Multivariate Rational Inattention*

Jianjun Miao[†] Jieran Wu[‡] Eric Young[§]

January 18, 2019

Abstract

We study optimal control problems in the multivariate linear-quadratic-Gaussian framework under rational inattention and show that the multivariate attention allocation problem can be reduced to a dynamic tracking problem in information theory. We propose a general solution method to solve this problem by using rate distortion function and semidefinite programming and derive the optimal form and dimension of signals without strong prior restrictions. We provide generalized reverse water-filling solutions for some special cases. Applying our method to solve three multivariate economic models, we obtain some results qualitatively different from the literature.

Keywords: Rational Inattention, Endogenous Information Choice, Rate Distortion Functions, Dynamic Tracking Problem, Optimal Control

JEL Classifications: C6, D8, E2.

*We thank Yulei Luo, Lin Peng, Giacomo Rondina, Gianluca Violante, Pengfei Wang, Wei Xiong, He Yang, and Tao Zha for helpful comments. We also thank the participants at the AFR Summer Institute of Economics and Finance for comments.

[†]Department of Economics, Boston University, miaoj@bu.edu

[‡]Academy of Financial Research and College of Economics, Zhejiang University, jw5ya@zju.edu.cn

[§]Department of Economics, University of Virginia, ey2d@virginia.edu

1 Introduction

Humans have limited capacity to process information when making decisions. People often ignore some pieces of information and pay attention to some others. In seminal contributions, Sims (1998, 2003) formalizes limited attention as a constraint on information flow and models decision-making with limited attention as optimization subject to this constraint. Such a framework for rational inattention (RI) has wide applications in economics as surveyed by Sims (2011) and Maćkowiak, Matějka, and Wiederholt (2018). Despite the rapid growth of this literature, most theories and applications are limited to univariate models.

Multivariate RI models are difficult to analyze both theoretically and numerically, especially in dynamic settings. Because many economic decision problems involve multivariate states and multivariate choices, it is of paramount importance to make progress in this direction as Sims (2011) points out. Our paper contributes to the literature by developing a framework for analyzing multivariate RI problems in a linear-quadratic-Gaussian (LQG) control setup.[1] The LQG control setup has a long tradition in economics and can deliver analytical results to understand economic intuition. It is also useful to derive numerical solutions for approximating nonlinear dynamic models (Kydland and Prescott (1982)). We formulate the LQG control problem under RI in both finite- and infinite-horizon setups as a problem of choosing both control and information structure. The decision maker observes a noisy signal about the unobserved controlled states. The signal vector is a linear transformation of the states plus a noise. The signal dimension, the linear transformation, and the noise covariance matrix are all endogenously chosen subject to the information-flow constraint.

Our second contribution is to develop an efficient three-step solution procedure. The first step is to apply the certainty equivalence principle to derive the optimal control. This step follows from the standard control literature. The second step is to transform the control problem under RI into a tracking problem under RI, in which only the information structure needs to be solved. The tracking objective is to minimize the weighted mean squared error with the weighting matrix endogenously derived from the first step. In the last step we transform the tracking problem under RI into a rate distortion problem in information theory (Cover and Thomas (2006)). We then use the semidefinite programming approach (Vandenberghe, Boyd, and Wu (1998) and Tanaka et al (2017)) to solve this problem to obtain the optimal information structure for any information-flow rate. This approach can be implemented using the publicly available semidefinite programming solver SDPT3 (Toh, Todd, and Tutuncu (1999) and Tutuncu, Toh, and Todd (2003)). This solver can handle optimization problems up to 100 dimensions accurately, robustly, and efficiently.

We are able to derive three sets of novel characterization results. First, we provide a full characterization for the static RI problem. We prove that the optimal information structure is

[1] See Sims (2006), Matějka and McKay (2015), and Caplin, Dean, and Leahy (2018) for static non-Gaussian RI models.

characterized by a generalized reverse water-filling solution. This result generalizes Theorem 10.3.3 in Cover and Thomas (2006, p. 314) by allowing for semidefinite weighting matrix in the tracking objective and correlated state uncertainty. In this case the decision maker attends more to the larger eigenvalue of the weighted prior covariance matrix, instead of the prior variance. We prove that the optimal signal dimension is equal to the rank of the weighting matrix when the information-flow rate is sufficiently large or the tracking error (distortion) is sufficiently small. The signal dimension decreases as the information-flow rate decreases. Since the rank of the weighting matrix cannot exceed the minimum of the state and control dimensions, the signal dimension can never exceed this minimum. We also show that initially independent states are ex post correlated when the weighting matrix is not identity as Sims (2011) notices using numerical examples.

Second, we prove that a similar generalized reverse water-filling solution and its implied properties also apply to the dynamic multivariate RI problem when the state transition matrix is diagonal with equal lag coefficients. This includes two special cases: (i) the state vector is serially independently and identically distributed (IID), conditional on a control, and (ii) all states are equally persistent AR(1) processes with correlated innovations.

Third, we prove that the optimal signal is one dimensional in the special case in which the rank of the weighting matrix is equal to one. This case happens when there is only one control variable, even though there are multiple states. The optimal signal is equal to the optimal control under full information plus a noise in the static case and in the dynamic case, in which the state transition matrix is diagonal with equal lag coefficients.

Based on extensive numerical experiments for the general dynamic models, we find that some of the preceding results still apply. For example, the signal dimension cannot exceed the minimum of the state and control dimensions and decreases as the information-flow rate decreases. Moreover, both the weight of the state in tracking variables and its innovation variance are important for attention allocation. Notably the persistence of the state, instead of innovation variance, plays a dominant role in allocating attention and determining the responsiveness to shocks.

Our third contribution is to apply our results to three economic problems. Our first application is the price setting problem studied by Maćkowiak and Wiederholt (2009), in which there are two exogenous state variables representing two sources of uncertainty. The profit-maximizing price is equal to a linear combination of the two shocks. We ignore the general equilibrium effect and just focus on the decision problem. Because there is only one choice variable (i.e., price), the optimal signal is one dimensional and can be normalized as the profit-maximizing price plus a noise in the dynamic case when the persistence is the same for all states. This result implies that the initial price responses to a shock with the same size to different states are the same, independent of the state's innovation variance, when the profit-maximizing price puts equal weight on the states.

By contrast, when the persistence is not identical, the optimal signal is still one dimensional and equal to a linear combination of the states, but it does not take the preceding form. We find that the firm attends more to a more persistent state by allocating a larger relative weight to that state in the signal. Moreover, the initial responses are larger to a more persistent state independent of its innovation variance relative to the other state within an economically reasonable range of parameter values.

The results above are in sharp contrast to those reported by Maćkowiak and Wiederholt (2009). They assume that the firm receives a separate signal for a different source of uncertainty. While they justify this signal independence assumption by bounded rationality and tractability, this assumption is not innocuous because it is suboptimal for the original RI problem and also leads to some qualitatively different predictions.

Our second application is the consumption/saving problem analyzed by Sims (2003), in which there is an endogenous state variable (wealth) and two exogenous persistent state variables (income shocks). We find that the optimal signal is still one dimensional even for a very large information-flow rate. The initial responses of consumption are larger for a more persistent income shock, independent of its innovation variance relative to other shocks. This result is consistent with that in the pricing example, but in sharp to Sims's (2003) finding. His Figures 7 and 8 show that the initial consumption responses to the less persistent income shock with a larger innovation variance is larger. Moreover, he assumes that the optimal signal is three dimensional.

Our last application is the firm investment problem in which the firm makes both tangible and intangible capital investment. We find that the signal dimension drops from two to one as the information-flow rate gradually decreases. Sims (1998, 2003) argues that an information-flow constraint can substitute for adjustment costs in a dynamic optimization problem. Our numerical results show that this constraint can generate inertia and delayed responses of investment to shocks, just like capital adjustment costs. Moreover, we find that RI combined with capital adjustment costs can generate hump-shaped investment responses.

We now discuss the related literature. Sims (2003) is the first paper that studies multivariate RI models in the LQG setup.[2] His framework and solution methods are different from ours. He formulates the LQG RI problem without explicitly reference to the signal structure. His solution procedure consists of two steps. The first step is to use the certainty equivalence principle to derive the optimal control. This step is the same as ours. The second step is to transform the control problem under RI into a tracking problem under RI.[3] This problem can be further transformed into a problem of choosing an optimal conditional covariance matrix for the state vector. This is a deterministic nonlinear convex optimization problem, which is the same as in our second step.

[2]Sims (2011) provides a summary of his approach. Luo (2008), Luo and Young (2010), and Luo, Nie and Young (2015) follow Sims's approach closely, but mainly focus on univariate models. Sims's approach is correct in the univariate case.

[3]We find that the weighting matrix presented in Sims (2003) is incorrect because it may not be semidefinite.

Sims suggests to first derive the first-order conditions for this problem, ignoring the no-forgetting constraint. After solving the first-order conditions using a nonlinear equation solver, check whether the no-forgetting constraint binds. If it does not bind, then we obtain the solution. Otherwise, apply a Cholesky decomposition to perform a change of variable. Derive the first-order condition for the new variable and solve this first-order condition using a nonlinear equation solver. Sims (2003) suggests the optimal signal is typically equal to the state vector plus a noise. The impulse responses are generated by the Kalman filter based on this signal vector. He does notice the possibility that the signal vector is only equal to a linear combination of a subset of state variables. But he does not offer an explicit solution.

Solving first-order conditions using nonlinear equation methods is numerically inefficient and not robust. The convergence is highly sensitive to the initial condition. More importantly, Sims's approach does not solve for the optimal information structure. His choice of the signal as the state plus noise is typically suboptimal for multivariate RI problems so that the impulse response functions generated by that signal vector are incorrect. Based on our theoretical and numerical results discussed earlier, we find that the no-forgetting constraint often binds and the signal dimension does not exceed the minimum of the control and state dimensions.

Because of the difficulty of the multivariate RI models, researchers often make simplifying assumptions. For example, Peng (2005), Peng and Xiong (2006), Maćkowiak and Wiederholt (2009, 2015), Van Niewerburgh and Veldkamp (2010), and Zorn (2018) impose the signal independence assumption or some restriction on the signal form. An undesirable implication of this assumption is that initially independent states remain ex post independent. Mondria (2010) and Kacperczyk, Van Niewerburgh, and Veldkamp (2016) remove this assumption in static finance models. The former paper considers only two independent assets (states), while the latter studies the case of many assets given some invertibility restriction on the signal form. Except for Maćkowiak and Wiederholt (2009, 2015) and Zorn (2018), all these papers study static models.

In independent work Fulton (2018) analyzes similar multivariate RI problems in the static case and derives results similar to our generalized reverse water-filling solution. Fulton (2017) discusses dynamic RI tracking problems and proposes an approximation method for the special case of low information costs (or high information-flow rate).[4] Maćkowiak, Matějka, and Wiederholt (2018) study a dynamic tracking problem with one control and one exogenous state, which follows an ARMA(p,q) process. They also briefly discuss the extension to the case with multiple exogenous states, but still with one control. Consistent with our result, the optimal signal is one dimensional. Our approach is different from those in these three papers and applies to general dynamic LQG control problems under RI with both multiple states and multiple controls.

[4]We would like to thank Gianluca Violante for pointing out Fulton's papers to us, when we presented a preliminary version of our paper in a conference in the summer of 2018.

2 LQG Control Problems with Rational Inattention

We start with a finite-horizon linear-quadratic control problem under rational inattention. Let the n_x-dimensional state vector x_t follow the linear dynamics

$$x_{t+1} = A_t x_t + B_t u_t + \epsilon_{t+1}, \; t = 0, 1, ..., T, \tag{1}$$

where u_t is an n_u dimensional control variable and ϵ_{t+1} is a serially independent Gaussian random vector with mean zero and covariance matrix W_t. The matrix W_t is positive semidefinite, denoted by $W_t \succeq 0$.[5] The state transition matrix A_t and the control coefficient matrix B_t are deterministic and conformable. The state vector x_t may contain both exogenous states such as AR(1) shocks and endogenous states such as capital.

Suppose that the decision maker does not observe the state x_t perfectly, but observes a multi-dimensional noisy signal s_t about x_t given by

$$s_t = C_t x_t + v_t, \; t = 0, 1, ..., T, \tag{2}$$

where C_t is a conformable deterministic matrix and v_t is a serially independent Gaussian random variable with mean zero and covariance matrix $V_t \succ 0$. Assume that x_0 is a Gaussian random variable with mean $\bar{x}_0$ and covariance matrix Σ_0. The random variables ϵ_t, v_t, and x_0 are all mutually independent for all t. The decision maker's information set at date t is generated by $s^t = \{s_0, s_1, ..., s_t\}$. The control u_t is measurable with respect to s^t.

Suppose that the decision maker is boundedly rational and has limited information-processing capacity. He faces the following information-flow constraint[6]

$$\sum_{t=0}^{T} I\left(x_t; s_t | s^{t-1}\right) \leq \kappa, \tag{3}$$

where $\kappa > 0$ denotes the information-flow rate or channel capacity and $I\left(x_t; s_t | s^{t-1}\right)$ denotes the conditional (Shannon) mutual information between x_t and s_t given s^{t-1},

$$I\left(x_t; s_t | s^{t-1}\right) \equiv H\left(x_t | s^{t-1}\right) - H\left(x_t | s^t\right).$$

Here $H(\cdot|\cdot)$ denotes the conditional entropy operator.[7] Let $s^{-1} = \varnothing$. Intuitively, entropy measures uncertainty. At each time t, given past information s^{t-1}, observing s_t reduces uncertainty about x_t. The total uncertainty reduction from time 0 to time T is measured by the expression on the

[5] We use the conventional matrix inequality notations: $W \succ (\succeq) \widetilde{W}$ means that $W - \widetilde{W}$ is positive definite (semidefinite) and $W \prec (\preceq) \widetilde{W}$ means $W - \widetilde{W}$ is negative definite (semidefinite).

[6] This constraint is equivalent to $I\left(x^T; s^T\right) \leq \kappa$. One may consider the per period capacity by dividing by $1/T$. Sims (2011) introduces discounting to the conditional mutual information in (3). Our approach applies to this formulation with suitable changes.

[7] See Cover and Thomas (2006) or Sims (2011) for the definitions of entropy, conditional entropy, mutual information, and conditional mutual information.

left-hand side of the inequality in (3). The decision maker can process information by choosing the information structure represented by $\{C_t, V_t\}_{t=0}^T$ for the signal s_t, but the rate of uncertainty reduction is limited by an upper bound κ. For example, if C_t is equal to the identity matrix and $V_t = 0$ for all t, then the decision maker fully observes x_t and hence $I\left(x_t; s_t|s^{t-1}\right) = \infty$, violating the information-flow constraint (3).

Notice that the choice of $\{C_t, V_t\}_{t=0}^T$ implies that the dimension of the signal vector s_t and the correlation structure of the noise v_t are endogenous and may vary over time. One may imagine the decision maker makes decisions sequentially. He first chooses the information structure $\{C_t, V_t\}_{t=0}^T$ and then selects a control $\{u_t\}_{t=0}^T$ adapted to $\left\{s^t\right\}$ to maximize an objective function. Suppose that the objective function is quadratic. We are ready to formulate the decision maker's problem as follows:

Problem 1 *(Finite-horizon LQG problem under RI)*

$$\max_{\{u_t\},\{C_t\},\{V_t\}} -\mathbb{E}\left[\sum_{t=0}^{T} \beta^t \left(x_t' Q_t x_t + u_t' R_t u_t + 2x_t' S_t u_t\right) + \beta^{T+1} x_{T+1}' P_{T+1} x_{T+1} \middle| s_0\right]$$

subject to (1), (2), and (3).

The parameter $\beta \in (0, 1]$ represents the discount factor. The deterministic matrices Q_t, R_t, and S_t for all t and P_{T+1} are conformable and exogenously given. For the infinite-horizon stationary case, we set $T \to \infty$ and remove the time index for all exogenously given matrices A_t, B_t, Q_t, R_t, and S_t. We also replace (3) by

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} I\left(x_t; s_t|s^{t-1}\right) \leq \kappa, \tag{4}$$

where

$$s_t = Cx_t + v_t \tag{5}$$

and v_t is a serially independent Gaussian random variable with mean zero and covariance matrix $V \succ 0$. The interpretation is that the average rate of uncertainty reduction per period is limited by $\kappa > 0$. We now formulate the infinite-horizon problem as follows:

Problem 2 *(Infinite-horizon stationary LQG problem under RI)*

$$\max_{\{u_t\},C,V} -\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \left(x_t' Q x_t + u_t' R u_t + 2x_t' S u_t\right)\right]$$

subject to (4), (5), and

$$x_{t+1} = Ax_t + Bu_t + \epsilon_{t+1}, \tag{6}$$

for $t \geq 0$. Here the expectation is taken with respect to the unconditional stationary distribution.

In applications, it may also be more convenient to consider the following relaxed problems.

Problem 3 *(Finite-horizon relaxed LQG problem under RI)*

$$\max_{\{u_t\},\{C_t\},\{V_t\}} \quad -\mathbb{E}\left[\sum_{t=0}^{T}\beta^t\left(x_t'Q_tx_t+u_t'R_tu_t+2x_t'S_tu_t\right)+\beta^{T+1}x_{T+1}'P_{T+1}x_{T+1}\middle| s_0\right] \\ -\lambda\sum_{t=0}^{T}I\left(x_t;s_t|s^{t-1}\right)$$

subject to (1) and (2).

Problem 4 *(Infinite-horizon stationary relaxed LQG problem under RI)*

$$\max_{\{u_t\},C,V} -\mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t\left(x_t'Qx_t+u_t'Ru_t+2x_t'Su_t\right)\right]-\lambda\limsup_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T}I\left(x_t;s_t|s^{t-1}\right)$$

subject to (5) and (6) for $t=0,1,...,$ where the expectation is taken with respect to the unconditional stationary distribution.

In these two problems $\lambda>0$ can be interpreted as the Lagrange multiplier associated with the information-flow constraint or the shadow price (cost) of the information flow. We will focus our analysis on Problem 1. The solution to Problem 2 is the limit of that to Problem 1. Problems 3 and 4 can be similarly analyzed and are easier to solve because the information-flow constraint is removed from the optimization.

2.1 Full Information Case

Before analyzing Problem 1, we first present the solution in the full information case, in which the decision maker observes x_t perfectly. The solution can be found in the textbooks by Ljungqvist and Sargent (2004) and Miao (2014). Suppose that $P_{T+1}\succeq 0$, $R_t\succ 0$, and

$$\begin{bmatrix} Q_t & S_t \\ S_t' & R_t \end{bmatrix}\succeq 0$$

for all $t=0,1,...,T$. Then the value function takes the form

$$v_t^{FI}\left(x_t\right)=-x_t'P_tx_t-g_t, \tag{7}$$

where $P_t\succeq 0$ and satisfies

$$\begin{aligned} P_t \;&=\; Q_t+\beta A_t'P_{t+1}A_t \\ &\quad -\left(\beta A_t'P_{t+1}B_t+S_t\right)\left(R_t+\beta B_t'P_{t+1}B_t\right)^{-1}\left(\beta B_t'P_{t+1}A_t+S_t'\right), \end{aligned} \tag{8}$$

and g_t satisfies

$$g_t=\beta\mathrm{tr}\left(P_{t+1}W_t\right)+\beta g_{t+1},\;\; g_{T+1}=0,$$

for $t = 0, 1, ..., T$. Here $\mathrm{tr}(\cdot)$ denotes the trace operator.

