

**Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.**

**Mechanics of linear quadratic Gaussian rational inattention
tracking problems**

Chad Fulton

2017-109

Please cite this paper as:

Fulton, Chad (2017). "Mechanics of linear quadratic Gaussian rational inattention tracking problems," Finance and Economics Discussion Series 2017-109. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2017.109>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

Mechanics of linear quadratic Gaussian rational inattention tracking problems

Chad Fulton *

Abstract

This paper presents a general framework for constructing and solving the multi-variate static linear quadratic Gaussian (LQG) rational inattention tracking problem. We interpret the nature of the solution and the implied action of the agent, and we construct representations that formalize how the agent processes data. We apply this infrastructure to the rational inattention price-setting problem, confirming the result that a conditional response to economics shocks is possible, but casting doubt on a common assumption made in the literature. We show that multiple equilibria and a social cost of increased attention can arise in these models. We consider the extension to the dynamic problem and provide an approximate solution method that achieves low approximation error for many applications found in the LQG rational inattention literature.

JEL Classification: D81, D83, E31

Keywords: Rational inattention, information acquisition, signal extraction

*chad.t.fulton@frb.gov. The views expressed in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System, or anyone else in the Federal Reserve System.

1 Introduction

Models incorporating rational inattention, in which agents faced with limited information processing capacity optimally allocate their attention across various economic shocks, can accommodate a wide range of behavior that deviates from the rational expectations baseline. They have been used to explain the sluggish responses to shocks observed for many macroeconomic time series, they imply behavior similar to standard logit models when applied to discrete choice problems, and they can result in discrete behavior by agents even when the underlying economic shocks that influence the agent are continuously distributed.¹ Despite their appeal, the technical challenges are such that explicit solutions have not been found for most problems. In this paper we derive an explicit solution for and give a comprehensive account of a foundational model: a multivariate static problem in which all shocks are Gaussian and the objective function of the agent is quadratic. These so-called static linear quadratic Gaussian problems are the most tractable class of rational inattention problems, but, even so, a full solution has been previously unknown. In addition, the model considered in this paper serves as an important special case of more complex dynamic models, and has been used to establish baseline results and provide intuition in many applications. Along these lines, much of the analysis and interpretation that we will develop in this paper will extend to the dynamic case.

Our first step is to lay a firm groundwork, since a variety of ways even to formulate the problem have arisen. We begin by writing down our preferred formulation, following Sims (2003) and Sims (2010), and explaining its relation to the classic signal extraction problem. In short, an agent must choose the optimal posterior covariance matrix for a vector of shocks given a loss function and subject to a constraint on how much uncertainty can be

¹ For sluggishness in macroeconomic series, see the price-setting model of Maćkowiak and Wiederholt (2009), the permanent income model of Sims (2003), or the numerous references contained in Sims (2010). For rational inattention as applied to discrete choice models, see Matějka and McKay (2015) or Steiner et al. (2017). For discrete actions in continuous settings, see Jung et al. (2015).

reduced relative to their prior. Our formulation can include an arbitrary number of shocks, potentially correlated, and can incorporate the information constraint in terms of a fixed quantity of information processing capacity or a fixed marginal cost associated with processing additional information. Throughout the paper, we clarify the relationship between this and alternative statements of the problem. In particular, we will take a closer look at the often-used formulation in which agents choose the noise variance of "signals" received by the agent, which, we will argue, can encourage misleading comparisons with the signal extraction problem.

After establishing the problem, we immediately present the general solution in two theorems. We show that the crucial element in constructing and understanding the solution lies in recognizing that the agents are not just choosing *how much* posterior uncertainty about shocks is optimal, they are also choosing the *form* of the posterior uncertainty. An illuminating example of this is given in Sims (2010): if a rationally inattentive agent wishes to track the sum of n random variables, then they will process information so as to make their posterior uncertainty about those random variables negatively correlated, even if the variables themselves are independent. We show how to construct what we call the canonical synthetic shocks (or just "canonical shocks"), specific linear combinations of the original, or "fundamental", shocks that capture the optimal form of posterior uncertainty chosen by a rationally inattentive agent. Understanding these canonical shocks is the key to solving the problem and understanding the implications of the solution, and their careful definition is one contribution of this paper.

While the fundamental shocks that exist as part of the formulation of the economic model may appear natural to the modeler, we argue that it is instead the canonical shocks that are natural for the agent within the model. We show that the canonical shocks represent the separate and distinct elements of uncertainty that actually matter to the agent. In fact, the solution to the problem is exactly constructed by transforming the problem into the

“canonical space”, and we provide a straightforward intuition of this by geometrizing the problem in terms of ellipsoids representing uncertainty. Then, given the solution, the agent’s action - their posterior estimate of each individual component of the canonical shock - turns out to be a simple Bayesian update, a weighted average of the agent’s prior for that component and their understanding of the incoming data. Moreover, using the canonical shocks we can construct a representation of the incoming data as understood by the agent that gives an intuitive sense of how the agent produced their posterior through information processing.

While the form these representations take is consistent with the concept of an “observation” or “signal” as in a signal extraction problem, a crucial point is that any given representation is simply a device that assists us in characterizing the agent’s decision. Representations are not unique, and we show how to construct the class of representations that would be valid for a given problem. We characterize the useful subset of these representations as “feasible”, and show that all feasible representations are only transformations of the representation constructed in terms of the canonical shocks. Importantly, we show that whereas this canonical representation always exists, in most cases there does not exist a feasible representation in terms of the “fundamental” shocks. This underscores that while the fundamental shocks may be of interest to the modeler, they are not the objects of interest to the agent. Finally, we present the “representation form” of the problem, and show that it is less useful than the canonical form. We also describe the related form of the problem, mentioned above, in which agents choose the noise variance of “signals”, and we show how issues can arise through the incautious application of this last formulation.

As an application, we consider the rational inattention price-setting problem of Maćkowiak and Wiederholt (2009). We start by showing how to cast the static case of their problem, including their “independence assumption”, in the terms of this paper and then solve it along with three new formulations that we introduce. In contrast to the involved derivations

that previous papers have often had to rely on, the exact solution to the general problem that we derive in this paper yields the results immediately. In comparing these solutions, we find that the key result of Maćkowiak and Wiederholt (2009) - a conditional response to different types of fundamental shocks - survives dropping the independence assumption, and we also present new results, including the introduction of multiple equilibria and the possibility that additional information processing capacity actually increases social costs.²

The more general dynamic RI-LQG tracking problem remains unsolved by the methods of this paper. Despite this, many key concepts - including the canonical synthetic shocks, the agent as a Bayesian updater, and our treatment of representations - do apply in the dynamic problem. We present this problem and show that the sequential application of the static solution combined with iteration of the dynamic transition equation approximates the full dynamic solution, and that the approximation error will be low as long as the parameter capturing the marginal cost of attention is close to zero. Since this condition holds in most existing applications of dynamic RI-LQG tracking problems in the literature, we conclude that the static approximation is a useful tool, particularly since no analytic solution so far exists and numerical solutions can be difficult to obtain.

This paper is most closely related to Sims (2003) and Sims (2010), to which we owe our basic formulation for the class of RI-LQG tracking problems. Additionally, in these two papers can be found the seeds of many of the concepts we make explicit and fully develop here for the static case. This paper is also related and complementary to Matejka et al. (2017), as both of our papers provide explicit solutions for special cases of the dynamic RI-LQG tracking problem. Whereas we consider the static version of the problem with multiple targets and arbitrary correlations and present an approximate solution in the dynamic case, they consider the dynamic problem with a single ARMA(p,q) target.

² The latter result recalls Morris and Shin (2002), except that here the incompleteness of information is endogenous.

2 Preliminaries

Here we introduce a few mathematical results related to information theory and generalized eigenvalue problems; these will be used throughout the rest of the paper.

2.1 Information theory

It is most transparent to introduce the concepts of information theory for the case of discrete random variables, and so in what follows we will let X and Y denote random variables with probability mass functions P_X and P_Y . For the results in the paper, we will be making use of an extension to the continuous case known as differential entropy. Although this extension is broadly consistent with discrete case, there are subtleties that must be accounted for; we point out a few examples of this below.

2.1.1 Entropy

The basic quantity in information theory is entropy, a measure of the uncertainty associated with a random variable. Entropy is defined as:³

$$h(X) = -E[\log(P(X))]$$

Entropy is typically measured in “bits”, where a bit is the quantity of uncertainty associated with a Bernoulli trial with probability of success $p = 0.5$. Thus a bit is a quantification of the uncertainty resolved by the realization of a single coin flip. We can also define joint entropy $h(X, Y) = -E[\log(P(X, Y))]$ and conditional entropy $h(X | Y) = -E[\log(P(X | Y))]$. Conditional entropy can be thought of as the uncertainty about X that remains after

³ Often entropy of a discrete random variable is denoted $H(X)$ and entropy of a continuous random variable, known as differential entropy, is denoted $h(X)$. To simplify notation, we will use $h(\cdot)$ for both cases.

observing Y . The “chain rule” of entropy states $h(X, Y) = h(X) + h(Y|X)$; in words, the uncertainty about X and Y together is the uncertainty about X plus the uncertainty about Y that remains after observing X .

If $X \perp Y$, then X does not resolve any uncertainty about Y and so $h(Y|X) = h(Y)$. Then by the chain rule $h(X, Y) = h(X) + h(Y)$, so that the uncertainty about X and Y together is just the sum of the uncertainty about X and Y separately. In the degenerate case that $X = Y$, observing X fully resolves the uncertainty about Y . If X and Y are discrete, then $h(Y|X) = 0$ and so $h(X, Y) = h(X) + h(Y|X) = h(X)$. However, in the continuous case, $h(Y|X) = -\infty$; this is an example of one subtlety that arises in information theory when moving from discrete to continuous random variables.

2.1.2 Mutual Information

Mutual information is a measure of the information about one random variable contained in another. Formally:

$$I(X; Y) = h(X) - h(X|Y)$$

This can be understood as the quantity of uncertainty about X resolved after the observation of Y . For example, if $X \perp Y$, then $h(X|Y) = h(X)$ and $I(X; Y) = 0$. This is true in both the discrete and continuous cases.

At the other extreme, if $X = Y$ then in the discrete case $h(X|Y) = 0$ and so $I(X; Y) = h(X)$. Thus observing Y resolves all uncertainty about X , and since the “quantity” of uncertainty about X is given by the entropy $h(X)$, this is also the quantity of mutual information. However, in the continuous case, $h(X|Y) = -\infty$ so that $I(X; Y) = \infty$. In fact, this too is an intuitive result reflecting the fact that a continuous random variable can take on an uncountably infinite number of values. By mapping each possible value to a

“message” of arbitrary content, it is clear that we can transmit as much information as we like through the realization of a continuous random variable.

2.1.3 Information theoretic results

Here we state some well-known properties of entropy and mutual information; see for example Cover and Thomas (2006) for details.

Property 1: Entropy is invariant under translation. Let W, X be arbitrary random vectors and let c be a W -measurable function. Then:

$$h(X + c(W) | W) = h(X | W)$$

Corollary: Mutual information is invariant under translation by a constant.

Property 2: Conditioning weakly reduces entropy. Let W, X be arbitrary random vectors. Then:

$$h(X) \geq h(X|W)$$

Property 3: Mutual information is invariant under invertible transformations. Let W, X, Y be arbitrary random vectors and let f, g be bijective functions. Then:

$$I(X; Y | W) = I(f(X), g(Y) | W)$$

Corollary: As a consequence of properties 1 and 3, if F, G are nonsingular conformable matrices and c, d are constants, then:

$$I(X; Y | W) = I(FX + c, GY + d | W)$$

Property 4: Let X be a random vector, and consider all possible distributions for X such that $\text{Var}(X) = P$ is fixed. Then the differential entropy is maximized when X is jointly Gaussian.

Property 5: Let X, Y be jointly Gaussian random vectors of dimension n , possibly conditional on some information \mathcal{I}_- , and let $\text{Var}(X | \mathcal{I}_-) = P_-$ and $\text{Var}(X | \mathcal{I}_-, Y) = P_+$. Then:

$$\begin{aligned} h(X | \mathcal{I}_-) &= \frac{1}{2} \log_b |2\pi e P_-| \\ h(X | \mathcal{I}_-, Y) &= \frac{1}{2} \log_b |2\pi e P_+| \\ I(X, Y | \mathcal{I}_-) &= h(X | \mathcal{I}_-) - h(X | \mathcal{I}_-, Y) = \frac{1}{2} (\log_b |P_-| - \log_b |P_+|) \end{aligned}$$

We have not specified the base of the logarithm in Property 5, since different bases simply correspond to different measures of mutual information; for example, if the base is 2 then mutual information is measured in bits, whereas if the base is e then mutual information is measured in nats.

2.2 Generalized eigenvalue problems

The generalized eigenvalue problem for two matrices A, B is to find scalars λ_i and vectors r_i such that the following equation holds:⁴

$$(A - \lambda_i B)r_i = 0, \quad i = 1, \dots, n$$

In what follows, we will be interested in the specialization in which A, B are symmetric positive semidefinite matrices. In fact, we will usually consider cases in which B is positive definite, and then since B is nonsingular it is easy to see that left multiplication by B^{-1}

⁴ In this section, we use the notation λ and Λ differently than we will in the rest of the paper.

yields a standard eigenvalue problem $(B^{-1}A - \lambda_i I)r_i = 0$. However it turns out that applying this transformation often obscures the form of the solution since $B^{-1}A$ may not be positive semidefinite and is generally not even symmetric. The matrix $A - \lambda B$ is often referred to as a matrix pencil and denoted by the pair (A, B) .

For positive semidefinite matrices A, B , the generalized eigenvalue problem can be solved via simultaneous diagonalization of A, B by congruence. We state this result as a lemma.

Lemma 1:⁵ If A and B are real symmetric positive semidefinite matrices of order n and $\text{rk}(B) = r$, then:

- a. There exists a nonsingular matrix S such that $B = S'(I_r \oplus 0_{n-r})S$ and $A = S'\Lambda S$, in which Λ is nonnegative diagonal and $\text{rk}(A) = \text{rk}(\Lambda)$.
- b. Defining $R \equiv S^{-1} = \begin{bmatrix} r_1 & \dots & r_n \end{bmatrix}$ and $\Lambda = \text{diag}(\{\lambda_i\}_{i=1}^n)$, the pairs (λ_i, r_i) solve the generalized eigenvalue problem associated with the matrix pencil (A, B) . The scalars λ_i are called generalized eigenvalues and the vectors r_i are called generalized right eigenvectors.
- c. If B is positive definite, there is a unique factorization $M'M = B$, where M is nonsingular. Defining $L = M^{-1}$, we can compute the eigendecomposition $Q\Lambda Q' = L'AL$. Then this matrix Λ along with $S = Q'M$ satisfy (a) and (b).

An important element of generalized eigenvalue problems is that the matrix containing generalized eigenvectors is not orthogonal with respect to the usual inner product, i.e. in general $R'R \neq I$. However, if B is positive definite, we can define a valid inner product induced by B as $\langle x, y \rangle_B$. That the generalized eigenvectors are B -orthogonal, i.e. that $R'BR = I$, follows directly from part (a) of the lemma.

Although the generalized eigenvalue problem will be crucial in several ways in the solution to the rational inattention problem considered in this paper, one important use can be

⁵ Proofs of all results in this paper are given in Appendix A.

immediately shown to simplify the mutual information of Gaussian random vectors.

Property 6: Let $X, Y, \mathcal{I}_-, P_-, P_+$ all be defined as in Property 5. Then we can write:

$$I(X, Y | \mathcal{I}_-) = \frac{1}{2} \sum_{i=1}^n \log_b \frac{1}{n_i}$$

where n_i denote the generalized eigenvalues of the matrix pencil (P_+, P_-) .⁶ Importantly, this property applies to both static and dynamic rational inattention problems.

3 Problem

Rational inattention problems fall into the larger class of problems in which agents must make decisions under imperfect information. In classical imperfect information problems, the information structure of the economy is often exogenously imposed. The rational inattention approach, introduced by Sims (2003), is one way to endogenize information imperfections as the rational behavior of agents that face constraints on the extent to which they can process and translate information into actions, even in the case that the information itself is freely available.

3.1 Exogenous information imperfections

We begin by briefly describing the classical signal extraction problem, one of the most common models of imperfect information, in which the characteristics of the signal and noise are exogenous. This is valuable because it will turn out that the rational inattention problem can be cast in the form of specific signal extraction problems. However, as we will show below, the signal extraction formulation of the rational inattention problem is not unique. A more fundamental representation of the rational inattention problem is in terms

⁶ We use the notation n_i instead of λ_i in order to make a notational connection with the following sections.

of a generalization of signal extraction problems known as tracking problems, which we also briefly introduce. This will allow us, in the next section, to describe the specific application to rationally inattentive tracking problems, and to present the problem and solution in the further special case known as the linear quadratic Gaussian (LQG) case.

3.1.1 Signal extraction problems

Given an unknown random vector⁷ of interest α and a given observation vector $y = h(\alpha, \varepsilon)$, where ε is an independent random vector representing contaminating noise and h is some measurable function, a signal extraction problem is to select a second function $a(y)$ such that the expected distance between α and $a(y)$ is “small” according to some distance, or loss, function d . The signal extraction problem can be formulated as:

$$\min_{a(y)} \int d(\alpha, a(y)) f(\alpha|y) d\alpha$$

If loss is quadratic in $\alpha - a(y)$, so that the problem is to minimize the (weighted) mean square error, then the solution is well known to be the conditional expectation $a(y) = E[\alpha | y]$. If it is also the case that y and α are jointly Gaussian, then it is similarly well known that the conditional expectation is a linear function, $a(y) = a_0 + Ky$.

The well-known Kalman filter recursively solves a dynamic version of the signal extraction problem in which the loss is quadratic, all variables are jointly Gaussian, and the vector of interest α follows a linear transition law. This case is referred to as a linear quadratic Gaussian (LQG) filtering problem. Because the static signal extraction problem introduced above is a special case of the recursive problem, we will also refer to it as an LQG signal extraction problem.

⁷ We derive all results in terms of random vectors, but everything remains valid for the 1-dimensional random variable case.

3.1.2 Tracking problems

To more clearly formulate the rational inattention problem and its solution below, we distinguish between a signal extraction problem and a “tracking” problem. Here, a tracking problem is a generalization of a signal extraction problem in which an observation vector is not a given fundamental component. Instead, the problem is:

$$\min_{f()} \int d(\alpha, a) f(\alpha, a | \mathcal{I}) da d\alpha$$

such that f is a valid joint density function for (α, a) and is potentially conditional on some given prior information set \mathcal{I} . We refer to α as the “target” or “state” and a as the “action”.⁸ In the case that the loss is quadratic in $\alpha - a$ and the variables are jointly Gaussian, we refer to this as an LQG tracking problem.

If there are no constraints, then the solution is to choose f such that $a = \alpha$ with probability 1. Then f is degenerate and expected losses are zero. To specify an interesting tracking problem, some constraint must be added. For example, the signal extraction problem above is a specialization of the tracking problem in which a constraint is placed on the form of a , so that a must be a measurable function of an exogenous observation y .

For what follows, it is notationally convenient to rewrite the tracking problem as $\min_a E[d(\alpha, a) | \mathcal{I}]$ where it is understood that the expectation is with respect to the joint distribution of (α, a) conditional on the marginal distribution of α and the prior information \mathcal{I} , and that the minimization is either over that joint distribution directly or, equivalently, over the conditional distribution of $a | \alpha, \mathcal{I}$.

⁸ It may be useful to have in mind some sport in which a player must track the position of a ball (the target) in order to place their foot so that it will meet the ball (their action). Their action depends on where they perceive the ball to be, and they wish to make that perception as close as possible to where the ball actually is.

3.2 Endogenizing imperfect information through rational inattention

In rational inattention problems, all information is generally supposed to be freely observable save for a constraint on the information processing capacity of the agent. If the relevant information can be expressed as a random vector α , then we will write the agent's perception of that information after processing as a_+ . Because the agent wishes to make a_+ as close to α as possible given some constraint, this is often naturally formulated in terms of a tracking problem, and so we will refer to the α as the target and a_+ as the action.

The constraint in a rational inattention problem is formalized using the mutual information between target and action, $I(\alpha, a_+)$. As described above, this quantification of "information processed" has various desirable properties and a natural interpretation: it is the quantity of uncertainty about the target resolved by the agent in the process of taking their action. There are two primary ways of formulating this constraint. The first allows agents a fixed processing capacity κ and requires that $I(\alpha, a_+) \leq \kappa$; we will refer to this as the "fixed capacity" or "fixed κ " formulation. The second allows agents to access any amount of information processing capacity at a fixed marginal cost λ^* ; we will refer to this as the "fixed marginal cost" or "fixed λ " formulation. As we will show below, these approaches lead to largely similar statements of the problem and solution, but they have different implications in comparative statics exercises.

3.3 Rational inattention tracking problems

The rational inattention tracking problem is:⁹

$$\min_{a_+} E [d(\alpha, a_+) | \mathcal{I}_-] + \lambda^* I(\alpha, a_+ | \mathcal{I}_-)$$

⁹ See Sims (2010) for more details regarding this formulation of the problem.

where λ^* is interpreted either as a cost parameter or as a Lagrange multiplier for a constraint $I(\alpha, a_+ | \mathcal{I}_-) \leq \kappa$; these interpretations correspond respectively to the fixed marginal cost and fixed capacity constraints introduced earlier, and we will provide an explicit solution for each case. Note that here and in what follows we will denote the prior information set as \mathcal{I}_- and the action as a_+ to emphasize the processing of new information. The function $I(\alpha, a_+ | \mathcal{I}_-)$ is the conditional Shannon mutual information, introduced above.

In general this is a difficult problem to solve. However, if the loss is quadratic and α is Gaussian, then an analytic solution exists. As described in Sims (2003) and Sims (2010), a solution to this problem makes $(\alpha, a_+) | \mathcal{I}_-$ jointly Gaussian and we can write $\alpha = a_+ + \eta$, where $a_+ \perp \eta$. Writing $\alpha | \mathcal{I}_- \sim N(a_-, P_-)$, we can then specify the components:

$$\begin{aligned} a_+ | \mathcal{I}_- &\sim N(a_-, P_- - P_+) \\ \eta | \mathcal{I}_- &\sim N(0, P_+) \\ \alpha | \mathcal{I}_-, a_+ &\sim N(a_+, P_+) \end{aligned}$$

Then it is clear that this optimal action is a conditional expectation: $a_+ = E[\alpha | \mathcal{I}_+]$, where \mathcal{I}_+ denotes the posterior information, with $\mathcal{I}_- \subseteq \mathcal{I}_+$. This of course immediately recalls the solution to the signal extraction problem. A crucial point to note at this stage, however, is that we have not been explicit about the contents of the posterior information set, and we have made no mention of an observation or signal vector. In fact, we will develop the complete formulation and solution to this problem with no mention of such a vector, and the fact that we can do this makes the tracking problem, rather than the signal extraction problem, fundamental. Nonetheless, an analogy with the signal extraction problem can be useful as an aid to interpretation, and so we will make the analogy precise and draw out its strengths and weaknesses as we proceed.

