Mechanics of static quadratic Gaussian rational inattention tracking problems

Chad Fulton *

Abstract

This paper presents a general framework for constructing and solving the multivariate 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 our approach to a price-setting problem and a portfolio choice problem - two popular rational inattention models found in the literature for which simplifying assumptions have thus far been required to produce a tractable model. In contrast to prior results, which have been limited to cases that restrict the number of underlying shocks or their correlation structure, we present general solutions. In each case, we show that imposing such restrictions impacts the form and interpretation of solutions and implies suboptimal decision-making by agents.

JEL Classification: D81, D83, E31

Keywords: Rational inattention, information acquisition, price setting, portfolio choice

*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 economics 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 it has been used to establish baseline results and provide intuition in many applications.

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. This formulation can include an arbitrary number of shocks, potentially correlated, and can incorporate its information constraint in terms of a fixed quantity of information processing capacity or a fixed marginal cost associated with processing additional information.

[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 Matêjka and McKay (2015) or Steiner et al. (2017). For discrete actions in continuous settings, see Jung et al. (2015).

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 ellisoids representing uncertainty. Then, given the solution, the agents 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 agents prior for that component and their understanding of the incoming data. Moreover, using the canonical shocks we can construct representations 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, an important point is that any given representation is simply a device that assists us in characterizing the agents 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.

As a first application, we consider rational inattention price-setting problems similar to those from Maćkowiak and Wiederholt (2009). We first describe the general case of the model including arbitrary correlations between underlying shocks and present the solution. This is the first general solution of this model that we are aware of. Second, we show how to modify the problem to incorporate the independence assumption of Maćkowiak and Wiederholt (2009) and present the solution in this case. We show that the introduction of the independence assumption prevents the agent from achieving the first-best outcome and that these two formulations yield qualitatively different feasible representations.

As a second application, we consider the problem of portfolio choice by rationally inattentive investors, along the lines of Mondria (2010) and Kacperczyk et al. (2016). Whereas the former is restricted to the case with two independent assets and the later is restricted to the case in which the signal structure is exogenously given, the methods developed in this paper allow us to solve for the fully optimal allocation of attention and portfolio choice for N assets with arbitrary correlations. Importantly, we show that the signals chosen by fully optimizing agents are not compatible with the exogenously given signals considered in Kacperczyk et al. (2016).

This paper is most closely related to Sims (2003) and Sims (2010), to which we owe our basic formulation for the class of rational inattention linear quadratic (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, they consider the dynamic problem with a single ARMA(p,q) target.

2 Problem

Rational inattention problems fall into the larger class of problems in which agents must make decisions under imperfect information. In classical imperfect information settings, the information structure of the economy is often exogenously imposed; for example, signal extraction problems assume particular observation vectors are given to agents. 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 though the information itself is assumed to be freely available.

To construct the the problem we are interested in, we start by defining static tracking problems. 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 $\min_{a_+} E[d(\alpha, a_+) \mid \mathcal{I}_-]$, where $d(\cdot)$ is a loss function, $\mathcal{I}_-$ is a prior information set, and the minimization is taken to be over the joint density of (α, a_+). We refer to α as the target or state and a_+ as the action.[2] The agent wishes to make the action as close as possible, on average, to the target. We call a static tracking problem linear quadratic Gaussian (LQG) if the loss function is quadratic and all variables are jointly Gaussian.[3] This problem is trivial unless some constraint is added, and the signal extraction problem is a special case with the constraint that the action be a measurable function of a given observation vector.

The constraint in rational inattention tracking problems is formalized using the mutual information between target and action, possibly conditional on some prior information set, $I(\alpha, a_+ \mid \mathcal{I}_-)$. Mutual information quantifies reduction in entropy; intuitively, it measures how much uncertainty about the target must have been resolved by the agent in the process of taking a particular action. It has a number of desirable properties including the fact that regardless of the dimensions of the random vectors, mutual information is always a scalar.[4] There are two primary ways of formulating the rational inattention constraint. The first allows agents a fixed processing capacity κ and requires that $I(\alpha, a_+ \mid \mathcal{I}_-) \leq \kappa$; 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. These approaches lead to largely similar statements of the problem and solution, but they will have different implications in comparative statics exercises. The rational inattention tracking problem is then $\min_{a_+} E\left[d(\alpha, a_+) \mid \mathcal{I}_-\right] + \lambda^* I(\alpha, a_+ \mid \mathcal{I}_-)$ where λ^* is interpreted either as a cost parameter or as a Lagrange multiplier for a constraint $I(\alpha, a_+ \mid \mathcal{I}_-) \leq \kappa$.

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 $(\alpha, a_+) \mid \mathcal{I}_-$ jointly Gaussian; our task is to use this insight

[2] 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] The linear part of LQG refers to a linear transition law in dynamic versions of the problem.

[4] A more detailed discussion of the properties of mutual information and its motivation in economic contexts can be found in Sims (2003) or Sims (2010).

to find an explicit formulation of a_+. To do so, we note that the joint distribution is such that can decompose $\alpha = a_+ + \eta$, where $a_+ \perp \eta$. If $\alpha \mid \mathcal{I}_- \sim N(a_-, P_-)$, then we can then specify each component: $a_+ \mid \mathcal{I}_- \sim N(a_-, P_- - P_+)$ and $\eta \mid \mathcal{I}_- \sim N(0, P_+)$. As a result, we can write $\alpha \mid \mathcal{I}_-, a_+ \sim N(a_+, P_+)$. Finally, by replacing the specific conditioning variables with a generic posterior information set $\mathcal{I}_+ \supseteq \mathcal{I}_-$, this allows us to consider the optimal action as a conditional expectation: $a_+ = E[\alpha \mid \mathcal{I}_+]$. The matrix P_+ ties down the posterior information set and so also the action, and we will be able to reformulate the problem and solution in terms of it.

It is tempting to use the fact that a solution to the RI-LQG tracking problem is a conditional expectation to draw an analogy with the LQG signal extraction problem, for which the solution is well known to be a conditional expectation. It is important to remember, however, that while an observation vector is a fundamental part of signal extractions problems, in the rational inattention setting there is no need to posit such a vector; we will define and solve the problem without doing so. The idea of a observation received by the agent that, along with their prior, generates the posterior information set can be useful for interpretation, but the concept is purely notional: by assumption, a rationally inattentive agent *observes* all data in its entirety, but may not *process* it fully. We discuss this in detail later.

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, given our quadratic loss, we can write $E[d(\alpha, a_+) \mid \mathcal{I}_-] = E[(\alpha - a_+)'W(\alpha - a_+) \mid \mathcal{I}_-] = tr(WP_+)$ where W is a positive semidefinite matrix defining the loss function. Second, the mutual information between conditionally Gaussian random vectors has a relatively simple form, $I(\alpha, a_+ \mid \mathcal{I}_-) = \frac{1}{2}\left(\log_b |P_-| - \log_b |P_+|\right)$.[5] 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.