The optimal control is

$$u_t = -F_t x_t, \tag{9}$$

where

$$F_t = \left(R_t + \beta B_t' P_{t+1} B_t\right)^{-1} \left(S_t' + \beta B_t' P_{t+1} A_t\right). \tag{10}$$

For the infinite horizon case, all exogenous matrices are time invariant. As $T \to \infty$, we obtain the infinite-horizon solution under some standard stability conditions. The value function is given by

$$v^{FI}(x_t) = -x_t' P x_t - g,$$

where $P \succeq 0$ and satisfies

$$P = Q + \beta A' P A - \left(\beta A' P B + S\right) \left(R + \beta B' P B\right)^{-1} \left(\beta B' P A + S'\right),$$

and

$$g = \frac{\beta}{1-\beta} \mathrm{tr}\,(PW).$$

The optimal control is given by

$$u_t = -F x_t, \tag{11}$$

where

$$F = \left(R + \beta B' P B\right)^{-1} \left(S' + \beta B' P A\right).$$

2.2 Tracking Problem

We solve Problem 1 in three steps. In the first step we observe that Problem 1 is a standard LQG problem under partial information for fixed $\{C_t, V_t\}_{t=0}^{T}$. Thus the usual certainty equivalence principle holds. This implies that the optimal control is given by

$$u_t = -F_t \widehat{x}_t, \tag{12}$$

where $\widehat{x}_t \equiv \mathbb{E}\left[x_t | s^t\right]$ denotes the estimate of x_t given information s^t. The state under the optimal control satisfies

$$x_{t+1} = A_t x_t - B_t F_t \widehat{x}_t + \epsilon_{t+1}. \tag{13}$$

By the Kalman filtering formula, $\widehat{x}_t$ follows the dynamics

$$\widehat{x}_t = \widehat{x}_{t|t-1} + \Sigma_{t|t-1} C_t' \left(C_t \Sigma_{t|t-1} C_t' + V_t\right)^{-1} \left(s_t - C_t \widehat{x}_{t|t-1}\right), \tag{14}$$

$$\widehat{x}_{t|t-1} = \left(A_{t-1} - B_{t-1} F_{t-1}\right) \widehat{x}_{t-1}, \tag{15}$$

where $\widehat{x}_{t|t-1} \equiv \mathbb{E}\left[x_t|s^{t-1}\right]$ with $\widehat{x}_{0|-1} = \bar{x}_0$ and $\Sigma_{t|t-1} \equiv \mathbb{E}\left[\left(x_t - \widehat{x}_{t|t-1}\right)\left(x_t - \widehat{x}_{t|t-1}\right)' |s^{t-1}\right]$ with $\Sigma_{0|-1} = \Sigma_0$. Moreover

$$\Sigma_{t|t-1} = A_{t-1}\Sigma_{t-1|t-1}A'_{t-1} + W_{t-1}, \tag{16}$$

$$\Sigma_{t|t} = \left(\Sigma^{-1}_{t|t-1} + \Phi_t\right)^{-1}, \tag{17}$$

for $t = 0, 1, ..., T$, where $\Sigma_{t|t} \equiv \mathbb{E}\left[\left(x_t - \widehat{x}_t\right)\left(x_t - \widehat{x}_t\right)' |s^t\right]$ and Φ_t denotes the "signal-to-noise ratio" defined by

$$\Phi_t = C'_t V_t^{-1} C_t \succeq 0$$

for $t = 0, 1, ..., T$.

We now solve for $\{C_t, V_t\}_{t=0}^T$ in the remaining two steps. In step 2 we show that $\{C_t, V_t\}_{t=0}^T$ is determined by a tracking problem. We present all technical proofs in Appendix A.

Proposition 1 *Suppose that the assumptions in Section 2.1 are satisfied. Then the optimal control for Problem 1 is given by (12) and the optimal* $\{C_t, V_t\}_{t=0}^T$ *solves the following problem:*

$$\min_{\{C_t\},\{V_t\}} \sum_{t=0}^{T} \mathbb{E}\left[(x_t - \widehat{x}_t)'\Omega_t(x_t - \widehat{x}_t) \mid s_0\right] \tag{18}$$

subject to (2), (3), and (13), where

$$\Omega_t = \beta^t F'_t(R_t + \beta B'_t P_{t+1} B_t)F_t \succeq 0. \tag{19}$$

The intuition behind this result is as follows. Since the information-flow constraint cannot help in the optimization, we know that $\mathbb{E}\left[v_t^{FI}(x_t)|s^t\right] > \widehat{v}_t(\widehat{x}_t)$. Following Sims (2003), we solve for $\{C_t, V_t\}_{t=0}^T$ by minimizing $\mathbb{E}\left[v_0^{FI}(x_0)|s_0\right] - \widehat{v}_0(\widehat{x}_0)$. In other words, we choose information structure so as to bring expected utility from the current date onward as close as possible to the expected utility value under full information. The proposition shows that $\mathbb{E}\left[v_0^{FI}(x_0)|s_0\right] - \widehat{v}_0(\widehat{x}_0)$ is equal to the expression in (18).

Notice that the matrix Ω_t is positive semidefinite because $R_t \succ 0$ and $P_{t+1} \succeq 0$. Since F_t is an n_u by n_x dimensional matrix, rank(Ω_t) does not exceed the minimum of the dimension n_x of the state vector and the dimension n_u of the control vector. Thus it is possible that Ω_t is singular. If $n_x \geq n_u$ and F_t has full column rank, then rank$(\Omega_t) = n_u$. If $n_x < n_u$ and F_t has full row rank, then rank$(\Omega_t) = n_x$.

In the infinite-horizon stationary case, suppose that the initial state is drawn from the stationary distribution, and so we use the stationary Kalman filter. We can then compute the limit

$$\lim_{T\to\infty}\sum_{t=0}^{T}\mathbb{E}\left[(x_t - \widehat{x}_t)'\Omega_t(x_t - \widehat{x}_t)\right] = \frac{1}{1-\beta}\mathbb{E}\left[(x_t - \widehat{x}_t)'\Omega(x_t - \widehat{x}_t)\right],$$

where the expectation is taken for the stationary distribution and

$$\Omega = F'(R + \beta B'PB)F. \tag{20}$$

We have the following result.

Proposition 2 *Suppose that the assumptions in Section 2.1 hold. The optimal control for Problem 2 is given by $u_t = -F\widehat{x}_t$, and the optimal C and V solve the following problem*

$$\min_{C,V} \quad \mathbb{E}\left[(x_t - \widehat{x}_t)'\Omega(x_t - \widehat{x}_t)\right]$$

subject to (4), (5), and

$$x_{t+1} = Ax_t - BF\widehat{x}_t + \epsilon_{t+1}, \tag{21}$$

where $\Omega \succeq 0$ is given by (20).

The previous two propositions state that an optimal information structure in the control problem under RI can be found by solving tracking problems. It is important that the weighting matrices Ω_t and Ω must be positive semidefinite, which ensure that the tracking problems are well-defined.

2.3 Rate Distortion Function

In the final step of our solution procedure, we transform the tracking problem into a simpler equivalent problem. We will draw connection to the engineering literature on information theory (Cover and Thomas (2006)). In this literature the objective function in (18) measures the distance between sequences x^T and $\widehat{x}^T$. It is a distortion measure for the source random sequence x^T and its estimate $\widehat{x}^T$. Define the function $D_T(\kappa)$ as the minimized distortion for all rates $\kappa > 0$ such that $D_T(\kappa)$ is finite. This function is called a distortion rate function in information theory and is equal to the minimized objective function in (18) in our context.

A closely related concept is the (information) rate distortion function $\kappa_T(D)$, defined as the minimized information-flow rate for all distortions $D > 0$ such that $\kappa_T(D)$ is finite. It is determined by the following problem in our context.

Problem 5 *(Finite-horizon rate distortion problem)*

$$\kappa_T(D) \equiv \min_{\{C_t,V_t\}_{t=0}^T} \sum_{t=0}^{T} I\left(x_t; s_t | s^{t-1}\right) \tag{22}$$

subject to (1), (2), and

$$\sum_{t=0}^{T} \mathbb{E}\left[(x_t - \widehat{x}_t)'\Omega_t(x_t - \widehat{x}_t) \mid s_0\right] \leq D. \tag{23}$$

It can be checked that both $\kappa_T(D)$ and $D_T(\kappa)$ are decreasing and convex functions. Thus, for any $\kappa > 0$, we can find a $D > 0$ such that the solution to the rate distortion problem in (22) gives the solution to the distortion rate problem or the tracking problem in (18) (Cover and Thomas (2006)). The rate distortion problem is easier to solve numerically because the complex information-flow constraint is moved into the objective function so that the problem has a semidefinite programming representation as shown in Section 3.[8] We thus focus our analysis on the problem in (22).

To solve this problem, we compute the mutual information[9]

$$\begin{aligned} I\left(x_t; s_t | s^{t-1}\right) &= H\left(x_t | s^{t-1}\right) - H\left(x_t | s^t\right) \\ &= \frac{1}{2}\log\det\left(A_{t-1}\Sigma_{t-1|t-1}A'_{t-1} + W_{t-1}\right) - \frac{1}{2}\log\det\left(\Sigma_{t|t}\right) \end{aligned}$$

for $t = 1, 2, .., T$, and

$$I\left(x_0; s_0 | s^{-1}\right) = H\left(x_0\right) - H\left(x_0 | s_0\right) = \frac{1}{2}\log\det\left(\Sigma_0\right) - \frac{1}{2}\log\det\left(\Sigma_{0|0}\right)$$

for $t = 0$, where the functions $H(\cdot)$ and $H(\cdot|\cdot)$ denotes entropy and conditional entropy operators, and $\det(\cdot)$ denotes the determinant operator. Moreover, we compute

$$\sum_{t=0}^{T} \mathbb{E}\left[(x_t - \widehat{x}_t)'\Omega_t(x_t - \widehat{x}_t) \mid s_0\right] = \sum_{t=0}^{T} \operatorname{tr}\left(\Omega_t \Sigma_{t|t}\right).$$

Thus the rate distortion problem becomes

$$\begin{aligned} \kappa_T(D) = \min_{\{\Phi_t \succeq 0\}, \{\Sigma_{t|t}\}} \quad & \frac{1}{2}\log\det\left(\Sigma_0\right) - \frac{1}{2}\log\det\left(\Sigma_{0|0}\right) \\ & + \sum_{t=1}^{T} \frac{1}{2}\log\det\left(A_{t-1}\Sigma_{t-1|t-1}A'_{t-1} + W_{t-1}\right) - \frac{1}{2}\log\det\left(\Sigma_{t|t}\right) \end{aligned} \tag{24}$$

subject to

$$\Sigma_{t|t} = \left[\left(A_{t-1}\Sigma_{t-1|t-1}A'_{t-1} + W_{t-1}\right)^{-1} + \Phi_t\right]^{-1}, \quad t = 1, 2, ..., T, \tag{25}$$

$$\Sigma_{0|0} = \left(\Sigma_0^{-1} + \Phi_0\right)^{-1}, \tag{26}$$

$$\sum_{t=0}^{T} \operatorname{tr}\left(\Omega_t \Sigma_{t|t}\right) \leq D. \tag{27}$$

Equation (25) follows from (16) and (17). Instead of choosing $\{C_t\}$ and $\{V_t\}$ directly, we view problem (24) as a standard optimal control problem with $\Sigma_{t|t}$ as the state variable and Φ_t as the control variable. After obtaining a solution for $\{\Phi_t\}_{t=0}^{T}$ to this problem, we can recover $\{C_t\}$ and $\{V_t\}$ from the following result.

[8]The relaxed Problems 3 and 4 are also easier to solve than the original Problems 1 and 2.

[9]The usual base for logarithm in the entropy formula is 2, in which case the unit of information is a "bit." In this paper we adopt natural logarithm, in which case the unit is called a "nat."

Proposition 3 *Given an optimal sequence $\{\Phi_t\}_{t=0}^T$ determined from problem (24), an optimal information structure $\{C_t, V_t\}_{t=0}^T$ satisfies $\Phi_t = C_t' V_t^{-1} C_t$. A particular solution is that $V_t = diag\left(\varphi_{it}^{-1}\right)_{i=1}^{m_t}$ and the m_t columns of $n_x \times m_t$ matrix C_t' are orthonormal eigenvectors for all positive eigenvalues of Φ_t, denoted by $\{\varphi_{it}\}_{i=1}^{m_t}$. The optimal dimension of the signal vector s_t is equal to $rank\,(\Phi_t) = m_t \leq n_x$.*

This proposition shows that the optimal information structure $\{C_t, V_t\}_{t=0}^T$ is not unique and can be computed by the singular-value decomposition. The optimal signal can always be constructed such that the components in the noise vector v_t of the signal s_t are independent. However, the optimal signal components are in general not independent in the sense that the matrix C_t may not be diagonal or invertible. The signal independence is a common assumption employed in the literature (e.g., Maćkowiak and Wiederholt (2009)), but we show that this assumption can be restrictive and lead to suboptimal solutions.

3 Semidefinite Programming

In this section we provide efficient methods to solve the tracking or rate distortion problems in both finite- and infinite-horizon settings. We will adopt the semidefinite programming approach recently proposed by Tanaka et al (2017). The key idea of this approach is that a dynamic rate distortion problem can be transformed into a determinant maximization problem subject to linear matrix inequality constraints. The latter problem is convex and can be solved efficiently by standard optimization methods developed in the mathematics literature (Vandenberghe, Boyd, and Wu (1998)). This approach can be implemented using the publicly available semidefinite programming solver SDPT3 described in Appendix C (Toh, Todd, and Tutuncu (1999) and Tutuncu, Toh, and Todd (2003)).[10] We also provide some characterization results in this section.

3.1 Finite-Horizon Problems

We follow the procedure in Tanaka et al (2017) closely, which consists of the following three steps. The first step is to transform (24) into an optimization problem in terms of $\left\{\Sigma_{t|t}\right\}$ only. We eliminate the control $\{\Phi_t\}$ and replace (25) and (26) by linear inequality constraints

$$0 \prec \Sigma_{t|t} \preceq A_{t-1}\Sigma_{t-1|t-1}A_{t-1}' + W_{t-1},\ t = 1, 2, ..., T, \tag{28}$$

and

$$0 \prec \Sigma_{0|0} \preceq \Sigma_0. \tag{29}$$

Equations (28) and (29) give the no-forgetting constraints discussed by Sims (2003, 2011). Intuitively, observing signals s_t over time will reduce uncertainty about the state. Sims (2003, 2011)

[10]This solver applies interior point methods implemented by a primal-dual infeasible-interior-point algorithm.

recommends to use a Cholesky decomposition as the new choice variable to handle the no-forgetting constraints for the infinite-horizon model studied in Section 3.3. We do not follow his approach.

Now we obtain the following equivalent problem

$$\min_{\{\Sigma_{t|t}\}} \quad \frac{1}{2}\log\det\left(\Sigma_0\right) - \frac{1}{2}\log\det\left(\Sigma_{0|0}\right) \tag{30}$$
$$+\sum_{t=1}^{T}\frac{1}{2}\log\det\left(A_{t-1}\Sigma_{t-1|t-1}A_{t-1}' + W_{t-1}\right) - \frac{1}{2}\log\det\left(\Sigma_{t|t}\right)$$

subject to (27), (28), and (29). The eliminated variable Φ_t can be recovered through

$$\Phi_t = \Sigma_{t|t}^{-1} - \left(A_{t-1}\Sigma_{t-1|t-1}A_{t-1}' + W_{t-1}\right)^{-1}, \ t = 1, 2, ..., T, \tag{31}$$

and

$$\Phi_0 = \Sigma_{0|0}^{-1} - \Sigma_0^{-1}. \tag{32}$$

The second step is to rewrite the objective function in (30) by regrouping terms as a sum of the initial cost $\frac{1}{2}\log\det\left(\Sigma_0\right)$, the final cost $\frac{1}{2}\log\det\left(\Sigma_{T|T}\right)$, and period costs

$$\frac{1}{2}\log\det\left(A_t\Sigma_{t|t}A_t' + W_t\right) - \frac{1}{2}\log\det\left(\Sigma_{t|t}\right),$$

for $t = 0, 1, ..., T-1$. The matrix determinant lemma (Theorem 18.1.1 in Harville (1997)) implies that the preceding expression is equal to

$$\frac{1}{2}\log\det W_t - \frac{1}{2}\log\det\left(\Sigma_{t|t}^{-1} + A_t'W_t^{-1}A_t\right)^{-1}.$$

Due to the monotonicity of the determinant function, this expression is equal to the optimal value of

$$\min_{\Pi_t} \quad \frac{1}{2}\log\det W_t - \frac{1}{2}\log\det \Pi_t$$

subject to

$$0 \prec \Pi_t \preceq \left(\Sigma_{t|t}^{-1} + A_t'W_t^{-1}A_t\right)^{-1}. \tag{33}$$

In the final step we apply the matrix inversion formula to rewrite (33) as

$$0 \prec \Pi_t \preceq \Sigma_{t|t} - \Sigma_{t|t}A_t'\left(W_t + A_t\Sigma_{t|t}A_t'\right)^{-1}A_t\Sigma_{t|t},$$

which is equivalent to

$$\begin{bmatrix} \Sigma_{t|t} - \Pi_t & \Sigma_{t|t}A_t' \\ A_t\Sigma_{t|t} & W_t + A_t\Sigma_{t|t}A_t' \end{bmatrix} \succeq 0, \quad \Pi_t \succ 0,$$

by the Schur complement property. Note that this is a linear matrix inequality condition. We summarize the analysis above in the following result.

Proposition 4 *Suppose that $W_t \succ 0$ and $\Omega_t \succeq 0$, for $t = 0, 1, ..., T$. Then the optimal information structure $\{C_t, V_t\}_{t=0}^T$ for problem (22) or (24) can be constructed by solving the following determinant maximization problem with decision variables $\left\{\Sigma_{t|t}, \Pi_t\right\}_{t=0}^T$:*

$$\kappa_T(D) = \min \quad -\sum_{t=0}^{T} \frac{1}{2} \log \det \Pi_t + c \tag{34}$$

subject to (27), (28), (29),

$$\begin{bmatrix} \Sigma_{t|t} - \Pi_t & \Sigma_{t|t} A_t' \\ A_t \Sigma_{t|t} & W_t + A_t \Sigma_{t|t} A_t' \end{bmatrix} \succeq 0, \quad t = 0, 1, ..., T-1, \tag{35}$$

$$\Sigma_{T|T} = \Pi_T, \quad \Pi_t \succ 0, \ t = 0, 1, ..., T, \tag{36}$$

where

$$c = \frac{1}{2} \log \det (\Sigma_0) + \sum_{t=1}^{T} \frac{1}{2} \log \det W_{t-1}.$$

The optimal sequence $\{\Phi_t\}_{t=0}^T$ is obtained from (31) and (32). The optimal information structure $\{C_t, V_t\}_{t=0}^T$ is given in Proposition 3.

The key difference between the problem above and that in Tanaka et al (2017) is that we have a single distortion constraint in (23) or (27), while they have a sequence of distortion constraints. The constraint (29) is also absent in their optimization problem. Moreover, our distortion constraint is derived from a LQG control problem, while their constraints are imposed exogenously from information theory. Following their arguments, we can show that there always exists an optimal solution to the problem in Proposition 4. Moreover, it is a strictly convex optimization problem and hence the optimal solution is unique.

The assumption of $W \succ 0$ can be restrictive in economic applications. This assumption implies that there must be a nontrivial random shock to each state transition equation (1). It is possible that there is no random shock to the state transition equation for some state variables. For example, we typically assume that the capital stock k_t follows the law of motion $k_{t+1} = (1 - \delta) k_t + I_t$, where $\delta > 0$ denotes the depreciation rate and I_t denotes investment. To get around this issue, one may introduce a depreciation or capital destruction shock often used in the literature. See Section 4.3 for the details. Alternatively, we allow $W \succeq 0$ and present a result similar to Proposition 4 in Appendix B. We need to impose a new assumption that A is invertible. This assumption can also be restrictive. For example, it rules out the case in which an IID shock is used as a state variable. This shock may represent a component of the TFP shock that enters the profit function in a firm's price setting problem or investment problem analyzed in Section 4.

