which is one of the main motivations for introducing this mechanism into macroeconomic models.

Without the no foresight constraint, the agent chooses to respond to the disturbance in advance. This can be seen in the fact that the coefficients marked with asterisks are nonzero during the period of anticipation. Despite the fact that the agent responds to the disturbance in advance, the peak response in the action still occurs after the peak response in the target. Moreover, the effect of the disturbance on the optimal action is also more persistent than its effect on the target. The impact response (at period zero) is higher with foresight, and remains higher for a number of periods (eventually, these lines do cross).

The reason it is optimal to respond to the disturbance in advance is that the agent would like to allocate attention *by frequency*. Since each frequency contains information about all time periods (past, present, and future), doing this requires the agent to receive information about future values of the target process. In this example, the spectral density of the target process has a large low-frequency component. With limited attention, the agent implements an endogenous low-pass filter, which only retains frequencies below a certain threshold.

7 Application

This section examines the importance of the no foresight constraint in a model of the real effects of nominal disturbances. The model is based on Woodford (2003), and illustrates a version of the hypothesis of Phelps (1969) and Lucas (1972) that purely nominal disturbances can have temporary real effects due to imperfect perception of aggregate conditions.

According to the model, the monopolistically competitive supplier of a good i chooses the price p_{it} at which the good is sold so as to maximize expected discounted profits. Up to a log-linear approximation, the supplier's optimal choice is to set the price equal to its best estimate of a linear combination between the aggregate price level $p_t = \int p_{it} di$ and aggregate nominal expenditure, q_t,

$$p_{it} = E_{it}[(1 - \xi)p_t + \xi q_t]. \tag{32}$$

The parameter $0 < \xi < 1$ controls the degree of strategic complementarity or substitutability in pricing: when $\xi < 1$ prices are strategic complements, whereas when

$\xi > 1$ they are strategic substitutes. In the case that $\xi = 1$, prices are neither complements nor substitutes, and the supplier's optimal price depends only on its expectation of aggregate nominal expenditure. To close the model, nominal expenditure is taken to be an exogenous stationary Gaussian process,

$$q_t = \sum_{s=0}^{\infty} \delta_s \varepsilon_{t-s},$$

where $\{\varepsilon_t\}$ is a sequence of independent normal random variables with zero mean and unit variance. These variables represent structural disturbances that originate in the activities undertaken by the monetary authority.

When suppliers perfectly perceive the aggregate price level and the level of nominal expenditure, the optimality condition (32) reduces to

$$p_{it} = (1 - \xi)p_t + \xi q_t.$$

Integrating across i, this implies that $p_t = q_t$, which indicates that purely nominal disturbances are reflected one-for-one in aggregate prices. Real output, defined as $y_t \equiv q_t - p_t$, is completely unaffected; nominal disturbances are neutral.

On the other hand, when suppliers can only imperfectly perceive aggregate variables, it is possible for nominal disturbances to have real effects. To model suppliers' imperfect perceptions of current conditions, suppose that each supplier solves a tracking problem with rational inattention. Woodford (2003) appeals to rational inattention as a motivation for imperfect perception, but nevertheless specifies agents' information sets exogenously. Maćkowiak, Matějka, and Wiederholt (2018) extend this model to explicitly allow for endogenous information choice, but do so in the time domain and do not explore the consequences of removing the no foresight constraint. To map the supplier's problem into the framework of the paper, define the target variable

$$x_t \equiv (1 - \xi)p_t + \xi q_t.$$

Up to a log-quadratic approximation of the supplier's objective function, maximizing discounted profits is equivalent to minimizing the size of tracking errors according to the quadratic loss function

$$E[(x_t - p_{it})^2].$$

The choice of p_{it} is made subject to an information processing constraint of the form

$$\lim_{T\to\infty} \frac{1}{T} I((x_{t+1}, \dots, x_{t+T}), (p_{i,t+1}, \dots, p_{i,t+T})) \leq \kappa,$$

which says that the supplier can only process information about the target variable up to rate $\kappa > 0$. Finally, the no foresight constraint takes the form

$$\lim_{T\to\infty} I(\varepsilon_{t+1},\ldots,\varepsilon_{t+T}),(p_{i,t},p_{i,t-1},\ldots)|(\varepsilon_t,\varepsilon_{t-1},\ldots)) = 0.$$