Specification of a_+ as a conditional expectation has not fully solved the problem, but it has reduced the optimization space and it will allow us to present a simpler formulation. First,

we can simplify $E[d(\alpha, a_+) | \mathcal{I}_-] = E[(\alpha - a_+)'W(\alpha - a_+) | \mathcal{I}_-] = \text{tr}(WP_+)$ where W is a positive semidefinite matrix defining the loss function. Second, from Property 5, we have $I(\alpha, a_+ | \mathcal{I}_-) = \frac{1}{2} (\log_b |P_-| - \log_b |P_+|)$.¹⁰ Finally, for notational convenience we write $\lambda = \lambda^*/(2 \ln b)$ to eliminate a constant term from this form of the information constraint, and we will often refer to λ as the marginal cost of attention.

This leads us to what might be termed the canonical formulation of the static rational inattention linear quadratic Gaussian (RI-LQG) tracking problem. This formulation is a static version of the dynamic problems described in Sims (2003) and Sims (2010).

Definition 1: The static RI-LQG tracking problem represented by the tuple (W, a_-, P_-) is:

$$\begin{aligned} \min_{P_+} \text{tr}(WP_+) + \lambda(\ln |P_-| - \ln |P_+|) & \quad (1) \\ \text{s.t. } \alpha | \mathcal{I}_- \sim N(a_-, P_-) & \\ P_+ \geq 0 & \\ P_- - P_+ \geq 0 & \end{aligned}$$

where the notation $P_- - P_+ \geq 0$ indicates that the difference of these matrices must be positive semidefinite. We will generally assume that the target α is an $n \times 1$ vector distributed $N(\bar{\alpha}, \Omega)$ where $\text{rank } \Omega = n$. Finally, we will refer to α and a_+ as the “fundamental” target and action, since we will extensively deal also with transformations of these vectors that we will call “synthetic” targets and actions.

We have thus reduced the problem from optimization over the space of random variables to optimization over the cone of positive semidefinite matrices, and we note that any solution P_+ determines a specific information set \mathcal{I}_+ that will be described in more detail below.

The problem as stated has two “positive semidefiniteness” constraints. The first requires

¹⁰ We have left the base of the logarithm unspecified here; in examples we will generally assume information to be measured in bits.

that P_+ is a valid covariance matrix. Given that P_- is full rank, the objective function grows without bound as the smallest eigenvalue of P_+ goes to zero, so it is clear that in any solution P_+ will be positive definite and this first constraint will not be binding in practice.

The second constraint, sometimes termed the “no-forgetting” constraint, is often binding, and it will turn out that handling that case is central to the full solution of the problem. This latter constraint is necessary because the problem trades off posterior uncertainty among the components of the target, so if the loss matrix W assigns little weight to some component then it can be optimal to assign that component more posterior uncertainty than existed prior uncertainty. Because the introduction of new information cannot achieve this result, the constraint is necessary. Mechanically, this constraint guarantees that our formulation of $a_+ | \mathcal{I}_-$, above, is valid.

4 Solution

In this section, we describe the solution to the static RI-LQG tracking problem presented above in Definition 1. To begin with, we will work with the fixed marginal cost formulation, and then show the extension to the fixed capacity case.

4.1 Solution to the static LQG-RI tracking problem

It is easy to check that the first order condition to the problem yields:

$$P_+^{-1} = W/\lambda \tag{2}$$

We cannot generally write the first order condition in terms of P_+ , because we have not required W to be nonsingular.¹¹ Despite this, if the positive semidefiniteness constraints

¹¹ For this reason it is sometimes more convenient to work in terms of precision matrices rather than covariance matrices. However, when possible we will present results in terms of covariance matrices.

are not binding, then this yields the solution to the static RI-LQG tracking problem. In the general case when the constraints may be binding, particularly the no-forgetting constraint, the solution is more complex. Before presenting the full solution in Theorem 1, some preliminaries are provided in Lemma 2.

Lemma 2: Assume that the loss matrix W is positive semidefinite and the prior covariance matrix P_- is positive definite. Then considering the matrix pencil (W, P_-^{-1}) we have the following results:

- a. The Cholesky factor $LL' = P_-$ is nonsingular, so that $M = L^{-1}$ exists.
- b. Define $V = L'WL$. This matrix is positive semidefinite, and its eigendecomposition can be written $QDQ' = V$.
- c. The matrix pencil can be simultaneously diagonalized by congruence so that $W = S'DS$ and $P_-^{-1} = S'IS$, where $S = Q'M$.
- d. The generalized eigenvalues of the matrix pencil, denoted d_i , are the diagonal elements of the matrix D . It will be convenient to always arrange the generalized eigenvalues in nonincreasing order.
- e. The generalized right eigenvectors of the matrix pencil, denoted r_i , are the columns of the matrix $R = S^{-1}$.

Theorem 1: The solution to the fixed marginal cost static RI-LQG tracking problem is given by:

$$P_+ = RN^+R' \quad (3)$$

where N^+ is a diagonal matrix with entries n_i^+ . These diagonal elements are defined by $n_i^+ = 1/\delta_i^+$, where $\delta_i^+ = \max\{d_i/\lambda, 1\}$ and d_i and R are as defined in Lemma 2. As a consequence of assuming that the generalized eigenvalues d_i are in nonincreasing order,

the values n_i^+ will be in nondecreasing order. In the following two corollaries, we state an even more explicit solution for the useful special case in which the loss matrix is rank one and we show how the elements of the solution are related to a matrix pencil of interest, (P_+, P_-) .

Corollary 1: If the loss matrix is rank one then we can decompose it as $W = ww'$, with w an $n \times 1$ vector, and the solution to the fixed marginal cost static RI-LQG tracking problem can be written:

$$P_+ = P_- - \frac{1 - n_1^+}{\|L'w\|^2} P_- W P_-$$

Corollary 2: Let P_+ denote the posterior covariance matrix solving the static RI-LQG tracking problem and let s'_i denote the i -th row of the matrix S , defined in Lemma 2. Then n_i^+ is the generalized eigenvalue of the matrix pencil (P_+, P_-) associated with the left generalized eigenvector s'_i .

In order to solve the fixed capacity version of the problem, it is useful to first define a new quantity r as the integer such that $d_r > \lambda \geq d_{r+1}$ and define $d_0 = \infty$ and $d_{n+1} = -\infty$ to encompass degenerate and full rank solutions.

Theorem 2: The solution to the fixed capacity static RI-LQG tracking problem with κ measured in base b is as given in Theorem 1, except that λ is interpreted as a shadow cost. The value of λ that solves the problem is:

$$\lambda = \left[b^{-2\kappa} \prod_{i=1}^r d_i \right]^{\frac{1}{r}} \quad (4)$$

as long as $\kappa > 0$ and is undefined otherwise. The quantity b is the base of the logarithm that defines the unit of information ($b = 2$ if information is measured in bits), and the quantity r is defined as above, but now is determined in concert with λ . The procedure for computing

r and λ is as follows:

- a. Set $r = n$
- b. Compute λ according to equation (4), given r .
- c. If $d_i > \lambda$, $i = 1, \dots, r$ then this pair (r, λ) describes the solution. Otherwise, set $r = r - 1$ and repeat from step 2.

Corollary: For the fixed capacity static RI-LQG tracking problem:

- a. The shadow cost λ is monotonic decreasing in κ , for $\kappa \in (0, \infty)$.
- b. The quantity r is nondecreasing in κ .

4.1.1 Canonical synthetic target

Before proceeding with implications of these theorems, we first define a new random vector that is instrumental in understanding the solution to the static RI-LQG tracking problem.

Definition 2: We define the canonical synthetic target (briefly the canonical target) as the vector $\beta_c = S\alpha$, where S is the matrix of left generalized eigenvectors from the second Corollary to Theorem 1.

The canonical synthetic target is a transformation of the target vector into a new set of coordinates. The importance of this transformation and insight into the new coordinate space is given in the next lemma.

Lemma 3: The canonical synthetic target β_c , satisfies the following:

- a. $\beta_c | \mathcal{I}_+ \sim N(b_{c,+}, N^+)$ where $b_{c,+} = Sa_+$.
- b. $\beta_c | \mathcal{I}_- \sim N(b_{c,-}, I)$ where $b_{c,-} = Sa_-$.
- c. $E[(\alpha - a_+)'W(\alpha - a_+) | \mathcal{I}_-] = E[(\beta_c - b_{c,+})'D(\beta_c - b_{c,+}) | \mathcal{I}_-]$
- d. $I(\alpha, a_+ | \mathcal{I}_-) = I(\beta_c, b_{c,+} | \mathcal{I}_-)$

e. $I(\beta_c, b_{c,+} | \mathcal{I}_-) = \sum_{i=1}^n I(\beta_{i,c}, b_{i,c,+} | \mathcal{I}_-)$ where $\beta_c = (\beta_{1,c}, \dots, \beta_{n,c})'$

Parts (c) and (d) demonstrate that the objective function can be rewritten entirely in terms of β_c . It is because of these results that we call β_c a “synthetic” target. As we will show later, there are many transformations that allow us to reformulate the problem in terms of a variety of synthetic target vectors. Parts (b), (c), and (e) demonstrate that the elements of the canonical synthetic target are separable with respect to prior uncertainty, the loss function, and mutual information; this is the essence of the new coordinate space and, because such a vector can always be constructed, we refer to this as the *canonical* synthetic target. Moreover, part (a) demonstrates that the elements of the canonical synthetic target remain separable in the posterior.

Part (c) furnishes us an intuition for the generalized eigenvalues d_i : they define the loss function as associated with the canonical synthetic target. Because D is diagonal, the element d_i captures the full loss associated with the element $\beta_{i,c}$, and we thus refer to the elements d_i as the canonical loss weights.

We are now in a position to state some results following from Theorems 1 and 2. These results will equally apply to the fixed λ or fixed κ formulations, unless otherwise noted.

4.1.2 Rank of the solution

Definition 3: We refer to r as the rank of the solution to the static RI-LQG tracking problem, and we say that the solution is full rank if $r = n$.

Lemma 4:

- a. $r = \text{rk}(P_- - P_+)$, so the solution is full rank if and only if the no-forgetting constraint is not binding. If the solution is full rank, then the solution is given by the first-order condition.
- b. $r \leq \text{rk}(W)$, so if W is singular then the solution cannot be full rank.

- c. r is the number of elements for which the loss in utility caused by increased uncertainty, as measured by the canonical loss weight d_i , is greater than the marginal cost of additional attention, as measured by λ .
- d. r is the number of elements in the canonical synthetic target for which the agent processes new information.
- e. In the fixed κ formulation, if $\kappa > 0$ and $\text{rk}(W) \geq 1$, then $r \geq 1$. This is in contrast to the fixed λ case, which may have $r = 0$ even if W is full rank.

4.1.3 Information capacity allocations

Definition 4: The total quantity of information capacity used by the agent, measured in base b , is:

$$\kappa \equiv I(\alpha, a_+ | \mathcal{I}_-) = \frac{1}{2} \sum_{i=1}^r \log_b \frac{1}{n_i^+} \quad (5)$$

where we could have also used n as the upper limit of summation, since for $i > r$, $\log \frac{1}{n_i^+} = \log 1 = 0$. Alternatively, given the definition of λ from Theorem 2, we can also write:

$$\kappa = \frac{1}{2} \left[\sum_{i=1}^r \log_b d_i - r \log_b \lambda \right]$$

These formulas are equivalent (although in the latter formula we cannot use n as the upper limit of summation), and so this latter formula is also valid in the fixed λ formulation.

Definition 5: The information capacity allocated to processing the i -th element of canonical synthetic target β_c is:

$$\kappa_i \equiv I(\beta_{i,c}, b_{i,c,+} | \mathcal{I}_-) = \frac{1}{2} \log_b \frac{1}{n_i^+} = \begin{cases} \frac{\kappa}{r} + \log_b \left[\frac{\sqrt{d_i}}{\prod_{j=1}^r \sqrt{d_j}^{1/r}} \right] & i = 1, \dots, r \\ 0 & i = r + 1, \dots, n \end{cases} \quad (6)$$

The last formulation suggests a straightforward intuition describing the allocation of capacity: first, each element is given an equal amount of attention (the κ/r term), and then attention is added (subtracted) if the square root of canonical loss weight for that element is higher (lower) than the geometric mean across all elements that are considered. Note that this result is in terms of the canonical synthetic target, and this intuition does not extend to the original (fundamental) target. Given this definition, we can also write $\kappa = \sum_{i=1}^r \kappa_i$, where we could again use either r or n as the upper limit of summation.

Unfortunately, there is generally no straightforward measure of the information capacity allocated to processing an individual element of the fundamental target α . This is because it is not straightforward to decompose mutual information for random vectors exhibiting correlation. However, we can introduce an approximate measure.

Definition 6: An approximate measure of the information capacity allocated to the i -th element of the fundamental target α , measured in base b , is the following component-wise mutual information:

$$k_i \equiv I(\alpha_i, a_+ | \mathcal{I}_-) = \frac{1}{2} \log_b \left(\frac{P_{ii,-}}{P_{ii,+}} \right) \quad (7)$$

where, for example $P_{ii,-}$ is the (i, i) -th element of the matrix P_- . This quantity computes the information about the i -th element of the target that is contained in the full action a_+ , and it ignores the effect of correlation in the prior and the posterior. Note that generally $\sum_{i=1}^n k_i \neq \kappa$ and, moreover, the sum does not provide either an upper or lower bound for κ .

Lemma 5: If both W and P_- are diagonal matrices, then component-wise mutual information k_i is equal to both the information capacity allocated to processing the i -th element of the fundamental target α and the i -th element of the canonical synthetic target β_c , so that $k_i = \kappa_i$ and $\sum_{i=1}^n k_i = \kappa$.

4.1.4 Illustration: separable target

The solution to the static RI-LQG tracking problem is easiest to understand when the elements of the canonical target happen to be oriented in the same directions as the elements of the fundamental target. In practice, this situation primarily occurs when W and P_- are both diagonal, because in this case the fundamental target vector is already separable with respect to prior uncertainty, the loss function, and mutual information. For this reason, we describe a target associated with diagonal W and P_- as separable. We will first demonstrate the relatively simple solution in the separable case, and then emphasize that this same logic also applies to general case, except in terms of the canonical target rather than the fundamental target.

To fix notation, we will assume P_- is a positive definite diagonal matrix with elements $\sigma_{i,-}^2$ and that W is a positive semidefinite diagonal matrix with elements w_i^2 . For convenience, we will assume that $w_1^2 \sigma_{1,-}^2 \geq \dots \geq w_n^2 \sigma_{n,-}^2$ (we can always re-order the elements of α to make this true). Application of Lemma 2 is trivial in this case since $V = L'WL$ is already diagonal, so that the generalized eigenvalues are simply $d_i = w_i^2 \sigma_{i,-}^2$. This formula implies that the canonical loss weights d_i can be interpreted as “loss-weighted volatility”.¹² The associated right generalized eigenvectors are $r_i = \sigma_{i,-} e_i$ where e_i is the i -th element of the standard basis.

We will examine the solution in the fixed marginal cost case and note that these results apply also to the fixed capacity formulation of the problem when the shadow cost λ is computed as described in Theorem 2. We suppose that the rank of the solution is r , so that λ is a fixed parameter satisfying $d_r > \lambda \geq d_{r+1}$. From Theorem 1, it is easy to see that P_+ will also be a diagonal matrix, and we denote its i -th diagonal element as $\sigma_{i,+}^2$. Then the

¹² This interpretation as “loss-weighted volatility” is still broadly true in the more general case, but the relationships are more complex due to interaction effects

full solution is:

$$\sigma_{i,+}^2 = \begin{cases} \lambda/w_i^2 & i = 1, \dots, r \\ \sigma_{i,-}^2 & i = r + 1, \dots, n \end{cases}$$

The first order condition would have set $\sigma_{i,+}^2 = \lambda/w_i^2$ for $i = 1, \dots, n$. This is infeasible, since we defined r such that $\lambda/w_{r+1} \geq \sigma_{i,-}^2$ and so this would suggest more posterior uncertainty for elements $r + 1, \dots, n$ than there existed prior uncertainty - the agent would have “forgotten” information they previously knew. In this case, it is straightforward to impose the constraint, setting $\sigma_{i,+}^2 = \sigma_{i,-}^2$ for $i = r + 1, \dots, n$.

This case admits a simple formula for the information capacity allocated to each element:

$$k_i = \kappa_i = \begin{cases} \frac{1}{2}(\log_b w_i^2 + \log_b \sigma_{i,-}^2 - \log_b \lambda) & i = 1, \dots, r \\ 0 & i = r + 1, \dots, n \end{cases}$$

More attention is paid to elements of the target that are more important (in terms of loss weight) or that are associated with more prior uncertainty, and as the marginal cost of attention falls, (weakly) more attention will be paid to every element. For those elements that receive no attention from the agent according to this result, it is easy to see in the previous result that, as one would expect, posterior uncertainty is equal to prior uncertainty. If the no-forgetting constraint were not enforced, these elements would be associated with *negative* capacity allocations.

This section applies directly to cases in which fundamental target itself is separable so that the loss and prior covariance matrices are diagonal. This will generally not be the case, but from Lemma 3 we know that these conditions will always be satisfied for the canonical target. This means that the above analysis, which is easy to understand, can still be applied in the general case, so long as it is cast in terms of the canonical target.

4.1.5 Comparative statics

We now consider how the solution changes as individual parameters vary, holding everything else constant. Mathematically, these exercises can be relatively straightforward given the explicit formulas we derived for posterior uncertainty and attention allocations, but the intuition can be obscured due to the presence of binding constraints and the somewhat opaque character of the generalized eigendecomposition. For this reason, in this section we will only briefly describe the general effects on posterior uncertainty of a change in each type of parameter and will then focus on illustrating important behavior using two specific examples.

There are three types of parameters in the model: (1) the parameter associated with the information constraint, λ or κ , (2) the elements of W describing the loss function, and (3) the elements of P_- describing prior uncertainty. The effect of a change in the first type can be understood by focusing only on the marginal, or shadow, cost parameter λ , as a consequence of the Corollary to Theorem 2. It is easy to see from Theorem 1 that an increase (decrease) in the marginal cost of attention always weakly increases (decreases) posterior uncertainty for every element of the target.

For the second and third types of parameters, it is difficult to achieve a simple presentation of the wide variety of effects possible, as these parameters affect both the generalized eigenvalues and the generalized eigenvectors, and so affect the definition of the canonical target. Rather than attempt it, we instead consider the effect of a change in one of the canonical loss weights d_i , with the justification that this captures all possible effects for a given canonical target.

For the first time, here the formulation of the information constraint has a material effect on results. If the problem is formulated with a fixed marginal cost of attention, then an increase in the canonical loss weight associated with the i -th element of the canonical target,

d_i , weakly decreases posterior uncertainty associated with that element, but leaves posterior uncertainty associated with the other elements unchanged. If the problem is instead formulated with a fixed capacity, then an increase in d_i still weakly decreases posterior uncertainty for that element, but now weakly increases posterior uncertainty for all other elements. In the latter case, the increase in d_i makes it optimal to pay more attention to the i -th component, but attention must be reallocated from elsewhere to achieve that. In the former case, the agent simply pays to allocate additional attention, and the end result is an increase in the total quantity of information processed.

Illustration We now illustrate these results using two specific examples. The baseline parameterizations are as follows:

$$\begin{array}{l}
 \text{Example (a)} \\
 \\
 \text{Example (b)}
 \end{array}
 \quad
 \begin{array}{l}
 W^{(a)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_-^{(a)} = \begin{bmatrix} 1.5 & 0 & 0 \\ 0 & 1.4 & 0 \\ 0 & 0 & 0.8 \end{bmatrix} \\
 \\
 W^{(b)} = \begin{bmatrix} 1.5 & 0 & 0 \\ 0 & 1.4 & 0 \\ 0 & 0 & 0.8 \end{bmatrix}, \quad P_-^{(b)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
 \end{array}$$

These examples are relatively easily to understand because they are separable, and they are relatively easy to contrast because they share the same canonical loss weights. This allows us to highlight those differences caused by different loss matrices separately from those differences caused by different levels of prior uncertainty. While example (a) might initially appear more plausible than example (b) - since it may seem particularly unrealistic that the prior covariance matrix be the identity - it is example (b) that will be more useful in understanding more complex models. This is because any static RI-LQG tracking problem will be in the form of example (b) when it is cast in terms of its canonical target.

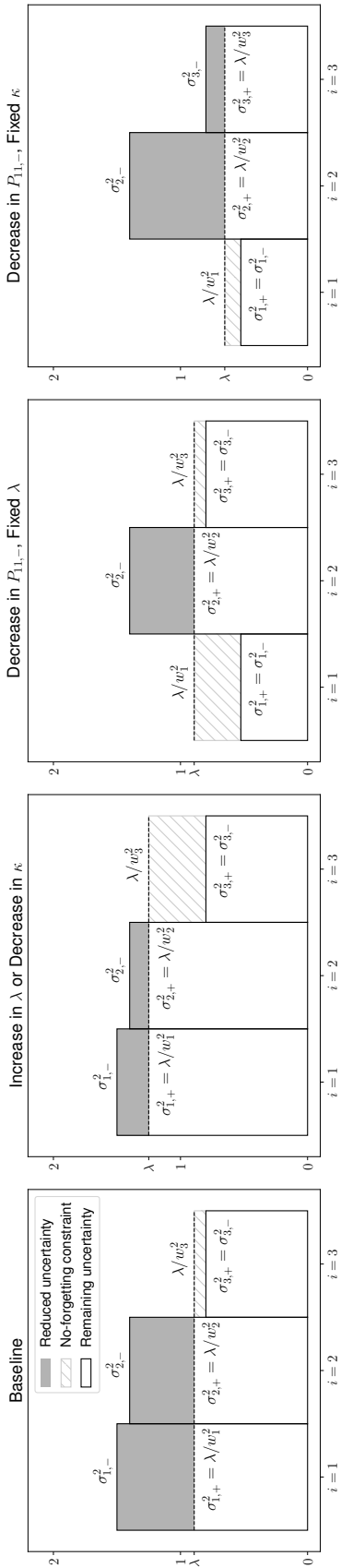


Figure 1: Comparative statics exercises for example (a)

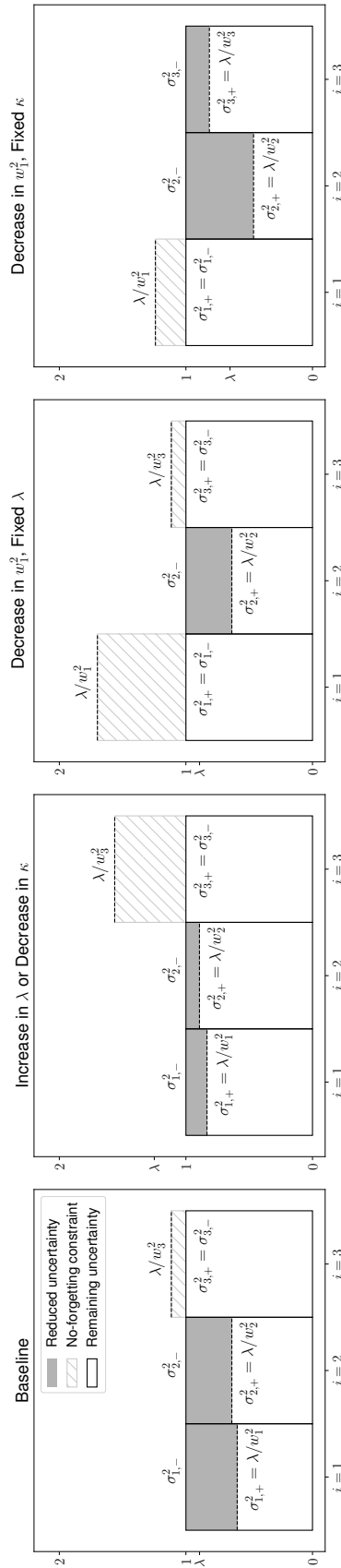


Figure 2: Comparative statics exercises for example (b)

Fig. 1 and Fig. 2, corresponding respectively to examples (a) and (b), each contain four panels depicting prior uncertainty and optimal posterior uncertainty.¹³ In both figures, the panel at the far left depicts the baseline case, while the three other panels depict specific deviations from that baseline case. In both figures, the second panel from the left depicts the effect of an increase in the marginal cost of attention (or equivalently a decrease in available capacity). The third and fourth panels depict a decrease in the canonical loss weight associated with the first element of the target, under the fixed marginal cost formulation in the third panel and under the fixed capacity formulation in the fourth panel. The two examples differ in how this decrease in the canonical loss weight is achieved - in example (a) we consider a decrease in prior uncertainty associated with the first element of the target, while in example (b) we consider a decrease in the loss weight associated with the first element.

