Multivariate Rational Inattention*

Jianjun Miao† Jieran Wu‡ Eric Young§

First version: December 2018
This version: February 2020

Abstract

We study optimal control problems in the multivariate linear-quadratic-Gaussian framework under rational inattention. We propose a general solution method to solve this problem using semidefinite programming and derive the optimal signal structure without strong prior restrictions. We analyze both the transition dynamics of the optimal posterior covariance matrix and its steady state. We characterize the optimal information structure for some special cases and develop numerical algorithms for general cases. Applying our methods to solve three multivariate economic models, we obtain some results qualitatively different from the literature.

Keywords: Rational Inattention, Endogenous Information Choice, Tracking Problem, Optimal Control, Entropy, Semidefinite Program

JEL Classifications: C61, D83, E21, E22, E31.

*We thank Lars Hansen, Yulei Luo, Lin Peng, Giacomo Rondina, Gianluca Violante, Pengfei Wang, Michael Woodford, Wei Xiong, He Yang, and Tao Zha for helpful comments. We also thank the participants at a seminar in Boston University and the AFR Summer Institute of Economics and Finance for comments. Please address all correspondences to Jianjun Miao.

†Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215. Email: 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 have been 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 the 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 period-by-period capacity constraints. Alternatively, the information choice incurs discounted (Shannon entropy) information costs measured in utility units.

Our second contribution is to develop an efficient three-step solution procedure. The first step is to derive the full information solution and the second step is to apply the certainty equivalence principle and the separation principle to derive the optimal control under an exogenous information structure. These two steps follow from the standard control literature. The third step is to solve for the optimal information structure under RI. We will focus on the formulation with discounted information costs and the analysis for the formulation with period-by-period capacity constraints is similar.

Like Sims (2011), we show that solving for the optimal information structure is equivalent to solving for the sequence of optimal posterior covariance matrices for the state vector. It seems natural to solve this sequence using dynamic programming. The difficulty is that this problem may not be convex and the choice variable must be a positive semidefinite matrix. Moreover, the RI problem involves no-forgetting constraints which are matrix inequality constraints. To tackle these

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

issues, we adopt the semidefinite programming (SDP) approach in the mathematics and engineering literature, which is the mathematical tool to study optimization over positive semidefinite matrices (Vandenberghe and Boyd (1996), Vandenberghe, Boyd, and Wu (1998), and Tanaka et al (2017)). We first transform the original dynamic programming problem into an auxiliary convex dynamic program and then derive a SDP representation. To facilitate an efficient and robust numerical implementation, we construct the representation as a disciplined convex program (DCP) (Grant (2004) and Grant, Boyd and Ye (2006)). A DCP must conform to the DCP ruleset so that it can be easily verified as convex and solvable in a computer. DCPs can be numerically solved using the powerful software CVX (Grant and Boyd (2008) and CVX Research, Inc. (2012)), which is freely available from the internet (http://cvxr.com/cvx/).[2]

The mathematics and engineering literature typically focuses on static SDP. We contribute to the literature by studying dynamic SDP. For the infinite-horizon case, such a dynamic program does not give a contraction mapping. Nevertheless, we use the method of value function iteration to show that the sequence of value functions for the truncated finite-horizon problems converges to the infinite-horizon value function. We can then derive the optimal sequence of posterior covariance matrices and the limiting steady state. As is well known, the method of value function iteration can be numerically slow especially for high dimensional problems. We thus borrow the idea in Sims (2003) to simplify the solution for the steady-state posterior covariance matrix.

We study a problem that minimizes the steady-state welfare loss including the information cost under RI. The solution is analogous to the golden-rule capital stock in the optimal growth model, and thus called the golden-rule information structure to differentiate it from the steady-state solution discussed earlier. For a pure tracking problem, the golden-rule solution is the same as that studied by Sims (2003). However, for a general control problem, we must take care of the initial state, which is drawn from an endogenous steady-state distribution. The literature has mistakenly ignored this initial value problem.[3]

We provide a characterization of the golden-rule information structure for the case in which 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. The first special case also gives the solution for the static RI problem studied by Fulton (2018) and Kőszegi and Matějka (2019). Our solution generalizes the static reverse water-filling solution studied in Theorem 10.3.3 of Cover and Thomas (2006, p. 314). We characterize the optimal signal dimension and show that it weakly decreases with the information cost parameter. For tracking problems, we prove that the optimal signal is one dimensional if the rank of the weighting matrix in the loss

[2]CVX supports two free SDP solvers, SeDuMi (Sturm (1999)) and SDPT3 (Toh, Todd, and Tutuncu (1999) and Tutuncu, Toh, and Todd (2003)). These solvers use the primal-dual interior-point method.