For the relaxed problem under RI in Problem 3, we can use a similar three-step procedure. In particular, the optimal control is given by $u_t = -F_t \widehat{x}_t$ in step 1 by the certainty equivalence principle. In step 2 we derive the relaxed tracking problem under RI in which the information-flow constraint is removed. In step 3 we use the semidefinite programming approach to transform this problem into the following determinant maximization problem:

$$\min_{\left\{\Sigma_{t|t},\Pi_t\right\}_{t=0}^T} \sum_{t=0}^{T} \operatorname{tr}\left(\Omega_t \Sigma_{t|t}\right) - \lambda \sum_{t=0}^{T} \frac{1}{2} \log\det \Pi_t + \lambda c$$

subject to (28), (29), (35), and (36). Using the solution to this problem and equations (31) and (32), we determine the optimal sequence $\{\Phi_t\}_{t=0}^T$. Proposition 3 gives the optimal information structure $\{C_t, V_t\}_{t=0}^T$.

After obtaining the solutions for $\{F_t, \Sigma_t, C_t, V_t\}$, we use the system of equations (1), (2), (9), (14), and (15) to generate impulse responses and simulations of the model.

3.2 Static Case

For the static case with $T = 0$, problem (30) or (34) becomes

Problem 6 *(Static rate distortion problem)*

$$\kappa_0(D) = \min_{\Sigma} \frac{1}{2} \log\det(\Sigma_0) - \frac{1}{2} \log\det \Sigma \tag{37}$$

subject to

$$tr(\Omega\Sigma) \leq D,\ 0 \prec \Sigma \preceq \Sigma_0. \tag{38}$$

Here we use Σ to denote $\Sigma_{0|0}$ and Ω to denote Ω_0 only in this subsection, without risk of confusion. Notice that Σ_0 is the prior covariance matrix of the $n_x \times 1$ random vector x_0. The decision maker receives a multi-dimensional signal $s_0 = Cx_0 + v_0$, where v_0 is a normal random vector with mean zero and covariance matrix V. The optimal control is $u_0 = -R_0^{-1} S_0' x_0$.

When Ω is an identity matrix and Σ_0 is diagonal, the problem admits the standard reverse water-filling solution analyzed by Cover and Thomas (2006). The general case with $\Omega \succeq 0$ and $\Sigma_0 \succ 0$ is nontrivial.[11] Fulton (2018) presents a similar result to ours, but our proof given in the appendix is different and much simpler than his.

Before stating our result, we introduce some notations. Let $\Sigma_0^{\frac{1}{2}} \succ 0$ denote the positive definite square root of Σ_0. Then the positive semidefinite matrix $\Sigma_0^{\frac{1}{2}} \Omega \Sigma_0^{\frac{1}{2}}$ admits an eigendecomposition $\Sigma_0^{\frac{1}{2}} \Omega \Sigma_0^{\frac{1}{2}} = U \Omega_d U'$, where U is an orthonormal matrix and $\Omega_d \equiv diag(d_1, ..., d_{n_x})$ is a diagonal matrix with $d_i \geq 0$, $i = 1, ..., n_x$, denoting the eigenvalues of the positive semidefinite matrix $\Sigma_0^{\frac{1}{2}} \Omega \Sigma_0^{\frac{1}{2}}$.

[11] The case of $\Omega \succ 0$ is easier to handle by a change of variable.

Proposition 5 *Suppose that $\Omega \succeq 0$ and $\Sigma_0 \succ 0$. Then the optimal solution to Problem 6 is given by*

$$\Sigma = \Sigma_0^{\frac{1}{2}} U \widehat{\Sigma} U' \Sigma_0^{\frac{1}{2}}, \tag{39}$$

where $\widehat{\Sigma} \equiv diag\left(\widehat{\Sigma}_1, ..., \widehat{\Sigma}_{n_x}\right)$ with

$$\widehat{\Sigma}_i = \begin{cases} 1, & \text{if } D \geq \sum_{i=1}^{n_x} d_i, \\ \min\left(1, \frac{1}{\alpha d_i}\right), & \text{if } 0 < D < \sum_{i=1}^{n_x} d_i. \end{cases}$$

Here the Lagrange multiplier $\alpha > 0$ is chosen such that $\sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i = D$. The rate distortion function is given by

$$\kappa_0(D) = \max\left\{0, -\frac{1}{2}\sum_{i=1}^{n_x} \log \widehat{\Sigma}_i\right\}.$$

The optimal information structure satisfies

$$C'V^{-1}C = \Sigma_0^{-\frac{1}{2}} U \;\; diag\{\max(0, \alpha d_i - 1)_{i=1}^{n_x}\} \, U' \Sigma_0^{-\frac{1}{2}}.$$

The intuition behind this proposition is best understood for the special case in which $\Sigma_0 = \text{diag}(\sigma_1^2, ..., \sigma_{n_x}^2)$ is diagonal and Ω is identity, where $\sigma_i^2 > 0$ for all i. In this case $d_i = \sigma_i^2 > 0$ and $U = I$. If the distortion is large enough such that $D \geq \sum_{i=1}^{n} \sigma_i^2$, the decision maker will not process any information to allocate attention to any source of uncertainty so that $\widehat{\Sigma} = I$, $\Sigma = \Sigma_0$, and $\kappa_0(D) = 0$. If the distortion is small enough such that $0 < D < \sum_{i=1}^{n_x} d_i$, the decision maker will process information to reduce uncertainty such that the posterior variance for the ith source of uncertainty is reduced to $\min\left(\sigma_i^2, 1/\alpha\right)$.

The decision maker allocates attention to the sources of uncertainty with high prior variances σ_i^2 according to a decreasing order so that the posterior variances are reduced to $1/\alpha < \sigma_i^2$. This process continues until the distortion constraint binds at some level of variance. The decision maker will not allocate any attention to the source of uncertainty with variances below that level.

In the general case d_i is the ith eigenvalue of the weighted prior covariance matrix and $\widehat{\Sigma}$ may be interpreted as a scaling factor for these eigenvalues. The attention is allocated according to a decreasing order of $\{d_i\}$, instead of prior variances. High eigenvalues d_i are scaled down by the factor $\widehat{\Sigma}_i = 1/(\alpha d_i) < 1$ for small distortions.

Proposition 5 also shows that, even though the prior information is independent (Σ_0 is diagonal), when Ω is not an identity matrix, the posterior variance Σ is not diagonal in general. This means that rational inattention induces ex post correlation of uncertainty across initially independent information (Sims (2011)).

The following result characterizes the signal dimension.

Corollary 1 *Suppose that* $\Omega \succeq 0$ *and* $\Sigma_0 \succ 0$ *in Problem 6. If* $0 < D < \min_i \{md_i > 0\}$ *where* $m \equiv$ *rank*(Ω) *, then the optimal signal dimension is equal to* m*. The signal dimension (weakly) decreases as* D *increases if positive eigenvalues* $d_i > 0$ *are not identical.*

This corollary states that when the distortion is sufficiently small or when the information-flow rate is sufficiently large, the decision maker processes the signal with the highest dimension, which is equal to rank(Ω) . By (19) it does not exceed the minimum of the numbers of controls and states, n_u and n_x. As the distortion gradually increases or the information-flow rate decreases, the decision maker processes less information so that the signal dimension decreases.

We now turn to the static relaxed problem under RI given by

$$\min_{\Sigma}\ \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\log\det(\Sigma_0) - \frac{\lambda}{2}\log\det\Sigma \tag{40}$$

subject to $0 \prec \Sigma \preceq \Sigma_0$, where $\lambda > 0$ is interpreted as the shadow price or cost of information. We can similarly prove the following result.

Proposition 6 *Suppose that* $\Omega \succeq 0$ *and* $\Sigma_0 \succ 0$*. Then the optimal solution to the static relaxed RI problem (40) is given by (39), where* $\widehat{\Sigma} \equiv diag\left(\widehat{\Sigma}_1, ..., \widehat{\Sigma}_{n_x}\right)$ *with*

$$\widehat{\Sigma}_i = \min\left(\frac{\lambda}{2d_i}, 1\right),$$

$$C'V^{-1}C = \Sigma_0^{-\frac{1}{2}}U\ diag\left\{\max\left(0, \frac{2d_i}{\lambda} - 1\right)_{i=1}^{n_x}\right\}U'\Sigma_0^{-\frac{1}{2}}.$$

The signal dimension is equal to the number of d_i *such that* $d_i > \lambda/2$ *and decreases as* λ *increases.*

In the special case with $2d_i > \lambda > 0$ for all i, we have $\widehat{\Sigma}_i = \lambda/(2d_i)$ for all i and hence $\Sigma^{-1} = 2\Omega/\lambda$. Moreover, the signal dimension is n_x and the no-forgetting constraint does not bind.

3.3 Infinite-horizon Stationary Problems

For the infinite-horizon stationary problem, all matrices A_t, B_t, W_t, C_t, and V_t are time invariant. We consider the stationary Kalman filter for which $\Sigma_{t|t} \to \Sigma$,

$$\widehat{x}_t = \widehat{x}_{t|t-1} + \left(A\Sigma A' + W\right)C'\left[C\left(A\Sigma A' + W\right)C' + V\right]^{-1}\left(s_t - C\widehat{x}_{t|t-1}\right), \tag{41}$$

$$\widehat{x}_{t|t-1} = (A - BF)\,\widehat{x}_{t-1},\ \widehat{x}_{0|-1} = \bar{x}_0, \tag{42}$$

and

$$\Sigma = \left[\left(A\Sigma A' + W\right)^{-1} + \Phi\right]^{-1}, \tag{43}$$

where $\Phi = C'V^{-1}C \succeq 0$.

Then we can derive that

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} I\left(x_t; s_t|s^{t-1}\right) = \frac{1}{2}\log\det\left(A\Sigma A' + W\right) - \frac{1}{2}\log\det\left(\Sigma\right),$$

and

$$\mathbb{E}\left[(x_t - \widehat{x}_t)'\Omega(x_t - \widehat{x}_t)\right] = \text{tr}(\Omega\Sigma),$$

where Ω is given by (20). Thus the distortion rate problem is given by

$$D(\kappa) \equiv \min_{\Sigma\succ 0,\ \Phi\succeq 0} \quad \text{tr}(\Omega\Sigma) \tag{44}$$

subject to (43) and

$$\frac{1}{2}\log\det(A\Sigma A' + W) - \frac{1}{2}\log\det\Sigma \leq \kappa.$$

This problem is similar to that analyzed by Sims (2003) with three differences. First he removes the choice variable Φ by replacing (43) by the following no-forgetting constraint without affecting the optimal solution for Σ :

$$A\Sigma A' + W \succeq \Sigma. \tag{45}$$

Second, we derive the matrix Ω from a more general LQG control problem, which is positive semidefinite. This guarantees that minimization problem always has a solution. Finally, we derive the optimal information structure $\{C, V\}$ using $\Phi = C'V^{-1}C$, while Sims (2003) assumes that the signal vector is given by $s_t = x_t + \xi_t$, where he calls ξ_t the information-processing-induced measurement error.

The corresponding rate distortion problem becomes:

Problem 7 *(Infinite-horizon stationary rate distortion problem)*

$$\kappa(D) \equiv \min_{\Sigma\succ 0} \frac{1}{2}\log\det(A\Sigma A' + W) - \frac{1}{2}\log\det\Sigma \tag{46}$$

subject to (45) and

$$tr(\Omega\Sigma) \leq D. \tag{47}$$

As in Section 3.1, we use the semidefinite programming approach to solve this problem.

Proposition 7 *Suppose that $W \succ 0$ and $\Omega \succeq 0$. Then the optimal solution to Problem 7 can be obtained from solving the following semidefinite programming problem:*

$$\kappa(D) = \min_{\Pi\succ 0, \Sigma\succ 0} -\frac{1}{2}\log\det\Pi + \frac{1}{2}\log\det W$$

subject to (45), (47), and

$$\begin{bmatrix} \Sigma - \Pi & \Sigma A' \\ A\Sigma & A\Sigma A' + W \end{bmatrix} \succeq 0. \tag{48}$$

The optimal Φ is determined by (43) and the optimal information structure $\{C, V\}$ satisfies $\Phi = C'V^{-1}C$.

Because of the difficulty of the dynamic multivariate RI problems, characterization results are rarely available in the literature. We are able to derive an analytical result for the special case in which all states are equally persistent in the sense that $A = \rho I$ with $|\rho| < 1$. We find that the stationary RI problem also admits a generalized reverse water-filling solution similar to that in Proposition 5 for the static case.

Let us introduce some notations used in the result below. Let $W^{\frac{1}{2}} \succ 0$ denote the positive definite square root of W. Then the positive semidefinite matrix $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}$ admits an eigendecomposition $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}} = U\Omega_d U'$, where U is an orthonormal matrix and $\Omega_d \equiv \text{diag}(d_1, ..., d_{n_x})$ is a diagonal matrix with $d_i \geq 0$, $i = 1, ..., n_x$, denoting the eigenvalues of the positive semidefinite matrix $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}$.

Proposition 8 *Suppose that $\Omega \succeq 0$, $W \succ 0$, and $A = \rho I$ with $|\rho| < 1$ in Problem 7. Then the optimal long-run conditional covariance matrix for the state is given by*

$$\Sigma = W^{\frac{1}{2}} U \widehat{\Sigma} U' W^{\frac{1}{2}},$$

where $\widehat{\Sigma} \equiv diag\left(\widehat{\Sigma}_1, ..., \widehat{\Sigma}_{n_x}\right)$ with

$$\widehat{\Sigma}_i = \begin{cases} \frac{1}{1-\rho^2}, & \text{if } D \geq \left(1-\rho^2\right)^{-1} \sum_{i=1}^{n_x} d_i, \\ \min\left(\frac{1}{1-\rho^2}, \widehat{\Sigma}_i^*\right), & \text{if } 0 < D < \left(1-\rho^2\right)^{-1} \sum_{i=1}^{n_x} d_i. \end{cases}$$

Here the Lagrange multiplier α and $\widehat{\Sigma}_i^*$, $i = 1, 2, ..., n_x$, are the unique positive solution to the following system of $n_x + 1$ equations

$$\left[\rho^2 \left(\widehat{\Sigma}_i^*\right)^2 + \widehat{\Sigma}_i^*\right]^{-1} = \alpha d_i, \; i = 1, ..., n_x; \; D = \sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i.$$

The rate distortion function is given by

$$\kappa(D) = \max\left\{0, \frac{1}{2}\sum_{i=1}^{n_x} \log\left(\rho^2 + \frac{1}{\widehat{\Sigma}_i}\right)\right\}.$$

The optimal information structure $\{C, V\}$ satisfies

$$C'V^{-1}C = W^{-\frac{1}{2}} U \; diag\left\{\max\left(0, \alpha d_i \left[1 - \left(1-\rho^2\right)\widehat{\Sigma}_i^*\right]\right)_{i=1}^{n_x}\right\} U' W^{-\frac{1}{2}}.$$

If the distortion is too large, no information is processed so that Σ is the same as the state's unconditional covariance matrix $W/\left(1-\rho^2\right)$. For small distortions, the high eigenvalues d_i of the weighted state noise covariance matrix are scaled down by the factor $\widehat{\Sigma}_i$. In the special IID case we have $A = 0$ and $x_{t+1} = Bu_t + \epsilon_{t+1}$. The proposition above is reduced to Proposition 5 in the static case. We can also verify that, in the special scalar case with $n_x = 1$ and $\Omega = 1$, this proposition is reduced to Proposition 4 of Maćkowiak and Wiederholt (2009). Specifically, using their notations and base 2 logarithm, set $W = d_1 = a^2$, $C = 1$, and $\kappa(D) = \kappa_j > 0$. We can derive $\widehat{\Sigma}_1 = \widehat{\Sigma}_1^* = D/a^2$, $D = a^2/\left(2^{2\kappa_j} - \rho^2\right)$, and $s_t = x_t + v_t$, where the variance of v_t is

$$V = \frac{2^{2\kappa_j} a^2}{\left(2^{2\kappa_j} - 1\right)\left(2^{2\kappa_j} - \rho^2\right)}.$$

The following result is analogous to the static case:

Corollary 2 *Suppose that $\Omega \succeq 0$, $W \succ 0$, and $A = \rho I$ with $|\rho| < 1$ in Problem 7. If $D > 0$ is sufficiently small, then the optimal signal dimension is equal to $m \equiv rank(\Omega)$. The signal dimension (weakly) decreases as D increases if positive eigenvalues $d_i > 0$ are not identical.*

For the general case, we are unable to derive analytical results, but Proposition 7 offers a useful formulation to implement an efficient numerical procedure using semidefinite programming. After obtaining the solutions for $\{F, \Sigma, C, V\}$, we use the system of equations (5), (11), (21), (41), and (42) to generate impulse responses and simulations of the model. By contrast, Sims (2003, p. 679) adopts a different system in which he assumes that the signal vector is given by $s_t = x_t + \xi_t$. Our numerical results in the next section show that this signal vector may be suboptimal for multivariate problems. It is optimal if the signal-to-noise ratio Φ defined in (43) is nonsingular. In this case the no-forgetting constraint (45) does not bind. Whenever this constraint binds, Φ is singular and the signal dimension is less than the number of state variables.

Using a similar procedure, we can derive the optimal control $u_t = -F\widehat{x}_t$ for the relaxed problem under RI in Problem 4. The optimal conditional covariance matrix Σ and information structure $\{C, V\}$ are determined by the following program:

$$\min_{\Pi, \Sigma \succ 0} \ \operatorname{tr}(\Omega\Sigma) - \frac{\lambda}{2}\log\det\Pi + \frac{\lambda}{2}\log\det W$$

subject to (45) and (48). We can derive results similar to Propositions 7 and 8, and Corollary 2. In particular the critical parameter is λ instead of D. For space limitation we omit the details.

4 Applications

In this section we study three applications to illustrate our results. We analyze a pure tracking problem in the first application and dynamic control problems in the other two. In the first application there are two exogenous states and one control. In the second application there are one endogenous and two exogenous states and one control. In the last application there are two endogenous and two exogenous states and two controls.

4.1 Price Setting

We first consider a single firm's price setting problem adapted from Maćkowiak and Wiederholt (2009).[12] Unlike their model we do not consider the general equilibrium price feedback effect. We focus on an infinite-horizon stationary tracking problem under RI.

Let the profit-maximizing price satisfy $p_t^* = a' x_t$, where $a = (a_1, a_2)'$ and $x_t = (x_{1t}, x_{2t})'$ is an VAR(1) process $x_t = A x_{t-1} + \epsilon_t$, where $A = \text{diag}(\rho_1, \rho_2)$ with $|\rho_1| < 1$, $|\rho_2| < 1$, and $\epsilon_t \equiv (\epsilon_{1t}, \epsilon_{2t})'$ is a Gaussian white noise with covariance matrix $W \succ 0$. The firm does not observe any states and receives a signal vector $s_t = C x_t + v_t$ as in (5). Let $\widehat{p}_t = \mathbb{E}\left[p_t^* | s^t\right]$, $\widehat{x}_t = \mathbb{E}\left[x_t | s^t\right]$, and $\Omega = aa'$. The firm solves the following tracking problem

$$\min_{C,V} \mathbb{E}\left[(p_t^* - \widehat{p}_t)^2\right] = \mathbb{E}\left[(x_t - \widehat{x}_t)' \Omega (x_t - \widehat{x}_t)\right]$$

subject to the information flow constraint (4).