[5] For a derivation of this result, see Cover and Thomas (2006), for example. We have left the base of the logarithm unspecified here since it only determines the units in which information is measured; often the base is set to 2, in which case information is measured in bits.

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

$$\min_{P_+} tr(WP_+) + \lambda(\ln|P_-| - \ln|P_+|) \tag{1}$$

$$\text{subject to } \alpha \mid \mathcal{I}_- \sim N(a_-, P_-),\ P_+ \geq 0,\ P_- - P_+ \geq 0$$

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 $N(\mu, \Omega)$ where 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.

The problem as stated has two positive semidefiniteness constraints. The first requires that P_+ is a valid covariance matrix; given the objective function, 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 $a_+ \mid \mathcal{I}_-$, above, is valid.

3 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.

We first introduce a lemma that is key to understanding the solution to both the static and dynamic RI-LQG tracking problems. It shows that it is possible to transform any such problem so that the information processing by the agent can be decoupled into independent components. We will develop these components at length below, and they will define what we call the canonical synthetic target. In the static case considered here, we will be able to construct a closed form representation of the transformation that produces the decoupling for the posterior covariance matrix optimally chosen by a rationally inattentive agent. In the dynamic case the methods used in this paper cannot be directly applied to construct a closed form; nonetheless, if given an optimal posterior covariance matrix (for example from a numerical solution), Lemma 1 still applies and many of the tools of this paper can be used to understand the solution.[6]

Lemma 1: Let X, Y be jointly Gaussian random vectors of dimension n, possibly conditional on some information $\mathcal{I}_-$, and let $Var(X \mid \mathcal{I}_-) = P_-$ and $Var(X \mid \mathcal{I}_-, Y) = P_+$. Then:

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

where $I(\cdot)$ denotes conditional mutual information and the elements n_i are generalized eigenvalues of the matrix pencil (P_+, P_-).

[6] A previous version of this paper, Fulton (2017), provides more details on the dynamic problem and presents an approximate solution method that is good in many practically relevant cases.

3.1 Solution to the static LQG-RI tracking problem

It is easy to check that the first order condition to Definition 1 yields:

$$(P_+^{\text{FOC}})^{-1} = W/\lambda \tag{2}$$

We generally write the first order condition in terms of the precision rather than covariance matrix, because we have not required W to be nonsingular.[7] Only if the positive semidefiniteness constraints are not binding does this yield the solution to the static RI-LQG tracking problem; generally the solution is more complex. Before presenting the full solution in Theorems 1 and 2, some preliminaries are provided in Lemma 2.

Lemma 2: The matrix pencil (W, P_-^{-1}), consisting of the loss matrix and prior covariance matrix from Definition 1, can be simultaneously diagonalized as $W = S'DS$ and $P_-^{-1} = S'IS$, where $S = Q'M$, QDQ' is the eigendecomposition of $L'WL$, L is the Cholesky factor of P_-, and $M = L^{-1}$. Denote r_i as the i-th column of $R = S^{-1}$ and d_i as the i-th diagonal element of D; then d_i is a generalized eigenvalue of the matrix pencil associated with the generalized eigenvector r_i.

For the remainder of this paper, we will always arrange the generalized eigenvalues of Lemma 2 in nonincreasing order.

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

$$P_+ = RN^+R' \tag{3}$$

where N^+ is a diagonal matrix with entries n_i^+. These 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.

[7] 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.

As a consequence of assuming that the generalized eigenvalues d_i are in nonincreasing order, the values n_i^+ will be in nondecreasing order.

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:

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

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 $d_i > \lambda$, $i = 1, \ldots, r$ then this pair (r, λ) describes the solution. Otherwise, set $r = r - 1$ and repeat from step (b).

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.

3.1.1 Canonical synthetic target

Before interpreting 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: The canonical synthetic target (briefly the canonical target) is the vector $\beta_c = S\alpha$, where S is the matrix of left generalized eigenvectors from the second Corollary to Theorem 1, defined in Lemma 2.

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. $\beta_c \mid \mathcal{I}_+ \sim N(b_{c,+}, N^+)$ where $b_{c,+} = Sa_+$.

b. $\beta_c \mid \mathcal{I}_- \sim N(b_{c,-}, I)$ where $b_{c,-} = Sa_-$.

c. $E[(\alpha - a_+)'W(\alpha - a_+) \mid \mathcal{I}_-] = E[(\beta_c - b_{c,+})'D(\beta - b_{c,+}) \mid \mathcal{I}_-]$

d. $I(\alpha, a_+ \mid \mathcal{I}_-) = I(\beta_c, b_{c,+} \mid \mathcal{I}_-)$

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

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.

Definition 3: We refer to the value r for which $d_r > \lambda \geq d_{r+1}$ is satisfied as the rank of the solution to the static RI-LQG tracking problem, and we say that the solution is full rank if $r = n$.

This terminology is appropriate because r is the number of generalized eigenvalues n_i^+ greater than one and so $r = \text{rk}(N^+ - I) = \text{rk}(P_- - P_+)$. This also implies that 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. Finally it is not hard to see that $r \leq \text{rk}(W)$, so that the solution can only be full rank if the loss matrix is nonsingular.

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

$$\kappa \equiv I(\alpha, a_+ \mid \mathcal{I}_-) = \frac{1}{2}\sum_{i=1}^{r} \log_b \frac{1}{n_i^+} = \frac{1}{2}\left[\sum_{i=1}^{r} \log_b d_i - r \log_b \lambda\right] \tag{5}$$

Definition 5: The information capacity allocated to processing the i-th element of canonical synthetic target β_c is $\kappa_i \equiv I(\beta_{i,c}, b_{i,c,+} \mid \mathcal{I}_-)$, and

$$\kappa_i = \frac{1}{2}\log_b \frac{1}{n_i^+} = \begin{cases} \frac{\kappa}{r} + \log_b\left[\frac{\sqrt{d_i}}{\prod_{j=1}^{r} \sqrt{d_j}^{1/r}}\right] & i = 1, \ldots, r \\ 0 & i = r+1, \ldots, n \end{cases} \tag{6}$$

The last formulation suggests a straightforward intuition describing the allocation of capacity by rationally inattentive agents: 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 write $\kappa = \sum_{i=1}^{r} \kappa_i$.

Unfortunately, there is generally no obvious 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:

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

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 4: 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 $\sum_{i=1}^n k_i = \kappa$.

3.1.2 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. This situation primarily occurs in practice 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. The solution in this case is relatively straightforward. More importantly, the intuition extends to general case when cast in terms of the canonical target, because this target was constructed exactly to satisfy separability.

To fix notation, let P_- be positive definite diagonal with elements $\sigma_{i,-}^2$, let W be positive semidefinite diagonal with elements w_i^2, and assume that $w_1^2\sigma_{1,-}^2 \geq \cdots \geq w_n^2\sigma_{n,-}^2$.[8] Application of Lemma 2 is trivial in this case, and 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.[9] The associated right generalized eigenvectors are $r_i = \sigma_{i,-}e_i$

[8] This is without loss of generality, since we can always re-order the elements of α to make this true.