The only difference relative to the problem studied in previous sections of this paper is the fact that the target variable x_t is itself a function of an endogenous variable, p_t. Solving the model therefore involves solving a fixed-point problem in the law of motion for this variable. To that end, write the equilibrium law of motion for the aggregate price level as

$$p_t = \sum_{s=-\infty}^{\infty} \varphi_s \varepsilon_{t-s}. \tag{33}$$

Note that this conjecture allows the price level to depend on future structural disturbances. With the no foresight constraint, it will be a result that $\varphi_s = 0$ for all $s < 0$. However, because I am also interested in characterizing the solution when the no foresight constraint is removed, it is necessary to allow for the more general case here. With this conjecture, the dynamics of the target variable are given by

$$x_t = \sum_{s=-\infty}^{\infty} ((1-\xi)\varphi_s + \xi\delta_s)\varepsilon_{t-s} \equiv \sum_{s=-\infty}^{\infty} a_s \varepsilon_{t-s}, \tag{34}$$

with the convention that $\delta_s = 0$ for all $s < 0$. Given this law of motion for the target, the solution to the supplier's tracking problem can be written as

$$p_{it} = \sum_{s=-\infty}^{\infty} b_s \varepsilon_{t-s} + \zeta_{it}, \tag{35}$$

where ζ_{it} is an idiosyncratic noise term, which is independent of the target process. Integrating this expression across suppliers, assuming that the idiosyncratic noise is cross-sectionally uncorrelated,

$$p_t = \sum_{s=-\infty}^{\infty} b_s \varepsilon_{t-s}. \tag{36}$$

In order for this to be an equilibrium, it must be that the perceived law of motion in (33) coincides with the actual law of motion implied by (36), i.e. $\varphi_s = b_s$ for all s. In the frequency domain, this means that the function $\varphi(\lambda)$ solves the nonlinear fixed-point equation

$$\boxed{\varphi(\lambda) = b(\lambda)} \tag{37}$$

where $b(\lambda)$ is viewed as an implicit function of $\varphi(\lambda)$.

This description of the equilibrium suggests a simple iterative algorithm for computing the solution numerically in the frequency domain.[16]

(1) Given $\varphi^{(k)}(\lambda)$, compute the target frequency response function $a^{(k)}(\lambda) = (1 - \xi)\varphi^{(k)}(\lambda) + \xi\delta(\lambda)$ as in (34).

(2) Given $a^{(k)}(\lambda)$, solve the supplier's tracking problem, either with the no foresight constraint as in Section (4) or without the no foresight constraint as in Section (5), to obtain the optimal value of $b^{(k)}(\lambda)$ in (35), and set $\varphi^{(k+1)}(\lambda) = b^{(k)}(\lambda)$.

(3) Repeat steps (1) and (2) until $||\varphi^{(k+1)}(\lambda) - \varphi^{(k)}(\lambda)||$ is acceptably low.

Figure (3) plots the impulse responses of the price level p_t and real output y_t to a nominal disturbance u_t. Nominal expenditure is assumed to follow the AR(2) process analyzed in Section (5). The top row shows the responses in the case that prices are neither strategic complements nor strategic substitutes ($\xi = 1$). In this case, the target variable is exogenous and suppliers only need to track nominal expenditure. As a result, the upper left panel exactly corresponds to the responses in Figure (2). With foresight, suppliers optimally anticipate the nominal disturbance and choose to start increasing their prices before it takes place (asterisks, upper left panel). Since nominal expenditure itself has not changed during the period of anticipation, this generates a *fall in real output* (asterisks, upper right panel). Once the disturbance occurs, real output increases because prices have not increased by as much as nominal expenditure, due to finite processing capacity. The magnitude of the peak response is attenuated somewhat with foresight, since in that case suppliers are permitted to allocate their attention optimally both to information about the past and the future.