The optimal price under RI is given by

$$\widehat{p}_t = \mathbb{E}\left[p_t^* | s^t\right] = \mathbb{E}\left[a' x_t | s^t\right] = a' \widehat{x}_t, \tag{49}$$

where $\widehat{x}_t$ satisfies the Kalman filter:

$$\widehat{x}_t = (I - KC) A \widehat{x}_{t-1} + K s_t, \tag{50}$$

for $t \geq 0$, with $\widehat{x}_{-1} = 0$ and $\{C, V\}$ being the solution to the above minimization problem. Here $K \equiv (A\Sigma A' + W) C' \left[C (A\Sigma A' + W) C' + V\right]^{-1}$ is the Kalman gain.

We are able to derive an analytical result for the special case in which all states have the same persistence parameter ρ.[13]

Proposition 9 *Consider the infinite-horizon stationary rate distortion problem in (46) with $p_t^* = a' x_t$, $\Omega = aa'$, $W \succ 0$, and $A = \rho I$ ($|\rho| < 1$). If $0 < D < \left(1 - \rho^2\right)^{-1} \left\|W^{1/2} a\right\|^2$, then $\kappa(D) > 0$, the optimal signal is one dimensional and can be normalized as*[14]

$$s_t = p_t^* + \left\|W^{1/2} a\right\| v_t, \tag{51}$$

the optimal variance of v_t satisfies

$$V^{-1} = \frac{\left\|W^{1/2} a\right\|^4}{\rho^2 D^2 + \left\|W^{1/2} a\right\|^2 D} \left[1 - \frac{\left(1 - \rho^2\right) D}{\left\|W^{1/2} a\right\|^2}\right] > 0,$$

[12] See Woodford (2003, 2009) for related pricing models.

[13] When $\rho = 0$, Proposition 9 is reduced to the IID case, which is also the static case by setting $W = \Sigma_0$, studied by Fulton (2018).

[14] We use $\|\cdot\|$ to denote the Euclidean norm.

and the optimal conditional variance matrix Σ *for* x_t *is given by*

$$\Sigma = \frac{W}{1-\rho^2} - W\Omega W \left\| W^{1/2} a \right\|^{-4} \left[\left(1-\rho^2\right)^{-1} \left\| W^{1/2} a \right\|^2 - D \right].$$

If $D \geq \left(1-\rho^2\right)^{-1} \left\| W^{1/2} a \right\|^2$, *then* $\kappa(D) = 0$ *and the firm does not process any information so that* $\Sigma = W / \left(1-\rho^2\right)$. *The rate distortion function is given by*

$$\kappa(D) = \max\left\{0, \frac{1}{2}\log\left(\rho^2 + \left\| W^{1/2} a \right\|^2 / D\right)\right\}.$$

This proposition implies that the impulse responses of prices to the ith source of shock are relatively larger if and only if that source carries a larger weight a_i as shown in equation (50). The price responses are the same when $a_1 = a_2$. This result is independent of the dimension of states and the innovation covariance matrix W. By contrast, Maćkowiak and Wiederholt (2009) assume that the firm receives one signal about one shock and the two signals are independent. They argue that this assumption is reasonable in practice, even though it does not lead to optimal decisions. They show that the price is more responsive to the shock with a higher variance even when $\rho_1 = \rho_2$ and $a_1 = a_2$.

Now suppose that $\rho_1 \neq \rho_2$. Based on numerical solutions for a wide range of parameter values, we find that the optimal signal is still one dimensional, but it cannot take the normalized form of the profit-maximizing price plus a noise. Maćkowiak and Wiederholt (2009) informally argue that the optimal signal should always take that normalized form. Our numerical results show that this claim is incorrect when $\rho_1 \neq \rho_2$. In particular, let $C = (C_1, C_2)$ for the one dimensional signal $s_t = Cx_t + v_t$. Normalize $a_1 = C_1 = 1$. We find $C_2 \neq a_2$ when $\rho_1 \neq \rho_2$.

It follows from equations (49) and (50) that the initial impact of the two shocks on the optimal price under RI is determined by two effects. First, the learning effect is reflected by the term $a'K$ due to signal extraction. Second, the attention allocation effect is reflected by the optimal choice of Σ, C, and V. Given one dimensional signal, the initial response is determined by $a'KC$. Thus the initial response to the second shock is larger than to the first shock if and only if $C_2 > C_1 = 1$, or if and only if the firm pays more attention to the second shock. To see the impact quantitatively, set $\rho_1 = 0.95$, $\rho_2 = 0.70$, $a_1 = a_2 = 1$, $\kappa = 0.46$, and $W = \text{diag}\left(0.02^2, 0.15^2\right)$ as baseline values. We may interpret the first (second) shock as the aggregate (idiosyncratic) shock. Aggregate shocks are typically more persistent, but less volatile than idiosyncratic shocks.

Figure 1 presents the rate distortion function and the signal noise variance as a function of the distortion. The rate distortion function is convex and decreasing in κ. We can then find the distortion or tracking error D for any information-flow rate $\kappa > 0$. Comparative statics of D can be equivalently translated to that of κ. Figure 1 also shows that the signal noise variance increases with the distortion or decreases with κ.

[Insert Figure 1 Here]

The top two panels of Figure 2 plot the impulse response functions for three values of κ. We find that the responses under RI are dampened and delayed, compared to the full information case. When the channel capacity κ is larger, the firm can process more information and hence the price is more responsive to shocks. We also find that the initial responses to the aggregate shock are larger than those to the idiosyncratic shock for all values of κ, because the firm attends more to the more persistent aggregate shock.

[**Insert Figures 2-3 Here**]

To further investigate the role of persistence and innovation variance, the bottom two panels of Figure 2 plot the impact of ρ_1. We adjust σ_1 to hold the unconditional variance $\sigma_1^2/\left(1-\rho_1^2\right)$ of the aggregate shock fixed as in Maćkowiak and Wiederholt (2009).[15] We also hold the other baseline values fixed. We find that, as we gradually increase the aggregate shock persistence ρ_1, the initial response to the aggregate shock increases, while the initial response to the idiosyncratic shock declines, suggesting that the firm shifts attention to the aggregate shock away from the idiosyncratic shock. Moreover, the initial response to the aggregate shock is larger (smaller) than that to the idiosyncratic shock whenever $\rho_1 > (<) \rho_2$, even though the aggregate shock has a much smaller innovation variance. When $\rho_1 = \rho_2$, the responses to the two sources of shocks are identical given $a_1 = a_2$, regardless of innovation variances, as Proposition 9 shows.

For example, when ρ_1 increases from 0.5 to 0.95 the conditional covariance matrix Σ of the states changes from

$$\begin{bmatrix} 0.0298 & -0.0152 \\ -0.0152 & 0.0263 \end{bmatrix} \text{ to } \begin{bmatrix} 0.0176 & -0.0146 \\ -0.0146 & 0.0260 \end{bmatrix}.$$

Thus the conditional variances of both shocks decrease and the two shocks are less negatively conditionally correlated. The learning effect causes $a'K$ to increase from 0.5894 to 0.6520. In the meantime, the firm shifts more attention to the aggregate shock as C_2 decreases from 1.0576 to 0.8920 for fixed $C_1 = 1$. The attention allocation effect dominates so that the response to the idiosyncratic shock $a'KC_2$ declines.

The top two panels of Figure 3 plot the impact of idiosyncratic volatility σ_2, holding all other parameter values fixed at the baseline values. We find that as σ_2 becomes larger the optimal price under RI is more responsive initially to both sources of shocks, but the responses are not sensitive and the price reverts back to the steady state faster. By contrast, Maćkowiak and Wiederholt (2009) find that the price is less responsive to the aggregate shock and more responsive to the idiosyncratic shock because the firm shifts attention to the idiosyncratic shock away from the aggregate shock. In our model without the signal independence assumption, there is a spillover of one source of uncertainty into the other. The learning effect causes $a'K$ to increase as σ_2 increases so that the price is more responsive to both sources of shocks. In the meantime the firm allocates

[15] Our results do not change significantly if we hold σ_1 fixed.

more attention to the idiosyncratic shock in that C_2/C_1 increases, but still $C_2 < C_1$. Overall the learning effect dominates.

We experiment with a wide range of economically reasonable values for $\rho_1 \in (0,1)$, $\rho_2 \in (0,1)$, $\sigma_1 \in (0, 0.5)$, and $\sigma_2 \in (0, 0.5)$ and find our preceding results are robust. Notably the persistence plays a dominant role. The optimal price under RI is always more responsive to the shock with higher persistence.[16] For very large values of σ_2 given $\rho_1 > \rho_2$, it is possible that the price is slightly more responsive to the second shock than to the first shock. However we cannot draw a general conclusion because our numerical solutions cannot exhaust all parameter values and because numerical errors are large for extreme parameter values.

So far we have fixed $a_1 = a_2 = 1$ for all previous numerical experiments. We now consider the impact of a_2 in the bottom two panels of Figure 3, holding all other parameter values fixed at the baseline values. We find that as a_2 increases from 0.5 to 5, the initial responses to the second shock increase, while the initial responses to the first shock decrease. This result reflects the intuition that the firm allocates more attention to the shock that carries a relatively larger weight in its objective function and hence the price is more responsive to that shock.

4.2 Consumption/Saving

In this subsection we study a consumption/saving problem similar to those in Hall (1978), Sims (2003), and Luo (2008). A household maximizes its quadratic utility over a consumption process $\{c_t\}$:

$$-\frac{1}{2}\mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t\left(c_t-\bar{c}\right)^2\right]$$

subject to the budget constraint

$$w_{t+1} = (1+r)\left(w_t - c_t\right) + y_{t+1},\ t \geq 0,$$

where $\bar{c}$ is a bliss level of consumption, w_t is wealth, and y_t is income. For simplicity we suppose $\beta(1+r) = 1$. Suppose that income y_t consists of two persistent components and a transitory component:

$$\begin{aligned} y_t &= \bar{y} + z_{1,t} + z_{2,t} + \epsilon_{y,t}, \\ z_{1,t} &= \rho_1 z_{1,t-1} + \eta_{1,t}, \\ z_{2,t} &= \rho_2 z_{2,t-1} + \eta_{2,t}, \end{aligned}$$

where $\bar{y}$ is average income and innovations $\epsilon_{y,t}$, η_{1t}, and η_{2t} are mutually independent Gaussian white noises with variances σ_y^2, σ_1^2, and σ_2^2. The two persistent components $z_{1,t}$ and $z_{2,t}$, and the

[16]Our results are more robust for the relaxed problem formulation with constant cost of information flow. In particular, in the large σ_2 limit, the initial price responses to the second shock catch up and converge to those to the first shock when $\rho_1 > \rho_2$. Such an analysis is available upon request.

transitory component $\epsilon_{y,t}$ may capture aggregate, local, and individual income uncertainties. The state vector is $x_t = (w_t, z_{1,t}, z_{2,t})'$ plus a constant state 1.

By the certainty equivalence principle, it is straightforward to show that optimal consumption under RI is given by

$$c_t = \frac{\bar{y}}{1+r} + \frac{r}{1+r}\left(\widehat{w}_t + \frac{\rho_1}{1+r-\rho_1}\widehat{z}_{1,t} + \frac{\rho_2}{1+r-\rho_2}\widehat{z}_{2,t}\right),$$

where $\widehat{x}_t = \mathbb{E}\left[x_t|s^t\right]$. We need to use numerical methods to solve for the optimal information structure $\{C, V\}$ for the signal vector $s_t = Cx_t + v_t$. Set the same parameter values as in Sims (2003): $\beta = 0.95$, $\rho_1 = 0.97$, $\rho_2 = 0.90$, $\sigma_y^2 = 0.01$, $\sigma_1^2 = 0.0001$, and $\sigma_2^2 = 0.003$.

[Insert Figure 4 Here]

We find that the optimal signal vector s_t is one dimensional and $C = [1, \ \ 11.7433, \ \ 5.8978]$ and $V = 0.0079$ for $\kappa = 1.5746$.[17] Thus the household processes information about a linear combination of all three state variables with the more persistent shock $z_{1,t}$ having the largest weight. As κ decreases, this weight increases slightly, while the weight on the less persistent shock $z_{2,t}$ barely changes. Moreover, the signal noise variance increases significantly. Intuitively, the household pays more attention to the more persistent income shock and the signal becomes more noisy when the information capacity decreases. We also find that the signal is always one dimensional even for a very large κ when the distortion is close to zero. This is in contrast to Sims's (2003) finding that the signal is three dimensional when κ is sufficiently large (see his Figures 7 and 8).

Figure 4 plots the impulse response functions for consumption to 1% shocks to the three true income components and one signal noise, starting from zero consumption. The flat lines correspond to the responses for the full information case. Under RI, the consumption responses to all three true component income shocks are damped initially, and then gradually rise permanently to high levels. Intuitively, the rationally inattentive household responds to shocks sluggishly. Lower consumption early leads to higher wealth. The extra savings earn a return $1 + r$ and allow the household to accumulate higher wealth to fund higher consumption later. We also find that the initial response is larger for a more persistent income shock given the same κ. And the initial responses to all true income shocks are larger when κ is larger. Unlike the income shocks, the noise shock causes consumption to rise immediately and then gradually decreases over time.

Our numerical results are different from that reported by Sims (2003). His Figures 7 and 8 show that the initial consumption response to the less persistent income shock is larger. He argues that this is because the innovation to this shock has a larger variance. Based on a wide range of parameter values, we find that the initial response to the more persistent income shock is always larger, independent of the innovation variance. This result is consistent with our earlier finding for the pricing example.

[17] We normalize $C_1 = 1$ for all cases.

4.3 Firm Investment

We finally solve a firm's investment problem subject to convex adjustment costs. The firm chooses two types of capital investment to maximize its discounted present value of dividends:

$$\max_{\{I_{1,t},I_{2,t}\}} \mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t d_t\right]$$

subject to

$$\begin{aligned} d_t &= \exp\left(z_t+e_t\right)k_{1,t}^{\alpha}k_{2,t}^{\theta}-I_{1,t}-I_{2,t}-\frac{\phi_1}{2}\left(\frac{I_{1,t}}{k_{1,t}}-\delta_1\right)^2 k_{1,t}-\frac{\phi_2}{2}\left(\frac{I_{2,t}}{k_{2,t}}-\delta_2\right)^2 k_{2,t} \\ &\quad -\tau\left(\exp\left(z_t+e_t\right)k_{1,t}^{\alpha}k_{1,t}^{\theta}-\chi I_{2,t}\right), \end{aligned}$$

where $d_t, k_{1,t}, k_{2,t}, I_{1,t}$, and $I_{2,t}$ denote dividends, tangible capital, intangible capital, tangible capital investment, and intangible capital investment, respectively. The parameters satisfy $\delta_1, \delta_2, \alpha, \theta, \tau \in (0,1)$, $\alpha+\theta<1$, and $\phi_1, \phi_2 > 0$. The variables z_t and e_t represent persistent and temporary Gaussian TFP shocks, $z_t = \rho z_{t-1} + \epsilon_{z,t}$. We include taxation of corporate profits because a key distinction between the two types of capital is that a fraction χ of intangible investment is expensed and therefore exempt from taxation. The capital evolution equations are

$$\begin{aligned} k_{1,t+1} &= \left(1-\delta_1\right)k_{1,t}+I_{1,t}+\epsilon_{1,t+1}, \\ k_{2,t+1} &= \left(1-\delta_2\right)k_{2,t}+I_{2,t}+\epsilon_{2,t+1}, \end{aligned}$$

where $\epsilon_{1,t+1}$ and $\epsilon_{2,t+1}$ represent depreciation or capital quality shocks. Suppose that $\epsilon_{z,t}$, e_t, $\epsilon_{1,t}$, and $\epsilon_{2,t}$ are mutually independent Gaussian white noises with variances σ_z^2, σ_e^2, σ_1^2, and σ_2^2.

To solve this problem numerically, we first approximate the firm's objective function by a quadratic function in the neighborhood of the nonstochastic steady state. We then obtain a linear-quadratic control problem with the state vector $x_t = \left(z_t, e_t, \tilde{k}_{1,t}, \tilde{k}_{2,t}\right)'$ plus a constant state 1, where $\tilde{k}_{i,t}$, $i=1,2$, denotes the deviation from the steady state. From this problem we can derive the decision rules and the weighting matrix Ω in the tracking problem in which the relevant state vector is x_t. For the no adjustment cost case under full information, the linearized optimal decision rules are given by

$$\tilde{k}_{i,t+1} = \frac{k_i\rho}{1-\alpha-\theta}z_t + \epsilon_{i,t+1},$$

where k_i is the steady-state capital stock.

We now solve the transformed rate distortion problem using the semidefinite programming approach. We set baseline parameter values as in McGrattan and Prescott (2010): $\alpha = 0.26$, $\theta = 0.076$, $\delta_1 = 0.126$, $\delta_2 = 0$, $\tau = 0.35$, and $\chi = 0.5$. Set $\rho = 0.91$, $\sigma_z = \sigma_1 = \sigma_2 = 0.01$, and $\sigma_e = 0.1$. We choose $\beta = 0.9615$ to generate a 4 percent steady state interest rate. Following Saporta-Eksten and Terry (2018), we set the capital adjustment cost parameter values as $\phi_1 = 0.46$

and $\phi_2 = 1.40$. For these parameter values, the steady-state levels of capital are $k_1 = 0.98$ and $k_2 = 0.639$. Intangible capital is larger than tangible capital both because of the tax advantage and because it depreciates more slowly, which more than offsets the smaller production coefficient.

[Insert Figures 5 and 6 here.]

Since this model features two control variables and four state variables, we can study the non-trivial determination of the information structure. Figure 5 presents the rate distortion function and the signal dimension as we vary the distortion with and without adjustment costs. At sufficiently low distortion or high channel capacity the signal dimension is 2. However, if the distortion is high enough (or channel capacity falls below a critical value), the signal dimension drops by 1; the firm, in an attempt to conserve channel capacity, only processes a single signal. The critical amount of capacity is higher in the presence of adjustment costs.

To understand how the signal dimension changes, we display here the optimal signal structure for two values of κ with no adjustment costs: For $\kappa = 0.0203$

$$s_t = -0.81 z_t + 0.24 \tilde{k}_{1,t} + 0.53 \tilde{k}_{2,t} + v_t,$$

where the variance of v_t is 1.348, and for $\kappa = 1.0309$

$$s_t = \begin{bmatrix} -0.87 z_t + 0.44 \tilde{k}_{1,t} + 0.21 \tilde{k}_{2,t} \\ 0.03 z_t + 0.48 \tilde{k}_{1,t} - 0.88 \tilde{k}_{2,t} \end{bmatrix} + v_t,$$

where the covariance matrix of v_t is $\mathrm{diag}(0.005, 0.015)$. With adjustment costs, the optimal signal structure is similar.

Note that in neither case does the signal depend on e_t, the transitory productivity shock; since e_t does not affect the value-maximizing level of investment under full information, there is no point using information capacity to learn about it. Thus rational inattention does not explain why investment responds to transitory shocks in the data documented by Saporta-Eksten and Terry (2018). If the information structure is exogenously given as in a standard signal extraction problem, then firms would be confused about the source of a productivity change; as a result, they would respond to transitory shocks. However, since value-maximizing investment is independent of the transitory shock e_t, if the firm can choose the allocation of attention, it will ignore the transitory shock completely.