In each panel, each bar outlined in black represents uncertainty associated with one element of the target. The height of the bar represents prior uncertainty, the dashed lines represent the level of posterior uncertainty suggested by the first order condition, the unshaded portion represents the optimal level of posterior uncertainty, and the shaded portion represents the reduction in uncertainty due to information processing. For some elements, there is a hatched region in place of the shaded region; in these cases, the first order condition suggested too high a level of posterior uncertainty, and the no-forgetting constraint became binding so that no information was processed. The hatched region represents the infeasible proposed enlargement of uncertainty.

In example (a), since the loss weight for each element of the target is equal to one, the proposed posterior uncertainty for each is simply equal to λ , which is set to be about 0.9 in the baseline case for this illustration. As shown in the first panel, this is feasible for the first two elements, which have relatively high prior uncertainty, but is not feasible for the third element, for which prior uncertainty is already lower than the given value of λ . In the second panel, we consider increasing λ , and this has straightforward effects: posterior

¹³ The solution process visualized in Fig. 1 is commonly known as “reverse water filling”.

uncertainty for the first two elements rises, while posterior uncertainty for the third element cannot rise any further. In the third panel, we consider, relative again to the baseline case, the effect of decreasing prior uncertainty associated with the first element while assuming that the model is formulated with a fixed marginal cost of attention. Because this change does not affect the loss weight, it does not affect the proposed level of uncertainty, which is still equal to λ . In fact, if we had only slightly reduced the prior uncertainty, it would not have changed the solution at all. However, in this case the reduction in prior uncertainty is so great that the no-forgetting constraint begins to bind. As described above, this has no effect on the solution for the second or third elements. In the last panel, we again consider, relative to the baseline case, the same decrease in prior uncertainty, but this time assuming that the model is formulated with a fixed capacity; the results clearly differ from those in the previous panel. Because the reduction in prior uncertainty makes it easier for the agent to achieve any desired level of posterior uncertainty, this has the effect of reducing the shadow cost of attention. While the no-forgetting constraint still begins to bind for the first element, in this case posterior uncertainty falls for both of the other elements, and in fact the no-forgetting constraint ceases to bind for the third element.

Even though the specifics of the solutions differ in example (b), there are qualitatively similar results from the comparative statics exercises. The two main differences are, first, since the loss weights differ, the first order condition will propose different levels of posterior uncertainty for each element of the target and, second, since prior uncertainty is the same, the no-forgetting constraint will bind at the same point for each element. The qualitative similarities are apparent in the second, third, and fourth panels: in response to an increase in λ , posterior uncertainty rises for each element; in response to a decrease in the canonical loss weight for the first element under a fixed marginal cost of attention, the no-forgetting constraint binds for the first element while the solutions for the other two elements remain unchanged; and in response to the same decrease under a fixed capacity, posterior uncertainty falls for the other two elements.

It is not an accident that these two examples share qualitative results; they were designed so that example (b) is simply example (a) recast in terms of its canonical target. In general, it may be quite difficult to interpret the solution in terms of the fundamental target, while it will always be easy to do so in terms of the canonical target. Simplifications achieved by considering the problem in terms of the canonical target will arise in every subsequent section of this paper.

4.2 Geometric interpretation of the static RI-LQG tracking problem and solution

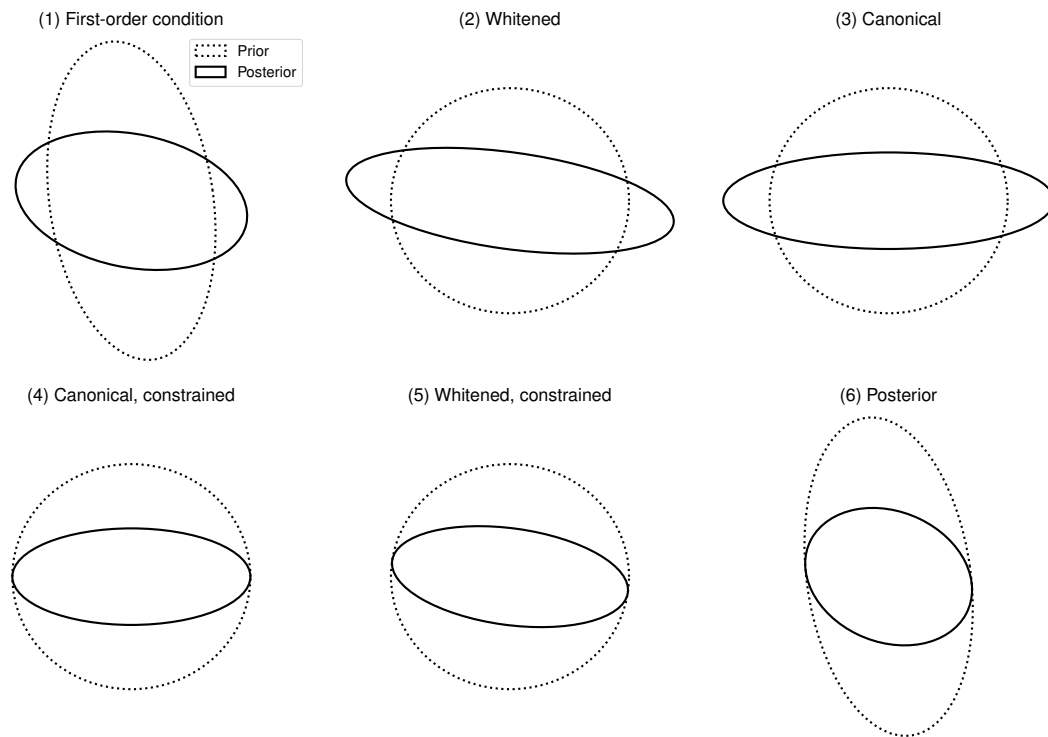


Figure 3: Geometrization of Theorem 1 using ellipsoids

In this section, we use a geometrical approach to interpret the problem and the nature of the solution given in Theorems 1 and 2. This is especially helpful in understanding the solution when the loss and prior covariance matrices are not diagonal. The general idea is

to take advantage of the geometrization of positive definite matrices, specifically covariance matrices, as ellipsoids.

The iso-density loci of the prior and posterior conditional distributions of the fundamental target form ellipsoids defined by the prior and posterior covariance matrices, and these ellipsoids can be interpreted as regions of uncertainty about the target, conditional on the prior or posterior information set. The volume of the ellipsoid defined by a positive definite matrix P is $V_P = |P| \times V_s$ where V_s defines the volume of an n -dimensional unit sphere. Iso-density ellipsoids with greater volume are associated with larger covariance matrices and increased uncertainty. The ratio of prior volume to posterior volume is given by $\frac{V_-}{V_+} = \frac{|P_-| \times V_s}{|P_+| \times V_s}$. Taking logs and dividing by two, we see that

$$\frac{1}{2}(\log_b(V_-/V_+)) = \frac{1}{2}(\log_b |P_-| - \log_b |P_+|) = I(\alpha, a_+ | \mathcal{I}_-)$$

Thus the information constraint can be understood in terms of the relative volumes of the prior and posterior ellipsoids. Under the fixed capacity formulation, the information constraint limits the volume of the ellipsoid describing posterior uncertainty in terms of the prior volume: if $I(\alpha, a_+ | \mathcal{I}_-) \leq \kappa$, then $V_- \geq V_+ \geq \frac{1}{2^{2\kappa}} V_-$. Similarly, under a fixed marginal cost, the total cost equals the marginal cost times a function of the ratio of volumes. The two positive semidefiniteness constraints in Definition 1 can also be understood in terms of the prior and posterior ellipsoids. The constraint $P_+ \geq 0$ simply requires that the posterior ellipsoid be well-defined. The no-forgetting constraint $P_- - P_+ \geq 0$ requires that the posterior ellipsoid be weakly contained within the prior ellipsoid. If the posterior ellipsoid extended beyond the prior ellipsoid in any direction, that would correspond to “forgetting” information previously known.

Formally, an ellipsoid defined by a positive definite covariance matrix P can be fully described in terms of its eigendecomposition. Its eigenvectors determine the directions of the ellipsoid’s principal axes, and its eigenvalues are proportional to the squares of the

semi-axis lengths. Because the determinant of a matrix is the product of its eigenvalues, the volume of an ellipsoid is invariant to its rotation, and for this reason, the information constraint depends only on the eigenvalues of the prior and posterior. The no-forgetting constraint, however, depends also on the eigenvectors.

A preliminary step in the proof of Theorem 1 was to establish that in the static RI-LQG tracking problem it will always be optimal for the eigenvectors of P_+ to coincide with the eigenvectors of a particular transformation of the loss matrix W . This result fixes the rotation of the posterior ellipsoid; what remains is to select its eigenvalues. The first-order condition proposes setting the eigenvalues equal to the eigenvalues of the inverse loss matrix scaled by λ . If the no-forgetting constraint is not binding then this fixes the semi-axis lengths and completes the solution. If the latter constraint does bind, however, the problem is more difficult because the posterior ellipsoid is usually not concentric with the prior ellipsoid; that is, they usually do not usually share eigenvectors. If the ellipsoids were concentric, then imposing the no-forgetting constraint would be straightforward: simply “pull in” the ends of each posterior principal axis that extend beyond the prior. This straightforward case is actually the situation when both W and P_- are diagonal, as we showed above. In the general case, however, it is not obvious which axes to “pull in”, or by how much.

This problem is solved by simultaneous diagonalization, which generates new coordinates under which the prior and posterior ellipsoids are not only concentric but are aligned with the standard axes. In fact, in the new coordinate space the prior is an n -dimensional unit sphere. The matrix S from Lemma 2 is the change of basis matrix implementing the transformation to the new coordinates, and the ellipsoids of uncertainty in the transformed space correspond to covariance matrices associated with the canonical synthetic target. The simple “pulling-in” approach can be implemented in the new coordinate space, and the solution in the original space can be found simply by reversing the transformation. This is possible because relative volumes are preserved by this transformation and the no-forgetting

constraint is satisfied in the original space if and only if it is satisfied in the transformed space.

We visualize the geometrical interpretation of the problem and solution in the six panels of Fig. 3. Panels (1) and (6), (2) and (5), and (3) and (4) represent prior and proposed posterior ellipsoids in three different coordinate spaces. In all panels, the dotted ellipsoids represent the prior in the given space and the solid ellipsoids represent a candidate posterior. The three upper panels represent the infeasible posterior proposed by the first order condition, and the three lower panels represent the feasible, constrained, posterior.

Panel (1) displays the prior ellipsoid and the proposed posterior ellipsoid satisfying the first order condition, $P_+ = \lambda W^{-1}$, in the standard basis. It is clear that the no-forgetting constraint is not satisfied, since the proposed posterior extends beyond the prior. Panel (2) represents the same prior and proposed posterior ellipsoids as in Panel (1), but after an intermediate transformation has been applied. This transformation will be called the “whitening” transformation and will be described in more detail later. The underlying coordinate space is called the “whitened” space.

Panel (3) again represents the same prior and proposed posterior ellipsoids, but now after the transformation to the canonical synthetic target has been applied. Accordingly, we call the underlying coordinate space the “canonical” coordinate space for this problem. In this panel, the problem has not been solved (since the posterior still extends beyond the prior), but the ellipsoids are concentric and are aligned with the standard axes, making the imposition of the constraints straightforward.

Panel (4) remains in the canonical coordinate space, but now displays the constrained posterior resulting from the pulling-in operation applied to the x semi-axis, as described above. The no-forgetting constraint is now satisfied. Panels (5) and (6) simply reverse the transformation to return to the original coordinate space. In Panel (6), the solid ellipsoid now represents the posterior covariance matrix that solves the static RI-LQG tracking problem,

with the no-forgetting constraint now satisfied.

4.3 The action solving the static RI-LQG tracking problem

In the previous sections, we noted that the optimal action is a conditional expectation, $a_+ = E[\alpha \mid \mathcal{I}_+]$, but presented the solution in terms of P_+ . In this section we provide a few important results concerning the structure and interpretation of the action a_+ itself, although two preliminary steps in this section will be asserted for the time being and will only be proved in later sections. First, we will write $\hat{\alpha}$ to denote the agent's understanding of the target based solely on incoming data; this will only be fully formalized later. Second, we present the result, derived later, that the agent's optimal action can be written as a weighted average of their prior and their understanding of the incoming data:

$$a_+ = (I - K)a_- + K\hat{\alpha} \tag{8}$$

where the weight matrix is $K = I - P_+P_-^{-1}$. This equation shows that, as usual in the LQG imperfect information setting, our agent is a Bayesian updater, but now because the agent is rationally inattentive, the weight matrix is not given but is selected. One important insight from this equation is that there are two channels through which the agent's action is driven away from the target. The first is that since the agent incompletely processes the incoming data, their understanding of the target is less than perfect, and so part of their action will be based on contaminating noise. The second is that even after receiving updated information, the rationally inattentive agent still places weight on their prior because they take into account their limited understanding of the incoming data.

Two limiting cases provide some intuition. First, as information becomes perfect, we have both $K \rightarrow I$ and $\hat{\alpha} \rightarrow \alpha$, so that the agent puts all weight on their understanding of the incoming data, and moreover their understanding is correct. When no information is

collected, $K \rightarrow 0$ and $\hat{\alpha}$ becomes diffuse, so that no weight is placed on incoming data and the action is equal to the prior. More general results are difficult in terms of the fundamental target α , because in general K will not be diagonal. As usual, however, things are more straightforward in terms of the canonical synthetic target β_c .

To motivate the use of the canonical coordinate space in interpreting the action, notice that we can rewrite $K = R(I - N^+)S$. The rows of the matrix S , s'_i , are the left generalized eigenvectors of (P_+, P_-) associated with generalized eigenvalues n_i^+ , and it is not hard to see that those rows are also the left eigenvectors of K associated with eigenvalues $1 - n_i^+$. The elements of the canonical target β_c are the linear combinations of α defined by these left eigenvectors s'_i . Taken together, the elements of the canonical target are exactly those random variables for which Bayesian updating by a rationally inattentive agent occurs independently. This is formalized in Lemma 6.

Lemma 6: The components of $b_{c,+} = E[\beta_c | \mathcal{I}_+]$, which we call the canonical synthetic action (briefly the canonical action), are:

$$b_{i,c,+} = n_i^+ b_{i,c,-} + (1 - n_i^+) \hat{\beta}_{i,c} \quad (9)$$

where $n_i^+ \in [0, 1]$ and $b_{c,-} = E[\beta_c | \mathcal{I}_-]$, and where the agent's understanding of the canonical target, $\hat{\beta}_{i,c}$, is defined as

$$\hat{\beta}_{i,c} \equiv y_{i,c} = \beta_{i,c} + \varepsilon_{i,c}, \quad \varepsilon_{i,c} \sim N(0, (1/n_i^+ - 1)^{-1})$$

if $n_i^+ \in [0, 1)$ and is diffuse if $n_i^+ = 1$. The noise term $\varepsilon_{i,c}$ is a mechanism to formalize the effects of inattention. The alternative notation used here, $y_{i,c}$, will connect this lemma with our definition of representations, introduced later. Importantly, while it will turn out that $\hat{\beta}_{i,c}$ will always correspond to what we term a feasible representation, so that we will always be justified in writing it as $y_{i,c}$, the same is not true of $\hat{\alpha}$.

In the canonical coordinate space, therefore, the Bayesian updating is straightforward: the action is a simple weighted average of the prior for that component and a noise-contaminated version of the canonical target. We now relate these results to the action associated with the fundamental target.

Theorem 3: The (fundamental) action that solves the static RI-LQG tracking problem for either the fixed marginal cost or fixed capacity case is:

$$a_+ = Rb_{c,+} \tag{10}$$

where $b_{c,+}$ is the canonical action and R is the matrix of right generalized eigenvectors defined in Lemma 2. Although this Theorem is in a sense trivial - a straightforward application of the definition of the canonical target - it is important because it formalizes the construction of a_+ .

4.3.1 Bias, variance, and responsiveness

In general, we know that rationally inattentive individuals will not respond perfectly to incoming data, but we can use the updating equation in the canonical space, given in Lemma 6, to provide a sharper comparison with the perfect information situation. Above, we described two channels driving the action away from the target; formally, these are, first, that a rationally inattentive agent introduces contaminating noise, since $\varepsilon_{i,c} \neq 0$, and, second, that a rationally inattentive agent chooses to be partially unresponsive, since $n_i^+ \neq 0$. By contrast, a perfectly informed agent has both of these equal to zero. By viewing the action $b_{i,c,+}$ as the rational inattention *estimator* of $\beta_{i,c}$, we can say that the variance of the estimator is due to the former channel, while the bias of the estimator is due to the latter. To justify this terminology, we define the bias and variance of the rational inattention action

as:

$$\begin{aligned}
E[b_{i,c,+} - \beta_{i,c} \mid \mathcal{I}_-, \beta_{i,c}] &= n_i^+(b_{i,c,-} - \beta_{i,c}) && \text{Bias} \\
\text{Var}(b_{i,c,+} \mid \mathcal{I}_-, \beta_{i,c}) &= (1 - n_i^+)^2 \text{Var}(\varepsilon_{i,c}) && \text{Variance}
\end{aligned}$$

The bias is generally nonzero unless the target is degenerate ($\beta_{i,c} \equiv b_{i,c,-}$) or information is perfect ($n_i^+ = 0$). The variance is nonzero unless information is perfect or the agent collects no information at all ($n_i^+ = 1$). The bias describes the extent to which the rational inattention action will differ from the target on average. The quantity n_i^+ is the proportion of the unexpected part of the incoming data to which the agent is unresponsive, and so the quantity $1 - n_i^+$ can be interpreted as the responsiveness of the agent. It is difficult to meaningfully extend these results to the fundamental target in a general way, other than by mechanically referencing Theorem 3.

4.3.2 Linear combinations of the target

It can be useful to explore arbitrary linear combinations of the action, $w'\alpha$ where w is an $n \times 1$ vector of weights, and it is easy to do so. Applying Theorem 3, we can compute any linear combination as $w'a_+ = \gamma'b_{c,+}$, where $\gamma' = w'R$ are the weights in the canonical space. One reason that this is interesting is that the loss function is often constructed exactly to minimize the weighted mean square error of one or more such linear combinations. Supposing that we are interested in n linear combinations defined by w_1, \dots, w_n with weights ξ_1, \dots, ξ_n , then the loss function is:

$$\sum_{i=1}^n \xi_i E[(w_i'\alpha - w_i'a_+)'(w_i'\alpha - w_i'a_+) \mid \mathcal{I}_-]$$

and this can be rewritten in the standard form $E[(\alpha - a_+)'W(\alpha - a_+) \mid \mathcal{I}_-]$ by setting $W = \sum_{i=1}^n \xi_i w_i w_i'$.

A special case that is often of interest occurs when an agent is only interested in tracking one specific linear combination $p = w'\alpha$, so that their loss function is $E[(p - p_+)^2 | \mathcal{I}_-]$. This can be written in the standard form using the rank one loss matrix $W = ww'$. Although the action solving the static RI-LQG tracking problem is a_+ , the agent is only interested in the synthetic posterior $p_+ = w'a_+$. We can of course compute this using Theorem 3, but in this case we can actually derive a more explicit solution. Using the first Corollary to Theorem 1 it is easy to show that w' is a left eigenvector of K and therefore the target of interest p is simply a scalar multiple of the canonical target. This result is very intuitive: the agent chooses to track exactly the object of interest. Finally, it is straightforward to show that the posterior collapses to $p_+ = n_1^+ p_- + (1 - n_1^+) \hat{p}$, so that the ultimate form of the solution is a simple Bayesian update in terms of the object of interest.

In this rank one case, we can simply characterize the sense in which uncertainty is reduced between prior and posterior. Since w' is proportional to the only generalized eigenvector associated with a nonzero eigenvalue, it follows that any vector orthogonal to w' is in the null space of K . Writing w^\perp as a vector orthogonal to w , it is not hard to show that $w'P_+w < w'P_-w$ and that $w^{\perp'}P_+w^\perp = w^{\perp'}P_-w^\perp$. The general version of this result for the rank n case is that uncertainty is only reduced for the space spanned by the canonical targets $\beta_{i,c}$ to which attention is actually allocated, i.e. for which $n_i^+ < 1$.

4.3.3 Illustration: rank one case

To illustrate the rank one case, we consider the example in section 3.2.3 of Sims (2010) in which an agent is supposed to be tracking a variable $y_t = \sum_{i=1}^n z_{it}$ subject to a fixed marginal cost of attention λ , where $z_{it} \sim N(0, \omega^2)$, independent across i and t . Since this problem is identical at each time period t , we can sequentially apply the static solution described here, and we assume that the agent's prior is just the unconditional distribution, so that $z_t | \mathcal{I}_{t-1} \sim N(0, \omega^2 I)$ for all t . While Sims (2010) gives the general form of the

solution to this problem, as a consequence of Theorem 1 we can easily derive the exact formula.