What is interesting about these results is that the fall in output before the nominal disturbance occurs is not the result of any expected deterioration in real conditions. Suppose that suppliers become aware that the monetary authority is planning to stimulate nominal expenditure in the future, perhaps by paying attention to speeches given by central bankers, and as a result real activity falls today. Is it correct to conclude from this sequence of events that suppliers must see central bank plans for future stimulus as a "signal" that the central bank expects real conditions to

[16] A Matlab script that implements this algorithm with the no foresight constraint is included in Appendix (C) for illustration.

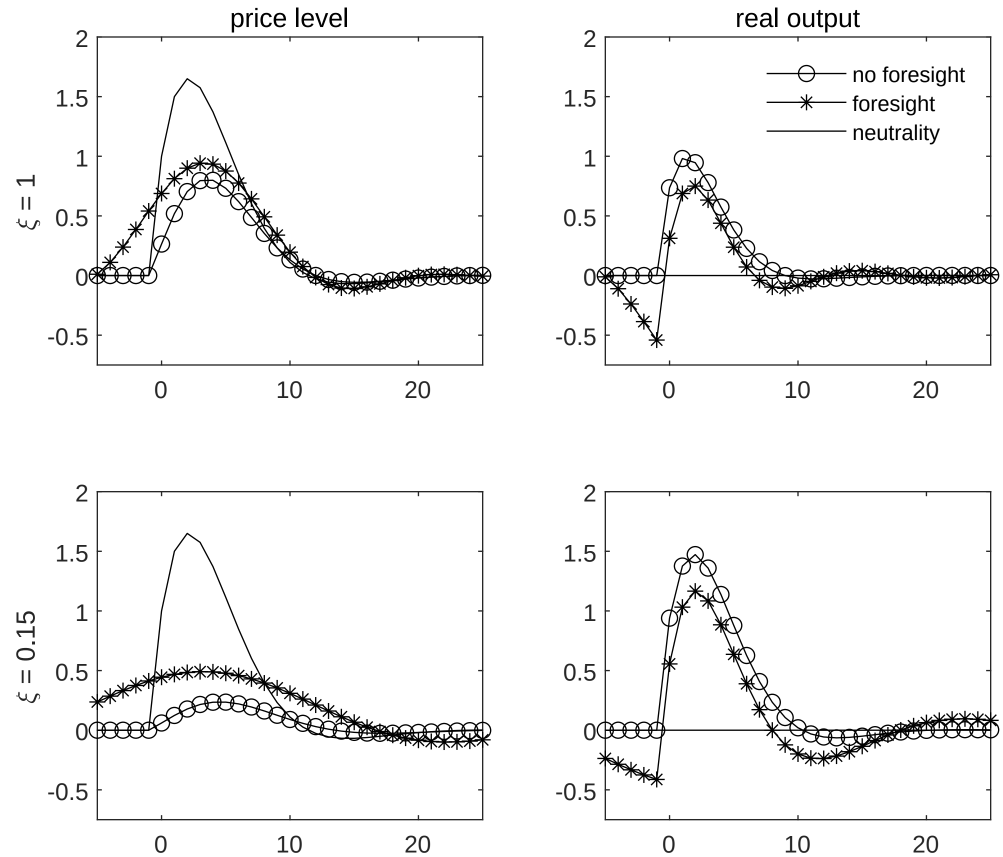

Figure 3: A model of rationally inattentive pricing with and without foresight. This figure plots impulse responses with respect to the disturbance ε_t. Nominal expenditure is an AR(2) of the form $q_t = 1.5q_{t-1} - 0.6q_{t-2} + \varepsilon_t$, $\kappa = 0.1$, and $\xi \in \{1, 0.15\}$.

worsen?[17] The model presented here suggests a different possible interpretation. Real activity falls today not because suppliers expect a *contractionary* real disturbance in the future, but because they expect an *expansionary* nominal disturbance, and so optimally begin raising prices in advance.