[3]We are extremely grateful to Chris Sims for pointing out this issue to us.

function is equal to one. The optimal signal is equal to the target under full information plus a noise.

Our third contribution is to apply our results to three economic problems. Our first application is the price setting problem adapted from Maćkowiak and Wiederholt (2009), in which there are two exogenous state variables representing two sources of uncertainty. We first ignore the general equilibrium price feedback effect and just focus on the decision problem as in Sims (2011). The profit-maximizing price is equal to a linear combination of the two shocks. We derive numerical solutions under RI similar to those in Sims (2011). We then study a general equilibrium problem in which the endogenous aggregate price level affects individual firms' profit-maximizing prices.

We derive an efficient numerical algorithm to solve this problem. Because solving for an equilibrium needs to repeatedly solve an individual firm's RI problem, we focus on the golden-rule information structure to save computation time. We approximate the equilibrium price by an ARMA process and derive a state space representation for a firm's tracking problem under RI. We find that the optimal signal dimension is less than the state dimension. This result violates the signal independence assumption in Maćkowiak and Wiederholt (2009), which assumes 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. In particular, given our optimal signal structure, a firm is confused about the sources of shocks and hence there is a volatility spillover effect: An increase in the volatility of one source of the shock causes the firm to raise price responses to both sources of shocks.

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 one dimensional for many information cost parameter values. 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 different from 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 when the information cost parameter is sufficiently large. Sims (1998, 2003) argues that RI can substitute for adjustment costs in a dynamic optimization problem. Our numerical results show that RI 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 introduces multivariate LQG RI models with information-flow constraints and focuses on the golden-rule solution only.[4] Sims (2011) studies the formulation with discounted information costs and formulates the LQG RI problem without explicit reference to the signal structure. His solution procedure consists of two steps. His first step is essentially the same as our first two steps. His second step is to transform the control problem under RI into a problem of choosing a sequence of optimal posterior covariance matrices for the state vector. Sims (2011) proposes to solve for the steady state as the limit point of the optimal sequence. He outlines a method based on first-order conditions when the no-forgetting constraints do not bind and recommends to use the Cholesky decomposition when they bind without providing a detailed analysis. Sims (2003) suggests that 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 nonuniqueness of the optimal signal and 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.

While Sims's methods are insightful, numerically solving first-order conditions for optimization problems with matrix inequality constraints is nontrivial. Our approach using the software CVX can handle such problems efficiently and robustly. Moreover, 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 constraints often bind and the signal dimension does not exceed the minimum of the control and state dimensions.

Because of the difficulty of solving 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. Under this assumption, solving for the optimal information structure is equivalent to solving for the noise covariance (or precision) matrix for the signal. An undesirable implication 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.

Under our formulation, both the linear transformation and the noise covariance matrix in the

[4] Luo (2008), Luo and Young (2010), and Luo, Nie and Young (2015) follow Sims's approach closely, but mainly focus on univariate models.

signal form must be endogenously chosen. In addition to the attention allocation effect emphasized in the literature, the learning effect induced by the linear transformation of states is also important for decision making because the linear transformation determines how the decision maker collects different sources of information by combining different states. Linear combination of states can cause the decision maker to be confused about different sources of uncertainty, thereby generating a spillover effect.

In independent work Fulton (2018) and Kőszegi and Matějka (2019) analyze 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).[5] Maćkowiak, Matějka, and Wiederholt (2018) study a dynamic tracking problem with one control and one exogenous state, which follows an ARMA 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.

Our paper is also related to other studies that are not in the discrete-time LQG framework. This literature is growing. Recent papers include Steiner, Stewart, and Matějka (2017), Dewan (2018), Hébert and Woodford (2018), and Zhong (2019). Miao (2019) studies continuous-time LQG RI problems, which require different mathematical tools. He focuses on the golden-rule steady state, but he does not study transitional dynamics and many economic examples in this paper.