To set up the problem in terms of our Definition 1, the fundamental target is the vector z_t and the loss matrix is $W = \iota \iota' = \mathbf{1}_{n \times n}$ (an $n \times n$ matrix of ones), where $\iota = (1, 1, \dots, 1)'$ is a vector of weights defining y_t as a linear combination of z_t . The prior covariance matrix is $P_- = \omega^2 I$. The canonical loss weights are $d_1 = n\omega^2$ and $d_i = 0$ for $i = 2, \dots, n$. This implies that $n_1^+ = \min(\lambda/n\omega^2, 1)$ and $n_i^+ = 1$ for $i = 2, \dots, n$. Applying the first Corollary to Theorem 1, we conclude that:

$$P_+ = \omega^2 (I - (1 - n_1^+) (1/n) \mathbf{1}_{n \times n})$$

This agrees with the solution in Sims (2010), except that we are able to be more explicit the term $(1 - n_i^+)$. As described above, we have also formalized Sims' remark that the variance of any linear combination $w'z_t$ that is uncorrelated with $\iota'z_t$ will not be reduced, regardless of the cost λ . This is easy to see here, because ι' is the only generalized eigenvector s'_i associated with a generalized eigenvalue for which it is possible that $n_i^+ < 1$.

4.4 Transformations of the static RI-LQG tracking problem

In previous sections, we have extensively used a specific transformation to construct what we call the canonical synthetic target. This transformation is particularly useful because it simplifies the problem while preserving important relationships, especially the information and no-forgetting constraints. However, this is not the only possible transformation of the problem, and so we provide a more general result here.

Definition 7: Consider a static RI-LQG tracking problem defined by the tuple (W, a_-, P_-) , referred to as the reference problem. Let B be a nonsingular $n \times n$ matrix. We define the B -transformed static RI-LQG tracking problem, corresponding to the B -synthetic target

$\beta = B\alpha$, as:

$$\begin{aligned} \min_{O_+} \text{tr}(VO_+) + \lambda(\ln |O_-| - \ln |O_+|) & \quad (11) \\ \text{s.t. } \beta \mid \mathcal{I}_- \sim N(b_-, O_-) & \\ O_+ \geq 0 & \\ O_- - O_+ \geq 0 & \end{aligned}$$

where $V = B^{-1'}WB^{-1}$, $O_- = BP_-B'$ and $b_- = Ba_-$. We represent the B -transformed problem by the tuple (B, W, a_-, P_-) , and note that this definition encompasses the standard formulation given by Definition 1, which can be included here by setting B to the identity matrix. Note also that any B -transformed problem can be written as an independent problem (I, V, b_-, O_-) , although this eliminates connection to the reference problem.

Theorem 4: If a matrix O_+ solves the B -transformed static RI-LQG tracking problem (B, W, a_-, P_-) , then the matrix $P_+ = B^{-1}O_+B^{-1'}$ solves the reference static RI-LQG tracking problem (W, a_-, P_-) .

We can use this result to redefine the canonical target.

Definition 8: Let S be the matrix defined in Lemma 2. Then the S -transformed problem (S, W, a_-, P_-) is called the canonical form of the reference problem and the S -synthetic target is exactly the canonical synthetic target given in Definition 2, $\beta_c \equiv S\alpha$.

There are two other synthetic targets that it will be useful to formally define.

Definition 9: Let $M'M = P_-^{-1}$. Then the M -transformed problem (M, W, a_-, P_-) is called the whitened form of the reference problem and the M -synthetic target is called the whitened synthetic target.

Definition 10: Let $ZXZ' = W$ be the eigendecomposition of W . Then the Z -transformed problem (Z, W, a_-, P_-) is called the eigendecomposition form of the reference problem

and the Z -synthetic target is called the eigendecomposition synthetic target.

Since the product of two nonsingular matrices is again nonsingular, we can chain transformations together, and still apply Theorem 4 to the product of the transformation matrices.

Lemma 7: If B and C are nonsingular $n \times n$ matrices, then the CB -transformed problem (CB, W, a_-, P_-) is equal to the C -transformation of the B -transformed problem.

This allows us to give further insight into the canonical form of the static RI-LQG tracking problem.

Lemma 8: The canonical form of the reference problem is equivalent to the transformed problem achieved by first applying the whitening transformation to the reference problem and then subsequently applying the eigendecomposition transformation to the resultant whitened problem.

Lemma 8 is a formalization of the geometrical steps visualized in Fig. 3.

5 Representations

Although we have continually described a_+ as a conditional expectation, we have so far left the posterior information set \mathcal{I}_+ vague and focused instead on the posterior covariance matrix P_+ , and we have also purposely presented both the problem and solution with no mention of the “observation” or “signal” vectors that are commonly used in the rational inattention literature. In this section, we finally consider the posterior information set and discuss what we call “representations” of the information processed by agents. We first pursue these issues qualitatively and then formalize them using an algebraic approach.

Definition 11: For a static RI-LQG tracking problem (W, a_-, P_-) with solution P_+ and the corresponding action a_+ , we define a representation as any random vector y_+ that generates the solution, i.e. for which $E[\alpha \mid \mathcal{I}_-, y_+] = a_+$. An innovation representation, denoted v_+ ,

is any representation that additionally satisfies $E[v_+ | \mathcal{I}_-] = 0$.

We think that “representation” is a natural term to capture the essence of these vectors, particularly because they are not fundamental to the static RI-LQG tracking problem and because there are many vectors that satisfy the definition. When we provide a formal derivation, we will show that the most useful subset of representations correspond to a noise-contaminated version of some synthetic target. The synthetic targets express the fundamental target in different coordinate systems, and this is also the role representations play, except that representations express the agent’s imperfect understanding after processing new data.

Lemma 9: The action a_+ is a representation, since $a_+ = E[\alpha | \mathcal{I}_-, a_+]$. Thus we can refer to the action a_+ as the agent’s “perception” of the target.

The term “perception” seems natural to use when discussing a_+ as it relates to the agent’s *understanding* of the target α whereas the term “action” seems natural when discussing how the agent *uses* the solution of the rational inattention problem in the context of a larger economic problem. However, both terms refer to the same object, the conditional expectation a_+ .

In the rational inattention literature, what we refer to here as representations are often instead referred to as “observations” or as “signals”, and the rational inattention problem is often formulated in terms of selecting the noise covariance matrix corresponding to a specific form of a signal vector, rather than in terms of selecting the posterior covariance as we have done. This approach can be valid, and in fact we will later show how to reformulate the static RI-LQG tracking problem in similar terms. However, we argue that using the terms “observation” or “signal” can create ambiguities because they conjure up certain connotations that may not be natural in the rational inattention context. As we develop formal definitions of representations and the posterior information set, we will make concrete these concerns.

5.1 Posterior information set and feasible representations

In this section, we use an algebraic approach to describe the posterior information set and the space of representations available to the agent.

Since the optimal action is a conditional expectation and since all variables are jointly Gaussian, a_+ is the linear projection of α onto a vector space. We take \mathcal{V} to be the vector space of Gaussian random vectors of dimension n equipped with the inner product $\langle X, Y \rangle = E[XY']^{14}$, and identify \mathcal{W}_- and \mathcal{W}_+ to be the subspaces of \mathcal{V} defined by the information sets \mathcal{I}_- and \mathcal{I}_+ . This implies that $\mathcal{W}_- \subseteq \mathcal{W}_+$. Now, recalling that we can write $\alpha = a_+ + \eta$, we have $\alpha \in \mathcal{V}$, $a_+ \in \mathcal{W}_+$, and $\eta \in \mathcal{W}_+^\perp$, where \mathcal{W}_+^\perp is the orthogonal complement to \mathcal{W}_+ in \mathcal{V} . Thus $\alpha = a_+ + \eta$ is a decomposition into orthogonal subspaces.

Our goal is to isolate only the new information collected by the agent, and formally we want to construct a subspace \mathcal{W}_* that is the orthogonal complement of \mathcal{W}_- in \mathcal{W}_+ . The first step is to pick a subspace \mathcal{W}_y such that $\mathcal{W}_+ = \mathcal{W}_- \oplus \mathcal{W}_y$. If we let $\{1, v_-\}$ be an orthogonal basis for \mathcal{W}_- and take as given a basis $\{y_+\}$ for \mathcal{W}_y , then $\{1, v_-, y_+\}$ will be a basis for \mathcal{W}_+ .¹⁵ However, because we did not require $\mathcal{W}_- \perp \mathcal{W}_y$, this latter basis will generally not be orthogonal, and thus \mathcal{W}_y , and so also the basis vector y_+ , contains a component of information already known by the agent. However, we can construct an orthogonal basis $\{1, v_-, v_+\}$ by applying the Gram-Schmidt process, so that $v_+ = y_+ - \text{proj}_{\mathcal{W}_-} y_+$. This v_+ is now orthogonal to \mathcal{W}_- and so only contains new information; thus we have defined the space we want as $\mathcal{W}_* = \text{span}(v_+)$. This allows us to write a_+ as an orthogonal

¹⁴ More precisely, this $\langle X, Y \rangle$ is the Gram matrix consisting of the component-wise inner products of the random vectors X, Y .

¹⁵ The dimensions of \mathcal{W}_- and \mathcal{W}_* are not essential to this section, and could be made to be any number greater than zero.

decomposition:

$$a_+ = \underbrace{\text{proj}_{\mathcal{W}_-} \alpha}_{a_-} + \underbrace{\text{proj}_{\mathcal{W}_*} \alpha}_{a_* \equiv K_v v_+}$$

where a_- is the prior mean, now interpreted as the projection of α on prior information, a_* is the projection of α on new information, and the projection matrix is $K_v = \langle \alpha, v_+ \rangle [\langle v_+, v_+ \rangle]^{-1}$. In this way, we have decomposed posterior information, defined by \mathcal{W}_+ , into purely prior information, in \mathcal{W}_- , and purely new information, in \mathcal{W}_* .

As suggested by the notation, the vectors y_+ will correspond to the representations introduced in the previous section and the vectors v_+ will correspond to innovation representations. However, we still do not have an operational definition of y_+ , \mathcal{W}_* , or \mathcal{W}_+ . To remedy this, we consider an arbitrary random vector $y \in \mathcal{V}$. Denoting the space spanned by α as \mathcal{V}_α and its orthogonal complement in \mathcal{V} as \mathcal{V}_α^\perp , we can perform an orthogonal decomposition $y = \text{proj}_{\mathcal{V}_\alpha} y + \text{proj}_{\mathcal{V}_\alpha^\perp} y$. Since α is a basis element of \mathcal{V}_α we can write $\text{proj}_{\mathcal{V}_\alpha} y \equiv Z\alpha$ where Z is some conformable matrix, and we will denote $\zeta \equiv \text{proj}_{\mathcal{V}_\alpha^\perp} y$. We can then construct $v_+ = y - \text{proj}_{\mathcal{W}_-} y$, and note that $\text{proj}_{\mathcal{W}_-} y = Za_- + \text{proj}_{\mathcal{W}_-} \zeta$. We define $\varepsilon \equiv \zeta - \text{proj}_{\mathcal{W}_-} \zeta$ and $\Lambda \equiv \langle \varepsilon, \varepsilon \rangle$, and we note that both $\varepsilon \perp \alpha$ and $\varepsilon \perp \mathcal{W}_-$. This allows us to operationalize an innovation representation as $v_+ = Z\alpha + \varepsilon - Za_-$.

We can now explicitly compute $\langle \alpha, v_+ \rangle = P_- Z'$ and $\langle v_+, v_+ \rangle = ZP_- Z' + \Lambda$ so that $K_v = P_- Z' (ZP_- Z' + \Lambda)^{-1}$. Notice that, given the prior, the innovation representation v_+ and the space \mathcal{W}_* are completely defined by the pair (Z, Λ) , as is K_v . Furthermore, from any such pair we can define a representation $y_+ = Z\alpha + \varepsilon$ for which $y_+ - \text{proj}_{\mathcal{W}_-} y_+ = v_+$. The last step is to specify the matrices Z and Λ that correspond to valid representations. To do so, we note that $P_+ = \langle \alpha, \alpha - a_+ \rangle$ and it is not hard to show that this yields $P_+ = P_- - P_- Z' (ZP_- Z' + \Lambda)^{-1} ZP_-$. Applying the matrix inversion lemma to this equation,

we arrive at:

$$Z'\Lambda^{-1}Z = P_+^{-1} - P_-^{-1}$$

Any pair (Z, Λ) that satisfies this equation, along with Λ positive semidefinite (since it results from an inner product), describes what we will call a feasible representation. It is in this way that the choice of P_+ in the static RI-LQG tracking problem defines \mathcal{W}_+ and thereby defines \mathcal{I}_+ . We now define a slightly more general concept of representation and give a formal definition of a feasible representation.

Definition 12: For a static RI-LQG tracking problem with solution P_+ , a representation of dimension m is defined as a tuple (d, Z, Λ^{-1}) such that:

- a. d is an $m \times 1$ vector that is constant with respect to the prior information set
- b. Z is an $m \times n$ matrix with full row rank¹⁶
- c. Λ^{-1} is an $m \times m$ positive semidefinite matrix
- d. The equation $Z'\Lambda^{-1}Z = P_+^{-1} - P_-^{-1}$ is satisfied

Because we only require Λ^{-1} positive semidefinite, such a representation cannot always be meaningfully written in terms of some target contaminated by a well-defined noise term.

We therefore introduce an additional condition:

- e. For some $0 < \ell \leq m$, we can write $E\Lambda^{-1}E' = \Lambda_{(\ell)}^{-1} \oplus 0_{(m-\ell, m-\ell)}$, where $\Lambda_{(\ell)}^{-1}$ is an $\ell \times \ell$ positive definite matrix, $0_{(m-\ell, m-\ell)}$ is an $m - \ell \times m - \ell$ matrix of zeros, and E is the product of elementary matrices that potentially implement row-swapping transformations.

A *feasible* representation is a representation that additionally satisfies condition (e). We can then define $E\Lambda E' = \Lambda_{(\ell)} \oplus \infty I_{(m-\ell)}$ and so any feasible representation can be written

¹⁶ It is not too difficult to expand this definition to include rank deficient Z , but these cases are not important for our purposes and including them would complicate the exposition that follows.

as a vector y_+ in the following form:

$$y_+ = d + Z\alpha + \varepsilon, \quad \varepsilon \sim N(0, \Lambda) \quad (12)$$

This definition is still somewhat loose, but it is understood that the agent simply does not process any updated data regarding the components of y_+ with infinite noise variance.¹⁷ We refer to a feasible representation as “proper” if $\ell = m$, so that Λ^{-1} is positive definite, and as improper if $\ell < m$, so that Λ^{-1} is only positive semidefinite. Since the block of y_+ with infinite noise variance corresponds to variables for which no data is processed by the agent, every improper representation can be made proper simply by eliminating the improper block and considering a reduced representation of dimension ℓ .

Definition 13: Given a feasible representation (d, Z, Λ^{-1}) , the reduced form of that representation is denoted $(d_{(\ell)}, Z_{(\ell)}, \Lambda_{(\ell)}^{-1})$, where ℓ and $\Lambda_{(\ell)}^{-1}$ are as defined in Definition 12 part (e), and $d_{(\ell)}$ and $Z_{(\ell)}$ contain the first ℓ rows of Ed and EZ , respectively. If the feasible representation is denoted y_+ , then its reduced form is simply the first ℓ rows of Ey_+ .

Since $\Lambda_{(\ell)}^{-1}$ is positive definite by construction, the reduced form of a feasible representation is proper. As a consequence, we can now give results for proper representations that automatically extend to the larger class of feasible representations through the reduced form of the latter.

A lower bound for the dimension of any representation is given in Lemma 10.

Lemma 10:

- a. The minimum dimension of any representation is the rank of the solution, so that

$$m \geq r.$$

- b. The dimension of any proper feasible representation is equal to the rank of the solu-

¹⁷ For example, even though the row and column interchange operations are well defined, constructing Λ as in the formula requires interpreting the product of zero and infinity as equal to zero.

tion, so that $m = r$.

Finally, we note that every feasible representation has a corresponding innovation representation that can be written as $v_+ = y_+ - E[y_+ | \mathcal{I}_-]$, and that every innovation representation is a feasible representation in its own right denoted by $(-Za_-, Z, \Lambda^{-1})$.

Implicit in the definition of a representation is the requirement that the no-formatting constraint be satisfied, since Λ^{-1} will be positive semidefinite if and only if $P_- - P_+$ is positive semidefinite. If we allowed the no-forgetting constraint to be violated, then we would have to admit representations containing noise terms with negative variances associated with one or more linear combinations. If the no-forgetting constraint is just satisfied, then according to our definition a representation does exist, but Λ^{-1} will be singular. This implies that any representation must include a noise term with infinite variance associated with one or more linear combinations. This is not invalid, since it simply implies that the agent processes no new information about those combinations, but it does compel us to make a distinction between feasible and infeasible representations. This is because if those linear combinations associated with infinite variance are not separable in the representation from those combinations associated with finite variance, a meaningful noise term cannot be constructed. For this reason, the feasible representations exactly formalize the ways in which one could meaningfully understand the processing of incoming data by a rationally inattentive agent. The infeasible representations are mathematically valid objects, but they do not provide insight into the mechanism of information processing by an agent.

We can now use the class of feasible representations to understand the solution to the static RI-LQG tracking problem as well as the corresponding action.

Theorem 5: Given a proper feasible representation (d, Z, Λ^{-1}) and associated innovation representation denoted v_+ , the solution to the static RI-LQG tracking problem can be writ-

ten as:

$$a_+ = a_- + K_v v_+ \tag{13}$$

$$P_+ = (I - K_v Z) P_-$$

where $K_v = P_- Z' (Z P_- Z' + \Lambda)^{-1}$.

These formulas will be familiar as the updating step of the Kalman filter, and, accordingly, as similar to the solution to the LQG signal extraction problem discussed above. Crucially, though, note that the signal extraction problem computes the optimal *unknown action* a_+ for a *given observation* y_+ . In our case, the solution to the static RI-LQG tracking problem yields a *given action* a_+ and we had to derive the corresponding *set of representations* that could be considered as generating it. This point is important because the fundamental for a rationally inattentive agent is the action itself, derived as a solution to the tracking problem, and it is unnecessary to posit an “observation” vector. While it is often useful to consider the problem as if the agent has processed the data as a particular representation, it must be remembered there are many such representations that would be equally valid.

5.1.1 Illustration: simplified vector space

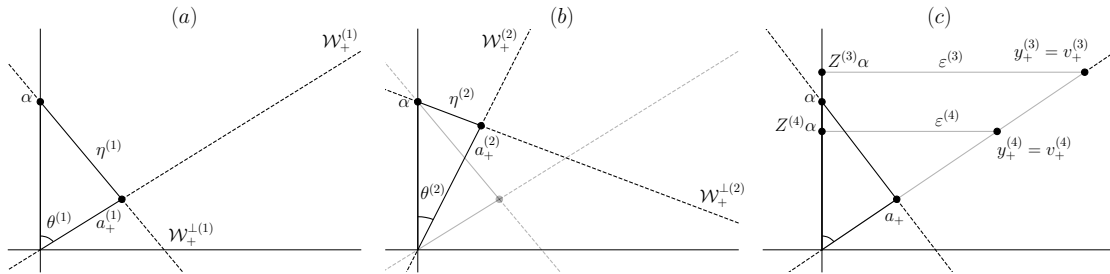


Figure 4: Visualization of the static RI-LQG tracking problem, solution, action, and representations in a simplified vector space

We can illustrate the algebraic approach using simplified vectors and vector spaces that

admit a graphical representation. In analogy with a univariate random variable, we consider a target α in the encompassing space $\mathcal{V} = \mathbb{R}^2$. The problem is then to find an action $a_+ \in \mathbb{R}^2$ that minimizes the (squared) Euclidean distance between target and action, $d(\alpha, a_+) = \langle \alpha - a_+, \alpha - a_+ \rangle$. We will make two simplifications. First, since this is analogous to a univariate problem, the loss matrix W is 1×1 , and we will normalize it to unity. Second, we will ignore prior information so that $a_- = 0$ and $\mathcal{W}_- = \{0\}$; this will imply that $\mathcal{W}_* = \mathcal{W}_+$.¹⁸

Our first step is as before: the form of any optimal action will be a projection on a subspace $\mathcal{W}_+ \subseteq \mathbb{R}^2$. We can then write the orthogonal decomposition $\alpha = a_+ + \eta$, where $\eta \in \mathcal{W}_+^\perp \subseteq \mathbb{R}^2$. The vector η represents tracking error, and the loss function can be interpreted as minimizing the length of the tracking error vector: $d(\alpha, a_+) = \langle \eta, \eta \rangle$, so this is the familiar sum of squared errors loss function. Now, the inner product concept in this simplified space is analogous to the concept of covariance in the full problem, and so we have $\langle \alpha, \alpha \rangle = P_-$, $\langle a_+, a_+ \rangle = P_- - P_+$, and $\langle \eta, \eta \rangle = P_+$. Thus, as before, our tracking objective is to minimize P_+ . The positive semidefiniteness constraints from the full problem are easily understood in this context as requiring valid action and error vectors (i.e. that these vectors must have nonnegative lengths).

For this illustration we will set $\alpha = \begin{bmatrix} 0 & 1 \end{bmatrix}'$, and in Fig. 4 (a) we show an example of vectors $a_+^{(1)}$, and $\eta^{(1)}$ that satisfy the definition of a_+ as a linear projection for some value $P_+^{(1)}$. We have also shown the corresponding subspaces $\mathcal{W}_+^{(1)}$ and $\mathcal{W}_+^{\perp(1)}$, and it is easy to see that a_+ is the projection of α onto \mathcal{W}_+ , while η is the residual. In Fig. 4 (b), we show a different set of action and error vectors that satisfy the above definition, but for a different value $P_+^{(2)}$. Because the length of $\eta^{(2)}$ is smaller, these new vectors must correspond to decreased posterior uncertainty: $P_+^{(2)} < P_+^{(1)}$. Since P_+ defines the length of η , it is easy to visualize how it is that P_+ specifies the vector space \mathcal{W}_+ and so ties down the posterior

¹⁸ We could extend the example to include a nontrivial prior, but it would require more complicated graphics that would obscure our primary goal.

information set \mathcal{I}_+ .

The remaining problem, analogous to Definition 1, is to select the optimal length of the error vector, P_+ , subject to either a constraint $\frac{1}{2} \log_b(P_-/P_+) \leq \kappa$ or a fixed cost λ of length reduction. It is interesting to note that in the case of univariate Gaussian random variables, the mutual information defining the analogous constraint can be written as $\frac{1}{2} \log_b 1/(1-\rho^2)$ where ρ denotes correlation. Here, correlation is analogous to the cosine of the angle between the target and action, defined by $\cos(\theta) = \frac{\langle \alpha, a_+ \rangle}{\|\alpha\| \|a_+\|}$. Thus, another way to write the constraint for this example would be in terms of the angle between action and target, to illustrate this, we have indicated the corresponding angles in Fig. 4 (a) and (b).