The bottom row in Figure (3) shows the same responses in the case when prices are strategic complements. The value of the strategic complementarity parameter is set to $\xi = 0.15$, as in Woodford (2003). The change has the effect of muting the overall response of the price level to nominal disturbances around the time it occurs.

[17] As in discussions of the "signaling" or "information" effects of monetary policy; see, e.g. Melosi (2016) and Nakamura and Steinsson (2018).

However, it *increases* the magnitude of the anticipation effects around one year in advance (asterisks, bottom left panel). The reason for this is that with greater strategic complementarity, the individual supplier's decision depends more on the endogenous aggregate price level. With rational inattention, suppliers pay attention only to the most important frequencies, which in this example are the lower frequencies. As a result, the price of an individual supplier has a relatively larger low-frequency component than the target variable. However, in equilibrium, the target variable depends on these prices, which means that the target variable endogenously places even *greater* weight on low-frequency movements. This self-reinforcing mechanism continues until an equilibrium is reached. In terms of real output, the overall more muted response of prices leads to overall larger real effects, while the larger anticipation effects around a year before the disturbance occurs leads to a larger fall in real output at that time (asterisks, lower right panel).

8 Conclusion

This paper has characterized the solution to the canonical dynamic rational inattention problem by formulating the problem in the frequency domain. In doing so, it has been important to recognize that the standard formulation of the problem imposes not one but two information constraints. The more easily overlooked of these is the no foresight constraint, which prevents the agent from paying attention to any information about future structural disturbances. Properly articulating this constraint in the frequency domain is the key to making the proposed solution procedure operational.

The paper has also analyzed the implications of removing the no foresight constraint, and has shown that it is generally sub-optimal for an agent to pay attention only to the current state. Instead, the agent does what macroeconomists often do when analyzing aggregate time series: apply a band-pass filter and only pay attention to the most relevant frequencies. Doing this involves processing information not just about the current state, but additional information that is correlated with future disturbances as well. In the context of a Phelps-Lucas model of the real effects of nominal disturbances, allowing for foresight suggests that anticipated *expansionary* nominal disturbances can have temporarily *contractionary* effects on real output. This is because suppliers optimally pay attention to information about future nominal disturbances, and begin raising prices in advance.

References

Brockwell, P. J. and R. A. Davis (2006). *Time Series: Theory and Methods.* New York, NY: Springer Science.

Chahrour, R. and K. Jurado (2020). Optimal Foresight. Working paper.

Conway, J. B. (1973). *Functions of One Complex Variable.* New York, NY: Springer-Verlag.

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing).* Hoboken, NJ: Wiley-Interscience.

Futia, C. A. (1981). Rational Expectations in Stationary Linear Models. *Econometrica 49*(1), 171–192.

Hansen, L. P. and T. J. Sargent (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control 2*, 7 – 46.

Huo, Z. and N. Takayama (2018). Rational Expectations Models with Higher Order Beliefs. Working paper.

Kasa, K. (2000). Forecasting the forecasts of others in the frequency domain. *Review of Economic Dynamics 3*(4), 726 – 756.

Kasa, K., T. B. Walker, and C. H. Whiteman (2014). Heterogeneous Beliefs and Tests of Present Value Models. *The Review of Economic Studies 81*(3), 1137–1163.

Kolmogorov, A. N. (1941). Stationary Sequences in Hilbert Space. *Byull. Moskov. Gos. Univ. Mat. 2*(6), 1–40.

Leeper, E. M., T. B. Walker, and S. S. Yang (2013). Fiscal foresight and information flows. *Econometrica 81*(3), 1115–1145.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory 4*(2), 103–124.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2018). Dynamic Rational Inattention: Analytical Results. *Journal of Economic Theory 176*, 650–692.

Maćkowiak, B. and M. Wiederholt (2009). Optimal Sticky Prices under Rational Inattention. *American Economic Review 99*(3), 769–803.

McDonald, R. A. and P. M. Schultheiss (1964). Information rates of gaussian signals under criteria constraining the error spectrum. *Proceedings of the IEEE 52*(4), 415–416.

Melosi, L. (2016). Signalling Effects of Monetary Policy. *The Review of Economic Studies 84*(2), 853–884.