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 Gaussian white noise with covariance matrix W_t. The matrix W_t is positive semidefinite, denoted by $W_t \succeq 0$.[6] 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-

[5] 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.

[6] 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).

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 Gaussian white noise with covariance matrix $V_t \succ 0$. Notice that we do not impose any other restriction on C_t or V_t.[7] In particular, C_t may not be an identity matrix or invertible. Assume that x_0 is a Gaussian random variable with mean $\bar{x}_0$ and covariance matrix $\Sigma_{-1} \succ 0$. 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 or she faces the following period-by-period capacity constraint[8]

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

where $\kappa > 0$ denotes the information-flow rate or 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.[9] 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 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 in each period is limited by an upper bound κ.

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. The decision maker makes decisions sequentially. He or she first chooses the information structure $\{C_t, V_t\}_{t=0}^T$ and then selects a control $\{u_t\}_{t=0}^T$ adapted to $\{s^t\}$ 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 with period-by-period capacity constraints)*

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

subject to (1), (2), and (3), where $\beta \in (0, 1]$ and the expectation is taken with respect to the joint distribution induced by the initial distribution for x_0 and the state dynamics (1).

[7]As will be clear later, the signal form in (2) is not restrictive and can be recovered from the optimal posterior covariance matrix for the state vector (see Proposition 1).

[8]We do not adopt the capacity constraint on the total information flows across periods because this formulation causes the dynamic inconsistency issue.

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

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. In applications it may be more convenient to consider the following relaxed problems with discounted information costs.

Problem 2 *(Finite-horizon LQG problem under RI with discounted information costs)*

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

subject to (1) and (2), where $\beta \in (0,1]$ and the expectation is taken with respect to the joint distribution induced by the initial distribution for x_0 and the state dynamics (1).

In this problem $\lambda > 0$ can be interpreted as the shadow price (cost) of the information flow. 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, S_t, and W_t. Under some stability conditions, the posterior distribution for x_t will converge to a long-run stationary distribution.

For simplicity we will focus our analysis on Problem 2 and its infinite-horizon limit as $T \to \infty$. We will discuss how we solve Problem 1 in Appendix B.

2.1 Full Information Case

Before analyzing Problem 2, 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 given a state x_t takes the form

$$v_t^{FI}(x_t) = -x_t'P_tx_t - \sum_{\tau=t}^{T}\beta^{\tau-t+1}\mathrm{tr}\left(W_\tau P_{\tau+1}\right), \tag{4}$$

where $P_t \succeq 0$ and satisfies

$$\begin{aligned} P_t \quad = \quad & Q_t + \beta A_t'P_{t+1}A_t \\ & - \left(\beta A_t'P_{t+1}B_t + S_t\right)\left(R_t + \beta B_t'P_{t+1}B_t\right)^{-1}\left(\beta B_t'P_{t+1}A_t + S_t'\right), \end{aligned} \tag{5}$$

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

The optimal control is

$$u_t = -F_tx_t, \tag{6}$$

where

$$F_t = \left(R_t + \beta B_t' P_{t+1} B_t\right)^{-1} \left(S_t' + \beta B_t' P_{t+1} A_t\right). \tag{7}$$

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

$$v^{FI}(x_t) = -x_t' P x_t - \frac{\beta}{1-\beta}\operatorname{tr}(WP),$$

where $P \succeq 0$ and satisfies

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

The optimal control is given by

$$u_t = -F x_t, \tag{8}$$

where

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

2.2 Control under Exogenous Information Structure

We solve Problem 2 in three steps. In the first step we derive the full-information solution as in Section 2.1. In the second step we observe that Problem 2 is a standard LQG problem under partial information when the information structure $\{C_t, V_t\}_{t=0}^T$ is exogenously fixed. Thus the usual separation principle and certainty equivalence principle hold. This implies that the optimal control is given by

$$u_t = -F_t \widehat{x}_t, \tag{9}$$

where $\widehat{x}_t \equiv \mathbb{E}\left[x_t | s^t\right]$ denotes the estimate of x_t given information s^t. Notice that the matrix F_t is determined by (7) in the full information case, which is independent of the information structure.