The solution to this problem is mechanically the same as in Theorems 1 and 2. Since we set $W = 1$, we have $P_+ = \min\{\lambda, P_-\}$. Then, given the form of a_+ and a solution P_+ , we can construct the vector space \mathcal{W}_+ and define the class of representations. To do so, consider an arbitrary $y \in \mathbb{R}^2$. We have $\mathcal{V}_\alpha = \text{span}\left(\begin{bmatrix} 0 & 1 \end{bmatrix}'\right) = \{Z \begin{bmatrix} 0 & 1 \end{bmatrix}' \mid Z \in \mathbb{R}\}$, so that we can write $y = Z\alpha + \zeta$, where $\zeta \in \text{span}\left(\begin{bmatrix} 1 & 0 \end{bmatrix}'\right) = \{c \begin{bmatrix} 1 & 0 \end{bmatrix}' \mid c \in \mathbb{R}\}$. Because we set $\mathcal{W}_- = \{0\}$, we must have $\text{proj}_{\mathcal{W}_-} y = 0$ so that $v_+ = y$ and $\varepsilon = \zeta$, with $\langle \varepsilon, \varepsilon \rangle = c^2 \equiv \Lambda$. Now, for a pair (Z, Λ) to be valid, it must satisfy $Z^2/\Lambda = 1/P_+ - 1/P_-$. For any solution P_+ , the right hand side is fixed, so that larger elements Z require larger Λ . Finally, for any valid pair (Z, Λ) , the associated innovation representation can be taken as a basis vector defining the subspace as $\mathcal{W}_* = \text{span}(v_+)$. In Fig. 4 (c), we plot representations $y_+^{(3)} = v_+^{(3)}$ and $y_+^{(4)} = v_+^{(4)}$ arising from two valid pairs $(Z^{(3)}, \Lambda^{(3)})$ and $(Z^{(4)}, \Lambda^{(4)})$. It is easy to see that any valid representation must lie in the subspace \mathcal{W}_+ and, conversely, the action a_+ will always be a projection onto the subspace spanned by a valid representation (and, more generally, also any prior information).

5.2 The fundamental and canonical representations

In this section, we present several important representations.

Definition 14: The fundamental representation is defined by $d = 0$, $Z = I$, and $\Lambda_f^{-1} = P_+^{-1} - P_-^{-1}$ and is denoted $(0, I, \Lambda_f^{-1})$. If the fundamental representation is feasible, we write it as:

$$y_f = \alpha + \varepsilon_f, \quad \varepsilon_f \sim N(0, \Lambda_f) \quad (14)$$

If the solution to the static RI-LQG tracking problem is full rank then the fundamental representation will be feasible, and also proper, but more generally it will usually be infeasible except in cases that exhibit a separation across the prior covariance and loss matrices that extends also to the posterior.

It is tempting to view the fundamental representation as the most straightforward representation, because it corresponds to the “true (fundamental) target plus white noise” concept often considered in the rational inattention literature. From the perspective of the agent, however, it is more natural to consider a representation based on the canonical synthetic target, because this latter target encapsulates the information of importance. Not only that, but the fundamental representation is often infeasible, whereas it will turn out that such a “canonical representation” will always be feasible.

Definition 15: The canonical representation is defined by $d_c = 0$, $Z_c = S$, and $\Lambda_c^{-1} = (N^+)^{-1} - I$ and is denoted $(0, Z_c, \Lambda_c^{-1})$. We write it as:

$$y_c = \beta_c + \varepsilon_c, \quad \varepsilon_c \sim N(0, \Lambda_c) \quad (15)$$

where $\beta_c = S\alpha$ is the canonical synthetic target. Because Λ_c^{-1} is diagonal, the canonical representation is always feasible.

The canonical representation corresponds to “true (canonical synthetic) target plus white noise”. While the fundamental target describes the shocks as they appear in the economy, the canonical target describes synthetic shocks as they matter - separately - to the agent. Because of this, it is conceivable how the agent could operationalize the solution to their problem in terms of this representation, by considering each component separately and choosing whether and how much to pay attention to each by adjusting the variance of the information processing noise.

Although the canonical representation is always feasible it is not always proper, because the agent may process no information about some components. However, by applying Definition 13, we can always construct a reduced canonical representation that is proper.

Definition 16: We write the reduced form of canonical representation as $(0, Z_r, \Lambda_r^{-1})$ and denote it by y_r .

This reduced canonical representation is perhaps the most useful representation, since it corresponds to the canonical target, contains a noise term with a finite diagonal covariance matrix, and can always be used to construct the action, by application of Lemma 6 and Theorem 3.

5.3 Representation form of the static RI-LQG tracking problem

We can now state an alternative form of the static RI-LQG tracking problem, which is in terms of selecting a representation rather than the posterior covariance.

Definition 17: The representation form of the static RI-LQG tracking problem is:

$$\begin{aligned}
& \min_{Z, \Lambda^{-1}} tr(WP_+) + \lambda(\ln|P_-| - \ln|P_+|) & (16) \\
& \text{s.t. } \alpha \mid \mathcal{I}_- \sim N(a_-, P_-) \\
& \Lambda^{-1} \geq 0 \\
& P_+ = (Z'\Lambda^{-1}Z + P_-^{-1})^{-1}
\end{aligned}$$

This formulation requires joint solution in Z and Λ^{-1} , and it is primarily of interest because many examples in the rational inattention literature use a similar form. One difficulty with this formulation is that the solution is not unique. For example, if (Z, Λ^{-1}) is a solution then so is $(XZ, (X\Lambda X')^{-1})$ for every nonsingular conformable matrix X . Possibly for this reason, this formulation of the problem is often split into two parts, and an optimal Z is solved for first. With some optimal Z fixed, an associated optimal Λ^{-1} can be solved for.

5.4 Representation form of the action

We can also now characterize the action in terms of specific representations, and derive the results we simply asserted when previously describing the action. From Theorem 5, for any feasible innovation representation $(-Za_-, Z, \Lambda^{-1})$ we have $a_+ = a_- + K_v v_+$, and we can extend this to any feasible representation by writing $v_+ = y_+ - Za_-$ so that $a_+ = (I - K_v Z)a_- + K_v y_+$. If Z has full row rank, then we can further write $a_+ = (I - K_v Z)a_- + K_v Z(\alpha + Z^- \varepsilon)$ where Z^- denotes the Moore-Penrose pseudo inverse. From this it is not hard to show that $K = I - P_+ P_-^{-1} = K_v Z$, and so by defining $\hat{\alpha} = \alpha + Z^- \varepsilon$, we have the formulation presented originally. We emphasize that even though $\hat{\alpha}$ is generally not a feasible representation, the action a_+ is always valid and the weight matrix K is always well-defined. This underscores once more the result that the representation of the data in an observation-like form is inessential to the solution of the static RI-LQG

tracking problem. If it happens that y_f is a feasible representation, then $\hat{\alpha} = y_f$, and we have $a_+ = (I - K)a_- + Ky_f$. However, we can always construct the action using some feasible representation. In particular, if we consider the canonical representation, we have $a_+ = (I - K_c Z_c)a_- + K_c y_c$ and it is easy to show that $K_c = R(I - N^+)$, and a little algebra brings us to either Lemma 6 or Theorem 3. Finally, these results could be easily rewritten in terms of the proper feasible representation y_r , since $1 - n_i^+ = 0$ for $i = r + 1, \dots, n$.

6 Application: rationally inattentive price-setting

In this section, we consider the model of rationally inattentive price-setters introduced by Maćkowiak and Wiederholt (2009), which we will refer to as MW. While our analytic results only extend to the static version of the problem, this has become a useful benchmark case. First, we apply the method derived in this paper to solve the problem formulated by MW in which a fixed capacity of attention is employed along with a restriction that amounts to requiring a diagonal posterior covariance. With the more general solution method now available to us, we can also consider three alternative formulations, and we discuss how the results change in each variant.

The basic setup considers a unit mass of monopolistically competitive firms indexed by i , each with identical profit function $\pi(P_{it}, P_t, Y_t, Z_{it})$ where P_{it} is the price of firm i 's differentiated good, P_t is the aggregate price level, Y_t is real aggregate demand, and Z_{it} is a firm-specific productivity shock. We assume an exogenous process for nominal aggregate demand, $Q_t = P_t Y_t$. Denote a second-order approximation to this profit function by $\tilde{\pi}(p_{it}, p_t, y_t, z_{it})$ where the lower case variables denote log-deviation from nonstochastic steady-state. The aggregate price is log-approximated as $p_t = \int_0^1 p_{it} di$. The optimal price

under perfect information is:

$$p_{it}^{\diamond} = \frac{\hat{\pi}_{14}}{|\hat{\pi}_{11}|} z_{it} + \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|} q_t + \left(1 - \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|}\right) p_t$$

where $\hat{\pi}_{ij}$ denotes a second partial derivative of the profit function evaluated at the non-stochastic steady-state. It is not hard to see that under perfect information, equilibrium yields $p_t = q_t$. To extend this to incorporate imperfect information, we follow MW in assuming that firms set prices to track the perfect information price but are rationally inattentive. Because the approximate profit function is quadratic, this is in the form of an RI-LQG tracking problem. Here we focus on the static case, and so we assume that z_{it} and q_t are Gaussian white noise with variances σ_z^2 and σ_q^2 , and that $z_{it} \perp q_t$.¹⁹ As described above, we know that the form of the action will be the conditional expectation²⁰ $p_{it}^* = E[p_{it}^{\diamond} | \mathcal{I}_+]$ and, given this form of the solution, the expected loss in profits due to setting a suboptimal price will be:

$$E[\tilde{\pi}(p_{it}^{\diamond}, p_t, y_t, z_{it}) - \tilde{\pi}(p_{it}^*, p_t, y_t, z_{it}) | \mathcal{I}_-] = \frac{|\hat{\pi}_{11}|}{2} E[(p_{it}^{\diamond} - p_{it}^*)^2 | \mathcal{I}_-]$$

Based on the form of the perfect information equilibrium aggregate price, we follow a guess-and-verify approach to solve for the imperfect information equilibrium, guessing that $p_t = \gamma q_t$.²¹ Then firm i will set their price according to:

$$p_{it}^* = E \left[\frac{\hat{\pi}_{14}}{|\hat{\pi}_{11}|} z_{it} + \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|} q_t + \left(1 - \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|}\right) \gamma q_t \mid \mathcal{I}_+ \right]$$

¹⁹ Given the generality of the solution method derived in this paper, it is no longer essential that z_{it} and q_t be independent, but we maintain this assumption for comparison with Maćkowiak and Wiederholt (2009).

²⁰ In the equation for the imperfect information case, Maćkowiak and Wiederholt (2009) condition on a vector of signals s_i^t . For the reasons described earlier in this paper, we use the more general posterior information set \mathcal{I}_+ .

²¹ Maćkowiak and Wiederholt (2009) write $p_t = \alpha q_t$, but we use γ in place of α to avoid confusion with the fundamental target.

This is a best response function given a particular γ , and the equilibrium solution represents a fixed point. For a given γ , the rational inattention problem can be written in the form of Definition 1, where the target vector is $\alpha_{it} = \begin{bmatrix} z_{it} & q_t \end{bmatrix}'$, and we can define a weight vector $w = (w_z, w_q)' = \left(\frac{\hat{\pi}_{14}}{|\hat{\pi}_{11}|}, \gamma + (1 - \gamma) \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|} \right)'$ so that the loss function is defined by the positive semidefinite matrix $W = \frac{|\hat{\pi}_{11}|}{2} w w'$; note that $\text{rank}(W) = 1$. We assume that agents have no special prior knowledge, so that $\alpha_{it} \mid \mathcal{I}_- = \alpha_{it} \sim N(0, \Omega)$, where $P_- = \Omega = \text{diag}\{\sigma_z^2, \sigma_q^2\}$.

At this point, MW make the additional assumption that firms must pay attention to z_{it} and q_t separately; following them, we refer to this as the independence assumption. We will expand on their results by considering four cases: with or without the independence assumption, and employing either fixed capacity or fixed marginal cost of attention. Since our framework immediately handles either a fixed capacity or a fixed marginal cost formulation, we need only now describe how to modify the problem and solution to impose the independence assumption.

It is most straightforward to introduce the independence assumption using the representation form of the problem given in Definition 17, because this assumption is most naturally interpreted as a limit on the form that representations (or “signals”, in their terminology) may take. The formalization of the independence assumption of MW requires that any representation form of the solution has both Z and Λ^{-1} as diagonal matrices. The implications for the posterior covariance matrix are easy to see by considering the equation $P_+^{-1} = Z' \Lambda^{-1} Z + P_-^{-1}$. Combined with the assumption that $z_{it} \perp q_t$, this requires that in any solution the posterior covariance matrix must be diagonal. However, it is also clear that the independence assumption does not put restrictions on the diagonal elements of P_+^{-1} . This can be stated as follows: the independence assumption restricts the eigenvectors of P_+ but not the eigenvalues, and, in this example, this amounts to requiring that $Q = I$, where Q is as defined in Lemma 2.

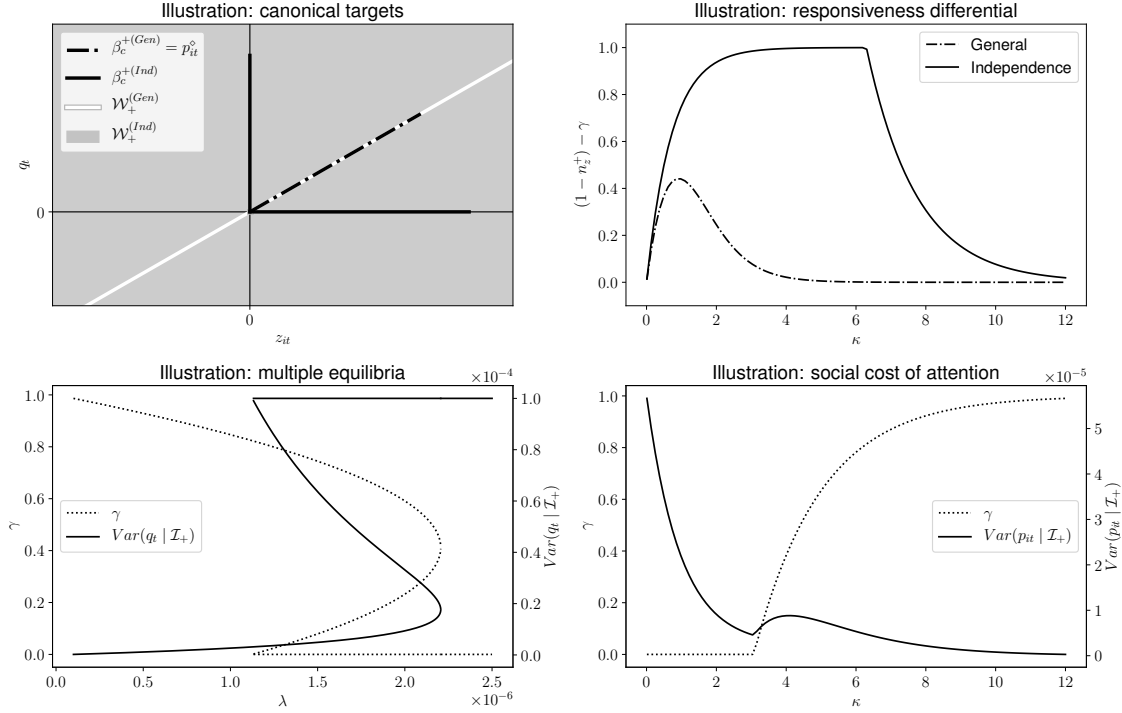


Figure 5: Illustrations of possible behavior in the rational inattention price-setting model

It might seem at first that our solution method cannot be applied here, because Q is the matrix of eigenvectors of $V = L'WL$, and it is clear that this matrix will not be diagonal, given the loss matrix derived above. However, the portion of the objective function that is responsible for the eigenvectors of the posterior covariance matrix is $tr(WP_+)$ and it is not hard to show that the independence assumption requires that this term be equal to $w_z^2 \sigma_z^2 n_1^+ + w_q^2 \sigma_q^2 n_2^+$. This suggests that we can simultaneously impose $Q = I$ while still applying the basic structure developed in this paper by employing a different loss matrix, $W_I = \text{diag}(w^2)$. For this particular example, this is a way of imposing the independence assumption while still allowing us to employ Theorems 1 and 2 to achieve the solution.

We now consider four cases. The first two consider the fixed marginal cost and fixed capacity formulations in the general case, while the second two proceed under the independence assumption. To conserve space, we relegate the details of the solutions to Appendix B, but broadly the solution involves two steps. We first take γ as given and solve the static

RI-LQG tracking problem. Then, since this yields attention allocations that themselves depend on γ , the second step is to solve the fixed point problem and find equilibrium values of γ . Although there are many interesting differences between these models, due to space constraints we focus here on only three: (1) the definition of the canonical synthetic target, (2) the responsiveness of agents to the two types of shocks, and (3) the values of γ in equilibrium. We are most interested in highlighting the difference between the general case and the case under the independence assumption; a few selected illustrations appear in Fig. 5 and will be discussed below.

The first, and most obvious, difference between these cases is the resultant canonical synthetic target. This is an important difference, because the components of this target define the *objects of attention* for rationally inattentive agents. In the general case, there is a single canonical target that consists of the optimal price, while under the independence assumption there are two canonical targets: the idiosyncratic and aggregate shocks (this latter fact was exactly the goal of the assumption).²² The canonical targets for each case are visualized in the upper left panel of Fig. 5. Both choices may appear reasonable, but while MW advocate for the independence assumption, we argue that the general case should be preferred; this issue is considered in detail in the next section. For now, we focus on an important practical effect, that the two cases lead to a qualitative difference in the form of the posterior uncertainty chosen by agents.

We showed above that uncertainty is only reduced for the space spanned by the canonical targets to which the agent actually pays attention. Here, under the independence assumption, this space can include both the idiosyncratic and aggregate shocks, but in the general case the space can only be a hyperplane corresponding to a single particular linear combination of the shocks. This difference in the dimensions of the posterior vector spaces is also apparent in the upper left panel of Fig. 5. The implication of this is that as the cost of

²² Technically even in the general case there are two components to the canonical synthetic target, but the agent never pays attention to the second component and it is defined only by its orthogonality to the first component.

attention falls to zero (or the capacity rises to infinity), under the independence assumption the agent will become fully informed about the idiosyncratic and aggregate shocks separately, but in the general case the agent will only become fully informed about the specific linear combination that is relevant to their economic problem, so that some uncertainty will remain about the shocks themselves. Thus the independence assumption yields suboptimal behavior for the agent, since they are acquiring costly information that they do not use. Specifically, for any given parameterization, posterior uncertainty about the optimal price and the objective function itself (both of which the agent wishes to reduce as much as possible) will be higher given the independence assumption, although posterior uncertainty about either the idiosyncratic or aggregate shocks individually will be higher in the general case.

The second difference between these cases that we consider is the implied responsiveness of rationally inattentive agents to shocks. To compute the responsiveness to a shock to the optimal price, we use the result (derived in the appendix) that for all the cases considered above, we can write the optimal posterior in the form

$$p_{it}^* = (1 - n_z^+)w_z z_{it} + \gamma q_t + \varepsilon$$

where $\gamma = (1 - n_q^+)w_q$ and ε is a mean zero noise term whose variance may differ under the various cases. For the general case, the solution imposes $n_z^+ = n_q^+$, whereas under the independent assumption they may differ. Recalling that the perfect information optimal price is $p_{it}^\diamond = w_z z_{it} + q_t$, we can measure the responsiveness of firms to idiosyncratic shocks as $(1 - n_z^+)$ and to aggregate shocks as γ . These values are between zero and one, and, since under perfect information both of these are equal to one, they describe the fraction of a shock reflected in the action of a rationally inattentive agent.

One feature of particular interest in MW and the related literature is whether it is possible that firms exhibit conditional responsiveness; that is, high responsiveness to idiosyncratic

shocks and low responsiveness to aggregate shocks. The key result of MW is that under the independence assumption, this can be achieved if firms pay close attention to idiosyncratic shocks (so that n_z^+ is close to zero) but they do not pay close attention to aggregate shocks (so that n_q^+ is close to one and hence γ is close to zero). In the general case, since $n_z^+ = n_q^+$ is imposed, it is more difficult to achieve this conditional response. Using a calibration based on that of MW, we compute the difference in responsiveness, $(1 - n_z^+) - \gamma$, across a range of values for the marginal cost and capacity parameters. We plot these values in the upper right panel of Fig. 5 for the fixed capacity formulations. Under the independence assumption, we find that an arbitrary difference can be achieved for some value of the marginal cost or fixed capacity parameters, and moreover that one can achieve any difference while also requiring that firms respond nearly perfectly to idiosyncratic conditions. This confirms the result of MW. In the general case, for this calibration, we find that the maximum difference is about 45 percentage points, and that this difference occurs when firms respond to about 75 percent of idiosyncratic shocks. This indicates that the independence assumption is not crucial to achieving a conditional response to shocks. However, since the contrast between attention paid to idiosyncratic and aggregate shocks is not as stark in the general case, this suggests that a richer price-setting model may be required to match empirical data on prices.²³

The final issue that we will consider is how the equilibrium values of γ vary over the cases. The term γ controls the strength of “coordination” across firms to aggregate shocks: if γ is high, then aggregate shocks have a high pass-through to all individual price-setting decisions, whereas if γ is low, aggregate shocks have a smaller impact. The primary result here is that for a given parameterization, γ will generally be lower under the independence assumption than it would be in the general case. This is because under the independence as-

²³ This is unsurprising, since the seminal model of Maćkowiak and Wiederholt (2009) was deliberately left relatively simple to expose the key mechanism. For example, Fulton (2015) demonstrates that even with the independence assumption, a more complex model is required to mitigate calibration issues that imply an implausible differential between the volatility of aggregate and idiosyncratic shocks.

sumption firms must pay attention to these shocks separately, and so part of the information collected is unused. The end result is that it is more costly for firms to pay attention to aggregate conditions. The two cases also display markedly different equilibrium behavior for γ : in the general case γ is monotonic nonincreasing in the marginal cost of attention and, at least for reasonable calibrations, there is a unique equilibrium; under the independence assumption, there are regions in which decreases in the marginal cost of attention actually decrease γ , and there are regions admitting multiple equilibria. This richer equilibrium behavior appears in the latter case because there are two components of the canonical target that end up receiving attention from the agent and so more complex interactions can arise.