[9] This interpretation as loss-weighted volatility is still broadly true in the more general case, but the

where e_i is the ith 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. Suppose that the rank of the solution is r, so that λ satisfies $d_r > \lambda \geq d_{r+1}$. 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 full solution is

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

The first order condition would have set $\sigma_{i,+}^{\text{FOC}^2} = \lambda/w_i^2$ for all of $i = 1, \ldots, n$, but this is infeasible, since we defined r such that $\lambda/w_{r+1}^2 \geq \sigma_{r+1,-}^2$ and so this would suggest more posterior uncertainty for elements $r+1, \ldots, n$ 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 $i = r+1, \ldots, n$. We can also simplify the formula describing information capacity allocations as

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

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, 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.

relationships are more complex due to interaction effects.

3.1.3 Comparative statics

We very briefly consider the way changes in parameters can affect the solution.[10] 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, and 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. We instead focus on 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. Here the formulation of the information constraint has a material effect on results. This is because an increase in d_i prompts the agent to pay more attention to the i-th component, and in the fixed capacity case this implies that the agent must reduce attention to the other components, whereas in the fixed marginal cost case the agent simply pays more and increases the total quantity of information processed.

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

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

[10] A set of more detailed comparative statics exercises on particular examples may be found in an earlier version of this paper.

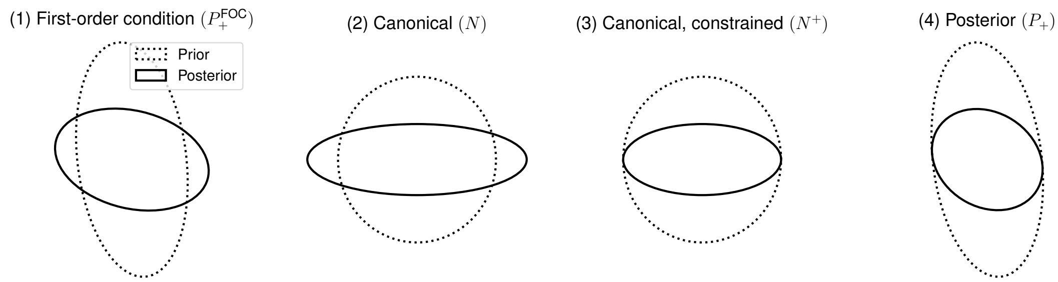

Figure 1: Geometrization of Theorem 1 using ellipsoids

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 as ellipsoids. In the case of covariance matrices of Gaussian distributions, these ellipsoids correspond to iso-density loci, and graphically represent uncertainty. The key to this illustration is that the information constraint and the two positive semidefiniteness constraints can be understood in terms of the properties of ellipsoids corresponding to the prior and posterior covariance matrices.

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. Thus 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 can compute $\frac{1}{2}(\log_b(V_-/V_+)) = \frac{1}{2}(\log_b |P_-| - \log_b |P_+|) = I(\alpha, a_+ \mid \mathcal{I}_-)$, so the information constraint can be understood in terms of the relative volumes of the prior and posterior ellipsoids. The positive semidefiniteness constraint $P_+ \geq 0$ simply requires that the posterior ellipsoid be well-defined, and the no-forgetting constraint $P_- - P_+ \geq 0$ requires that the posterior ellipsoid be weakly contained within the prior ellipsoid.

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 ellipsoids 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. In general the difficulty in solving the problem of Definition 1 is that the prior and posterior covariance ellipsoids are not concentric; that is the associated covariance matrices do not share eigenvectors. The solution given by Theorem 1 transforms the problem into the canonical space, where the transformed ellipsoids are concentric and the no-forgetting constraint can be imposed by pulling in the axes of the posterior ellipsoid that extend beyond the prior.

We visualize the geometrical interpretation of the problem and solution in the four panels of Fig. 1. Panels (1) and (4) represent the prior and proposed posterior ellipsoids in the standard coordinate space while panels (2) and (3) represent the problem in the transformed canonical space. The first two panels show an infeasible posterior, as proposed by the first order condition, while the last two panels show the feasible, constrained, posterior. Imposition of the constraints by the pulling in operation can be easily seen in the difference between panels (2) and (3).

3.3 The action solving the static RI-LQG tracking problem

Although we initially defined the optimal action through its joint distribution with the target, it can be useful to consider it as a conditional expectation, $a_+ = E[\alpha \mid \mathcal{I}_+]$. We take up this latter approach now so as to develop an intuitive understanding of the action of a rationally inattentive agent, taking as given the posterior information set $\mathcal{I}_+$. To facilitate this, we will, for now, assert two results that we develop in later sections. First, we will write $\hat{\alpha}$ to denote the agents understanding of the target based solely on incoming data; this will only be fully formalized (as a representation) later. Second, we present the result, derived later, that the agents 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. It further shows that are two channels through which the agents action is driven away from the target. First, 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. Second, 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.

Lemma 5: The components of $b_{c,+} = E[\beta_c \mid \mathcal{I}_+]$, which we call the canonical synthetic action (briefly the canonical action), are:

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

where $n_i^+ \in [0, 1]$ and $b_{c,-} = E[\beta_c \mid \mathcal{I}_-]$, and where the agents understanding of the canonical target, that generates the posterior information set, is defined as

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

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, Bayesian updating is straightforward: each component of the canonical action is a simple weighted average of the prior for that component and that component of the canonical target contaminated by noise due to rational inattention. To understand why this transformation effectively decouples the components of the action, notice that we can write $K = R(I - N^+)S$. 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. 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_+.

3.3.1 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 5, 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:

$$E\left[b_{i,c,+} - \beta_{i,c} \mid \mathcal{I}_-, \beta_{i,c}\right] = n_i^+(b_{i,c,-} - \beta_{i,c}) \qquad \text{Bias}$$
$$Var(b_{i,c,+} \mid \mathcal{I}_-, \beta_{i,c}) = n_i^+(1 - n_i^+) \qquad \text{Variance}$$

Thus the bias describes the extent to which the rational inattention action will differ from the target on average, while the variance gives the effect of contaminating noise. In general, the bias is nonzero unless information is perfect $(n_i^+ = 0)$, and the variance is nonzero unless information is perfect or the agent collects no information at all $(n_i^+ = 1)$.

We can take a similar approach to quantify the responsiveness of the agent, by considering the extent to which the action differs from the prior on average, $E[b_{i,c,+} - b_{i,c,-} \mid \mathcal{I}_-, \beta_{i,c}] =$

$(1 - n_i^+)(\beta_{i,c} - b_{i,c,-})$. The term $(\beta_{i,c} - b_{i,c,-})$ is the unexpected component of the target, and so the quantity $1 - n_i^+ \in [0, 1)$ captures the (average) responsiveness of the agent, as new information moves them away from their prior.