The intuition for the above signal forms can be better understood as follows. The variables the firm cares about are the TFP and the (linearized) marginal products of tangible and intangible capital ($mpk_{1,t}$ and $mpk_{2,t}$). We can then use a change of variables to express the signal vector s_t in terms of a linear transformation of the new vector $[z_t, e_t, mpk_{1t}, mpk_{2t}]'$. For example, the linear transformation is given by $C = \begin{bmatrix} 0.061 & 0.872 & -1.823 & -3.553 \end{bmatrix}$ for the case of no adjustment costs and $\kappa = 0.0203$. A positive persistent change in TFP z_t will be confused with a positive change in e_t, which explains why both of these coefficients are positive. Since the firm does not want investment to respond to e_t, it must correct the confusion by believing that the marginal products of capital fall, which in turn explains why these coefficients are negative. With higher capacity ($\kappa = 1.0309$) the firm will collect a two-dimensional signal with the new linear transformation of $[z_t, e_t, mpk_{1t}, mpk_{2t}]'$:

$$C = \begin{bmatrix} -0.024 & 0.850 & -2.557 & -1.718 \\ 0.108 & 0.142 & 1.682 & -5.005 \end{bmatrix}.$$

In this case the firm does not have to believe that both marginal products of capital fall. This explains why there is only one negative coefficient in the second row of C.

We now turn to the impulse responses of two types of capital investment to a positive 1% shock to the persistent TFP component displayed in Figure 6. Each panel of the figure includes the full information case as well as at least one case with sufficiently low κ such that the signal vector becomes one-dimensional. The top two panels show the case without adjustment costs. Under full information, in response to a positive persistent TFP shock, investment increases immediately and then falls back to the steady state following a path similar to the TFP shock. As the information-flow rate κ falls, the investment responses under RI become dampened and delayed – investment rises less on impact and remains above the steady state longer. If κ is sufficiently small, then the response is very small.

In the case with adjustment costs displayed in the bottom two panels, investment responses under RI are delayed further, and can become hump-shaped, a pattern not present in the full information case. The reason for the hump-shape is a horse race between two effects. Consider the response of tangible investment to a positive TFP shock z_t (bottom left panel). Value-maximizing investment under full information rises on impact and then gradually falls back to the steady state, but at a slower rate than the case without adjustment costs. Under rational inattention, since the firm does not know z_t with certainty, exactly how much investment has risen is unknown. Since the firm learns slowly and the capital adjustment is costly, it takes several periods before the firm knows the investment level it should have chosen on impact, which leads to a rising investment path. On the other hand, since z_t is mean reverting the value-maximizing level of investment is falling over time. Thus optimal investment under RI will eventually falls back to the steady state. Without adjustment costs, mean reversion is sufficiently fast such that learning is always behind, leading to monotonic but delayed responses. With adjustment costs, but without information-flow constraint, there is no hump-shaped investment response either.

Our results are similar to Zorn's (2018) findings, while his model has only one type of capital and assumes there is no capital quality shock.[18] He documents evidence that investment at the sectoral level displays a hump-shaped response to aggregate shocks and a monotonic response to

[18]Notice that this assumption is subtle. Without capital quality shocks $\epsilon_{i,t+1}$, $k_{i,t+1}$ is measurable with respect to date t information. The firm only needs to track the persistent shock z_t and the optimal signal is always one dimensional. Such an analysis is available upon request.

sectoral shocks. He shows that a model with both rational inattention and capital adjustment costs can deliver the two different types of responses. In contrast, models with just capital adjustment costs, models with just investment adjustment costs (Christiano, Eichenbaum, and Evans (2005)), and models with just rational inattention cannot match both types of impulse responses.

5 Conclusion

We have developed a framework to analyze multivariate RI problems in a LQG setup. We have proposed a three-step solution procedure to theoretically analyze and numerically solve these problems. We have provided generalized reverse water-filling solutions to some special cases. We have also applied our approach to three economic examples. Our analysis demonstrates that the existing solution approach proposed by Sims (2003, 2011) is flawed and not robust. Moreover, many simplifying assumptions adopted in the literature such as signal independence and exogenous signal structure are not innocuous. They lead to suboptimal behavior and some qualitatively different predictions from ours. While some simplifying assumptions may be justified by bounded rationality and deliver interesting results, removing these assumptions can generate new insights such as asset price comovement (Mondria (2010)). Our approach provides researchers' a useful toolkit to solve multivariate RI problems without simplifying assumptions and will find wide applications in economics and finance.

Appendix

A Proofs

Proof of Proposition 1: Let $\mathcal{F}_t$ denote the information set at time t under full information. The value function under full information is given by (7) and satisfies the Bellman equation:

$$v_t^{FI}(x_t) = -\left(x_t'Q_tx_t + u_t^{*'}R_tu_t^* + 2x_t'S_tu_t^*\right) + \beta\mathbb{E}\left[v_{t+1}^{FI}(x_{t+1}^*) \mid \mathcal{F}_t\right], \tag{A.1}$$

where $u_t^* = -F_tx_t$ denotes the optimal control under full information and

$$x_{t+1}^* = A_tx_t + Bu_t^* + \epsilon_{t+1}. \tag{A.2}$$

Let $\widehat{v}_t(\widehat{x}_t)$ denote the value function under RI. By the certainty equivalence principle,

$$\widehat{v}_t\left(\widehat{x}_t\right) = -\widehat{x}_t'P_t\widehat{x}_t - \widehat{d}_t,$$

where

$$\widehat{d}_{T+1} = tr\left(P_{T+1}\Sigma_{T+1|T+1}\right).$$

We have the Bellman equation

$$\widehat{v}_t(\widehat{x}_t) = -\mathbb{E}\left[x_t'Q_tx_t + u_t'R_tu_t + 2x_t'S_tu_t|s^t\right] + \beta\mathbb{E}\left[\widehat{v}_{t+1}(\widehat{x}_{t+1})|s^t\right]. \tag{A.3}$$

Since the information-flow constraint cannot help in the optimization, we know that $\mathbb{E}\left[v_t^{FI}\left(x_t\right)|s^t\right] > \widehat{v}_t(\widehat{x}_t)$. Following Sims (2003), we solve for $\{C_t\}$ and $\{V_t\}$ by minimizing $\mathbb{E}\left[v_0^{FI}\left(x_0\right)|s_0\right] - \widehat{v}_0(\widehat{x}_0)$. In other words, we choose information structure so as to bring expected utility from the current date onward as close as possible to the full information expected utility value. Using (A.1) and (A.3), we derive

$$\begin{aligned}\mathbb{E}\left[v_t^{FI}\left(x_t\right)|s^t\right] - \widehat{v}_t(\widehat{x}_t) &= \mathbb{E}\left[-u_t^{*'}R_tu_t^* - 2x_t'S_tu_t^* + u_t'R_tu_t + 2x_t'S_tu_t|s^t\right] \\ &\quad +\beta\mathbb{E}\left[v_{t+1}^{FI}(x_{t+1}^*) - \widehat{v}_t(\widehat{x}_{t+1})|s^t\right],\end{aligned}$$

where x_{t+1}^* is the state in (A.2). We rewrite this equation as

$$\begin{aligned}\mathbb{E}\left[v_t^{FI}\left(x_t\right)|s^t\right] - \widehat{v}_t(\widehat{x}_t) &= \mathbb{E}\left[-u_t^{*'}R_tu_t^* - 2x_t'S_tu_t^* + u_t'R_tu_t + 2x_t'S_tu_t|s^t\right] \\ &\quad +\beta\mathbb{E}\left[v_{t+1}^{FI}(x_{t+1}^*) - v_{t+1}^{FI}\left(x_{t+1}\right)|s^t\right] \\ &\quad +\beta\mathbb{E}\left[v_{t+1}^{FI}\left(x_{t+1}\right) - \widehat{v}_{t+1}(\widehat{x}_{t+1})|s^t\right],\end{aligned} \tag{A.4}$$

where x_{t+1} is the state in (13) when the control under RI $u_t = -F_t\widehat{x}_t$ is followed.

We compute that

$$\begin{aligned}x_{t+1}^{*'}P_{t+1}x_{t+1}^* &= (A_tx_t + B_tu_t^* + \epsilon_{t+1})'P_{t+1}(A_tx_t + B_tu_t^* + \epsilon_{t+1}) \\ &= \left(x_t'A_t'P_{t+1}A_tx_t + 2u_t^{*'}B_t'P_{t+1}A_tx_t + u_t^{*'}B_t'P_{t+1}B_tu_t^*\right) + ...,\end{aligned}$$

where we have omitted terms related to ϵ_{t+1} because they will be cancelled out later. Similarly,

$$x'_{t+1}P_{t+1}x_{t+1} = \left(x'_tA'_tP_{t+1}A_tx_t + 2u'_tB'_tP_{t+1}A_tx_t + u'_tB'_tP_{t+1}B_tu_t\right) + ...$$

Hence,

$$\begin{aligned} & \mathbb{E}\left[v^{FI}_{t+1}(x^*_{t+1}) - v^{FI}_{t+1}\left(x_{t+1}\right)|s^t\right] \\ = & \mathbb{E}\left[-x^{*\prime}_{t+1}P_{t+1}x^*_{t+1} + x'_{t+1}P_{t+1}x_{t+1}|s^t\right] \\ = & -\mathbb{E}\left[2(u^*_t - u_t)'B'_tP_{t+1}A_tx_t + u^{*\prime}_tB'_tP_{t+1}B_tu^*_t - u'_tB'_tP_{t+1}B_tu_t|s^t\right]. \end{aligned}$$

Using the first-order condition with respect to u^*_t, we have

$$S'_tx_t = -R_tu^*_t - \beta B'_tP_{t+1}A_tx_t - \beta B'_tP_{t+1}B_tu^*_t.$$

Substituting this equation into (A.4) yields

$$\begin{aligned} & \mathbb{E}\left[v^{FI}_t\left(x_t\right)|s^t\right] - \widehat{v}_t(\widehat{x}_t) - \beta\left\{\mathbb{E}\left[v^{FI}_{t+1}\left(x_{t+1}\right) - \widehat{v}_{t+1}(\widehat{x}_{t+1})|s^t\right]\right\} \\ = & \mathbb{E}\left[-u^{*\prime}_tR_tu^*_t + u'_tR_tu_t - 2x'_tS_t\left(u^*_t - u_t\right)|s^t\right] \\ & -\beta\mathbb{E}\left[2(u^*_t - u_t)'B'_tP_{t+1}A_tx_t + u^{*\prime}_tB'_tP_{t+1}B_tu^*_t - u'_tB'_tP_{t+1}B_tu_t|s^t\right] \end{aligned}$$

$$\begin{aligned} =&\mathbb{E}\left[-u^{*\prime}_tR_tu^*_t + u'_tR_tu_t + 2u^{*\prime}_tR_t(u^*_t - u_t)|s^t\right] \\ &+ 2\beta\mathbb{E}\left[\left(x'_tA'_tP_{t+1}B_t + u^{*\prime}_tB'_tP_{t+1}B_t\right)\left(u^*_t - u_t\right)|s^t\right] \\ &- \beta\mathbb{E}\left[2(u^*_t - u_t)'B'_tP_{t+1}A_tx_t + u^{*\prime}_tB'_tP_{t+1}B_tu^*_t - u'_tB'_tP_{t+1}B_tu_t|s^t\right] \\ =&\mathbb{E}\left[(u^*_t - u_t)'R_t(u^*_t - u_t)|s^t\right] + \beta\mathbb{E}\left[(u^*_t - u_t)'B'_tP_{t+1}B_t(u^*_t - u_t)|s^t\right] \\ =&\mathbb{E}\left[(u^*_t - u_t)'(R_t + \beta B'_tP_{t+1}B_t)(u^*_t - u_t)|s^t\right]. \end{aligned}$$

Substituting $u_t = -F_t\widehat{x}_t$ and $u^*_t = -F_tx_t$ into the last equation above, we obtain the recursive equation

$$\begin{aligned} \mathbb{E}\left[v^{FI}_t\left(x_t\right)|s^t\right] - \widehat{v}_t(\widehat{x}_t) \quad = \quad & \mathbb{E}_t\left[(x_t - \widehat{x}_t)'F'_t(R_t + \beta B'_tP_{t+1}B_t)F_t(x_t - \widehat{x}_t)|s^t\right] \\ & +\beta\left\{\mathbb{E}\left[v^{FI}_{t+1}\left(x_{t+1}\right) - \widehat{v}_{t+1}(\widehat{x}_{t+1})|s^t\right]\right\} \end{aligned}$$

for $t = 0, 1, ..., T$. The terminal condition is

$$\mathbb{E}\left[v^{FI}_{T+1}\left(x_{T+1}\right)|s^{T+1}\right] - \widehat{v}_{T+1}\left(\widehat{x}_{T+1}\right) = -\mathbb{E}\left[x'_{T+1}P_{T+1}x_{T+1}|s^{T+1}\right] + \mathbb{E}\left[x'_{T+1}P_{T+1}x_{T+1}|s^{T+1}\right] = 0.$$

Solving recursively we can derive that

$$\mathbb{E}\left[v^{FI}_0\left(x_0\right)|s_0\right] - \widehat{v}_0(\widehat{x}_0) = \sum_{t=0}^{T}\mathbb{E}\left[(x_t - \widehat{x}_t)'\Omega_t(x_t - \widehat{x}_t) \mid s_0\right],$$

where Ω_t is given by (19). Q.E.D.

Proof of Proposition 2: See the main text. Q.E.D.

Proof of Proposition 3: For simplicity we omit the time t subscript for all variables in the proof. By the spectral decomposition of a positive semidefinite matrix, there exists an $n_x \times n_x$ orthogonal matrix and a diagonal matrix Ψ such that $\Phi = U\Psi U'$. Let

$$\Psi = \begin{bmatrix} \widehat{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where $\widehat{\Psi}$ is an $m \times m$ diagonal matrix, $\widehat{\Psi} = \text{diag}(\varphi_1, ..., \varphi_m)$, and $\{\varphi_i\}_{i=1}^m$ are the positive eigenvalues of Φ. Clearly, $\text{rank}(\Phi) = m \leq n_x$. The matrix Φ can be factorized as $\Psi = \Delta'\widehat{\Psi}\Delta$, where $\Delta = \begin{bmatrix} I_m & \mathbf{0}_{m\times(n_x-m)} \end{bmatrix}$. Let $C = \Delta U'$ and $V = \widehat{\Psi}^{-1}$, completing the proof. Q.E.D.

Proof of Proposition 4: See the main text. Q.E.D.

Proof of Proposition 5: Recall the positive semidefinite matrix $\Sigma_0^{\frac{1}{2}}\Omega\Sigma_0^{\frac{1}{2}}$ admits an eigendecomposition $\Sigma_0^{\frac{1}{2}}\Omega\Sigma_0^{\frac{1}{2}} = U\Omega_d U'$. Let

$$\widehat{\Sigma} \equiv U'\Sigma_0^{-\frac{1}{2}}\Sigma\Sigma_0^{-\frac{1}{2}}U.$$

Then

$$\Sigma = \Sigma_0^{\frac{1}{2}}U\widehat{\Sigma}U'\Sigma_0^{\frac{1}{2}}, \quad \Sigma^{-1} = \Sigma_0^{-\frac{1}{2}}U\widehat{\Sigma}^{-1}U'\Sigma_0^{-\frac{1}{2}}.$$

We rewrite the objection function as

$$\begin{aligned} & \frac{1}{2}\log\det\Sigma_0 - \frac{1}{2}\log\det\Sigma \\ = & \frac{1}{2}\log\det\Sigma_0 - \frac{1}{2}\log\det\left(\Sigma_0^{\frac{1}{2}}U\widehat{\Sigma}U'\Sigma_0^{\frac{1}{2}}\right) \\ = & -\frac{1}{2}\log\det\left(U\widehat{\Sigma}U'\right) = -\frac{1}{2}\log\det\left(\widehat{\Sigma}\right). \end{aligned}$$

Using the rotation invariance property of the trace operator, we derive

$$\text{tr}\left(\Omega\Sigma\right) = \text{tr}\left(\Omega\Sigma_0^{\frac{1}{2}}U\widehat{\Sigma}U'\Sigma_0^{\frac{1}{2}}\right) = \text{tr}\left(U'\Sigma_0^{\frac{1}{2}}\Omega\Sigma_0^{\frac{1}{2}}U\widehat{\Sigma}\right) = \text{tr}\left(\Omega_d\widehat{\Sigma}\right).$$

We also show that

$$\begin{aligned} \Sigma \quad \preceq \quad & \Sigma_0 \Longleftrightarrow \left(\Sigma_0^{\frac{1}{2}}U\right)\widehat{\Sigma}\left(U'\Sigma_0^{\frac{1}{2}}\right) - \left(\Sigma_0^{\frac{1}{2}}U\right)\left(U'\Sigma_0^{\frac{1}{2}}\right) \preceq 0 \\ \Longleftrightarrow \quad & \left(\Sigma_0^{\frac{1}{2}}U\right)\left(\widehat{\Sigma} - I\right)\left(U'\Sigma_0^{\frac{1}{2}}\right) \preceq 0 \\ \Longleftrightarrow \quad & \widehat{\Sigma} \preceq I. \end{aligned}$$

Thus we can rewrite the static RI problem as

$$\kappa_0\left(D\right) = \min_{\widehat{\Sigma}} -\frac{1}{2}\log\det\left(\widehat{\Sigma}\right)$$

subject to

$$0 \prec \widehat{\Sigma} \preceq I, \ \operatorname{tr}\left(\Omega_d \widehat{\Sigma}\right) \leq D.$$

By the Hadamard inequality for positive definite matrices (Cover and Thomas, 2006, Theorem 17.9.2),

$$\det \widehat{\Sigma} \leq \prod_{i=1}^{n_x} \widehat{\Sigma}_{ii},$$

where $\widehat{\Sigma}_{ii}$ is the diagonal element of $\widehat{\Sigma}$. The equality holds if and only if $\widehat{\Sigma}$ is diagonal. Thus, if diagonal elements of $\widehat{\Sigma}$ are fixed, $\det \widehat{\Sigma}$ is maximized by setting all off-diagonal entries to zero. The optimal solution to the above problem must be diagonal. Let $\widehat{\Sigma} = \operatorname{diag}\left(\widehat{\Sigma}_i\right)_{i=1}^{n_x}$. Then the optimization problem becomes

$$2\kappa_0(D) = \min_{\widehat{\Sigma}_i} - \sum_{i=1}^{n_x} \log \widehat{\Sigma}_i$$

subject to

$$0 < \widehat{\Sigma}_i \leq 1 \text{ all } i,$$
$$\sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i \leq D. \tag{A.5}$$

If $d_i = 0$ for some i, then $\widehat{\Sigma}_i = 1$. If $d_i > 0$, then we use the Kuhn-Tucker condition to show that

$$\widehat{\Sigma}_i = \min\left(1, \frac{1}{\alpha d_i}\right),$$

where $\alpha > 0$ is the Lagrange multiplier associated with the distortion constraint. We then obtain the solution for Σ in the proposition. The optimal signal-to-noise ratio is given by

$$\begin{aligned}
\Phi &= \Sigma^{-1} - \Sigma_0^{-1} = \Sigma_0^{-\frac{1}{2}} U \widehat{\Sigma}^{-1} U' \Sigma_0^{-\frac{1}{2}} - \Sigma_0^{-1} \\
&= \Sigma_0^{-\frac{1}{2}} U \left(\widehat{\Sigma}^{-1} - I\right) U' \Sigma_0^{-\frac{1}{2}} \\
&= \Sigma_0^{-\frac{1}{2}} U \operatorname{diag}\left\{\max\left(0, \alpha d_i - 1\right)_{i=1}^{n_x}\right\} U' \Sigma_0^{-\frac{1}{2}}.
\end{aligned}$$

The dimension of the signal vector is equal to the rank of Φ, which is equal to the rank of the matrix $\operatorname{diag}\{\max\left(0, \alpha d_i - 1\right)_{i=1}^{n_x}\}$. This rank is less or equal to the rank of Ω_d or Ω. Q.E.D.