The state under the optimal control satisfies the dynamics

$$x_{t+1} = A_t x_t - B_t F_t \widehat{x}_t + \epsilon_{t+1}. \tag{10}$$

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

$$\widehat{x}_t = \widehat{x}_{t|t-1} + \Sigma_{t|t-1} C_t' \left(C_t \Sigma_{t|t-1} C_t' + V_t\right)^{-1}\left(s_t - C_t \widehat{x}_{t|t-1}\right), \tag{11}$$

$$\widehat{x}_{t|t-1} = \left(A_{t-1} - B_{t-1} F_{t-1}\right)\widehat{x}_{t-1}, \tag{12}$$

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 $\Sigma_{0|-1} = \Sigma_{-1}$ exogenously given. Moreover,

$$\Sigma_{t+1|t} = A_t \Sigma_t A_t' + W_t, \tag{13}$$

$$\Sigma_t = \left(\Sigma_{t|t-1}^{-1} + \Phi_t\right)^{-1}, \tag{14}$$

for $t=0,1,...,T,$ where $\Sigma_t \equiv \mathbb{E}\left[(x_t-\widehat{x}_t)(x_t-\widehat{x}_t)'|s^t\right]$ denotes the posterior covariance matrix given s^t and Φ_t denotes the signal-to-noise ratio (SNR) defined by $\Phi_t = C_t' V_t^{-1} C_t \succeq 0$, $t=0,1,...,T$.

We need the following lemma to derive the optimal information structure. Its proof and proofs of all other results are collected in Appendix A.

Lemma 1 *Under the optimal control policy in (9) for fixed information structure* $\{C_t, V_t\}_{t=0}^{T}$, *we have*

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=0}^{T}\beta^t\left(x_t'Q_tx_t+u_t'R_tu_t+2x_t'S_tu_t\right)+\beta^{T+1}x_{T+1}'P_{T+1}x_{T+1}\right]\\ =\;&\mathbb{E}\left[x_0'P_0x_0\right]+\sum_{t=0}^{T}\beta^{t+1}\mathrm{tr}\left(W_tP_{t+1}\right)+\sum_{t=0}^{T}\beta^t\mathrm{tr}\left(\Omega_t\Sigma_t\right),\end{aligned}$$

where

$$\Omega_t = F_t'(R_t+\beta B_t'P_{t+1}B_t)F_t \succeq 0. \tag{15}$$

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, the rank of Ω_t, denoted by $\mathrm{rank}(\Omega_t)$, 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 $\mathrm{rank}(\Omega_t)=n_u$. If $n_x<n_u$ and F_t has full row rank, then $\mathrm{rank}(\Omega_t)=n_x$.

2.3 Optimal Information Structure

In the final step of our solution procedure, we solve for the optimal information structure $\{C_t, V_t\}$. In doing so, we compute the mutual information[10]

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

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

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

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

Since $\{P_t\}$ is independent of the information structure and $\mathbb{E}\left[x_0'P_0x_0\right]$ is determined by the exogenous initial prior distribution, it follows from Lemma 1 that solving for the optimal information structure in Problem 2 is equivalent to solving for the optimal sequence of posterior covariance matrices for the state vector:

[10] 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."

Problem 3 *(Optimal information structure for Problem 2)*

$$\min_{\{\Sigma_t\}_{t=0}^T} \sum_{t=0}^{T} \beta^t \left[\operatorname{tr}(\Omega_t \Sigma_t) + \lambda I\left(x_t; s_t | s^{t-1}\right) \right]$$

subject to

$$I\left(x_t; s_t|s^{t-1}\right) = \frac{1}{2}\log\det\left(A_{t-1}\Sigma_{t-1}A_{t-1}' + W_{t-1}\right) - \frac{1}{2}\log\det(\Sigma_t),$$

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

$$\Sigma_t \preceq A_{t-1}\Sigma_{t-1}A_{t-1}' + W_{t-1}, \tag{16}$$

$$\Sigma_0 \preceq \Sigma_{-1}, \tag{17}$$

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

It follows from Lemma 1 and (4) that the expression $\sum_{t=0}^{T} \beta^t \operatorname{tr}(\Omega_t \Sigma_t)$ represents the expected welfare loss due to the limited information (i.e., the difference between the expected discounted utilities under full information and under limited information). The optimal information structure under RI minimizes the welfare loss plus the discounted information cost. Sims (2011) formulates an essentially identical problem for the infinite-horizon case as $T \to \infty$, except that there is a difference in constraints at date zero. The matrix inequalities (16) and (17) are called the no-forgetting constraints (Sims (2003, 2011)). They can be derived from (13) and (14) as the SNR Φ_t is positive semidefinite. After obtaining $\{\Sigma_t\}$, we can recover $\{\Phi_t\}$ and hence $\{C_t\}$ and $\{V_t\}$ from the following result.