3.3.2 Linear combinations of the action

It is now straightforward to understand arbitrary linear combinations of the action, $w'\alpha$ where w is an $n \times 1$ vector of weights. 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 $w_1, \ldots, w_n$ with weights $\xi_1, \ldots, \xi_n$, then the loss function is:

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

and this can be rewritten in the standard form $E\left[(\alpha - a_+)'W(\alpha - a_+) \mid \mathcal{I}_-\right]$ 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 $E[(p - p_+)^2 \mid \mathcal{I}_-]$. 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 compute this quantity 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, and the posterior collapses to $p_+ = n_1^+ p_- + (1 - n_1^+)\hat{p}$. 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 also 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 $w^{\perp\prime}P_{+}w^{\perp} = w^{\perp\prime}P_{-}w^{\perp}$. The general version of this result for the rank r 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 $n_i^+ < 1$.

3.3.3 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 agents prior is just the unconditional distribution, so that $z_t \mid \mathcal{I}_{t-1} \sim N(0, \omega^2 I)$ 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 $W = \iota\iota' = \underset{n\times n}{\mathbf{1}}$ (an $n \times n$ matrix of ones), where $\iota = (1, 1, \ldots, 1)'$ 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 $i = 2, \ldots, n$. This implies that $n_1^+ = \min(\lambda/n\omega^2, 1)$ and $n_i^+ = 1$ for $i = 2, \ldots, n$. Applying the first Corollary to Theorem 1, we conclude that:

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

This agrees with the solution in Sims (2010), except that we are able to be more explicit about 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 $n_i^+ < 1$.

3.4 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:

$$\min_{O_+} tr(VO_+) + \lambda(\ln|O_-| - \ln|O_+|) \tag{11}$$

$$\text{s.t. } \beta \mid \mathcal{I}_- \sim N(b_-, O_-)$$

$$O_+ \geq 0$$

$$O_- - O_+ \geq 0$$

where $V = B^{-1\prime} W B^{-1}$, $O_- = BP_-B'$ 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 $P_+ = B^{-1}O_+B^{-1\prime}$ solves the reference static RI-LQG tracking problem (W, a_-, P_-).

Theorem 4 provides the formal justification for the geometrical steps visualized in Fig. 1. Moreover, recalling that the canonical transformation was defined as $S = Q'M$, it is easy to see that the construction of the canonical synthetic target can be broken down into two iterative applications of Theorem 4: the first application applies the transformation M to construct a whitened synthetic target for which the prior independent, and the second application applies the transformation Q' to that whitened target in order to construct the canonical target, for which the posterior is independent. [11]

4 Representations

Although we have considered a_+ as a conditional expectation, we have so far left the posterior information set $\mathcal{I}_+$ vague and focused instead on solving the problem directly, in terms of the posterior covariance matrix P_+. In this section, we finally consider the posterior information set and discuss what we call representations of the information processed by agents.

Definition 8: For a static RI-LQG tracking problem (W, a_-, P_-) with solution P_+ and the corresponding action a_+, we define a representation to be 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 $E[v_+ \mid \mathcal{I}_-] = 0$.

Then the posterior information set is generated by the prior information set together with

[11] More details on the iterated application of transformations and on these transformations in particular can be found in an earlier version of this paper.

any valid representation. We think that representation is a natural term to capture the essence of these vectors, because although the rationally inattentive agent observes the full set of data, each representation is simply a description of a way in which the agent could have operationalized that incoming data to update their information. 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 agents imperfect understanding after processing new data.

Lemma 6: The action a_+ is a representation, since $a_+ = E[\alpha \mid \mathcal{I}_-, a_+]$. Thus we can refer to the action a_+ as the agents perception of the target.

The term perception seems natural to use when discussing a_+ as it relates to the agents *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, specific instances of what we refer to here as representations are often described as observations or as signals, and the rational inattention problem is often formulated in terms of selecting the noise covariance matrix corresponding to one of these specific signal vectors. 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, focusing on specific vectors can imply that the agent must be processing the data in a particular way that is not justified in the rational inattention context. Along the same lines, using the terms observation or signal can suggest inappropriate analogies with signal extraction problems.

4.1 The posterior information set

In this section, we take an algebraic approach to construct explicit formulas for valid representations, and show how the class of such representations is pinned down by the solution to the RI-LQG tracking problem P_+.

Since all variables are jointly Gaussian, considering a_+ as a conditional expectation is the same as viewing it as the linear projection of α onto a vector space $\mathcal{W}_+$ that is identified with the posterior information set. Similarly, the prior a_- is the projection of α onto a subspace $\mathcal{W}_- \subseteq \mathcal{W}_+$ that is identified with the prior information set. To construct the class of valid representations, we will proceed by isolating the purely new information as the orthogonal complement of $\mathcal{W}_-$ in $\mathcal{W}_+$, which we denote $\mathcal{W}_*$. This allows us to decompose $a_+ = \text{proj}_{\mathcal{W}_+}\alpha = \text{proj}_{\mathcal{W}_-}\alpha + \text{proj}_{\mathcal{W}_*}\alpha$, and we know that $\text{proj}_{\mathcal{W}_-}\alpha = a_-$, while we can write $\text{proj}_{\mathcal{W}_*}\alpha = K_v v_+$ where $v_+ \in \mathcal{W}_*$ and $K_v = \langle \alpha, v_+\rangle[\langle v_+, v_+\rangle]^{-1}$.

To identify the vectors $v_+ \in \mathcal{W}_*$, begin with any n-dimensional Gaussian random vector y, and decompose it into the subspace spanned by α and the perpendicular space as $y = Z\alpha + \zeta$. Now, if a vector is in $\mathcal{W}_*$ then it must be orthogonal to $\mathcal{W}_-$; to ensure this, apply the Gram-Schmidt process to construct $v_+ = y - \text{proj}_{\mathcal{W}_-} y = Z\alpha + \varepsilon - Za_-$, where $\varepsilon \perp \alpha$ and $\varepsilon \perp \mathcal{W}_-$. Define $\Lambda = \langle \varepsilon, \varepsilon\rangle$. For any such v_+, we can compute the above projection matrix $K_v = P_- Z'(ZP_- Z' + \Lambda)^{-1}$, and this defines a candidate $a_+^v = a_- + K_v v_+$. If v_+ is to be a valid action, it must be the case that a_+^v coincides with the solution to the static RI-LQG tracking problem, and this occurs when $P_+ = P_- - P_- Z'(ZP_- Z' + \Lambda)^{-1} ZP_-$. Any pair (Z, Λ) that satisfies this equation for the matrix P_+ that solves the static RI-LQG tracking problem, along with Λ positive semidefinite (since it results from an inner product), describes a valid representation, $y_+ = Z\alpha + \varepsilon$, as well as a valid innovation representation, $v_+ = Z\alpha + \varepsilon - Za_-$. It is in this way that the choice of P_+ in the static RI-LQG tracking problem defines $\mathcal{W}_+$ and thereby identifies $\mathcal{I}_+$. If we further require that

Λ is positive definite, then we can apply the matrix inversion lemma to arrive at:

$$Z'\Lambda^{-1}Z = P_+^{-1} - P_-^{-1} \tag{12}$$

and we will call representations that additionally satisfy this condition proper feasible representations.