Proof of Corollary 1: Let the Lagrange multiplier α satisfy $m/\alpha = D$, where m denotes the rank of Ω. Then given the condition in the corollary,

$$\frac{1}{\alpha d_i} = \frac{D}{m d_i} < \frac{\min_i \{d_i > 0\}}{d_i} < 1.$$

Thus $\widehat{\Sigma}_i = 1/\left(\alpha d_i\right)$ for all i satisfying $d_i > 0$, and $\widehat{\Sigma}_i = 1$ for all i with $d_i = 0$. Since $\alpha d_i > 1$ for all i satisfying $d_i > 0$, it follows from Proposition 5 that the signal dimension is equal to the rank of Ω.

For simplicity we suppose that d_1 is the unique smallest positive eigenvalue and $d_m \geq ... \geq d_2 > d_1 > 0$. We show that the signal dimension is equal to $m-1$ when

$$md_1 < D < (m-1)\, d_2 + d_1.$$

To show this, let α satisfy

$$d_1 + \frac{m-1}{\alpha} = D.$$

Then we can check that $\alpha d_1 < 1$ and $\alpha d_i > 1$ for $i = 2, ..., m$. Then the dimension of $\text{diag}\{\max(0, \alpha d_i - 1)_{i=1}^{n_x}\}$ is equal to $m-1$. It follows from Proposition 5 that the signal dimension is equal to $m-1$. By a similar procedure we can show that the signal dimension decreases as D increases. Q.E.D.

Proof of Proposition 6: As in the proof Proposition 5, we can rewrite the static relaxed RI problem as

$$\min_{\widehat{\Sigma}} \; \text{tr}\left(\Omega_d \widehat{\Sigma}\right) - \frac{\lambda}{2} \log\det\left(\widehat{\Sigma}\right)$$

subject to

$$0 \prec \widehat{\Sigma} \preceq I.$$

By the same argument, the optimal $\widehat{\Sigma}$ must be diagonal. Let $\widehat{\Sigma} = \text{diag}\left(\widehat{\Sigma}_i\right)_{i=1}^{n_x}$. Then the optimization problem becomes

$$\min_{\widehat{\Sigma}_i} \; \sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i - \frac{\lambda}{2} \sum_{i=1}^{n_x} \log \widehat{\Sigma}_i$$

subject to

$$\widehat{\Sigma}_i \leq 1 \text{ all } i.$$

Solving this problem yields

$$\widehat{\Sigma}_i = \min\left(\frac{\lambda}{2d_i}, 1\right).$$

The optimal signal-to-noise ratio is given by

$$\begin{aligned}
\Phi &= \Sigma^{-1} - \Sigma_0^{-1} = \Sigma_0^{-\frac{1}{2}} U \widehat{\Sigma}^{-1} U' \Sigma_0^{-\frac{1}{2}} - \Sigma_0^{-1} \\
&= \Sigma_0^{-\frac{1}{2}} U \left(\widehat{\Sigma}^{-1} - I\right) U' \Sigma_0^{-\frac{1}{2}} \\
&= \Sigma_0^{-\frac{1}{2}} U \; \text{diag}\left\{\max\left(0, \frac{2d_i}{\lambda} - 1\right)_{i=1}^{n_x}\right\} U' \Sigma_0^{-\frac{1}{2}}.
\end{aligned}$$

Thus the signal dimension is equal to the number of d_i such that $d_i > \lambda/2$ and decreases as λ increases. Q.E.D.

Proof of Proposition 7: See the main text. Q.E.D.

Proof of Proposition 8: As discussed in Section 3.3, we only need to solve the rate distortion problem (46). The matrix determinant lemma implies that

$$\begin{aligned} & \frac{1}{2}\log\det(A\Sigma A' + W) - \frac{1}{2}\log\det\Sigma \\ = \; & \frac{1}{2}\log\det W - \frac{1}{2}\log\det\left(\Sigma^{-1} + A'W^{-1}A\right)^{-1}. \end{aligned}$$

Thus this problem becomes

$$\kappa(D) = \min_{\Pi,\Sigma\succ 0} \frac{1}{2}\log\det W - \frac{1}{2}\log\det\Pi \tag{A.6}$$

subject to

$$\Pi = \left(\Sigma^{-1} + A'W^{-1}A\right)^{-1}, \tag{A.7}$$

$$A\Sigma A' + W \succeq \Sigma, \tag{A.8}$$

$$\mathrm{tr}(\Omega\Sigma) \leq D. \tag{A.9}$$

Recall the positive semidefinite matrix $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}$ admits an eigendecomposition $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}} = U\Omega_d U'$. Define matrices

$$\widehat{\Pi} = U'W^{-\frac{1}{2}}\Pi W^{-\frac{1}{2}}U,\ \widehat{\Sigma} = U'W^{-\frac{1}{2}}\Sigma W^{-\frac{1}{2}}U.$$

Then

$$\Pi = W^{\frac{1}{2}}U\widehat{\Pi}U'W^{\frac{1}{2}},\ \Sigma = W^{\frac{1}{2}}U\widehat{\Sigma}U'W^{\frac{1}{2}}.$$

As in the proof of Proposition 5, we can derive that

$$\frac{1}{2}\log\det W - \frac{1}{2}\log\det\Pi = -\frac{1}{2}\log\det\widehat{\Pi}.$$

Given $A = \rho I$, we can also show that equations (A.7), (A.8), and (A.9) are equivalent to

$$\widehat{\Pi}^{-1} = \widehat{\Sigma}^{-1} + \rho^2 I, \tag{A.10}$$

$$I \succeq \left(1-\rho^2\right)\widehat{\Sigma}, \tag{A.11}$$

$$\mathrm{tr}\left(\Omega_d\widehat{\Sigma}\right) \leq D. \tag{A.12}$$

Now the preceding problem is equivalent to

$$\kappa(D) = \min_{\widehat{\Pi},\widehat{\Sigma}} -\frac{1}{2}\log\det\widehat{\Pi}$$

subject to (A.10), (A.11), and (A.12). By the Hadamard inequality for positive definite matrices (Cover and Thomas, 2006, Theorem 17.9.2),

$$\det\widehat{\Pi} \leq \prod_{i=1}^{n_x}\widehat{\Pi}_i,$$

where $\widehat{\Pi}_i$ is the diagonal element of $\widehat{\Pi}$. The equality holds if and only if $\widehat{\Pi}$ is diagonal. Thus, if diagonal elements of $\widehat{\Pi}$ are fixed, $\det \widehat{\Pi}$ is maximized by setting all off-diagonal entries to zero. The optimal solution for $\widehat{\Pi}$ must be diagonal. By (A.10), $\widehat{\Sigma}$ is also diagonal

$$\widehat{\Sigma} = \operatorname{diag}\left(\left[\widehat{\Pi}_i^{-1} - \rho^2\right]^{-1}\right)_{i=1}^{n_x}. \tag{A.13}$$

Thus the problem is equivalent to

$$\kappa(D) = \min_{\widehat{\Pi}_i} -\frac{1}{2}\sum_{i=1}^{n_x} \log \widehat{\Pi}_i$$

subject to

$$\sum_{i=1}^{n_x} d_i \left[\widehat{\Pi}_i^{-1} - \rho^2\right]^{-1} \leq D$$

$$\left(1-\rho^2\right)\left[\widehat{\Pi}_i^{-1} - \rho^2\right]^{-1} \leq 1,\ i = 1, .., n_x.$$

Equivalently rewriting this problem in terms of $\widehat{\Sigma}_i$ using (A.13) yields

$$2\kappa(D) = \min_{\widehat{\Sigma}_i} \sum_{i=1}^{n_x} \log\left(\rho^2 + \frac{1}{\widehat{\Sigma}_i}\right)$$

subject to

$$\sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i \leq D,\quad 0 < \widehat{\Sigma}_i \leq \frac{1}{1-\rho^2},\quad i = 1, ..., n_x.$$

If $d_i = 0$, then $\widehat{\Pi}_i = 1$ or $\widehat{\Sigma}_i = 1/\left(1-\rho^2\right)$. If $d_i > 0$, then we use the Kuhn-Tucker condition to show that

$$\widehat{\Sigma}_i = \min\left(\frac{1}{1-\rho^2}, \widehat{\Sigma}_i^*\right), \tag{A.14}$$

where the Lagrange multiplier $\alpha > 0$ and $\widehat{\Sigma}_i^*, i = 1, ..., n_x$, are the unique positive solution to the system of $n_x + 1$ equations

$$\left[\rho^2\left(\widehat{\Sigma}_i^*\right)^2 + \widehat{\Sigma}_i^*\right]^{-1} = \alpha d_i,\ i = 1, ..., n_x, \tag{A.15}$$

$$D = \sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i. \tag{A.16}$$

To show the existence and uniqueness, we observe that there is a unique positive root $\widehat{\Sigma}_i^*$ to equation (A.15) when $\alpha, d_i > 0$ for each i :

$$\widehat{\Sigma}_i^* = \frac{1}{2\rho^2}\left(\sqrt{1 + \frac{4\rho^2}{\alpha d_i}} - 1\right).$$

This root decreases with α, approaches infinity as $\alpha \to 0$, and approaches zero as $\alpha \to \infty$. Thus, $\sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i$ approaches zero as $\alpha \to \infty$ and approaches $\sum_{i=1}^{n_x} d_i / \left(1-\rho^2\right)$ as $\alpha \to 0$. Moreover,

$\sum_{i=1}^{n_x} d_i\widehat{\Sigma}_i$ as a function of α decreases with α. Thus, when $0 < D < \sum_{i=1}^{n_x} d_i / \left(1-\rho^2\right)$, there is a unique solution for $\alpha > 0$ to equation (A.16) by the intermediate value theorem.

The optimal signal-to-noise ratio is given by

$$\begin{aligned}
\Phi &= \Sigma^{-1} - \left(\rho^2\Sigma + W\right)^{-1} \\
&= W^{-\frac{1}{2}}U\widehat{\Sigma}^{-1}U'W^{-\frac{1}{2}} - \left[\rho^2 W^{\frac{1}{2}}U\widehat{\Sigma}U'W^{\frac{1}{2}} + W\right]^{-1} \\
&= W^{-\frac{1}{2}}U\widehat{\Sigma}^{-1}U'W^{-\frac{1}{2}} - W^{-\frac{1}{2}}U\left[\rho^2\widehat{\Sigma} + I\right]^{-1}U'W^{-\frac{1}{2}} \\
&= W^{-\frac{1}{2}}U\left(\widehat{\Sigma}^{-1} - \left[\rho^2\widehat{\Sigma} + I\right]^{-1}\right)U'W^{-\frac{1}{2}} \\
&= W^{-\frac{1}{2}}U \text{ diag}\left\{\max\left(0, \alpha d_i\left[1 - \left(1-\rho^2\right)\widehat{\Sigma}_i^*\right]\right)\right\}U'W^{-\frac{1}{2}},
\end{aligned}$$

where the last equality follows from (A.14) and (A.15). Q.E.D.

Proof of Corollary 2: The proof is similar to that of Corollary 1. Suppose that $d_1 \leq d_2 \leq ... \leq d_m$ are the positive eigenvalues, where m is the rank of Ω. We want to show that $\widehat{\Sigma}_i^* < 1/\left(1-\rho^2\right)$ for $i = 1, 2..., m$, if $D > 0$ is sufficiently small. Then the rank of $\text{diag}\left\{\max\left(0, \alpha d_i\left[1-\left(1-\rho^2\right)\widehat{\Sigma}_i^*\right]\right)_{i=1}^{n_x}\right\}$ is m. It follows from Proposition 8 that the signal dimension is m.

Consider the system of $n_x + 1$ equations (A.15) and

$$D = \sum_{i=1}^{n_x} d_i\widehat{\Sigma}_i^*. \tag{A.17}$$

As we show in the proof of Proposition 8, the positive root $\widehat{\Sigma}_i^* > 0$ of equation (A.15) decreases with $\alpha > 0$. We write it as a decreasing function of α, $\widehat{\Sigma}_i^*(\alpha)$. It also decreases with d_i. Thus one of the largest roots in the set $\{\widehat{\Sigma}_i^*(\alpha) : i = 1, ..., m\}$ is $\widehat{\Sigma}_1^*(\alpha)$. Let α_1^* be the unique positive value such that

$$\widehat{\Sigma}_1^*(\alpha_1^*) = \frac{1}{1-\rho^2}.$$

Suppose that

$$0 < D < D_1^* \equiv \sum_{i=1}^{n_x} d_i\widehat{\Sigma}_i^*(\alpha_1^*). \tag{A.18}$$

Then we claim that the solution to equation (A.15) and (A.17) satisfies $\widehat{\Sigma}_i^* < 1/\left(1-\rho^2\right)$. The reason is that the equation

$$D = \sum_{i=1}^{n_x} d_i\widehat{\Sigma}_i^*(\alpha) \tag{A.19}$$

as a function of α has a unique positive root $\alpha > \alpha_1^*$, when (A.18) is satisfied. When $\alpha > \alpha_1^*$, we have $\widehat{\Sigma}_1^*(\alpha) < 1/\left(1-\rho^2\right)$. Thus $\widehat{\Sigma}_i^*(\alpha) < 1/\left(1-\rho^2\right)$ for $i = 1, 2, ..., m$ as desired.

Now suppose that $d_m \geq ... \geq d_2 > d_1 > 0$. When D is sufficiently large, the solution α to equation (A.19) decreases to a value smaller than α_1^*. Then $\widehat{\Sigma}_i^*(\alpha) > 1/\left(1-\rho^2\right)$. The second largest element in the set $\{\widehat{\Sigma}_i^*(\alpha) : i = 1, ..., m\}$ is $\widehat{\Sigma}_2^*(\alpha)$. By the previous argument, there is a critical value α_2^* associated with d_2 such that

$$\widehat{\Sigma}_2^*(\alpha_2^*) = \frac{1}{1-\rho^2}.$$

There is a critical value D_2^* such that, when $D_1^* < D < D_2^*$, the root of equation (A.19) satisfies $\alpha_2^* < \alpha < \alpha_1^*$. Then we have

$$\widehat{\Sigma}_2^*(\alpha) < \frac{1}{1-\rho^2}, \text{ but } \widehat{\Sigma}_1^*(\alpha) > \frac{1}{1-\rho^2}.$$

Thus $\widehat{\Sigma}_i^*(\alpha) < 1/\left(1-\rho^2\right)$ for $i = 2, 3, ..., n_x$. The rank of $\text{diag}\left\{\max\left(0, \alpha d_i\left[1-\left(1-\rho^2\right)\widehat{\Sigma}_i^*\right]\right)_{i=1}^{n_x}\right\}$ is $m-1$. By a similar procedure we can show that the signal dimension decreases as D increases. Q.E.D.

Proof of Proposition 9: We consider the general case in which x_t is an n_x-dimensional vector. Since $rank(\Omega) = 1$, we have $rank\left(W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}\right) = 1$. We claim that matrix $W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}$ has a unique positive eigenvalue $d_1 \equiv \left\|W^{1/2}a\right\|^2$ and an associated unit eigenvector $W^{\frac{1}{2}}a/\left\|W^{1/2}a\right\|$ where $\|\cdot\|$ denotes the Euclidean norm. To prove this claim we verify that

$$\begin{aligned} W^{\frac{1}{2}}\Omega W^{\frac{1}{2}}\frac{W^{\frac{1}{2}}a}{\left\|W^{1/2}a\right\|} &= \left(W^{\frac{1}{2}}a\right)\left(W^{\frac{1}{2}}a\right)'\frac{W^{\frac{1}{2}}a}{\left\|W^{1/2}a\right\|} = \left(W^{\frac{1}{2}}a\right)a'W^{\frac{1}{2}}\frac{W^{\frac{1}{2}}a}{\left\|W^{1/2}a\right\|} \\ &= \left(W^{\frac{1}{2}}a\right)\frac{\left\|W^{1/2}a\right\|^2}{\left\|W^{1/2}a\right\|} = \left\|W^{1/2}a\right\|^2\frac{W^{\frac{1}{2}}a}{\left\|W^{1/2}a\right\|}. \end{aligned}$$

Thus Ω_d has only one positive element $d_1 = \left\|W^{1/2}a\right\|^2$ and other diagonal elements $d_i = 0$ for $i = 2, ..., n_x$. Moreover, the optimal signal dimension is at most one.

Suppose that $0 < D < \left(1-\rho^2\right)^{-1}d_1$. By Propositions 3 and 8, we have

$$\begin{aligned} \widehat{\Sigma}_1 &= \min\left(\frac{1}{1-\rho^2}, \widehat{\Sigma}_1^*\right), \\ \widehat{\Sigma}_i &= \frac{1}{1-\rho^2},\ i = 2, ..., n_x, \end{aligned}$$

where α and $\widehat{\Sigma}_1^*$ are the unique positive solution to the following system of equations:

$$\frac{1}{\alpha d_1} = \rho^2\left(\widehat{\Sigma}_1^*\right)^2 + \widehat{\Sigma}_1^*, \quad D = d_1\widehat{\Sigma}_1^*.$$

We can solve explicitly

$$\alpha = \frac{d_1}{\rho^2 D^2 + d_1 D}, \quad \widehat{\Sigma}_1^* = \frac{D}{d_1}.$$

The optimal information structure $\{C, V\}$ satisfies

$$C'V^{-1}C = W^{-\frac{1}{2}}U\ \text{diag}\left\{\max\left(0, \alpha d_i\left[1-\left(1-\rho^2\right)\widehat{\Sigma}_i^*\right]\right)_{i=1}^{n_x}\right\}U'W^{-\frac{1}{2}}.$$

There is only one positive element in the inside diagonal matrix, which is

$$\alpha d_1 \left[1 - \left(1 - \rho^2\right) \widehat{\Sigma}_1^*\right] = \frac{d_1^2}{\rho^2 D^2 + d_1 D} \left[1 - \left(1 - \rho^2\right) \frac{D}{d_1}\right] > 0.$$

The optimal information structure corresponds to the positive eigenvalue's eigenvector and is given by

$$C' = W^{-\frac{1}{2}} \frac{W^{\frac{1}{2}} a}{\left\|W^{1/2} a\right\|} \Longrightarrow C = \frac{a'}{\left\|W^{1/2} a\right\|},$$

$$V^{-1} = \frac{d_1^2}{\rho^2 D^2 + d_1 D} \left[1 - \left(1 - \rho^2\right) \frac{D}{d_1}\right] > 0.$$

The optimal conditional covariance in the proposition follows from Proposition 8. In particular

$$\Sigma = W^{\frac{1}{2}} U \left[\begin{array}{cc} \frac{D}{d_1} & 0 \\ 0 & \frac{1}{1-\rho^2} I \end{array}\right] U' W^{\frac{1}{2}}.$$

Partition $U = [U_1, U_2]$ conformably, where $U_1 = W^{\frac{1}{2}} a / \left\|W^{1/2} a\right\|$. Then we have $U_1 U_1' + U_2 U_2' = I$. Thus

$$\Sigma = W^{\frac{1}{2}} \left[\frac{I}{1 - \rho^2} - U_1 U_1' \left(\frac{1}{1 - \rho^2} - \frac{D}{d_1}\right)\right] W^{\frac{1}{2}}.$$

Simplifying yields the expression in the proposition.

We can normalize C as $C = a'$ so that the normalized optimal signal is given by

$$s_t = a' x_t + \left\|W^{1/2} a\right\| v_t.$$

We then obtain (51).

If $D \geq \left(1 - \rho^2\right)^{-1} d_1$, we can check that $\widehat{\Sigma}_i = 1/\left(1 - \rho^2\right)$ for all i so that $\Sigma = \rho^2 \Sigma + W$ and the firm does not process any information.

Using Proposition 8 we can derive the rate distortion function. Q.E.D.

B The Case of $W \succeq 0$

Here we present the result for the rate distortion problem in the finite-horizon case. The results for the finite-horizon relaxed problem and for the infinite-horizon case are similar and omitted.