Miao, J., J. Wu, and E. Young (2020a). Macro-Financial Volatility under Dispersed Information. Working paper.

Miao, J., J. Wu, and E. Young (2020b). Multivariate Rational Inattention. Working paper.

Nakamura, E. and J. Steinsson (2018). High-Frequency Identification of Monetary Non-Neutrality: The Information Effect. *The Quarterly Journal of Economics 133*(3), 1283–1330.

Phelps, E. S. (1969). The new microeconomics in inflation and employment theory. *The American Economic Review 59*(2), 147–160.

Pinsker, M. S. (1954). Amount of information about a gaussian random process contained in a second process which is stationarily correlated with it. *Dokl. Akad. Nauk SSSR 99*(2), 213–216.

Rozanov, Y. A. (1967). *Stationary Random Processes*. San Francisco, CA: Holden-Day.

Sims, C. A. (1998). Stickiness. *Carnegie-Rochester Conference Series on Public Policy 49*, 317–356.

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

Taub, B. (1989). Aggregate fluctuations as an information transmission mechanism. *Journal of Economic Dynamics and Control 13*(1), 113–150.

Whiteman, C. H. (1983). *Linear Rational Expectations Models: A User's Guide.* Minneapolis, MN: University of Minnesota Press.

Woodford, M. (2003). Imperfect common knowledge and the effects of monetary policy. In P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford (Eds.), *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, pp. 25–58. Princeton, NJ: Princeton University Press.

Appendix

A Additional closed-form results

This section presents closed-form solutions to the canonical tracking problem in the case that the target process is an MA(1), AR(1), or AR(2) process. The first two can be derived as corollaries of Proposition (1) by taking appropriate limits; the third can be found by following the same steps described in Section (4). The result in Proposition (2) was first presented in Maćkowiak and Wiederholt (2009), and the result in Corollary (2) was first presented in Maćkowiak, Matějka, and Wiederholt (2018). In each case, I also present, as a corollary, a subjective signal that delivers the same optimal action.

Proposition 2. *When the target is an AR(1) process, so*

$$a(\lambda) = \sigma \frac{1}{(1-\rho e^{-i\lambda})}$$

with $\sigma > 0$ *and* $|\rho| < 1$*, the unique solution to Problem* (1) *is*

$$b(\lambda) = \sigma(1-e^{-2\kappa}) \frac{1}{(1-\rho e^{-i\lambda})(1-\rho e^{-2\kappa} e^{-i\lambda})}$$

$$c(\lambda) = \sigma \sqrt{e^{-2\kappa} \frac{(1-e^{-2\kappa})}{(1-\rho^2 e^{-2\kappa})}} \frac{1}{(1-\rho e^{-2\kappa} e^{-i\lambda})}.$$

Corollary 2. *The optimal action in Proposition* (2) *is equal to* $y_t = E[x_t | s^t]$ *when the subjective signal* s_t *is given by*

$$s_t = x_t + \varsigma_v v_t,$$

where $\{v_t\}$ is a sequence of independent normal random variables with zero mean and unit variance that is independent of $\{\varepsilon_t\}$, and

$$\varsigma_v = \sigma\sqrt{\frac{e^{-2\kappa}}{(1-e^{-2\kappa})(1-\rho^2 e^{-2\kappa})}}.$$

Proof. Define the signal

$$s_t \equiv \frac{(1-\rho e^{-2\kappa}L)}{(1-e^{-2\kappa})} y_t.$$

Since $|\rho e^{-2\kappa}| < 1$, this transformation does not change the agent's optimal estimate of the target. □