4.2 Feasible, proper, and reduced representations

In this section we formally define the feasible representations introduced in the previous section and describe their properties. However, before doing so we give a slightly more general definition of a representation.

Definition 9: 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 rank[12]

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 implementing row-swapping transformations.

[12] 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.

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 as a vector y_+ in the following form:

$$y_+ = d + Z\alpha + \varepsilon, \qquad \varepsilon \sim N(0, \Lambda) \tag{13}$$

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.[13] 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 feasible representation can be made proper simply by eliminating the improper block and considering a reduced representation of dimension ℓ.

Definition 10: 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 9 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 7.

Lemma 7:

a. The minimum dimension of any representation is the rank of the solution, so that $m \geq r$.

[13] 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.

b. The dimension of any proper feasible representation is equal to the rank of the solution, so that $m = r$.

Finally, we note that every feasible representation has a corresponding innovation representation that can be written as $v_+ = y_+ - E[y_+ \mid \mathcal{I}_-]$, and that every such 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-forgetting 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 definable 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 written as:

$$a_+ = a_- + K_v v_+ \tag{14}$$
$$P_+ = (I - K_v Z) P_-$$

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.

4.2.1 Illustration: simplified vector space

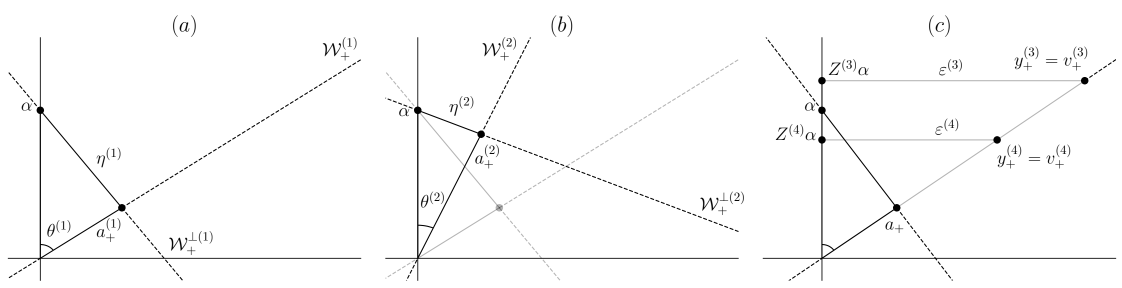

Figure 2: Visualization of the static RI-LQG tracking problem, solution, action, and representations in a simplified vector space

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 $\mathcal{W}_* = \mathcal{W}_+$.[14]

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. 2 (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. 2 (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

[14] 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. 2 (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 $\mathcal{V}_\alpha = \text{span}(\begin{bmatrix} 0 & 1 \end{bmatrix}') = \{Z \begin{bmatrix} 0 & 1 \end{bmatrix}' \mid Z \in \mathbb{R}\}$, so that we can write $y = Z\alpha + \zeta$, where $\zeta \in \text{span}(\begin{bmatrix} 1 & 0 \end{bmatrix}') = \{c \begin{bmatrix} 1 & 0 \end{bmatrix}' \mid c \in \mathbb{R}\}$. 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. 2 (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).

4.3 The fundamental and canonical representations

In this section, we present several important representations that can be used in practice to construct the action as described in Theorem 5.

Definition 11: 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:

$$y_f = \alpha + \varepsilon_f, \qquad \varepsilon_f \sim N(0, \Lambda_f) \tag{15}$$

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 special 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 12: 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:

$$y_c = \beta_c + \varepsilon_c, \qquad \varepsilon_c \sim N(0, \Lambda_c) \tag{16}$$

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 10, we can always construct a reduced canonical representation that is proper.

Definition 13: 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 5 and Theorem 3.

4.4 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 14: The representation form of the static RI-LQG tracking problem is:

$$\min_{Z,\Lambda^{-1}} tr(WP_+) + \lambda(\ln|P_-| - \ln|P_+|) \tag{17}$$

$$\text{subject to } \alpha \mid \mathcal{I}_- \sim N(a_-, P_-),\ \Lambda^{-1} \geq 0,\ P_+ = (Z'\Lambda^{-1}Z + P_-^{-1})^{-1}$$

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 Applications

5.1 Price setting

In this section, we illustrate the static RI-LQG tracking problem and solution by considering a stylized model of rationally inattentive price-setters similar to that introduced by Maćkowiak and Wiederholt (2009).

The general setup posits a firm that wishes to set their price p_{it} so as to track the profit-maximizing price. The profit-maximizing price is a linear combination of two fundamental shocks, one an idiosyncratic shock to productivity and one a shock to aggregate demand, written $p_{it}^* = w_z z_{it} + w_q q_t$. We suppose that the firm is rationally inattentive, so that it can only pay a limited amount of attention to the fundamental shocks, and that the firm trades off uncertainty between elements so as to minimize its loss, $E[d(p_{it}^*, p_{it}) \mid \mathcal{I}_-] = E[(p_{it}^* - p_{it})^2 \mid \mathcal{I}_-]$.[15] For this example, we suppose that the fundamental shocks are *iid*

[15] This is a simplification of the of Maćkowiak and Wiederholt (2009), since in their formulation the

multivariate Gaussian, and to simplify notation we write $p_{it}^* = w'\alpha$, where $w = (w_z, w_q)'$ and $\alpha = (z_{it}, q_t)'$. Then, since we know that firms will choose $p_{it} = E[p_{it}^* \mid \mathcal{I}_+]$ for some $\mathcal{I}_+$, we can rewrite the loss function as $tr(WP_+)$, where $W = ww'$ and $P_+ = Var(\alpha \mid \mathcal{I}_+)$. This problem is in the form of Definition 1.

We can apply Theorems 1 and 2 to compute the posterior covariance matrix chosen by the firm, and we can apply Lemma 5 and Theorem 3 to compute their action. Since $\text{rk}(W) = 1$, we know that the rank of the solution will be at most one, meaning that the firm will only pay attention to a single linear combination of the fundamental shocks. This linear combination is defined by $\beta_{1,c} = s_1'\alpha$, and the firms action, in terms of the canonical synthetic target, is $b_{1,c,+} = n_1^+ b_{1,c,-} + (1 - n_1^+)y_r$, where y_r is the reduced canonical representation. Here $s_1' = q_1'M$ and $n_1^+ = \min\{\lambda/d_1, 1\}$, where q_1 is the eigenvector of $V = L'WL$ associated with the largest eigenvalue, d_1. The firm will not incorporate new information about any other linear combination, so that $b_{i,c,+} = b_{i,c,-}$ for $i = 2, \ldots, n$.