Proposition 10 *Suppose that $W_t \succeq 0$ is singular for some t, $\Omega_t \succeq 0$, and rank$(A_t) = n_x$ for $t = 0, 1, ..., T$. Then the optimal information structure $\{C_t, V_t\}_{t=0}^T$ for problem (22) or (24) can be constructed by solving the following determinant maximization problem with decision variables $\left\{\Sigma_{t|t}, \Psi_t\right\}_{t=0}^T$:*

$$\kappa_T(D) = \min - \sum_{t=0}^{T} \frac{1}{2} \log \det \Psi_t + c$$

subject to (27), (28), (29) and the following constraints

$$\Sigma_{T|T} = \Psi_T, \quad \Psi_t \succ 0, \tag{B.1}$$

$$\begin{bmatrix} I - \Psi_t & M_t' \\ M_t & A_t \Sigma_{t|t} A_t' + W_t \end{bmatrix} \succeq 0, \tag{B.2}$$

for $t = 0, 1, 2, \ldots T$, *where* $W_t = M_t M_t'$, *and*

$$c = \frac{1}{2} \log \det \left(\Sigma_0 \right) + \sum_{t=1}^{T} \log \left| \det A_{t-1} \right|.$$

Proof: We can rewrite the objective function in (30) by regrouping terms as a sum of the initial cost $\frac{1}{2} \log \det (\Sigma_0)$, the final cost $\frac{1}{2} \log \det \left(\Sigma_{T|T} \right)$, and period costs

$$\frac{1}{2} \log \det \left(A_t \Sigma_{t|t} A_t' + W_t \right) - \frac{1}{2} \log \det \left(\Sigma_{t|t} \right),$$

for $t = 0, 1, \ldots, T-1$. Since W_t is singular, we apply the matrix determinant lemma by employing the decomposition $W_t = M_t M_t'$ with $M_t \succeq 0$. It follows that

$$\det \left(A_t \Sigma_{t|t} A_t' + W_t \right) = \det \left(I^{-1} + M_t' \left(A_t \Sigma_{t|t} A_t' \right)^{-1} M_t \right) \det(I) \det \left(A_t \Sigma_{t|t} A_t' \right).$$

Therefore, the period cost function can be written as

$$-\frac{1}{2} \log \det \left(I + M_t' \left(A_t \Sigma_{t|t} A_t' \right)^{-1} M_t \right)^{-1} + \log \left| \det A_t \right|,$$

provided that A_t is non-singular. Due to the monotonicity of the determinant function, this expression is equal to the optimal value of

$$\min_{\Psi_t} \quad \log \left| \det A_t \right| - \frac{1}{2} \log \det \Psi_t$$

subject to

$$0 \prec \Psi_t \preceq \left(I + M_t' \left(A_t \Sigma_{t|t} A_t' \right)^{-1} M_t \right)^{-1}, \tag{B.3}$$

where $\Psi_t \succ 0$. Now use the matrix inversion lemma to get

$$\left(I + M_t' \left(A_t \Sigma_{t|t} A_t' \right)^{-1} M_t \right)^{-1} = I - M_t' \left(A_t \Sigma_{t|t} A_t' + M_t M_t' \right) M_t.$$

By the Schur complement property, (B.3) is equivalent to

$$\begin{bmatrix} I - \Psi_t & M_t' \\ M_t & A_t \Sigma_{t|t} A_t' + W_t \end{bmatrix} \succeq 0. \tag{B.4}$$

Finally, summing up the sequence of period cost objectives subject to the constraint (27), (28), (29), and (B.4), we obtain the representation in the proposition. Q.E.D. □

To illustrate the application of this proposition, we consider the LQG control problem with VAR(p) state dynamics

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + ... + A_p x_{t-p} + B_0 u_t + \epsilon_t,$$

where $A_1, ...,$ and A_p are $n \times n$ matrices and ϵ_t is Gaussian white noise with covariance matrix $W_0 \succ 0$. We transform the state dynamics into VAR(1) form:

$$\overline{x}_t = A\bar{x}_{t-1} + Bu_t + \bar{\epsilon}_t,$$

where $\bar{x}_t = \left[x_t', \ x_{t-1}', \ ...x_{t-p+1}'\right]'$, $\bar{\epsilon}_t$ is a Gaussian white noise with covariance matrix W, and

$$A = \begin{bmatrix} A_1 & A_2 & ... & A_{p-1} & A_p \\ I_n & 0 & .... & 0 & 0 \\ 0 & I_n & .... & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & I_n & 0 \end{bmatrix}, \ B = \begin{bmatrix} B_0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ W = \begin{bmatrix} I_n \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} W_0 \begin{bmatrix} I_n & 0' & 0' & 0' & 0' \end{bmatrix}.$$

Now the problem fits in our general LQG RI framework. Notice that the covariance matrix of $\bar{\epsilon}_t$ satisfies $W \succeq 0$, but it is singular. So Proposition 4 or 7 does not apply. As long as A_p is invertible so that A is invertible, we can apply Proposition 10 to solve the model numerically.

C Numerical Implementation

For space limitation, we only describe the implementation of the infinite-horizon stationary RI optimization problem. As discussed in the main text, we only need to solve the rate distortion problem in Proposition 7. The relaxed problem and the finite-horizon problem can be similarly solved. We apply the software package SDPT3, version 4.0, which can solve semidefinite programming problems up to 100 dimensions efficiently and robustly (Toh, Todd, and Tutuncu (1999) and Tutuncu, Toh, and Todd (2003)). Instead of presenting the general form of the optimization problems that can be solved by this package, we focus on the form in which our problem can fit:

$$\max_{y \in \mathbb{R}^m, \ Z_j^s, \ Z^l} \quad b'y + \sum_{j=1}^{n_s} \nu_j^s \log\det(Z_j^s) \tag{C.1}$$

subject to

$$(\mathcal{A}_j^s)^T y + Z_j^s = c_j^s, \quad Z_j^s \in K_s^{s_j}, \quad j = 1, 2, ..., n_s, \tag{C.2}$$

$$\left(A^l\right)' y + Z^l = c^l, \quad Z_l \in \mathbb{R}_+^{n_l}, \tag{C.3}$$

where constraints (C.2) and (C.3) correspond to the linear matrix inequality and the linear vector inequality in applications.

Let us explain the notations in this problem. The choice variables are an m-dimensional real vector y, an s_j-dimensional positive semidefinite matrix Z_j^s, and an n_l-dimensional nonnegative vector Z^l. The set $K_s^{s_j}$ denotes the cone of positive semidefinite symmetric matrices of dimension s_j. In our application, Z_j^s and Z^l are slack variables that transform the matrix inequality constraints and linear inequalities into equalities. All other variables are exogenous input parameters. In particular, b is an m-dimensional vector, $\nu_s^j \geq 0$ for all j, c_j^s is an s_j-dimensional matrix, A^l is an m by n_l dimensional matrix, and c^l is an n_l-dimensional vector.

Let $\mathcal{S}^n$ denote the set of all symmetric matrices. Define a vectorization operator for a symmetric matrix $X \in \mathcal{S}^n$ as $\mathbf{svec} : \mathcal{S}^n \longmapsto \mathbb{R}^{n(n+1)/2}$,

$$\mathbf{svec}(X) = \left[X_{11}, \ \ \sqrt{2}X_{12}, \ \ X_{22}, \ \sqrt{2}X_{13}, ..., \ \sqrt{2}X_{1n}, ..., \ \ \sqrt{2}X_{n-1,n}, \ \ X_{n,n}\right]',$$

where the $\sqrt{2}$ is to make the operation isometry. We use the Matlab notation $[U; V]$ to denote the matrix obtained by appending V below the last row of U. Then we identfiy $\mathcal{A}_j^s$ with the following m by $s_j\left(s_j+1\right)/2$ matrix

$$A_j^s = \left[\mathbf{svec}(a_{j,1}^s)'; \ \mathbf{svec}(a_{j,2}^s)'; \ \ ....; \ \mathbf{svec}(a_{j,m}^s)'\right],$$

where $a_{j,1}^s, ..., a_{j,m}^s$ are model specific input symmetric coefficient matrices of dimension s_j associated with each element in y. The operation $(\mathcal{A}_j^s)^T y$ is then defined as

$$(\mathcal{A}_j^s)^T y = \sum_{k=1}^{m} a_{j,k}^s y_k.$$

To use the software SDPT3, we need to transform our optimization problem in Proposition 7 into the form in (C.1), (C.2), and (C.3) by the following four steps.

Step 1. Set the choice variable $y = [\mathbf{svec}(\Sigma); \ \ \mathbf{svec}(\Pi)]$ so that $m = n_x(n_x+1)$. The other choice variables are the slack variables Z_1^s, Z_2^s, and Z_3^s defined next. Set $s_j = n_x$.

Step 2. To derive the constraint (C.3), we consider (47) and write

$$\text{tr}(\Omega\Sigma) = (\mathbf{svec}(\Omega))'\mathbf{svec}(\Sigma).$$

Set $n_l = 1$, $A^l = [\mathbf{svec}(\Omega); \ \ \mathbf{0}]$, and $c^l = D$. Introduce $Z^l \geq 0$. Then (47) is in the form of (C.3).

Step 3. Derive the constraint (C.2). There are three linear matrix inequalities in our optimization problem. Set $n_s = 3$. We introduce three slack variables Z_1^s, Z_2^s, and Z_3^s. First, write the semidefinite constraint $\Pi \succ 0$ as

$$-\Pi + Z_1^s = 0, \ \ Z_1^s \in K_s^{n_x}.$$

and

$$-\Pi = \sum_{k=1}^{m} a_{1,k}^s y_k,$$

where $a^s_{1,k}$, $k = 1, ..., m$, are coefficient matrices. To understand the last equation, consider the simple 2 by 2 case:

$$-\Pi = -\begin{bmatrix} \pi_{12} & \pi_{12} \\ \pi_{12} & \pi_{22} \end{bmatrix} = -\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \pi_{11} - \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \left(\sqrt{2}\pi_{12}\right) - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \pi_{22},$$

where $\left[\pi_{11}, \sqrt{2}\pi_{12}, \pi_{22}\right]'$ are in the vector y. Set $c^s_1 = \mathbf{0}$.

Second, we can similarly write the no-forgetting constraint (45) as

$$\Sigma - A\Sigma A' + Z^s_2 = W, \;\; Z^s_2 \in K^{n_x}_s,$$

and

$$\Sigma - A\Sigma A' = \sum_{k=1}^{m} a^s_{2,k} y_k,$$

where $a^s_{2,k}$, $k = 1, ..., m$, are coefficient matrices. Set $c^s_2 = W$. The Matlab file $lmifun.m$ in our code computes this transformation and $\mathcal{A}^s_1$ and $\mathcal{A}^s_2$.

Third, we write (48) as

$$\begin{bmatrix} \Pi & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \Sigma & \Sigma A' \\ A\Sigma & A\Sigma A' + W \end{bmatrix} + Z^s_3 = \begin{bmatrix} 0 & 0 \\ 0 & W \end{bmatrix}, \;\; Z^s_3 \in K^{2n_x}_s,$$

and

$$\begin{bmatrix} \Pi & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \Sigma & \Sigma A' \\ A\Sigma & A\Sigma A' + W \end{bmatrix} = \sum_{k=1}^{m} a^s_{3,k} y_k,$$

where $a^s_{3,k}$, $k = 1, 2, ..., m$, are coefficient matrices. Set

$$c^s_3 = \begin{bmatrix} 0 & 0 \\ 0 & W \end{bmatrix}.$$

The Matlab file $lmifun3.m$ in our code computes this transformation and the matrix $\mathcal{A}^s_3$.

Step 4. Set input parameters in the objective function (C.1) as $b = \mathbf{0}$, $\nu^s_1 = 0.5$, $\nu^s_2 = \nu^s_3 = 0$. After solving for optimal $y = [\mathbf{svec}(\Sigma); \;\; \mathbf{svec}(\Pi)]$, we use the inverse operator of **svec**, **smat**, to obtain optimal Σ and Π. Since $\Pi = Z^s_1$ enters log det in the objective function, optimal $\Pi = Z^s_1$ will be positive definite because the singular case will lead the objective function to approach negative infinity. After solving for the optimal Π, we obtain the rate distortion function

$$\kappa(D) = -\frac{1}{2} \log \det \Pi + \frac{1}{2} \log \det W.$$

The Matlab file *RIdata.m* stores the problem input data in the above form and calls the main Matlab function for SDPT3, *sqlp.m*. The accuracy tolerance in terms of the relative duality gap and infeasibilities is 10^{-8}. Using the dynamic price setting model as an example, the Matlab file *dyprice.m* calls *RIdata.m* and computes the optimal solution for κ, Σ, and Φ, given the input A, Ω, W, and D. The Matlab file *sig.m* computes the optimal information structure C and V, given Φ. Finally, the code *IRFprice.m* computes the impulse response functions.

References

Caplin, Andrew, Mark Dean, and John Leahy, 2018, Rational Inattention, Optimal Consideration Sets, and Stochastic Choice, forthcoming in *Review of Economic Studies.*

Cover, Thomas M., and Joy A. Thomas, 2006, *Elements of Information Theory*, 2th edition, John Wiley & Sons, Inc.

Fulton, Chad, 2017, Mechanics of Linear Quadratic Gaussian Rational Inattention Tracking Problems, Finance and Economics Discussion Series 2017-109. Washington: Board of Governors of the Federal Reserve System.

Fulton, Chad, 2018, Mechanics of Static Quadratic Gaussian Rational Inattention Tracking Problems, working paper, Board of Governors of the Federal Reserve System.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans, 2005, Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy, *Journal of Political Economy* 113, 1-45.

Harville, David A., 1997, *Matrix Algebra from a Statistician's Perspective*, Springer, 1997, vol. 1.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2016, A Rational Theory of Mutual Funds' Attention Allocation, *Econometrica* 84, 571-626.

Kydland, Finn E., and Edward C. Prescott, 1982, Time to Build and Aggregate Fluctuations, *Econometrica* 50, 1345-70.

Ljungqvist Lars, and Thomas J. Sargent, 2004, *Recursive Macroeconomic Theory*, 2nd, the MIT Press, Cambridge, MA.

Luo, Yulei, 2008, Consumption Dynamics under Information Processing Constraints, *Review of Economic Dynamics* 11, 366-385.

Luo, Yulei, Jun Nie, and Eric R. Young, 2015, Slow Information Diffusion and the Inertial Behavior of Durables Consumption, *Journal of the European Economic Association* 13, 805-840.

Luo, Yulei, and Eric R. Young, 2010, Risk-sensitive Consumption and Savings under Rational Inattention, *American Economic Journal: Macroeconomics* 2, 281-325.

Maćkowiak, Bartosz, and Mirko Wiederholt, 2009, Optimal Sticky Prices under Rational Inattention, *American Economic Review* 99, 769–803.

Maćkowiak, Bartosz, and Mirko Wiederholt, 2015, Business Cycle Dynamics under Rational Inattention, *Review of Economic Studies* 82, 1502–1532.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, 2018, Dynamic Rational Inattention: Analytical Results, forthcoming in *Journal of Economic Theory.*

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, 2018, Survey: Rational Inattention, a Disciplined Behavioral Model, working paper, Sciences Po.

Matějka, Filip, and Alisdair McKay, 2015, Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model, *American Economic Review* 105, 272-298.

McGrattan, Ellen R., and Edward C. Prescott, 2010, Unmeasured Investment and the Puzzling US Boom in the 1990s, *American Economic Journal: Macroeconomics* 2, 88-123.

Miao, Jianjun, 2014, *Economic Dynamics in Discrete Time*, the MIT Press, Cambridge, MA.

Mondria, Jordi, 2010, Portfolio Choice, Attention Allocation, and Price Comovement, *Journal of Economic Theory* 145, 1837-1864.

Peng, Lin, 2005, Learning with Information Capacity Constraints, *Journal of Financial and Quantitative Analysis* 40, 307-329.

Peng, Lin, and Wei Xiong, 2006, Investor Attention, Overconfidence and Category Learning, *Journal of Financial Economics* 80, 563-602.

Saporta-Eksten, Itay, and Stephen J. Terry, 2018, Short-term Shocks and Long-term Investment, working paper, Boston University.

Sims, Christopher A., 1998, Stickiness, *Carnegie-Rochester Conference Series On Public Policy* 49, 317–356.

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

Sims, Christopher A., 2006, Rational Inattention: Beyond the Linear-Quadratic Case, *American Economic Review* 96, 158–163.

Sims, Christopher A., 2011, Rational Inattention and Monetary Economics, in *Handbook of Monetary Economics*, V. 3A, Benjamin M. Friedman and Michael Woodford (ed.), North-Holland.

Tanaka, T., K.K. Kim, P.A. Parrio, and S.K. Mitter, 2017, Semidefinite Programming Approach to Gaussian Sequential Rate-Distortion Trade-Offs, *IEEE Transactions on Automatic Control* 62, 1896-1910.

Toh, Kim-Chuan, Michael J. Todd, and R.H. Tutuncu, 1999, SDPT3 — A Matlab Software Package for Semidefinite Programming, *Optimization Methods and Software* 11, 545–581.

Tutuncu, R.H., Kim-Chuan Toh, and Michael J. Todd, 2003, Solving Semidefinite-quadratic-linear Programs Using SDPT3, *Mathematical Programming Series B*, 95, 189–217.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information Acquisition and Under-diversification, *Review of Economic Studies* 77, 571-626.

Vandenberghe, Lieven, Stephen Boyd, and Shao-Po Wu, 1998, Determinant Maximization with Linear Matrix Inequality Constraints, *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 2, pp. 499–533.

Woodford, Michael, 2003, Imperfect Common Knowledge and the Effects of Monetary Policy, In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, ed. Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford, 25–58. Princeton, NJ: Princeton University Press.

Woodford, Michael, 2009, Information-Constrained State-Dependent Pricing, *Journal of Monetary Economics* 56, S100-S124.

Zorn Peter, 2018, Investment under Rational Inattention: Evidence from US Sectoral Data, working paper, University of Munich.

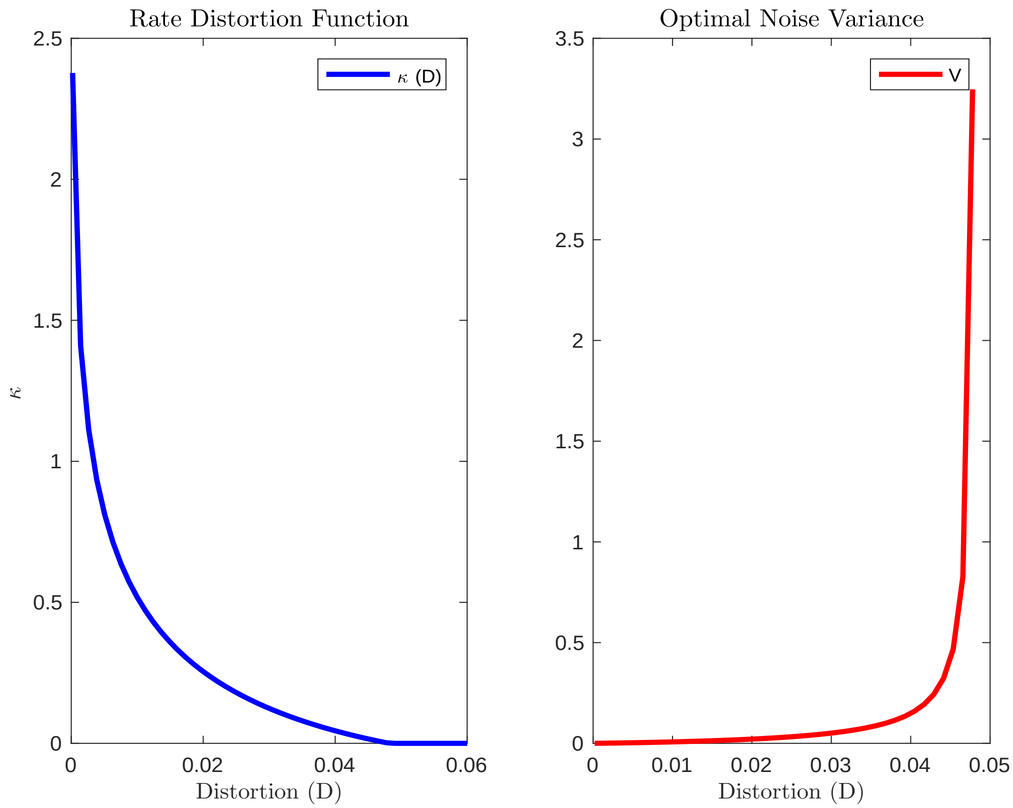

Figure 1: Rate distortion function and optimal noise variance for the price setting example.