Proposition 3. *When the target is an MA(1) process, so*

$$a(\lambda) = \sigma(1-\alpha e^{-i\lambda})$$

with $\sigma > 0$ and $|\alpha| < 1$, the unique solution to Problem (1) *is*

$$b(\lambda) = \sigma_b \frac{(1-\phi e^{-i\lambda})(1-\phi e^{-i\lambda})}{(1-r_1 e^{-i\lambda})}$$

$$c(\lambda) = \sigma_c \frac{(1-\phi e^{-i\lambda})}{(1-r_1 e^{-i\lambda})},$$

where

$$\sigma_b = \sigma\frac{\alpha r_1}{\phi^2} \quad \sigma_c = \sigma\sqrt{e^{-2\kappa}\frac{\alpha^2 r_1}{\phi^3}} \quad r_1 = \frac{\phi - \alpha(1-\phi^2)}{\alpha\phi},$$

and ϕ is the root of the cubic equation

$$\mathcal{P}(\phi) = \alpha\phi^3 + (1-\alpha^2(1+e^{-2\kappa}))\phi^2 - 3\alpha\phi + 2\alpha^2$$

which satisfies $|\phi| < 1$ and ensures that $|r_1| < 1$.

Corollary 3. *The optimal action in Proposition* (3) *is equal to $y_t = E[x_t|s^t]$ when the subjective signal s_t is given by*

$$s_t = \varepsilon_t - \phi\varepsilon_{t-1} + \varsigma_v v_t,$$

where

$$\varsigma_v = \frac{\sigma\sigma_c}{\sigma_b},$$

and ϕ, σ_b, and σ_c are defined as in Proposition (3).

Proof. Define the signal

$$s_t \equiv \frac{\sigma}{\sigma_b}\frac{(1-r_1L)}{(1-\phi L)}y_t.$$

Since $|r_1|< 1$ and $|\phi|< 1$, this transformation does not change the agent's optimal estimate of the target. □

Proposition 4. *When the target is an AR(2) process, so*

$$a(\lambda) = \sigma\frac{1}{(1-\rho_1L)(1-\rho_2L)},$$

with $\sigma > 0$, $|\rho_1|< 1$ *and* $|\rho_2|< 1$, *the unique solution to Problem* (1) *is*

$$b(\lambda) = \sigma_b\frac{(1-\phi e^{-i\lambda})(1-\phi e^{-i\lambda})}{(1-\rho_1e^{-i\lambda})(1-\rho_2e^{-i\lambda})(1-r_1e^{-i\lambda})(1-r_2e^{-i\lambda})}$$

$$c(\lambda) = \sigma_c\frac{(1-\phi e^{-i\lambda})}{(1-r_1e^{-i\lambda})(1-r_2e^{-i\lambda})},$$

where

$$\sigma_b = \frac{\sigma\omega\rho_1\rho_2}{\phi^2}(r_1+r_2+r_1^2r_2+r_2^2r_1-e^{-2\kappa}(\rho_1+\rho_2+\rho_1^2\rho_2+\rho_2^2\rho_1))$$

$$\sigma_c = \sqrt{e^{-2\kappa}\sigma_b\frac{\sigma\omega\rho_1\rho_2}{\phi}}$$

$$\omega = \frac{\phi}{\rho_1\rho_2(1+\phi(r_1+r_2)-\rho_1^2\rho_2^2e^{-2\kappa})-\phi(\rho_1+\rho_2)}$$

$$r_2 = \frac{\rho_1\rho_2e^{-2\kappa}}{r_1}$$

r_1 *is the root of the quadratic polynomial*

$$\mathcal{P}(r_1) = r_1^2 + \frac{(\rho_1+\rho_2)(\phi^2-e^{-2\kappa}\rho_1^2\rho_2^2)-\phi(1+\rho_1\rho_2)(1-e^{-2\kappa}\rho_1^2\rho_2^2)}{(1-\phi^2)\rho_1^2\rho_2^2}r_1 + \frac{(1-\phi)}{(1+\phi)\rho_1\rho_2}$$

that satisfies $|r_1|< 1$, *and* ϕ *is the root of the quartic polynomial*

$$\mathcal{P}(\phi) = (\rho_1+\rho_2)\phi^4 - (1+2\rho_1\rho_2+e^{-2\kappa}\rho_1^2\rho_2^2)\phi^3$$

$$+\rho_1\rho_2(1+e^{-2\kappa}(2\rho_1\rho_2+\rho_1^2\rho_2^2))\phi - e^{-2\kappa}\rho_1^2\rho_2^2(\rho_1+\rho_2)$$

which satisfies $|\phi|< 1$ *and ensures that* $|r_1|< 1$ *and* $|r_2|< 1$.