Because the loss matrix is rank one, we can compute the form of the eigenvector as $q_1 = L'w/\|L'w\|$ along with the associated eigenvalue $d_1 = \|L'w\|^2$. This implies that the the canonical synthetic target is defined by $s_1 = w/\|L'w\|$. Multiplying both sides of the equation describing the canonical synthetic action by $\|L'w\|$, we arrive at $p_{it} = n_1^+ p_{it-} + (1 - n_1^+)y_p$ where $y_p = p_{it}^* + \varepsilon_p$ is a proper feasible representation and $\varepsilon_p \sim N(0, \|L'w\|^2(\delta_1^+ - 1)^{-1})$. In words, the firm simply performs a Bayesian update on the object of interest, here the profit-maximizing price, and one representation of the way the agent processes the incoming data is profit-maximizing price plus noise. Importantly, we did not assume that the agent had access to a signal of this form. Instead, the agent optimally paid attention to the incoming data and the result is that a representation of this form exists.

weight w_q is determined in equilibrium. We abstract from this detail here to emphasize the rational inattention solution. A more detailed example that also solves for the equilibrium values can be found in an earlier version of this paper.

5.2 Price setting under an independence assumption

The model actually considered in Maćkowiak and Wiederholt (2009) is somewhat different than that above, because they assume that the fundamental shocks are independent and they impose the requirement that the firm pay attention to each fundamental shock separately; following them, we refer to this as the independence assumption. The former assumption implies that we can write $P_- = Var(\alpha \mid \mathcal{I}_-) = \text{diag}\{\sigma_z^2, \sigma_q^2\}$. The latter assumption is most easily formalized using the representation form of the problem given in Definition 14, since it limits the form that representations can take. In particular, it requires that in any solution both Z and Λ^{-1} must be diagonal.[16] Because of these restrictions, the problem does not immediately fall into the form of Definition 1, but we can modify it so that it does.

Using the equation $P_+^{-1} = Z'\Lambda^{-1}Z + P_-^{-1}$, it follows that, under the assumptions above, P_+ must be diagonal and the firms objective function, given the restrictions, is $E[(p_{it}^* - p_{it})^2 \mid \mathcal{I}_-] = w_z^2 Var(z_{it} \mid \mathcal{I}_+) + w_q^2 Var(q_t \mid \mathcal{I}_+)$. This objective function can be achieved within the framework of this paper by considering a different problem in which agents optimize given the alternative loss matrix $W_I \equiv \text{diag}\{w_z^2, w_q^2\}$. With this modification, $E[(p_{it}^* - p_{it})^2 \mid \mathcal{I}_-] = tr(W_I P_+)$, and so the problem is in the form required by Definition 1.

We can then solve the problem using the methods of this paper. One feature that is of particular interest is that since $\text{rk}(W_I) = 2$, under the independence assumption agents may pay attention to each fundamental shock separately. Furthermore, since the assumptions imposed $Q = I$, the canonical synthetic shocks are simply the fundamental shocks, standardized. Finally, the fundamental representation is valid in this case, so that we can interpret agents as observing the fundamental shocks plus noise.

It is important to notice that imposing the independence assumption prevents firms from

[16] The independence assumption by itself does not require that the posterior covariance matrix P_+ must be diagonal, although this will be the case under the assumption that the fundamental shocks are independent.

reaching the first-best solution, which we derived above in our first formulation of the price-setting problem. Maćkowiak and Wiederholt (2009) justified this by suggesting that it is unrealistic to suppose that firms actually have access to a signal of the form profit-maximizing price plus noise. However, as we have shown in this paper, the availability of a *representation* of this form does not require that firms be given any particular signal *externally*. It only reflects the fact that a rationally inattentive firm will *internally* process the available information to maximize its relevance to their object of interest. We argue, therefore, that the independence assumption should only be used when there is specific evidence that rationally inattentive agents are unable to *process* information freely.

5.3 Portfolio choice

As a second application, we consider the problem of optimal portfolio choice when investors are rationally inattentive. Although the problem faced by agents is different than that of Definition 1, it is a mark of the flexibility of our framework that we can use it to solve the problem and interpret the solution. This section is related to Admati (1985), Mondria (2010), and Kacperczyk et al. (2016). Admati (1985) solves an N asset portfolio choice problem given an exogenous information structure. Mondria (2010) extends this to allow for rational inattention, but is limited to the case with two assets. Kacperczyk et al. (2016) consider the N asset case in which rationally inattentive agents choose the precision of exogenously given signals, but they do not solve for the form of optimal signals. The framework introduced in this paper allows us to consider fully optimal attention allocation for N assets, including arbitrary correlation structures. Thus, this section can be understood either as extending Mondria (2010) to the N asset case, or as extending Kacperczyk et al. (2016) to include the optimal selection of signal vectors.

In brief, the model considered by these papers is as follows. There is a continuum of rationally inattentive investors that are tasked with selecting a portfolio of assets. There is a single riskless asset with return r along with N risky assets with returns given by the random vector $\alpha \sim N(\bar{a}, \Omega_a)$.[17] The common prior for asset returns is $\alpha \mid \mathcal{I}_{-} \sim N(a_{-}, P_{-})$, and investors must pay an attention cost in order to receive updated information. We denote investor js information set after processing their individual information about returns as $\mathcal{I}_{j+}$. The equilibrium price vector will also provide common information about returns, and we assume that agents observe these prices costlessly. We denote investor js information set after processing both individual information and prices as $\mathcal{I}_{jp+}$. The net supply of each risky asset is assumed to be a random vector, $z \sim N(\bar{z}, \Omega_z)$, with $z \perp \alpha$, and we assume that agents receive no information about net asset supply directly.

Each investor j has initial wealth ω_0 and maximizes mean-variance utility $U_j = \rho^{-1}E[\omega_j \mid \mathcal{I}_{jp+}] - 0.5\rho^{-2}Var(\omega_j \mid \mathcal{I}_{jp+})$ subject to the budget constraint $\omega_j = r\omega_0 + x_j r_e$, where ρ is the risk-aversion parameter, x_j denotes asset holdings, $r_e \equiv (\alpha - rp)$ is the vector of excess returns, and p denotes asset prices. Market clearing requires $\int x_j dj = z$.

The timing of the model is as follows. In period one, investors choose how to allocate their attention to process information about the returns of the risky assets, in period two investors choose portfolios, and in period three payoffs are realized. To solve the model we work backwards; we first solve the optimal portfolio decision for a given posterior information set, and second solve for the optimal posterior information set.

Portfolio selection problem

To solve the portfolio selection problem for a given allocation of attention by investors, we posit and later verify that the equilibrium price vector can be written as $p = A_0 + A_1\alpha + A_2 z$ where A_0, A_1, and A_2 are matrices that are not affected by the realizations of any random variables, and A_2 is nonsingular. In this case, the price vector will be multivariate Gaussian, and can be considered as a signal about asset returns, with the net asset supply acting as contaminating noise.[18] Since all variables are Gaussian, each investor will act so that their

[17] In Kacperczyk et al. (2016), the vector α would refer to the returns of what they term risk factors.