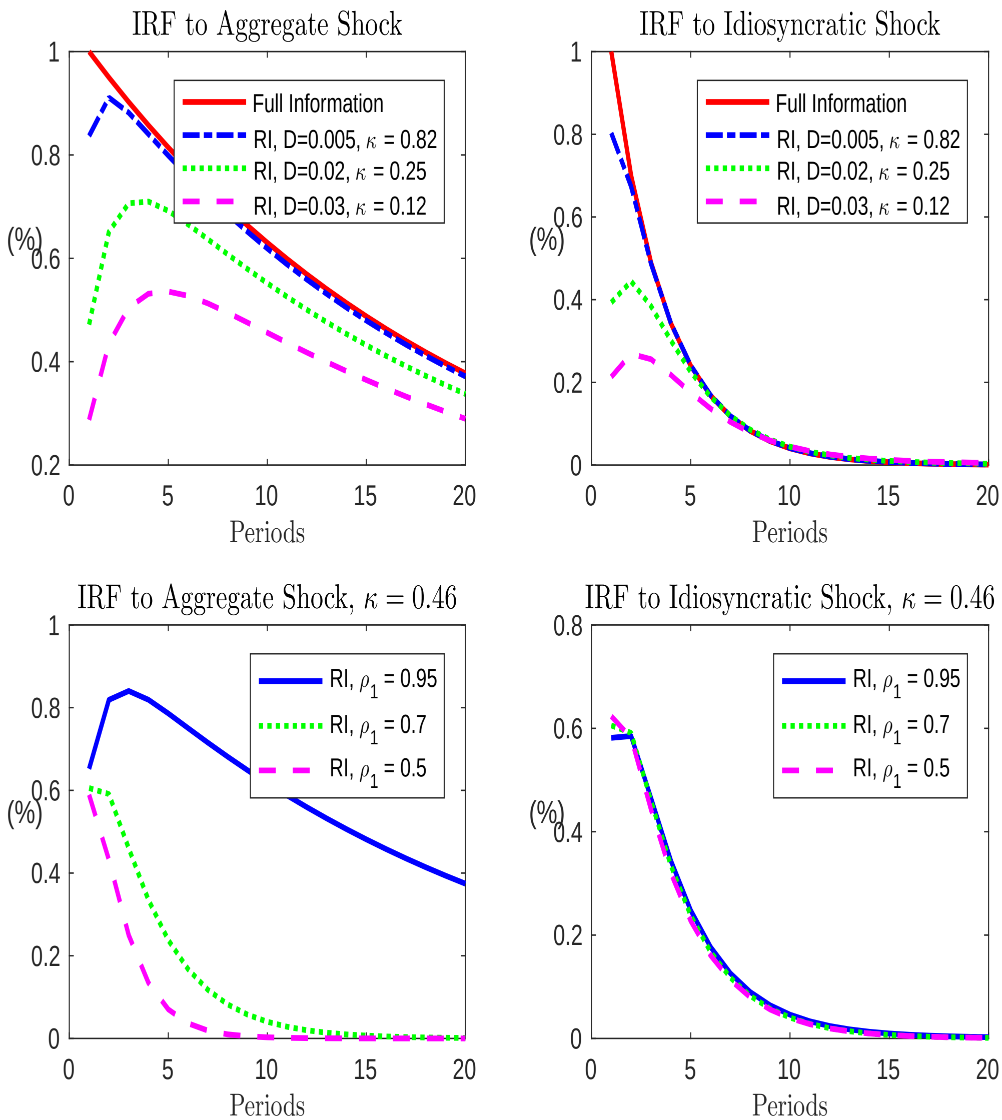

Figure 2: Impulse responses of prices to positive 1% shocks to aggregate and idiosyncratic states for various values of κ and ρ_1.

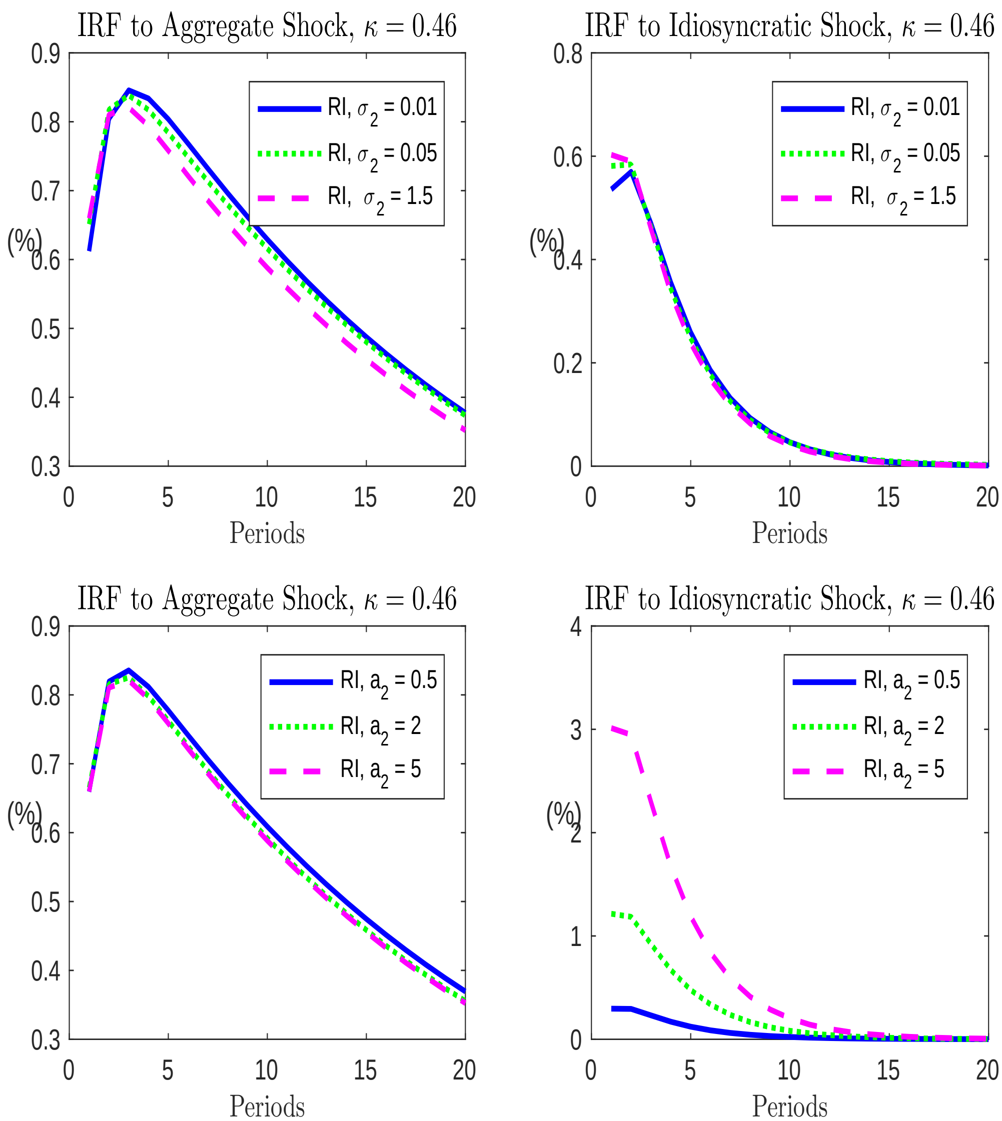

Figure 3: Impulse responses of prices to positive 1% shocks to aggregate and idiosyncratic states for various values of σ_2 and a_2.

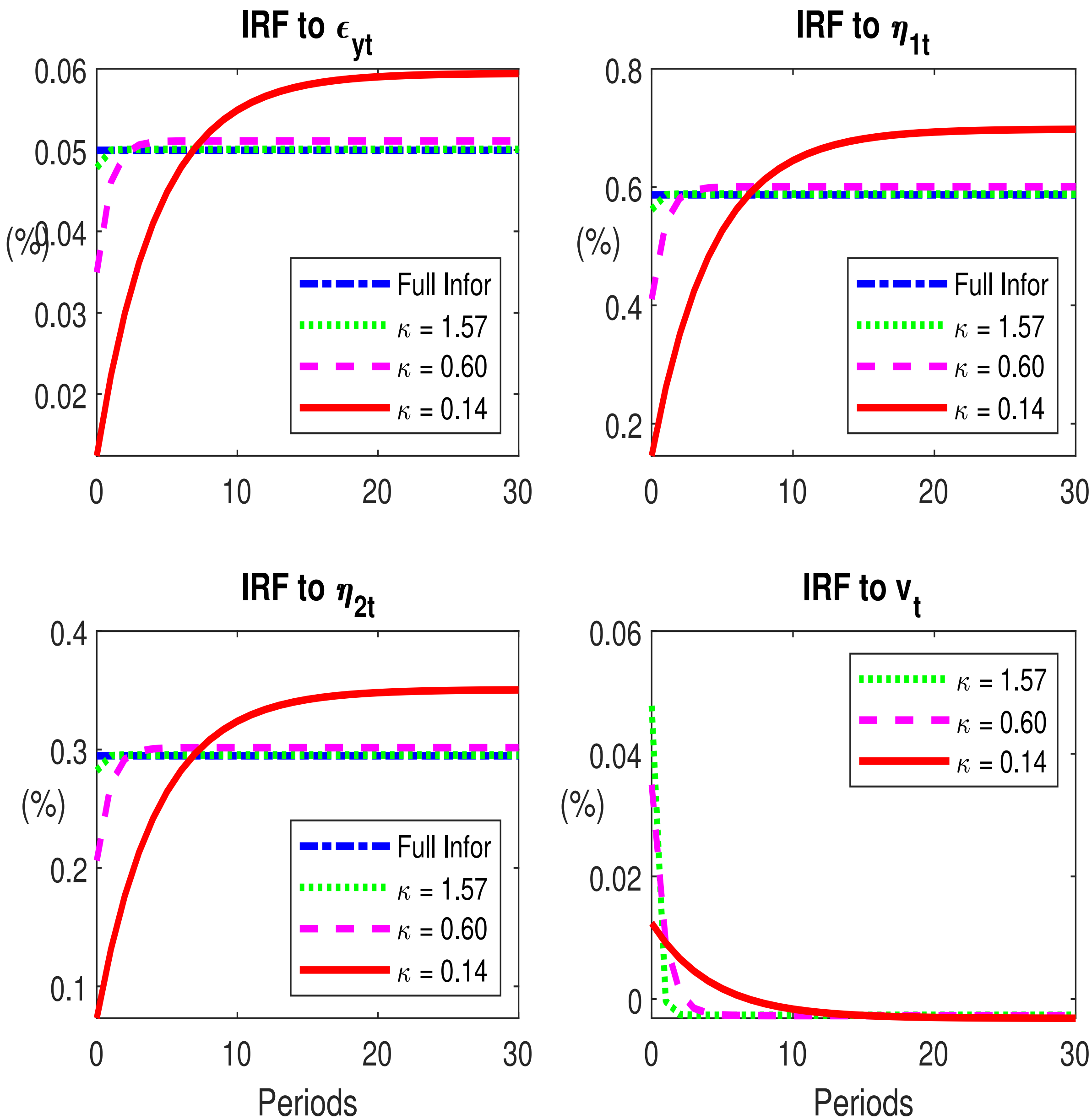

Figure 4: Impulse responses of consumption to positive 1% shocks to the transitory income component ($\epsilon_{y,t}$), two persistent income components ($\eta_{1,t}$ and $\eta_{2,t}$), and the signal noise (v_t).

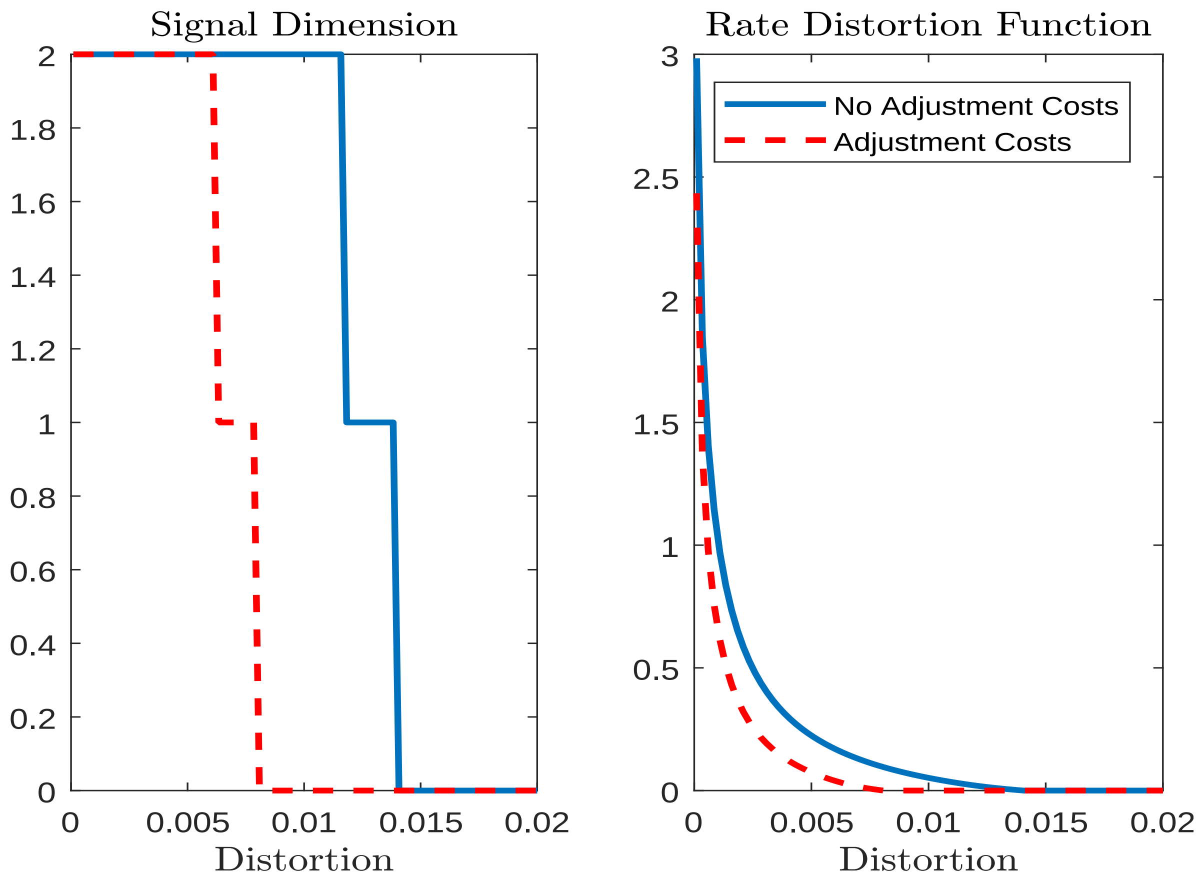

Figure 5: The signal dimension and rate distortion function for the investment example.

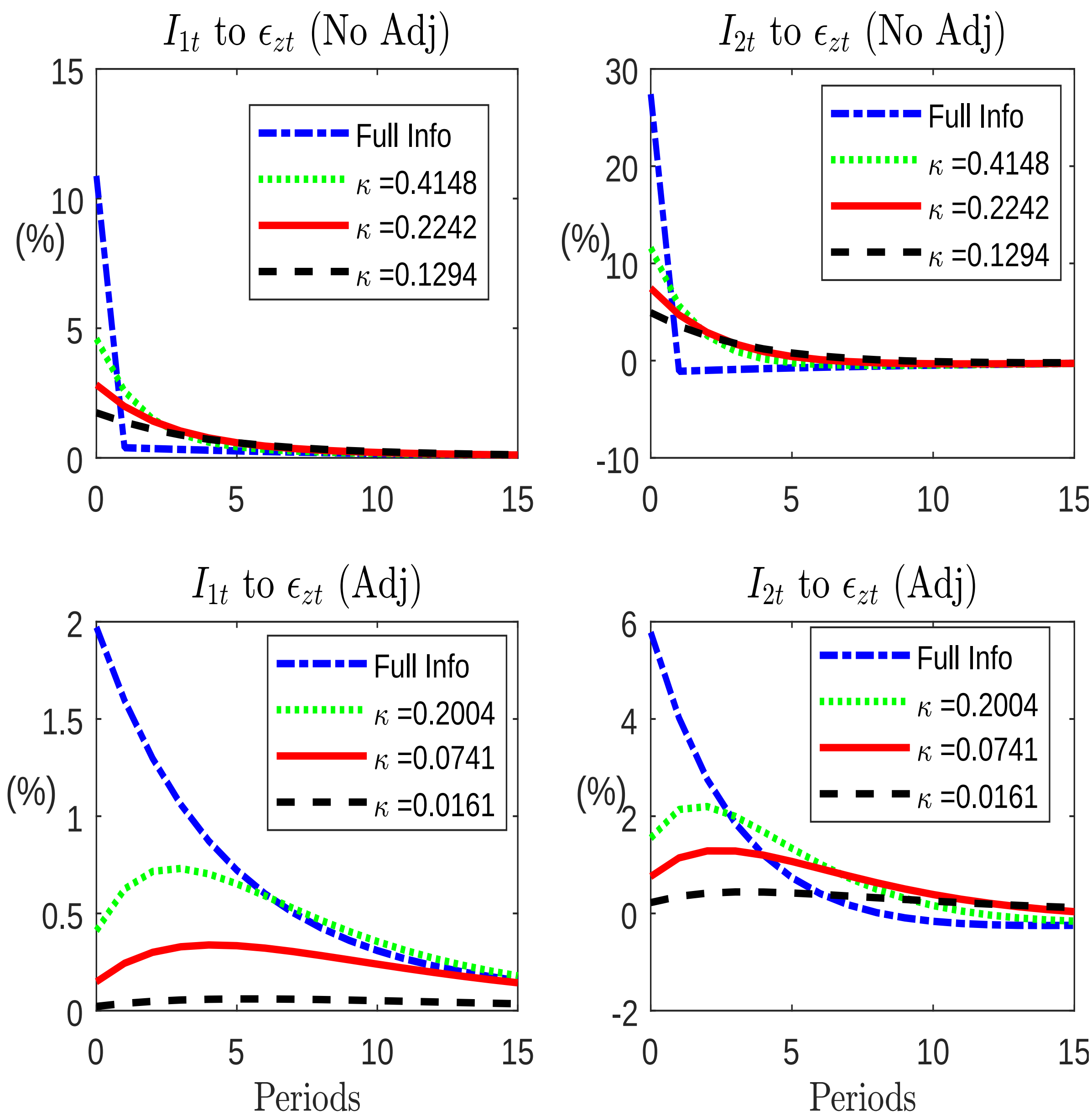

Figure 6: Impulse responses of tangible investment ($I_{1,t}$) and intangible investment ($I_{2,t}$) to a positive 1% persistent TFP shock for the case without adjustment costs (top two panels) and the case with adjustment costs (bottom two panels).