Corollary 4. *The optimal action in Proposition* (4) *is equal to* $y_t = E[x_t|s^t]$ *when the subjective signal* s_t *is given by*

$$s_t = x_t - \phi x_{t-1} + \varsigma_v v_t,$$

where $\{v_t\}$ *is a sequence of independent normal random variables with zero mean and unit variance that is independent of* $\{\varepsilon_t\}$*,*

$$\varsigma_v = \frac{\sigma\sigma_c}{\sigma_b},$$

and ϕ*,* σ_b*,* σ_c *are defined as in Proposition* (4).

Proof. Define the signal

$$s_t \equiv \frac{\sigma}{\sigma_b}\frac{(1-r_1L)(1-r_2L)}{(1-\phi L)}y_t.$$

Since $|r_1|< 1$, $|r_2|< 1$, and $|\phi|< 1$, this transformation does not change the agent's optimal estimate of the target. □

B Matlab functions

```
function [b,c] = track(a,kappa)
% ------------------------------------------------------------------------
% Solve the tracking problem of Sims (2003) in the frequency domain
%   a     = a(lambda), one row for each lambda in 2*pi*(0:(n-1))/n
%   kappa = information rate parameter
%   b     = b(lambda)
%   c     = c(lambda)
% ------------------------------------------------------------------------

% Initial guess
psi = zeros(size(a));

% Iterate to convergence
err = 1;
while err > 1e-10

    % Transformations
    d = a - psi;
```

```
    f = diag(d*d');

    % Multiplier
    [~,h] = wold(f);
    theta = real(exp(-2*kappa)*h(1).^2);

    % Singular part
    [new,~] = singular(a - theta.*d./f);

    % Update
    err = norm(new - psi);
    psi = new;
end

% Optimal weights
b     = d.*(1 - theta./f);
c2    = theta*(1 - theta./f);
[c,~] = wold(c2);

end

function [hf,ht] = wold(f)
% -------------------------------------------------------------------------
% Wold representation of f
% -------------------------------------------------------------------------
n           = length(f);
ma          = ifftshift(ifft(log(f)));
ma(1:(n/2)) = 0;
ma(n/2+1)   = 1/2*ma(n/2+1);
hf          = exp(fft(fftshift(ma)));
ht          = real(ifft(hf));
end

function [sf,st] = singular(f)
% -------------------------------------------------------------------------
% Singular part of f
% -------------------------------------------------------------------------
n                 = length(f);
st                = ifftshift(ifft(f));
st((n/2)+1:end,:) = 0;
sf                = fft(fftshift(st));
end
```

C Equilibrium model

```
% ------------------------------------------------------------------------
% Solve the model of Woodford (2003) with rational inattention (no foresight)
% ------------------------------------------------------------------------

% Parameters
xi    = 0.15;
kappa = 0.1;
sigma = 1;

% Frequency grid
n    = 300;
grid = (2*pi*(0:(n-1))/n).';

% Law of motion for nominal expenditure
df = sigma./(1 - 1.5.*exp(-1i.*grid) + 0.6.*exp(-1i.*grid.*2));

% Initial guess
plm = df;

% Iteration loop
err = 1;
while err > 1e-10

    % Law of motion for target
    af = (1-xi).*plm + xi.*df;

    % Solve tracking problem
    [alm,~] = track(af,kappa);

    % Update
    r   = norm(alm-plm);
    err = r*(err > r);
    plm = alm;
end

% Time domain
dt   = real(ifft(df));
phit = real(ifft(alm));
gt   = dt - phit;
```

```
% Plot results
hor = 10;
ind = [n-hor+1:n,1:hor+1].';
subplot(1,2,1);
plot(-hor:hor,[dt(ind),phit(ind)]);
legend('full information','rational inattention');
subplot(1,2,2);
plot(-hor:hor,[0.*ind,gt(ind)]);
```