[18] If net asset supply were non-stochastic, this would imply that agents could learn about asset returns

perception of the target, $a_+ = E[\alpha \mid \mathcal{I}_+]$, is also Gaussian, and we can apply a similar procedure to that of Amati (1985) to confirm that the posited price vector exists, where $A_2^{-1}A_1 \equiv \Pi = \rho \int (P_{j+}^{-1} - P_-^{-1})dj$ and $P_{j+} = Var(\alpha \mid \mathcal{I}_{j+})$. We can then compute expected excess returns as $E[r_e \mid \mathcal{I}_{j,+}] = \Theta^{-1}\bar{Z}/\rho$, where $\Theta^{-1} = (P_-^{-1} + \Pi\Omega_z^{-1}\Pi + \Pi/\rho)^{-1}$, and write the optimal portfolio as $x_j = \rho P_{jp+}^{-1} E[r_e \mid \mathcal{I}_{j,+}]$, where $P_{jp+} = Var(\alpha \mid \mathcal{I}_{jp+})$.

Attention allocation problem

Given the solution to the portfolio selection problem, we can now rewrite the attention allocation problem:[19]

$$\max_{P_{j+}^{-1}} \text{const.} + \frac{1}{2} tr(P_{j+}^{-1}W) - \lambda(\ln|P_-| - \ln|P_{j+}|) \tag{18}$$

$$\text{subject to } \alpha \mid \mathcal{I}_- \sim N(a_-, P_-),\ P_{j+} \geq 0,\ P_- - P_{j+} \geq 0$$

where the constant term is independent of the choice variable and the utility matrix $W = \Theta^{-1}(\Theta + \rho^{-2}(\Omega_z + \bar{z}\bar{z}') + \Pi/\rho)\Theta^{-1}$ is fixed from the perspective of an infinitesimal investor.

This problem has a similar form to that of Definition 1, but it has a key difference - whereas the objective function in Definition 1 is concave in the object of choice, in this problem the objective function is convex, so that the solution will be at a corner. Intuitively there are increasing returns to information precision (P_{j+}^{-1}), so that if information was available at a fixed marginal cost the agent would choose to allocate infinite attention. For this reason, we focus on the fixed capacity version of the problem (as do the previous papers), interpreting λ is a Langrange multiplier on the constraint $0.5(\ln|P_-| - \ln|P_{j+}|) \leq \kappa$.

To solve this, we follow similar steps to those in Lemma 2 and the proof of Theorem 1. The solution is $P_{j+}^{-1} = S'\Delta^+ S$, where $LL' = P_-$, $M = L^{-1}$, $S = Q'M$, Q is the matrix of eigenvectors from the eigendecomposition $QDQ' = V \equiv MWM'$, and $\Delta^+ =$

perfectly by observing prices.

[19] Details can be found in section 3 of Mondra (2010) or in section 1.2 of Kacperczyk et al. (2016).

$\text{diag}\{\delta_i^+\}_{i=1}^n$. Whereas in Theorem 1 the elements δ_i^+ were determined by the first-order conditions, here they are determined by corner conditions. In particular, the agent will allocate all available attention to the component with the largest generalized eigenvalue d_i, so that $\delta_1^+ = e^{2\kappa}$ and $\delta_i^+ = 1$ for $i = 2, \dots, n$.[20]

This solution can be understood using the tools of this paper. Investors choose to attend to only one dimension of asset returns (i.e. the rank of the solution is one), corresponding to the canonical synthetic target $\beta_{1,c} = s_1'\alpha$. Here, it is not that the other components of the target are unimportant, but rather that the marginal benefit from attending to this first linear combination is highest.

Equilibrium

In equilibrium, it must be the case that the posterior covariances P_{j+} chosen by investors in the attention allocation step for a given utility matrix W also generate that matrix W. Following the papers cited above, we focus here on symmetric equilibrium, in which each investor chooses the same posterior covariance matrix P_+.

The solution method derived in this paper considerably simplifies the description of equilibrium, because the matrix of generalized eigenvectors S is instrumental in decoupling the entire problem in terms of separate linear combinations of interest. First, it is straightforward to diagonalize $\Pi = \rho(P_+^{-1} - P_-^{-1}) = \rho S'(\Delta^+ - I)S$ and $\Theta = S'TS$ where $T = \Delta^+ + \rho^2(\Delta^+ - I)S\Omega_z^{-1}S'(\Delta^+ - I)$ is a diagonal matrix. Moreover, it is not hard to show that in a symmetric equilibrium, $\Theta = P_{p+}^{-1}$, so that S also can be used to diagonalize P_{p+}^{-1}. Finally, S is also instrumental in diagonalizing the utility matrix W. Letting $R = S^{-1}$, we can rewrite $W = R\left[T^{-1} + T^{-1}(\rho^{-2}Q'L'(\Omega_z + \bar{z}\bar{z}')LQ + (\Delta^+ - I))T^{-1}\right]R'$. In the

[20] If there are multiple generalized eigenvalues d_i that share the maximum value, then the investor is indifferent between allocating all attention to any one of those components. However, the agent will not divide their attention between them, but will select one to receive their full attention. As described in Kacperczyk et al. (2016), this is actually not a knife-edge case because there exists a $\bar{\kappa}$ such that, in equilibrium, if $\kappa > \bar{\kappa}$, then, for example, $d_1 = d_2$. While we will focus on symmetric equilibrium here, handling the case with indentical generalized eigenvalues is essential in understanding the class of asymmetric equilibrium that appear when $\kappa > \bar{\kappa}$.

attention allocation problem, we applied the eigendecomposition $QDQ' = V = MWM'$; for this to be valid, it must be the case that $D \equiv SWS'$ is diagonal. This can be achieved by setting Q to be the eigenvectors of $L'(\Omega_z + \bar{z}\bar{z}')L$.

In summary, any symmetric equilibrium is pinned down by setting Q and D according to:

$$QEQ' = L'(\Omega_z + \bar{z}\bar{z}')L$$
$$D = T^{-1}(T + \rho^{-2}E + \Delta^{+} - I)T^{-1}$$

However, it may be the case that a symmetric equilibrium does not exist. In the attention allocation problem we assumed that the eigenvalues d_i were arranged in nonincreasing order; the existence of a symmetric equilibrium, then, requires that the diagonal elements of the matrix D, as computed above, satisfy this requirement. It is not hard to show, however, that $\partial d_1/\partial \kappa < 0$, and this implies that there exists a capacity value $\bar{\kappa}$ such that $d_1 = d_2$.[21] For any capacity $\kappa > \bar{\kappa}$, constructing the symmetric equilibrium then yields the contradiction $d_1 < d_2$, implying that such an equilibrium does not exist in that case.[22] Asymmetric equilibria can be constructed numerically when $\kappa \geq \bar{\kappa}$, but we leave the analytical characterization as a problem for future research.

Representations and signals

Since the solution is not full rank, in general the fundamental representation will not be feasible. This is important because it implies that the signals chosen by optimizing agents are not compatible with the formulation of Kacperczyk et al. (2016), in which signals are related to the fundamental representation by a nonsingular transformation. Instead, we must consider the reduced canonical representation $y_r = s_1'\alpha + \varepsilon_r$ with $\varepsilon_r \sim N(0, (\delta_1^{+} - 1)^{-1})$, or some nonsingular transformation of it. To demonstrate this, we will show that the solution of Mondria (2010), who also considers optimal signal selection, is simply a

[21] In the context of Kacperczyk et al. (2016), this is formalized in their Lemma 2.