Multiple equilibria can arise in this model due to the combination of strategic complementarities and endogenous information choice. Here, if most agents are paying attention to aggregate shocks and set their prices accordingly, then the aggregate shock is actually relevant for every individual agent, whereas if few agents pay attention to aggregate shocks then the cost of ignoring them for any individual agent can be small. For the calibration we consider, there is a region of parameterizations for λ in the fixed marginal cost case under the independence assumption that implies three equilibria: a high equilibrium, a low equilibrium, and one in which γ is zero. We illustrate the equilibrium values of γ in this case along with the corresponding posterior uncertainty about the aggregate shock in the lower left panel of Fig. 5.

A social cost of increased attention can arise in this model for a similar reason. If no firm pays attention to aggregate shocks (so that in equilibrium $\gamma = 0$), then these shocks do not enter the optimal price-setting equation, and all the posterior uncertainty faced by an imperfectly informed agent is driven by the idiosyncratic shocks. As available attention rises, if firms start to pay attention to aggregate shocks (so that γ becomes nonzero), these shocks become relevant and this can result in an overall increase in posterior uncertainty. Because the expected loss in profits increases with posterior uncertainty, this makes all

firms worse off. This is illustrated in the lower right panel of Fig. 5.

As far as we are aware, rational inattention price-setting models incorporating multiple equilibria or a social cost of increased attention have so far not been considered in the literature - although there is of course a vast body of work dedicated to these issues in imperfect information contexts generally. We hope that the solution method derived in this paper may facilitate access to these interesting questions.

7 Modeling rational inattention problems

We have emphasized throughout this paper that caution must be used when making modeling decisions in rational inattention models on the basis of intuition derived from signal extraction models. In this section, we consider the extent to which the independence assumption, and other assumptions with similar implications, can be justified without relying on an inappropriate analogy.

The independence assumption, introduced in Maćkowiak and Wiederholt (2009) (MW), has seen increased use in the rational inattention literature in recent years. A portion of its appeal is surely because it made the rational inattention problem more tractable (and indeed many authors do not provide a justification for its use), although this concern is less relevant now that we have derived an exact solution in the static case. However, it was introduced by MW not only for convenience but because they argued that the general case of the model was implausible. As we showed above, the linear combination that defines the relevant canonical target in the general case is exactly the linear combination that generates the optimal price from the idiosyncratic and aggregate shocks; this implies that the canonical representation is exactly of the form “profit-maximizing price plus noise”. MW write that this “amounts to assuming that the decision maker can attend directly to the profit-maximizing price” and suggest “we think that, in most economic contexts, decision

makers cannot attend directly to the optimal decision ... The independence assumption is the simplest way of modeling the idea that decision making is about first paying attention to a variety of variables, and then combining these different pieces of information in a single decision”. In this section, we take the opposite position and argue that the use of the general case is justified, and that the appeal of the independence assumption comes exactly from intuition derived from signal extraction models that does not apply to rationally inattentive agents.

To consider this issue we will examine two related claims along the lines of those from MW: (1) a rationally inattentive agent should not have access to a representation of the form optimal action plus noise, and (2) there may be restrictions that prevent a rationally inattentive agent from processing information in an arbitrary way. The first claim addresses whether we should *require* restrictions on representations (for example the independence assumption), while the second claim addresses whether we should *allow* restrictions.

There is no doubt that the first claim is plausible in the context of a signal extraction model, in which case the word “representation” would be replaced with “observation”. It is certainly the case that datasets observed by agents are not usually in the form of their optimal decision. In a rational inattention context, however, the data observed by the agent is the *fundamental* target, and this target is indeed unrelated to the optimal action. The *canonical* target, which is generally related to the optimal action, is not given to the agent but is constructed by them as they solve their problem and process the incoming data as efficiently as possible. Thus the existence of a *representation* in the form optimal action plus noise is not suspicious, because it is exactly constructed by the agent to capture the most important aspects of available information. We therefore reject the first claim and suggest that the default position for rational inattention models should not include assumptions restricting the form of representations available to the agent.

The second claim is more complex, and in some ways it is an obviously true statement

- it would be foolish to argue that real economic agents have no restrictions to data processing other than cognitive capacity, particularly when the economic agent in question is a firm, composed of many individuals. However, assessment of this claim must take into account the context of the relevant abstraction proposed by the model. For the formulation laid down in Sims (2003) and Sims (2010), rational inattention models “do not subsume or claim to replace all previous economic models of costly information”; instead, the abstraction supposes the free availability of all underlying information so that any incompleteness of information is entirely due to inattention on the part of the agent. This is only straightforward in stylized examples and so some license must clearly be extended if the rational inattention approach is to be used for more complex situations and agents.

In designating a person as rationally inattentive, we model the person as a finite capacity channel through which information flows. The abstraction of the model ignores purely psychological quirks related to information processing so that agent’s behavior can be considered through the lens of optimization.²⁴ Since all relevant data is assumed to be freely available to the agent, imposition of the independence assumption would amount to the imposition of such a psychological quirk. Justification of such an assumption would presumably need to be done on a case-by-case basis, rather than as a general rule for rational inattention models. Ultimately, other frameworks, for example that of Woodford (2014), may be more natural for problems in which these issues represent serious concerns.

Firms, on the other hand, are composed of many individuals and decision making processes are often complex. It seems plausible that in designating a firm as rationally inattentive, what we mean is that the firm’s operations *generally* are conceived as a finite capacity channel through which information flows, as individual managers consult a variety of information sources to make a myriad of decisions. For the price-setting example, the in-depth study of firm behavior in Zbaracki et al. (2004) suggests something like this. In this case, would it not be reasonable to assume that firms have one group dedicated to understanding

²⁴ This point is made clear in footnote 1 of Sims (2003).

idiosyncratic shocks and a second group dedicated to understanding aggregate shocks, so that the independence assumption would be justified? We argue not. The key consideration for us is that these restrictions result in suboptimal outcomes as firms process costly but irrelevant information, and yet there is no particular barrier that prevents firms from structuring their decision-making in any way they please. While of course the actual structure of firms is influenced by many considerations (for example economies of scale), it does not seem clear that there is any general justification for the independence assumption. Despite this, it is undeniable that complex agents surely face some restrictions on the way that they process information, and in specific cases there may be evidence indicating some particular deviation from the baseline model. Therefore, while we advocate for applying the rational inattention model without *ad hoc* restrictions, we do not reject the second claim altogether.

8 Extension: dynamic RI-LQG tracking problems

The signal extraction and tracking problems can usually be extended as dynamic problems in a straightforward manner, especially in the LQG case where the target follows the linear transition law $\alpha_t = T\alpha_{t-1} + \eta_t$ with $\eta_t \sim N(0, \Omega)$. This nests the static case when $T = 0$. The dynamic signal extraction problem can be solved recursively by the Kalman filter, and a key feature is that at each stage the solution is given by a conditional expectation, which we will denote $a_{t|t} = E[\alpha_t | y^t]$, where y^t collects the (exogenously given) observations $\{y_\tau\}_{\tau \leq t}$. To construct a dynamic tracking problem, we again shed the exogenously imposed observation vector y_t , and we also now assume that the agent discounts the future at rate β , so that the problem is:

$$\min_{\{a_{t|t}\}_{t \geq 0}} E \left[\sum_{t=0}^{\infty} \beta^t d(\alpha_t, a_{t|t}) \mid \mathcal{I}_0 \right]$$

along with the transition equation. By introducing an information constraint at each time period we can construct a rational inattention problem and proceed in a similar fashion to the static case. Sims (2003) and Sims (2010) show that at each stage it will be optimal to set $a_{t|t} = E[\alpha_t | \mathcal{I}_t]$, and we have $I(\alpha_t, a_{t|t} | \mathcal{I}_{t-1}) = \frac{1}{2} (\log_b |P_t| - \log_b |P_{t|t}|)$ where $P_t = Var(\alpha_t | \mathcal{I}_{t-1})$ and $P_{t|t} = Var(\alpha_t | \mathcal{I}_t)$. By using the transition law to derive the predicted covariance matrix P_{t+1} , we can recursively define the dynamic RI-LQG tracking problem.

Definition 18: the dynamic RI-LQG tracking problem, denoted (W, a_-, P_-, T, Ω) , is:

$$\begin{aligned} \min_{P_{t|t}} \quad & tr(WP_{t|t}) + \lambda(\ln |P_-| - \ln |P_+|) + \beta\lambda \ln |P_{t+1}| & (17) \\ \text{s.t.} \quad & \alpha | \mathcal{I}_- \sim N(a_t, P_t) \\ & P_{t+1} = TP_{t|t}T' + \Omega \\ & P_{t|t} \geq 0 \\ & P_t - P_{t|t} \geq 0 \end{aligned}$$

Because the prior covariance matrix for time $t+1$, P_{t+1} , depends on the posterior at time t , the dynamic problem features linkages across time that do not appear in the static problem: decreasing uncertainty today makes it less costly to achieve a given level of uncertainty tomorrow. It is immediate that for any given marginal cost of attention, more attention will be allocated in the dynamic problem than would be allocated for the same problem with $T = 0$. Another feature of the dynamic problem that does not appear in the static problem is that achieving equilibrium, if one exists, may take several periods. This is because in the static problem, the prior was generally equal to the unconditional distribution so repetitions were generally identical, whereas in the dynamic problem the prior evolves from period to period and equilibrium is only reached when the prior P_t is equal to the predicted covariance P_{t+1} constructed using the optimal posterior $P_{t|t}$.

It is easy to check that the first order condition for the time t iteration of the dynamic problem is:

$$P_{t|t}^{-1} = W/\lambda + \beta T' P_{t+1}^{-1} T \quad (18)$$

Although the matrix $P_{t|t}$ that solves this equation generally cannot be given explicitly, it is not too hard to compute it numerically. However, as before, the first order condition only solves the problem if the constraints are not binding. Similar to the static case, while the first positive semidefiniteness constraint will always be satisfied, the no-forgetting constraint will usually be binding in the dynamic case. To understand why, first recall that Lemma 4, associated with the static problem, suggests that the rank of the solution, will always be less than the rank of the loss matrix; this result is not strictly true in the dynamic case, although the intuition is still usually valid. Now, in order to map dynamic targets into the form required by Definition 18, an augmented target usually has to be constructed in order to satisfy the requirement of a linear first order transition equation. Thus although the loss function is defined in terms of the original target, the loss matrix W is defined in terms of the augmented target; this generally introduces rows and columns of zeros, and the result is that the loss matrix for most problems is not full rank.

For example, consider tracking an AR(2) target $\alpha_t^o = \phi_1 \alpha_{t-1}^o + \phi_2 \alpha_{t-2}^o + \eta_t^o$. In order to put this into a form amenable to Definition 18, we write:

$$\alpha_t = \begin{bmatrix} \alpha_t^o \\ \alpha_{t-1}^o \end{bmatrix}, \quad T = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}, \quad \eta_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \eta_t^o, \quad W = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Thus the no-forgetting constraint will almost always bind for dynamic problems and so the first order condition will not provide the solution. Unfortunately, the method of Theorem 1 does not immediately help, because in the dynamic case the eigenvectors of $P_{t|t}$ cannot be completely decoupled from the eigenvalues. Finding a fully general analytic solution for

the dynamic RI-LQG tracking problem remains an open problem. Despite this, if a solution is found, for example numerically, many of the results derived in this paper can be applied, since they only depend on the conditional Gaussianity of prior and posterior. Analysis can still proceed based on the generalized eigenvalue problem associated with the matrix pencil (P_+, P_-) , where the generalized eigenvalues and left eigenvectors can be found by applying simultaneous diagonalization. Thus Proposition 6 is still valid and the canonical synthetic target is still well-defined, as is the rank of the solution and the associated definitions of information capacity allocations. The construction of the action as a projection on an appropriately defined vector space also continues to hold, as does the concept of feasible and proper representations. Of course, results that depend on the specific relation of the generalized eigenvectors and eigenvalues to the loss matrix, especially Theorems 1 and 2, do not extend to the dynamic case.

In the special case of a one-dimensional target, it is sometimes possible to construct the solution analytically, and for the important class of one-dimensional targets following an ARMA(p, q) process under a fixed capacity constraint, an analytic solution has been derived by Matejka et al. (2017). Although their setup is nominally different, their results can be stated in the terms introduced here; we describe only a few. First, the rank of the solution will always be $r = 1$, and so the no-forgetting constraint will always bind except possibly in the AR(1) case. Second, except for the AR(1) case, no solution admits a representation of the form $y_t^o = Z^o \alpha_t^o + \varepsilon_t^o$ where α_t^o is the one-dimensional ARMA(p, q) target, although in each case there still exists some $1 \times n$ matrix Z such that there exists a feasible representation $y_t = Z \alpha_t + \varepsilon_t$, where α_t is an augmented target constructed to satisfy the transition equation. Matejka et al. (2017) term this second result the “dynamic attention principle”.

The solution of Matejka et al. (2017) is very promising, but it does not apply to the general problem of Definition 18 in which agents must trade off between many target processes, and

there is no way to expand their method to multivariate series without imposing an *ad hoc* restriction like the independence assumption. One option is to solve the general problem numerically as was done in examples given in Sims (2003) and Sims (2010). Unfortunately, numerical optimization can prove difficult for even moderate sized systems. Instead, we advocate an approximation suggested by the first order conditions. It is not hard to see that a first order Taylor approximation in λ to the dynamic first order condition, around the point of perfect information ($\lambda = 0$), is equal to the static first order condition.²⁵ We suggest, then, that when the marginal cost of information is sufficiently low (or capacity is sufficiently high), iterated application of the static solution given in Theorems 1 and 2 along with the transition equation, starting from an arbitrary prior, will yield a good approximation of the full dynamic solution. Although we do not prove the result, it appears that such iterations always converge to an equilibrium. This can be a particularly attractive method because, in practice, most applications of RI-LQG problems have been associated with λ close to zero.²⁶ Finally, we note that for ARMA(p, q) targets, this approximation could also be justified by appealing to Proposition 7 of Matejka et al. (2017).

Of course this approximation only imperfectly captures the full solution to the general problem. Its key strength is that it takes into account intratemporal tradeoffs between target processes, while its key weakness is that it fails to take into account intertemporal tradeoffs: it ignores future benefits from reducing uncertainty today.²⁷ One outcome of this is it will tend to select higher levels of uncertainty than the analytic solution. For the same reason, it will also run into the no-forgetting constraint sooner (at a higher marginal cost or lower capacity) than would the analytic solution. In many cases of practical interest, however, this static approximation will be quite good, as we illustrate below.

²⁵ This approximation could alternatively be derived from an approximation around $\beta = 0$; it thus effectively imposes that individuals fully discount future uncertainty.

²⁶ Some care must be taken when considering the scale of λ , since it is actually the scale of λ *relative* to the eigenvalues of the loss matrix W (and, in some problems, potentially those of T and Ω) that matters.

²⁷ The approximation is not completely divorced from intertemporal issues since the prior, which influences the posterior in the static solution, evolves over iterations. It is also possible that the no-forgetting constraint may be imposed or lifted as the prior changes.

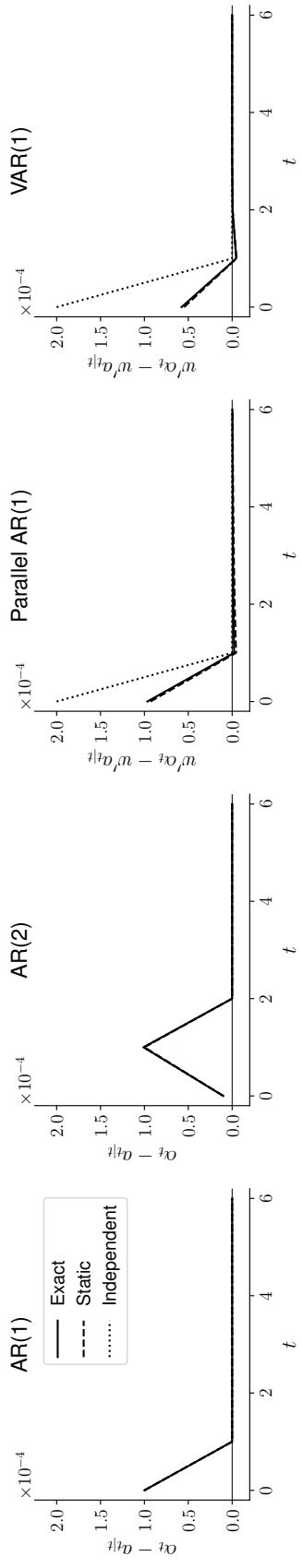


Figure 6: Dynamic unresponsiveness with $\lambda = 10^{-4}$

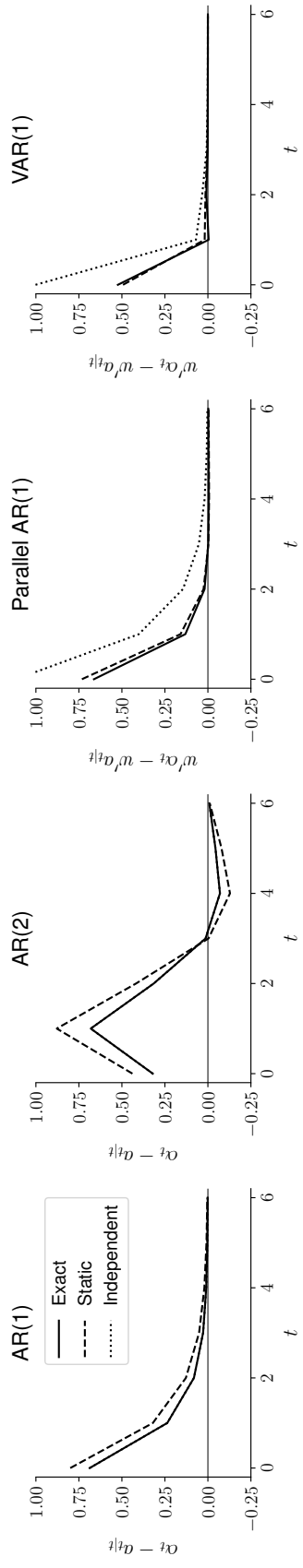


Figure 7: Dynamic unresponsiveness with $\lambda = 1$

		AR(1)		AR(2)		Parallel AR(1)			VAR(1)		
		κ	k_1	k_2	κ	k_1	k_2	κ	k_1	k_2	κ
$\lambda = 10^{-4}$	Exact	6.64	6.64	0.00	6.64	0.21	0.46	7.20	0.83	0.65	7.58
	Static	6.64	6.64	0.00	6.64	0.21	0.46	7.20	0.83	0.65	7.58
	Ind.	–	–	–	–	6.64	6.64	13.29	6.64	6.64	13.29
$\lambda = 1$	Exact	0.27	0.50	0.06	0.51	0.12	0.38	0.83	0.66	0.33	1.03
	Static	0.16	0.36	0.09	0.36	0.13	0.31	0.72	0.56	0.43	1.01
	Ind.	–	–	–	–	0.27	0.61	0.88	0.76	0.22	0.96

Table 1: Attention allocations in dynamic examples for the exact solution, the static approximation, and an approximation based on the independence assumption.

8.1 Illustration: goodness of approximation

In Table 1 we present a few examples demonstrating the accuracy of the static approximation (“Static”) compared to the exact solution of the problem (“Exact”) or the solution under the independence assumption (“Ind.”), and we consider four examples: an AR(1) process, an AR(2) process, a model with two separate (“parallel”) AR(1) processes, and a bivariate VAR(1) process. In the former two cases, there is only a single variable to track, while in the latter two cases, there are two variables to track and we assume that the agent wishes to track the sum of the two variables.²⁸ We apply the fixed marginal cost of attention formulation of the problem with $\beta = 0.99$, and consider two different costs: $\lambda = 10^{-4}$ and $\lambda = 1$. We report the approximate information capacity allocated to processing the i -th element of the fundamental target, k_i , as well as the total information processed, κ .²⁹

As could be expected, when the cost of information is low, the static approximation is practically identical to the full dynamic solution while when the cost is high, its performance degrades. Imposition of the independence assumption drives the solution away from the exact dynamic solution regardless of the information cost, but there is an especially marked difference when the information cost is low. This is because, as described above in the

²⁸ Since there is only a single variable in the AR(1) and AR(2) cases, the independence assumption does not change the problem there.

²⁹ In the AR(1) case, there is only a single variable of interest and $k_1 = \kappa$. In the AR(2) case, k_1 refers to the approximate information capacity allocated to processing the contemporaneous value of the target while k_2 refers to the approximate information capacity allocated to processing the lagged value.

prices example, a solution to the general problem will often reduce uncertainty only about certain relevant linear combinations of the target, whereas a solution under the independence assumption will reduce uncertainty about each target element separately.

The information capacity allocations in Table 1 provide one way of assessing the goodness of the static approximation, but an alternative method is to directly examine the final effect on the agent’s action. We will consider how a rationally inattentive agent responds to a one unit innovation in each of the four example models. In particular, in Fig. 6 and Fig. 7, we plot the difference between the true impulse response function of the model and the action taken by the agent.³⁰ We label this difference the “unresponsiveness” of the agent, because it captures the portion of the impulse that the agent does not respond to. These figures provide more evidence that the static approximation is very good when the marginal cost of information is close to zero, and is often still quite good when the marginal cost of information is relatively large.

One final interesting characteristic of these results can be found in the “Exact” solutions to the AR(2) problem. For this problem, the agent is only concerned with tracking the contemporaneous variable, and the lagged variable is associated with zero weight in the loss matrix. When the cost of information is low ($\lambda = 10^{-4}$), the approximate capacity allocated to processing the lagged variable is $k_2 \approx 0$. It might seem counterintuitive that as attention becomes more costly there is an *increase* in the approximate capacity allocated to processing the lagged variable, with $k_2 \approx 0.06$ when $\lambda = 1$. The reason for this can be found in the first order condition to the dynamic problem: when λ is small, the effects of transitional dynamics are dwarfed by the effect of W/λ , whereas when λ is larger they can become important. Ultimately, the effect of transitional dynamics can induce the agent to pay attention to variables that receive zero weight in the loss matrix as long as they help predict the variables that are of interest. This is why the agent pays more attention to the

³⁰ In all cases, we show the unresponsiveness to the object of interest; this does not necessarily correspond to one of the fundamental targets. For the AR(1) and AR(2) models we show the responsiveness to the contemporaneous value of process α_{1t} , while for the multivariate series we show the unresponsiveness $w'\alpha_t$.

second component as attention becomes more costly, and it is exactly an illustration of what Matejka et al. (2017) refer to as the “dynamic attention principle”.

9 Conclusion

In this paper, we describe the optimal allocation of attention by agents interested in tracking multiple economic shocks each of which provides valuable information subject to a limited ability to process incoming data. The key insight is that by constructing a transformation of the economic shocks, we can simplify the problem, facilitate the solution, and ease the interpretation of a wide variety of results. The transformed “canonical” shocks introduce a decoupling that captures the independent aspects of the economic shocks as they matter to the agent. Even in a complex multivariate setting with correlation between economic shocks, for each of the canonical shocks the agent acts as a simple Bayesian updater, giving some weight to the imperfectly processed incoming data while retaining some weight on their prior. We show how these canonical shocks define a representation of the incoming data that provides insight into how a rationally inattentive agent processes information. Throughout, we carefully examine the similarities and differences between the rational inattention problem and the classical signal extraction problem.