[22] For the case $N = 2$, this existence result is equivalent to the one given by Proposition 2 of Mondria (2010).

transformation of the reduced canonical representation.

While our solution above allows for N risky assets with arbitrary correlations, Mondria (2010) considers the special case in which there are two risky assets and in which there is no correlation within asset returns or within net asset supplies, so that $P_{-} = \text{diag}(\sigma_{r1}^2, \sigma_{r2}^2)$ and $\Omega_z = \text{diag}(\sigma_{z1}^2, \sigma_{z2}^2)$. Since $N = 2$, we can compute a closed form solution for the eigendecomposition $QEQ' = L'(\Omega_z + \bar{z}\bar{z}')L$. Let

$$Y \equiv L'(\Omega_z + \bar{z}\bar{z}')L = \begin{bmatrix} \sigma_{r1}^2(\sigma_{z1}^2 + \bar{z}_1^2) & \sigma_{r1}\sigma_{r2}\bar{z}_1\bar{z}_2 \\ \sigma_{r1}\sigma_{r2}\bar{z}_1\bar{z}_2 & \sigma_{r2}^2(\sigma_{z2}^2 + \bar{z}_2^2) \end{bmatrix} \equiv \begin{bmatrix} y_1 & y_{12} \\ y_{12} & y_2 \end{bmatrix}$$

Then the largest eigenvalue is $e_1 = 0.5\left(y_1 + y_2 + \sqrt{(y_2 - y_1)^2 + 4y_{12}^2}\right)$ and the associated eigenvector is $q_1 = (y_{12}, e_1 - y_1)'$. From this, we can compute the generalized eigenvector defining the reduced canonical representation as:

$$s_1 = M'q_1 = \begin{bmatrix} y_{12}/\sigma_{r1} \\ (e_1 - y_1)/\sigma_{r2} \end{bmatrix} \equiv \begin{bmatrix} s_{11} \\ s_{12} \end{bmatrix}$$

The signal vector considered by Mondria (2010) normalizes the loading on the first fundamental to be equal to one, and we can express that normalization using the representation $y_m = \alpha_1 + c_2\alpha_2 + \varepsilon_m$ where $\varepsilon_m \sim N(0, \Sigma^*)$. This can be constructed from the reduced canonical representation as $y_m = y_r/s_{11}$, and then it is immediate that $c_2 = s_{12}/s_{11}$ and $\Sigma^* = (s_{11}(\delta^+ - 1))^{-1}$. Straightforward algebra confirms that these values are identical to those reported in Proposition 2 of Mondria (2010).

6 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.

We use our method to setup and solve two important applications of rational inattention considered in the literature: the optimal price-setting problem and the optimal portfolio choice problem. We first show how to solve the static price-setting problem for the first best solution and how to interpret the solution. Next, we show how to incorporate the independence assumption, commonly used in the literature, and describe how this restriction prevents agents from achieving the first best outcome. Finally, we consider the optimal portfolio choice problem, and are able to solve the full N asset problem with arbitrary correlations, including the optimal selection of the form of signals received by agents. In doing so, we show that rationally inattentive agents will choose signals that are not compatible with those imposed in the previous literature.

7 Appendices

7.1 Appendix A: Proofs

7.1.1 Proof of Lemma 1

Simultaneously diagonalize $P_- = S'IS$ and $P_+ = S'NS$ as in Theorem 7.6.4 of Horn and Johnson (2012). Then:

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

7.1.2 Proof of Lemma 2

This is a straightforward application of Theorem 7.6.4 of Horn and Johnson (2012).

7.1.3 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:

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

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:

$$\begin{aligned}\mathcal{O} &= tr(WP_+) + \lambda \sum_{i=1}^n \ln \frac{1}{n_i} \\ &= tr(WLZNZ'L') - \lambda \sum_{i=1}^n \ln n_i \\ &= tr(Z'VZN) - \lambda \sum_{i=1}^n \ln n_i\end{aligned}$$

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 $QDQ' = V$), for any matrix N. Thus the optimal Z contains the eigenvectors of $V = L'WL$. This also implies that $X = S = Q'M$.

This allows us to further simply the objective function:

$$\begin{aligned}\mathcal{O} &= tr(Q'VQN) - \lambda \sum_{i=1}^n \ln n_i \\ &= tr(Q'(QDQ')QN) - \lambda \sum_{i=1}^n \ln n_i \\ &= tr(DN) - \lambda \sum_{i=1}^n \ln n_i \\ &= \sum_{i=1}^n d_i n_i - \lambda \sum_{i=1}^n \ln n_i\end{aligned}$$

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 $i = 1, \ldots, n$.

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

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

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 $n_i \to \infty$, 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:

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

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

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

7.1.4 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:

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

7.1.5 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 $s_i'RN^+ = n_i^+ s_i'R$.

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

7.1.6 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 $d_i > \lambda$ for $i = 1, \ldots, r$ and $d_i \leq \lambda$ for $i = r+1, \ldots, n$. 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 $i = 1, \ldots, r$, 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 $\kappa_i = \frac{1}{2} \log_b 1 = 0$.

7.1.7 Proof of Corollary to Theorem 2

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:

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

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 $(d_1, \ldots, d_r)$, and by assumption we have $d_1 \geq d_2 \geq \cdots \geq d_r \geq \cdots \geq d_n$. 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 the 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 for Part (a) and Part (b).

7.1.8 Proof of Lemma 3

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

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

Part (c):

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

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

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

7.1.9 Proof of Lemma 4

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 $(N^+)^{-1} = P_-P_+^{-1}$, or $\frac{1}{n_i^+} = \frac{P_{ii,-}}{P_{ii,+}}$.

7.1.10 Proof of Lemma 5

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 $E[\beta_{i,c} \mid \mathcal{I}_-] = b_{i,c,-}$ and Var$(\beta_{i,c} \mid \mathcal{I}_-) = 1$. 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.

7.1.11 Proof of Theorem 3

This follows directly from Definition 2 and Lemma 5.

7.1.12 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:

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

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 $P_- - P_+ \geq 0 \iff B(P_- - P_+)B' = O_- - O_+ \geq 0$.

7.1.13 Proof of Lemma 6

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

7.1.14 Proof of Lemma 7

Part (a):

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

Part (b):

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

7.1.15 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.

References

Admati, A. R. (1985). A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets. *Econometrica 53*(3), 629.

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory*. John Wiley & Sons.

Fulton, C. (2017). Mechanics of Linear Quadratic Gaussian Rational Inattention Tracking Problems. Technical Report 2017-109, Board of Governors of the Federal Reserve System (U.S.).

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.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2016). A Rational Theory of Mutual Funds' Attention Allocation. *Econometrica 84*(2), 571–626.

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

Matêjka, 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.

Mondria, J. (2010). Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory 145*(5), 1837–1864.

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.