We apply our solution method to solve the static version of the rational inattention price-setting problem, and find a richer set of equilibrium behavior than previously known, including multiple equilibria and a social cost of increased attention by agents. We show how our framework can be used to help inform rational inattention modeling decisions, and this leads us to argue that the “independence assumption”, often employed in rational inattention models to make the model tractable, imposes unjustifiable restrictions on agents. At the same time, the solution method developed in this paper all but eliminates the need for such an assumption in the static case. Finally, we describe how our solution to the static

problem can be used to approximate the solution to the dynamic problem, and moreover show that this approximation is quite good in many cases of practical interest.

10 Appendices

10.1 Appendix A: Proofs

10.1.1 Proof of Property 6

Simultaneously diagonalize $P_- = S'IS$ and $P_+ = S'NS$ as described in Lemma 1. Then:

$$\begin{aligned} I(X, Y \mid \mathcal{I}_-) &= \frac{1}{2} (\log_b |P_-| - \log_b |P_+|) \\ &= \frac{1}{2} (\log_b |S'IS| - \log_b |S'NS|) \\ &= \frac{1}{2} (\log_b |I| - \log_b |N|) \\ &= \frac{1}{2} \log_b |N^{-1}| \\ &= \frac{1}{2} \sum_{i=1}^n \log_b \frac{1}{n_i} \end{aligned}$$

10.1.2 Proof of Lemma 1

See Theorem 7.6.4 of Horn and Johnson (2012).

10.1.3 Proof of Lemma 2

This is a straightforward application of Lemma 1.

10.1.4 Proof of Theorem 1

Throughout this proof, the matrices L , M , V , D , and Q are as defined in Lemma 2. We note at the outset that we can assume without loss of generality that P_+ is positive definite, since if it were not the objective function would grow without bound.

Ignoring the no-forgetting constraint, simultaneously diagonalize P_+^{-1} and P_-^{-1} as:

$$P_+^{-1} = X' \Delta X$$

$$P_-^{-1} = X' I X$$

where $X = Z' M$ with $Z \Delta Z' = L' P_+^{-1} L$, and denote $\Delta = \text{diag}(\{\delta_i\}_{i=1}^n)$. Because P_+ is full rank, Δ is nonsingular and we can define $N = \Delta^{-1} = \text{diag}(\{n_i\}_{i=1}^n)$ where $n_i = 1/\delta_i$.

Denoting the objective function as \mathcal{O} , we can rewrite it using the above decomposition and applying Property 6 as:

$$\begin{aligned} \mathcal{O} &= \text{tr}(W P_+) + \lambda \sum_{i=1}^n \ln \frac{1}{n_i} \\ &= \text{tr}(W L Z N Z' L') - \lambda \sum_{i=1}^n \ln n_i \\ &= \text{tr}(Z' V Z N) - \lambda \sum_{i=1}^n \ln n_i \end{aligned}$$

Notice that the matrix of eigenvectors, Z , appears only in the first term. A standard result is that minimizing the first term over unitary matrices Z yields $Z = Q$ (recall that $Q D Q' = V$), for any matrix N . Thus the optimal Z contains the eigenvectors of $V = L' W L$. This also implies that $X = S = Q' M$.

This allows us to further simplify the objective function:

$$\begin{aligned} \mathcal{O} &= \text{tr}(Q' V Q) - \lambda \sum_{i=1}^n \ln n_i \\ &= \text{tr}(Q' (Q D Q') Q N) - \lambda \sum_{i=1}^n \ln n_i \\ &= \text{tr}(D N) - \lambda \sum_{i=1}^n \ln n_i \\ &= \sum_{i=1}^n d_i n_i - \lambda \sum_{i=1}^n \ln n_i \end{aligned}$$

We can also use the simultaneous diagonalization to simplify the no-forgetting positive semidefiniteness constraint. First, note that if $P_- - P_+ \geq 0$ if and only if $P_+^{-1} - P_-^{-1} \geq 0$. Then from above, $P_+^{-1} - P_-^{-1} = S'(\Delta - I)S$, and this is positive semidefinite if and only if $\Delta - I \geq 0$. Since Δ is diagonal and $N = \Delta^{-1}$, this condition is satisfied if and only if $\delta_i \geq 1$ or $n_i \leq 1$ for $i = 1, \dots, n$.

With this, the objective and the constraint can be separated into n isolated problems, each of which is of the form:

$$\min_{n_i} d_i n_i - \lambda \ln n_i \quad \text{s.t. } n_i \leq 1$$

If $d_i > 0$, then this is a convex objective function with a linear inequality constraint, so the solution, denoted by n_i^+ , is characterized by the Kuhn-Tucker conditions. The first order condition yields $n_i = \lambda/d_i$, and the full solution is:

$$n_i^+ = \begin{cases} \lambda/d_i & \lambda \leq d_i \\ 1 & \text{otherwise} \end{cases}$$

If $d_i = 0$, then the problem is $\min_{n_i} -\lambda \ln n_i$, and the solution sends $n_i \rightarrow \infty$, so that the constraint is binding and $n_i^+ = 1$.

Defining $\delta_i^+ = 1/n_i^+$ and $\Delta^+ = \text{diag}(\{\delta_i^+\}_{i=1}^n)$, we have solved for the optimal S and Δ that define P_+^{-1} , and in particular:

$$P_+^{-1} = S' \Delta^+ S$$

$$P_+ = R N^+ R'$$

Finally, if $d_i \geq \lambda \forall i$, then $\Delta^+ = D/\lambda$ and:

$$\begin{aligned}
 P_+^{-1} &= M'Q\Delta^+Q'M \\
 &= M'Q(D/\lambda)Q'M \\
 &= M'L'(W/\lambda)LM \\
 &= W/\lambda
 \end{aligned}$$

10.1.5 Proof of first Corollary to Theorem 1

Let $W = ww'$ and define $q = \frac{1}{\|L'w\|}L'w$. Then:

$$\begin{aligned}
 P_+^{-1} &= S'\Delta^+S \\
 &= P_-^{-1} + S'(\Delta^+ - I)S \\
 &= P_-^{-1} + (\delta_1^+ - 1)M'q_1q_1'M \\
 &= P_-^{-1} + (\delta_1^+ - 1)\frac{1}{\|L'w\|^2}W
 \end{aligned}$$

From above, we have:

$$\begin{aligned}
P_+^{-1} &= P_-^{-1} + (\delta_1^+ - 1)M'q_1q_1'M \\
&= M' [I + (\delta_1^+ - 1)q_1q_1'] M \\
P_+ &= L [I + (\delta_1^+ - 1)q_1q_1']^{-1} L' \\
&= L \left[I^{-1} - I^{-1}q_1 \left((\delta_1^+ - 1)^{-1} + q_1'I^{-1}q_1 \right)^{-1} q_1'I^{-1} \right] L' \\
&= L \left[I - \left((\delta_1^+ - 1)^{-1} + 1 \right)^{-1} q_1q_1' \right] L' \\
&= L \left[I - \left(\frac{\delta_1^+}{\delta_1^+ - 1} \right)^{-1} q_1q_1' \right] L' \\
&= P_- - \frac{\delta_1^+ - 1}{\delta_1^+} \frac{1}{\|L'w\|^2} P_- W P_- \\
&= P_- - (1 - n_1^+) \frac{1}{\|L'w\|^2} P_- W P_-
\end{aligned}$$

10.1.6 Proof of second Corollary to Theorem 1

We want to show that $s_i'(P_+ - n_i^+ P_-) = 0$ for each pair (s_i', n_i^+) .

From Lemma 2 we have $P_- = RIR'$, and from Theorem 1 we have $P_+ = RN^+R'$. Since $R = S^{-1}$, then $s_i'R$ is equal to a row vector with each element equal to zero except for the i -th element which is equal to 1, and so $s_i'RN^+ = n_i^+ s_i'R$.

$$\begin{aligned}
s_i'(P_+ - n_i^+ P_-) &= s_i'(RN^+R' - n_i^+ RIR') \\
&= (n_i^+ s_i'RR' - n_i^+ s_i'RR') \\
&= 0
\end{aligned}$$

10.1.7 Proof of Theorem 2

Since Definition 1 is valid for the fixed capacity problem, except with $\lambda^* = 2 \ln(b)\lambda$ interpreted as a Lagrange multiplier, the solution in Theorem 1 is valid in this case, but we must also derive the value of the Lagrange multiplier at the solution. To do so, note that the associated constraint is $\frac{1}{2}(\log_b |P_-| - \log_b |P_+|) \leq \kappa$ and, as in the proof of Theorem 1, we can rewrite it as:

$$\frac{1}{2} \sum_{i=1}^n \log_b \delta_i^+ \leq \kappa$$

In any solution, all processing capacity will be used, so that this constraint will hold with equality. Define r such that $d_i > 0$ for $i = 1, \dots, r$ and $d_i \leq \lambda$ for $i = r + 1, \dots, n$. Recall from Theorem 1 that $\delta_i^+ = 1$ for $i > r$, and so the constraint is:

$$\begin{aligned} \sum_{i=1}^r \log_b \delta_i^+ &= 2\kappa \\ \log_b \prod_{i=1}^r \frac{d_i}{\lambda} &= 2\kappa \\ \lambda^r &= b^{-2\kappa} \prod_{i=1}^r d_i \\ \lambda &= \left[b^{-2\kappa} \prod_{i=1}^r d_i \right]^{\frac{1}{r}} \end{aligned}$$

Since the choice of r depends on λ , we can compute r in the following way. Initialize $r = n$. First, compute the λ associated with r . If $d_i > \lambda$ for $i = 1, \dots, r$, then this is the solution. If $\exists d_i \leq \lambda$ with $i \leq r$, then set $r = r - 1$ and repeat these steps.

Notice that if $r = 1$, then $\lambda = 2^{-2\kappa} d_1$. As long as $\kappa > 0$ and $d_1 > 0$ (and recall that d_1 is the largest eigenvalue, so only in completely degenerate problems will $\kappa = 0$ or $d_1 = 0$), we will have $d_1 > \lambda$. Thus, except for degenerate problems, it will always be optimal to have $r \geq 1$.

Finally, for $i \leq r$, we have:

$$\begin{aligned}\delta_i^+ &= d_i \lambda^{-1} \\ &= b^{\frac{2\kappa}{r}} d_i \left[\prod_{j=1}^r d_j \right]^{-\frac{1}{r}}\end{aligned}$$

Taking logs, we define:

$$\kappa_i \equiv \frac{1}{2} \log_b \delta_i^+ = \frac{\kappa}{r} + \log_b \left[\frac{\sqrt{d_i}}{\prod_{j=1}^r \sqrt{d_j}^{1/r}} \right]$$

For $i > r$, we have $\delta_i^+ = 1$, so $\kappa_i = \frac{1}{2} \log_b 1 = 0$.

10.1.8 Proof of Corollary to Theorem 2

Part (a):

We want to show that $\frac{\partial \lambda}{\partial \kappa} < 0$. The only difficulty is accounting for the fact that r as a function of κ acts like a step function.

Our first step is to notice that if the change in κ does not change r , then we have:

$$\frac{\partial \lambda}{\partial \kappa} = \frac{-2\kappa}{r} (b^{(-2\kappa/r)-1}) \left[\prod_{i=1}^r d_i \right]^{1/r} < 0$$

Our second step is to show that if r is nondecreasing in κ (i.e. r and κ move (weakly) together), the result still holds. To see this, consider the two terms of λ separately.

- a. It is easy to see that $\frac{\partial 2^{-2\kappa/r}}{\partial \kappa} < 0$ and $\frac{\partial 2^{-2\kappa/r}}{\partial r} < 0$.
- b. The second term is the geometric mean of (d_1, \dots, d_r) , and by assumption we have $d_1 \geq d_2 \geq \dots \geq d_r \geq \dots \geq d_n$. An increase in r will therefore introduce into the geometric mean terms that are no larger than any of the existing terms; similarly, a

decrease in r will remove only the smallest existing terms. Thus, the term as a whole is nonincreasing in r . Since this term is independent of κ , we have our result.

Our final step is to show that r is nondecreasing in κ . This follows directly from the first step, above, and the algorithm for computing r . Consider an increase in κ . At any iteration of the algorithm, the proposed value for λ will be smaller than it was under the original value of κ , and so while the algorithm may terminate earlier, it certainly will not terminate later. The reverse is true for a decrease in κ . This yields the result.

Part (b):

See the last paragraph of the proof to part (a).

10.1.9 Proof of Lemma 3

Part (a): This follows directly from $\beta_c = S\alpha$ and $\alpha | \mathcal{I}_+ \sim N(a_+, P_+)$.

Part (b): This follows directly from $\beta_c = S\alpha$ and $\alpha | \mathcal{I}_- \sim N(a_-, P_-)$.

Part (c):

$$\begin{aligned}
E[(\beta_c - b_{c,+})' D(\beta - b_{c,+}) | \mathcal{I}_-] &= E[(\alpha - a_+)' S' D S(\alpha - a_+) | \mathcal{I}_-] \\
&= E[(\alpha - a_+)' M' Q D Q' M(\alpha - a_+) | \mathcal{I}_-] \\
&= E[(\alpha - a_+)' M' V M(\alpha - a_+) | \mathcal{I}_-] \\
&= E[(\alpha - a_+)' W(\alpha - a_+) | \mathcal{I}_-]
\end{aligned}$$

Part (d): This follows from parts (a) and (b) along with the fact that (P_+, P_-) and (N_+, I) share generalized eigenvalues. Alternatively, this follows from Property 3.

Part (e): This follows because $Var(\beta_c | \mathcal{I}_+) = N_+$ is a diagonal matrix.

10.1.10 Proof of Lemma 4

Part (a): The quantity r is the integer such that $n_r = \lambda/d_r < 1$ but $n_r = \lambda/d_{r+1} \geq 1$. Thus we have $n_i^+ > 1$ for $i = 1, \dots, r$ and $n_i^+ = 1$ for $i = r+1, \dots, n$, so that $\text{rk}(I - N^+) = r$. Then, since S is nonsingular, we have $\text{rk}(P_- - P_+) = \text{rk}(SP_-S' - SP_+S') = \text{rk}(I - N_+) = r$.

Part (b): If $\text{rk}(W) = \ell$, then each of $d_n, d_{n-1}, \dots, d_{\ell+1}$ must equal zero, and for any i such that $d_i = 0$, it must also be that $n_i^+ = 1$. Then $r = \text{rk}(I - N^+) \leq \ell = \text{rk}(W)$.

Part (c): This follows directly from the definition of δ_i^+ in Theorem 1.

Part (d): This follows from Lemma 3 part (e), since for each i such that $n_i^+ = 1$, we have $I(\beta_{i,c}, b_{i,c,+} \mid \mathcal{I}_-) = 0$.

Part (e):

In the fixed κ formulation, suppose that $r = 1$. Then $\lambda = b^{-2\kappa}d_1$, so that $d_1 > \lambda$. Thus the algorithm of Theorem 2 will always terminate at $r = 1$ if it did not terminate earlier.

In the fixed λ formulation, set $\lambda = d_1 + 1$. Then $r = 0$.

10.1.11 Proof of Lemma 5

If P_- is diagonal, then the Cholesky factor L is also diagonal. Along with W diagonal, this implies that $V = L'WL$ is diagonal, so that the matrix of eigenvectors Q is equal to the identity.

Then $S = Q'M = M$ and $P_+ = RN^+R' = LN^+L' = N^+P_-$. Rearranging, we get $(N^+)^{-1} = P_-P_+$. Rearranging, we get $\frac{1}{n_i^+} = \frac{P_{ii,-}}{P_{ii,+}}$.

10.1.12 Proof of Lemma 6

Let $\hat{\beta}_{i,c} = \beta_{i,c} + \varepsilon_{i,c}$ with $\varepsilon_{i,c} \sim N(0, (1/n_i^+ - 1)^{-1})$, as in the Lemma. Recall from Lemma 3 that $E[\beta_{i,c} | \mathcal{I}_-] = b_{i,c,-}$ and $\text{Var}(\beta_{i,c} | \mathcal{I}_-) = 1$. Then standard signal extraction formulas imply $b_{i,c,+} = b_{i,c,-} + K_c(\hat{\beta}_{i,c} - b_{i,c,-})$ where:

$$K_c = (1 + (1/n_i^+ - 1)^{-1})^{-1} = (1 - n_i^+)$$

Plugging this in yields the result.

10.1.13 Proof of Theorem 3

This follows directly from Definition 2 and Lemma 6.

10.1.14 Proof of Theorem 4

Let O_+ solve the B -transformed problem, and recall that we have B nonsingular. Now consider the objective function of the reference problem:

$$\begin{aligned} \mathcal{O} &= \text{tr}(WP_+) + \lambda(\log |P_-| - \log |P_+|) \\ &= \text{tr}(B'(B')^{-1}WB^{-1}BP_+) + \lambda(\log |BP_-B'| - \log |BP_+B'|) \\ &= \text{tr}(VBP_+B') + \lambda(\log |O_-| - \log |BP_+B'|) \end{aligned}$$

By considering $P_+ = B^{-1}O_+B'^{-1}$, it is clear that if O_+ is optimal for the B -transformed objective function, P_+ will be optimal for the reference problem, as long as the constraints are the same. To see that they are the same, notice that since B is nonsingular, $O_+ \geq 0 \iff P_+ \geq 0$ and $P_- - P_+ \geq 0 \iff B(P_- - P_+)B' = O_- - O_+ \geq 0$.

10.1.15 Proof of Lemma 7

This follows directly from the fact that the product of nonsingular matrices is nonsingular.

10.1.16 Proof of Lemma 8

This follows directly from Lemma 7 along with Definitions 8, 9, and 10

10.1.17 Proof of Lemma 9

This tautology follows directly from the definition of a_+ as a conditional expectation.

10.1.18 Proof of Lemma 10

Part (a):

Since $\text{rk}(Z) = m$, we have $r = \text{rk}(P_- - P_+) \leq \text{rk}(\Lambda^{-1}) \leq m$

Part (b):

Since $\text{rk}(Z) = m$ and $\text{rk}(\Lambda^{-1}) = m$, we also have $\text{rk}(Z'\Lambda^{-1}Z) = m$, but $\text{rk}(Z'\Lambda^{-1}Z) = \text{rk}(P_- - P_+) = r$.

10.1.19 Proof of Theorem 5

Given the innovation representation $v_+ = Z\alpha + \varepsilon - Za_-$ where $\varepsilon \sim N(0, \Lambda)$, we have that the posterior information set is $\mathcal{I}_+ = \mathcal{I}_- \cup \{v_+\}$, that $\alpha \mid \mathcal{I}_- \sim N(a_-, P_-)$, and that α and v_+ are jointly Gaussian. Theorem 5 is then simply a statement of the form of the conditional distribution of jointly Gaussian random vectors.

10.2 Appendix B: rationally inattentive price-setting

The rationally inattentive price-setting problem supposes that monopolistically competitive firms cannot pay perfect attention to the shocks that determine the optimal price for their differentiated good. In order to minimize the expected loss in (log quadratically approximated) prices, they choose how to allocate attention.

The basic rational inattention result of MW is that more attention is paid to shocks that are more important or more volatile. The former characteristic is captured in the loss function and the latter is captured by the shock's variance. In this paper, we have refined this result and shown that it should be applied to the "canonical synthetic shocks" rather than the original, or "fundamental", shocks. In MW, these two types of shocks are required to be identical, but in general this requirement imposes suboptimal behavior.

To understand why these shocks are treated differently, it is important to recall that a rationally inattentive agent has *access* to the complete data but, in optimally processing the information, they may choose only certain components of the data to pay attention to, with any remainder ignored. The agent's problem can be thought of as (1) selecting the components that matter to them, and (2) selecting the amount of attention to pay to each component. The canonical synthetic shocks provide exactly the decomposition that solves the former problem. It can be instructive to consider the inattentive agent as receiving a particular noisy signal of the data, but this is only appropriate in the context of the optimally chosen, canonical, shocks.

In the price-setting problem, even though the space of fundamental shocks is two-dimensional, the space of canonical synthetic shocks is only one dimensional, because as the agent processes information, there is only one variable that is of interest to them: the optimal price. The agent, assumed to have access to the complete data, only *processes* information about that relevant combination.

10.2.1 Setup

The firm's **profit function** is $\pi(P_{it}, P_t, Y_t, Z_{it})$, and the log quadratic approximation is $\tilde{\pi}(p_{it}, p_t, y_t, z_{it})$.

Aggregate demand is given by $Q_t = P_t Y_t$ or $q_t = p_t + y_t$

Under **perfect information**, optimal price-setting is:

$$p_{it}^{\diamond} = \frac{\hat{\pi}_{14}}{|\hat{\pi}_{11}|} z_{it} + \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|} q_{it} + \left(1 - \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|} z_{it}\right) p_t$$

In equilibrium $p_t = q_t$.

Under **rational inattention**, $p_{it}^* = E[p_{it}^{\diamond} | \mathcal{I}_t]$, and the objective is to minimize the loss in profits due to inattention. This can be written as:

$$\min \tilde{\pi}(p_{it}^{\diamond}, p_t, y_t, z_{it}) - \tilde{\pi}(p_{it}^*, p_t, y_t, z_{it})$$

and this can be simplified to:

$$\min \frac{|\hat{\pi}_{11}|}{2} E[(p_{it}^{\diamond} - p_{it}^*) | \mathcal{I}_t]$$

To ease notation, define $\zeta_0 = \frac{|\hat{\pi}_{11}|}{2}$, $\zeta_z = \frac{\hat{\pi}_{14}}{|\hat{\pi}_{11}|}$, and $\zeta_q = \frac{\hat{\pi}_{13}}{|\hat{\pi}_{11}|}$.

Guess and verify approach: guess that $p_t = \gamma q_t$. Then:

$$p_{it}^* = E[\zeta_z z_{it} + (\gamma + (1 - \gamma)\zeta_q) q_t | \mathcal{I}_t]$$

With this guess, the RI-LQG tracking problem can be defined by:

$$\alpha = \begin{bmatrix} z_{it} \\ q_t \end{bmatrix}, \quad W = \frac{|\hat{\pi}_{11}|}{2} ww', \quad w = \begin{bmatrix} w_z \\ w_q \end{bmatrix}$$

where $w_z = \zeta_z$ and $w_q = \gamma + (1 - \gamma)\zeta_q$.

10.2.2 Solutions

General solution In the general case, we proceed as usual. First we solve the fixed marginal cost problem, according to Theorem 1, and then we solve the fixed capacity problem, according to Theorem 2.

Fixed marginal cost

$$L'WL = \begin{bmatrix} \sigma_z & 0 \\ 0 & \sigma_q \end{bmatrix} \zeta_0 ww' \begin{bmatrix} \sigma_z & 0 \\ 0 & \sigma_q \end{bmatrix} = \zeta_0 \begin{bmatrix} \sigma_z w_z \\ \sigma_q w_q \end{bmatrix} \begin{bmatrix} \sigma_z w_z & \sigma_q w_q \end{bmatrix}$$

Let $q = \frac{L'w}{\|L'w\|} = \frac{1}{\sqrt{\sigma_z^2 w_z^2 + \sigma_q^2 w_q^2}} \begin{bmatrix} \sigma_z w_z & \sigma_q w_q \end{bmatrix}$ so that $\|q\| = \frac{\sqrt{\sigma_z^2 w_z^2 + \sigma_q^2 w_q^2}}{\sqrt{\sigma_z^2 w_z^2 + \sigma_q^2 w_q^2}} = 1$. Then we have:

$$L'WL = q \underbrace{\left(\zeta_0 (\sigma_z^2 w_z^2 + \sigma_q^2 w_q^2) \right)}_{\equiv d_1} q' = \underbrace{\begin{bmatrix} q & q^\perp \end{bmatrix}}_Q \underbrace{\begin{bmatrix} d_1 & 0 \\ 0 & 0 \end{bmatrix}}_D \underbrace{\begin{bmatrix} q' \\ q'^\perp \end{bmatrix}}_{Q'}$$

Now, $\delta_1^+ = \max\{d_1/\lambda, 1\}$, $n_1^+ = \min\{\lambda/d_1, 1\}$ and $\delta_2^+ = n_2^+ = 1$. The latter result implies the agent will never pay attention to a second component. This means that the rank of the solution will be at most $r = 1$, although it is possible that the agent will choose to not pay any attention at all ($r = 0$).

$$\text{Next, } S = Q'M = \begin{bmatrix} q' \\ q^{\perp'} \end{bmatrix} \begin{bmatrix} 1/\sigma_z & 0 \\ 0 & 1/\sigma_q \end{bmatrix} = \begin{bmatrix} w'/\|L'w\| \\ q^{\perp'}M \end{bmatrix} = \begin{bmatrix} s'_1 \\ s'_2 \end{bmatrix}.$$

Then we can compute the optimal posterior:

$$P_+^{-1} = S'\Delta^+S = \begin{bmatrix} s_1 & s_2 \end{bmatrix} \begin{bmatrix} \delta_1^+ & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s'_1 \\ s'_2 \end{bmatrix}$$

The weight matrix $K = I - P_+P_-^{-1} = R(1 - N^+)S$.

Of the left generalized eigenvectors s'_i of the matrix pencil (P_+, P_-) , only $s'_1 = w'/\|L'w\|$ is associated with a nonzero generalized eigenvalue. As described above, this vector is also a left eigenvector of the weight matrix K associated with the eigenvalue $1 - n_1^+$. Of course any scalar multiple of an eigenvector is also an eigenvector, so that we can write:

$$w'K = (1 - n_1^+)w'$$

The fundamental representation is:

$$y_f = \alpha_t + \varepsilon_f, \quad \varepsilon_f \sim N(0, \Lambda_f) \text{ with } \Lambda_f^{-1} = S'(\Delta_+ - I)S$$

This is not a feasible representation, because $S'(\Delta_+ - I)S$ is neither full rank nor can it be written in the form required by Definition 12. Instead, we can construct the canonical representation:

$$y_c = S\alpha_t + \varepsilon_c, \quad \varepsilon_c \sim N(0, \Lambda_c) \text{ with } \Lambda_c^{-1} = \Delta^+ - I$$

This is (as always) a feasible representation, but it is not proper since $\delta_2^+ - 1 = 0$ so that the error due to inattention has infinite variance for the second component. Thus we instead

use the reduced canonical representation:

$$y_r = s_1' \alpha_t + \varepsilon_r, \quad \varepsilon_r = \varepsilon_{1,c} \sim N(0, (\delta_1^+ - 1)^{-1})$$

the associated innovation representation is:

$$v_r = y_r - s_1' a_-$$

and the associated weight matrix is

$$K_r = s_1(1 + (\delta_1^+ - 1)^{-1})^{-1} = (1 - n_1^+) s_1$$

where $s_1' K_r = (1 - n_1^+)$ and $w' K_r = (1 - n_1^+) \|L' w\|$

Now we can construct the posterior:

$$\begin{aligned} a_+ &= a_- + K_r v_r \\ &= a_- + (1 - n_1^+) s_1 (s_1 \alpha_t + \varepsilon_r - s_1' a_-) \\ &= (I - (1 - n_1^+) s_1 s_1') a_- + (1 - n_1^+) s_1 s_1' \alpha_t + (1 - n_1^+) s_1 \varepsilon_r \end{aligned}$$

We can then construct the posterior of interest:

$$\begin{aligned}
p_{it}^* &= w' a_+ \\
&= w'(a_- + K_r v_r) \\
&= w' [(I - (1 - n_1)^+ s_1 s_1') a_- + (1 - n_1^+) s_1 s_1' \alpha_t + (1 - n_1^+) s_1 \varepsilon_r] \\
&= (w' - (1 - n_1)^+ w' s_1 s_1') a_- + (1 - n_1^+) w' s_1 s_1' \alpha_t + (1 - n_1^+) w' s_1 \varepsilon_r \\
&= (w' - (1 - n_1)^+ w') a_- + (1 - n_1^+) w' \alpha_t + (1 - n_1^+) \|L' w\| \varepsilon_r \\
&= (1 - (1 - n_1)^+) w' a_- + (1 - n_1^+) w' \alpha_t + (1 - n_1^+) \|L' w\| \varepsilon_r \\
&= n_1^+ p_- + (1 - n_1^+) p_{it}^\diamond + (1 - n_1^+) \|L' w\| \varepsilon_r
\end{aligned}$$

In this case, the prior is $a_- = 0$, and $p_{it}^\diamond = w' \alpha_t = w_z z_{it} + w_q q_t$. To find the aggregate price level, integrate over firms:

$$\begin{aligned}
p_t &= \int_I p_{it}^* di \\
&= \int_I (1 - n_1^+) w_z z_{it} di + \int_I (1 - n_1^+) w_q q_t di + \int_I (1 - n_1^+) \|L' w\| \varepsilon_r di \\
&= (1 - n_1^+) w_q q_t
\end{aligned}$$

Recall that our guess was $p_t = \gamma q_t$; this result confirms our guess, with $\gamma = (1 - n_1^+) w_q$. However, w_q is a function of γ , so the full solution yet requires solving for γ .

First, there is an equilibrium with $\gamma = 0$ if $\zeta_0(\sigma_z^2 \zeta_z^2 + \sigma_q^2 \zeta_q^2) \leq \lambda$. Since $w_q > 0$ regardless of γ , then $\gamma = 0$ requires $n_1^+ = 1$, i.e. it requires all agents to collect no information whatsoever. For this to be an equilibrium, it requires that $d_1 = \zeta_0(\sigma_z^2 w_z^2 + \sigma_q^2 w_q^2) \leq \lambda$. Since $w_z = \zeta_z$ and $w_q = \gamma + (1 - \gamma)\zeta_z$, requiring that $d_1 \leq \lambda$ when $\gamma = 0$ is equivalent to the condition given above.

We can find nonzero equilibria by solving for γ ; this is difficult to do analytically, but symbolic math software indicates that in the domain of interest, there is a unique real

solution along with a conjugate pair of complex solutions. Numerical solution methods find agreement with the unique real solution.

Fixed capacity

Given the solution to the fixed marginal cost formulation above, we can find the solution to the fixed capacity formulation of the problem by applying Theorem 2 to find the shadow marginal cost associated with capacity constraint.

From above, we know that $r \leq 1$, but because the problem is not degenerate we also know that $r \geq 1$. Thus it must be that $r = 1$ and so we can easily apply Theorem 2 to yield:

$$\lambda = 2^{-2\kappa} d_1$$

Therefore, $\delta_1^+ = \max\{d_1/\lambda, 1\} = \max\{2^{2\kappa}, 1\} = 2^{2\kappa}$, and so $n_1^+ = 2^{-2\kappa}$. Then:

$$\begin{aligned} \gamma &= (1 - n_1^+) w_q \\ &= (1 - 2^{-2\kappa})(\gamma + (1 - \gamma)\zeta_q) \\ &= \frac{(1 - 2^{-2\kappa})\zeta_q}{(1 - 2^{-1\kappa})\zeta_1 + 2^{-2\kappa}} \\ &= \frac{\zeta_q}{(2^{2\kappa} - 1)^{-1} + \zeta_q} \end{aligned}$$

The fixed capacity version is often easy to solve in the case $r = 1$ because it can tie down the posterior covariance matrix based only on the parameter κ . This was the case here, where we were able to substitute $n_1^+ = 2^{-2\kappa}$ whereas in the fixed marginal capacity case (and assuming an interior solution) we had $n_1^+ = \lambda/[\zeta_0(\sigma_z^2 w_z^2 + \sigma_q^2 w_q^2)]$.

Independence assumption We can also proceed here as usual, but using the alternate weight matrix $W_I = \text{diag}\{w_z^2, w_q^2\}$. We first apply Theorem 1 to solve the fixed marginal cost case and then apply Theorem 2 to solve the fixed capacity case.

Fixed marginal cost

$$L'WL = \begin{bmatrix} \sigma_z & 0 \\ 0 & \sigma_q \end{bmatrix} \zeta_0 \begin{bmatrix} w_z^2 & 0 \\ 0 & w_q^2 \end{bmatrix} \begin{bmatrix} \sigma_z & 0 \\ 0 & \sigma_q \end{bmatrix} = \zeta_0 \begin{bmatrix} \sigma_z^2 w_z^2 & 0 \\ 0 & \sigma_q^2 w_q^2 \end{bmatrix}$$

Then $Q = I$ and $d_i = \zeta_0 \sigma_i^2 w_i^2$ (where $i \in \{z, q\}$). As usual, we have:

$$\delta_i^+ = \max\{d_i/\lambda, 1\}$$

$$n_i^+ = \min\{\lambda/d_i, 1\}$$

The rank of the solution will be $r \in \{0, 1, 2\}$ because the agent may choose to pay attention to either, both, or neither of the components.

Now $S = Q'M = \begin{bmatrix} 1/\sigma_z & 0 \\ 0 & 1/\sigma_q \end{bmatrix} = \begin{bmatrix} s'_1 \\ s'_2 \end{bmatrix}$ with $s_i = \frac{1}{\sigma_i} e_i$ (and e_i is the i -th standard basis element). This implies that the canonical synthetic target is nothing more than a scaled version of the fundamental target; in fact, this was essentially the goal of the independence assumption.

We will abuse notation somewhat to now interpret the index as $i = 1, 2$ where $d_1 = \max\{d_z, d_q\}$ and $d_2 = \min\{d_z, d_q\}$; this accords with the usual practice of listing these generalized eigenvalues in nonincreasing order.

Then we can compute the optimal posterior:

$$P_+^{-1} = \begin{bmatrix} \delta_1^+/\sigma_1^2 & 0 \\ 0 & \delta_2^+/\sigma_2^2 \end{bmatrix}$$

or

$$P_+ = \begin{bmatrix} n_1^+ \sigma_1^2 & 0 \\ 0 & n_2^+ \sigma_2^2 \end{bmatrix}$$

The weight matrix is $K = I - P_+ P_-^{-1} = R(I - N^+)S$. Because R , S , and N^+ are diagonal, they commute, so that we have simply $K = (I - N^+)$. We can then write:

$$w'K = \begin{bmatrix} w_1 n_1^+ & w_2 n_2^+ \end{bmatrix}$$

The fundamental representation is:

$$y_f = \alpha_t + \varepsilon_f, \quad \varepsilon_f \sim N(0, \Lambda_f) \text{ with } \Lambda_f^{-1} = S'(\Delta^+ - I)S$$

This representation is feasible, because S and Δ^+ are diagonal and so the matrix will either be full rank or can be written as required by Definition 12. The associated innovation representation is:

$$v_f = y_f - a_-$$

We could still construct the canonical or reduced canonical representations in this case, although it is unnecessary for computing the action.

The action is:

$$\begin{aligned}
a_+ &= a_- + K v_f \\
&= (I - K)a_- + K y_f \\
&= (I - K)a_- + K \alpha_t + K \varepsilon_f \\
&= \begin{bmatrix} n_1^+ a_{1,-} + (1 - n_1^+) (\alpha_{1,t} + \varepsilon_{1,f}) \\ n_2^+ a_{2,-} + (2 - n_2^+) (\alpha_{2,t} + \varepsilon_{2,f}) \end{bmatrix}
\end{aligned}$$

and the posterior of interest is:

$$p_{it}^* = w' a_+ = \sum_{j \in \{z, q\}} (w_j n_j^+ a_{j,-} + w_i (1 - n_i^+) (\alpha_{i,t} + \varepsilon_{i,f}))$$

As before, the prior is $a_- = 0$; aggregating over firms yields:

$$p_t = \int_I p_{it}^* di = (1 - n_q^+) w_q q_t$$

Note that this is almost identical to the result in the general case, except that here the generalized eigenvalue n_q^+ is specific to the aggregate demand shock, whereas in the general case it corresponded to the synthetic shock the combined both the idiosyncratic and aggregate shocks.

Recall that $n_q^+ = \min\{\lambda/d_q, 1\}$ with $d_q = \zeta_0 \sigma_q^2 w_q^2$, and $w_q = \gamma + (1 - \gamma)\zeta_q$. Now, we can combine these results to solve for the equilibrium value of γ .

First, there is an equilibrium with $\gamma = 0$ if $\zeta_0 \sigma_q^2 \zeta_q^2 \leq \lambda$. This always corresponds to an agent paying no attention to aggregate conditions; however, the agent may still pay some attention to idiosyncratic conditions.

Otherwise, we can solve for γ :

$$\begin{aligned}\gamma &= (1 - n_q^+)w_q = \left(1 - \frac{\lambda}{\zeta_0\sigma_q^2w_q^2}\right)w_q \\ &= \dots \\ &= \frac{\pm\sqrt{\lambda(1 - \zeta_q) + \zeta_0\sigma_q^2/4} + (0.5 - \zeta_q)\sqrt{\zeta_0\sigma_q^2}}{(1 - \zeta_q)\sqrt{\zeta_0\sigma_q^2}}\end{aligned}$$

Both of these roots may be valid and may, moreover, coexist with the $\gamma = 0$ equilibrium, so that there may be as many as three equilibria.

Fixed capacity

As before, we know that in the fixed capacity case $r \geq 1$. This means that we have:

$$\lambda = \begin{cases} (2^{-2\kappa}d_1d_2)^{1/2} & r = 2 \\ \lambda = 2^{-2\kappa}d_1 & r = 1 \end{cases}$$

and recall that we have stipulated $d_1 \geq d_2$, where $d_i = \sigma_i^2w_i^2$. We have $r = 1$ if $d_2 \leq \lambda = 2^{-2\kappa}d_1$ and $r = 2$ otherwise; i.e. the agent will pay attention to both idiosyncratic and aggregate conditions as long as the canonical loss weights are relatively close together, and will pay attention to only one component if they are far enough apart.

Now we can compute δ_i^+ :

$$\begin{aligned}\delta_1^+ &= \max\left\{\frac{d_1}{\lambda}, 1\right\} = \begin{cases} 2^{2\kappa} & d_2 \leq \lambda \\ 2^\kappa \left(\frac{d_1}{d_2}\right)^{1/2} & d_2 > \lambda \end{cases} \\ &= \begin{cases} 2^{2\kappa} & d_2 \leq 2^{-2\kappa}d_1 \\ 2^\kappa \frac{\sigma_1w_1}{\sigma_2w_2} & d_2 > 2^{-2\kappa}d_1 \end{cases}\end{aligned}$$

$$\delta_2^+ = \max \left\{ \frac{d_2}{\lambda}, 1 \right\} = \begin{cases} 1 & d_2 \leq \lambda \\ 2^\kappa \left(\frac{d_2}{d_1} \right)^{1/2} & d_2 > \lambda \end{cases}$$

$$= \begin{cases} 1 & d_2 \leq 2^{-2\kappa} d_1 \\ 2^\kappa \frac{\sigma_2 w_2}{\sigma_1 w_1} & d_2 > 2^{-2\kappa} d_1 \end{cases}$$

To determine the equilibrium value of γ , note that we still have $p_t = (1 - n_q^+) w_q q_t$, but now there are three cases:

$$\delta_q^+ = \begin{cases} 2^{2\kappa} & d_z \leq 2^{-2\kappa} d_q \\ 2^\kappa \left(\frac{d_q}{d_z} \right)^{1/2} & d_q > d_z > 2^{-2\kappa} d_q \text{ or } d_z > d_q > 2^{-2\kappa} d_z \\ 1 & d_q \leq 2^{-2\kappa} d_z \end{cases}$$

we can restate the interior (middle) condition as $\min\{d_z, d_q\} > 2^{-2\kappa} \max\{d_z, d_q\}$ or as $2^{-2\kappa} d_z < d_q < 2^{2\kappa} d_z$, and then we have:

$$\delta_q^+ = \begin{cases} 2^{2\kappa} & d_q \geq 2^{2\kappa} d_z \\ 2^\kappa \left(\frac{d_q}{d_z} \right)^{1/2} & 2^{-2\kappa} d_z < d_q < 2^{2\kappa} d_z \\ 1 & d_q \leq 2^{-2\kappa} d_z \end{cases}$$

This gives bounds for d_q based on κ and d_z that determine whether aggregate conditions are paid attention to (the top two options) and, if so, whether idiosyncratic conditions are then also paid attention to (the middle option).

For an interior solution, we compute γ as:

$$\begin{aligned}
\gamma &= (1 - n_q^+)w_q \\
&= \left(1 - 2^{-\kappa} \frac{\sigma_z w_z}{\sigma_q w_q}\right) w_q \\
&= w_q - 2^{-\kappa} w_z \frac{\sigma_z}{\sigma_q} \\
&= \gamma + (1 - \gamma)\zeta_q - 2^{-\kappa} \zeta_z \frac{\sigma_z}{\sigma_q} \\
&= 1 - 2^{-\kappa} \frac{\sigma_q \zeta_z}{\sigma_q \zeta_q}
\end{aligned}$$

We must of course check that this γ is consistent with an interior solution.

Note: Our formulation is notationally different from MW, but we can rewrite it in their terms.

Given an interior solution, we can compute the loss weight as:

$$\begin{aligned}
w_q &= \gamma + (1 - \gamma)\zeta_q \\
&= \left(1 - 2^{-\kappa} \frac{\sigma_q \zeta_z}{\sigma_q \zeta_q}\right) + \left(2^{-\kappa} \frac{\sigma_q \zeta_z}{\sigma_q \zeta_q}\right) \zeta_q \\
&= 1 - (1 - \zeta_q) 2^{-\kappa} \frac{\sigma_q \zeta_z}{\sigma_q \zeta_q}
\end{aligned}$$

And we have an interior solution if:

$$\begin{aligned}
2^{-2\kappa} d_z &< d_q < 2^{2\kappa} d_z \\
2^{-2\kappa} \sigma_z^2 \zeta_z^2 &< \sigma_q^2 w_q^2 < 2^{2\kappa} \sigma_z^2 \zeta_z^2 \\
2^{-\kappa} &< \frac{\sigma_q w_q}{\sigma_z \zeta_z} < 2^\kappa
\end{aligned}$$

Now:

$$\begin{aligned}
\frac{\sigma_q w_q}{\sigma_z \zeta_z} &= \frac{\sigma_q}{\sigma_z \zeta_z} - (1 - \zeta_q) 2^{-\kappa} \frac{1}{\zeta_q} \\
&= \frac{\sigma_q}{\sigma_z \zeta_z} - \left(\frac{1}{\zeta_q} - 1 \right) 2^{-\kappa} \\
&= \frac{1}{\zeta_q} \frac{\sigma_q \zeta_q}{\sigma_z \zeta_z} - \left(\frac{1 - \zeta_q}{\zeta_q} \right) 2^{-\kappa} \\
&= \frac{1}{\zeta_q} \left(\frac{\sigma_q \zeta_q}{\sigma_z \zeta_z} - 2^{-\kappa} (1 - \zeta_q) \right)
\end{aligned}$$

Then we can write the condition as:

$$\begin{aligned}
2^{-\kappa} &< \frac{1}{\zeta_q} \left(\frac{\sigma_q \zeta_q}{\sigma_z \zeta_z} - 2^{-\kappa} (1 - \zeta_q) \right) < 2^\kappa \\
2^{-\kappa} \zeta_q + 2^{-\kappa} (1 - \zeta_q) &< \frac{\sigma_q \zeta_q}{\sigma_z \zeta_z} < 2^\kappa \zeta_q + 2^{-\kappa} (1 - \zeta_q)
\end{aligned}$$

or finally as:

$$2^{-\kappa} < \frac{\sigma_q \zeta_q}{\sigma_z \zeta_z} < 2^{-\kappa} + (2^\kappa - 2^{-\kappa}) \zeta_q$$

This is identical to MW's condition for an interior solution, which is:

$$\frac{\sigma_q \zeta_q}{\sigma_z \zeta_z} \in (2^{-\kappa}, 2^{-\kappa} + (2^\kappa - 2^{-\kappa}) \zeta_q)$$

References

- Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory*. John Wiley & Sons.
- Fulton, C. (2015). Optimal Prices in a Multisector Model under Rational Inattention.
- Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis*. Cambridge University Press.
- Jung, J., J.-h. Kim, F. Matejka, and C. A. Sims (2015). Discrete actions in information-constrained decision problems.
- Maćkowiak, B. and M. Wiederholt (2009). Optimal Sticky Prices under Rational Inattention. *The American Economic Review* 99(3), 769–803.
- Matějka, F. and A. McKay (2015). Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review* 105(1), 272–298.
- Matejka, F., M. Wiederholt, and B. Maćkowiak (2017). The rational inattention filter. Working Paper Series 2007, European Central Bank.
- Morris, S. and H. S. Shin (2002). Social Value of Public Information. *The American Economic Review* 92(5), 1521–1534.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics* 50(3), 665–690.
- Sims, C. A. (2010). Rational Inattention and Monetary Economics. Handbook of monetary economics, Elsevier.
- Steiner, J., C. Stewart, and F. Matějka (2017). Rational Inattention Dynamics: Inertia and Delay in Decision-Making. *Econometrica* 85(2), 521–553.

Woodford, M. (2014). Stochastic Choice: An Optimizing Neuroeconomic Model. *The American Economic Review* 104(5), 495–500.

Zbaracki, M. J., M. Ritson, D. Levy, S. Dutta, and M. Bergen (2004). Managerial and Customer Costs of Price Adjustment: Direct Evidence from Industrial Markets. *The Review of Economics and Statistics* 86(2), 514–533.