Proposition 1 *Given an optimal sequence* $\{\Sigma_t\}_{t=0}^T$ *determined from Problem 3, the optimal SNR is given by*

$$\Phi_0 = \Sigma_0^{-1} - \Sigma_{-1}^{-1}, \;\; \Phi_t = \Sigma_t^{-1} - \left(A_{t-1}\Sigma_{t-1}A_{t-1}' + W_{t-1}\right)^{-1}, \;\; t \geq 1.$$

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. Throughout the paper we will focus on the signal structure such that V_t is diagonal for each t. In this case C_t is unique up to a scalar constant and up to an interchange of rows. When C_t is scaled by a constant b, V_t is scaled by b^2. By the Kalman filter, the impulse responses to structural shocks to all state variables do not change, but the responses to noise shocks are scaled by $1/b$. Notice that optimal signals are in general not independent in the sense that the matrix C_t may not be diagonal or invertible. The signal independence assumption is widely adopted in the literature (e.g., Maćkowiak and Wiederholt (2009)), but we show that this assumption can be restrictive and lead to suboptimal solutions.

3 Dynamic Semidefinite Programming

In this section we focus on the analysis of Problem 3, which is not a trivial dynamic problem because the choice variables are positive semidefinite matrices and the constraints are matrix inequalities. We extend the semidefinite programming approach recently proposed by Tanaka et al (2017) for static programs to the dynamic case. We also provide some characterization results for some special cases.

3.1 Finite-Horizon Case

We use dynamic programming to study Problem 3 (Stokey and Lucas with Prescott (1989) and Miao (2014)). Let $\mathcal{V}_0(\Sigma_{-1})$ be the value function for Problem 3. Let $\mathcal{V}_t(\Sigma_{t-1})$ be the value function for the continuation problem in period $t \geq 1$ defined as

$$\mathcal{V}_t(\Sigma_{t-1}) = \min_{\{\Sigma_\tau\}_{\tau=t}^T} \sum_{\tau=t}^T \beta^{\tau-t} \left[\operatorname{tr}(\Omega_\tau \Sigma_\tau) + \lambda I\left(x_\tau; s_\tau | s^{\tau-1}\right)\right]$$

subject to

$$I\left(x_\tau; s_\tau | s^{\tau-1}\right) = \frac{1}{2}\log\det\left(A_{\tau-1}\Sigma_{\tau-1}A_{\tau-1}' + W_{\tau-1}\right) - \frac{1}{2}\log\det(\Sigma_\tau),$$

$$\Sigma_\tau \preceq A_{\tau-1}\Sigma_{\tau-1}A_{\tau-1}' + W_{\tau-1},$$

for $\tau = t, t+1, ..., T$.

The sequence of value functions $\mathcal{V}_t(\Sigma_{t-1})$ for $t \geq 0$ satisfies Bellman equations. But $\mathcal{V}_t(\Sigma_{t-1})$ may not be convex, as will become clear later. We thus solve an auxiliary convex problem. We formulate this problem as DCP so that we can apply the efficient software CVX. Specifically, in the last period T, consider

$$J_T(\Sigma_{T-1}) \equiv \min_{\Sigma_T \succ 0} \operatorname{tr}(\Omega_T \Sigma_T) - \frac{\lambda}{2}\log\det(\Sigma_T) \tag{18}$$

subject to (16) for $t = T$. Since the log-determinant function is strictly concave and (16) is a linear matrix inequality, the problem in (18) is a convex program and hence $J_T(\Sigma_{T-1})$ is also strictly convex in Σ_{T-1}.

In any period $t = 0, 1, ..., T-1$, consider the Bellman equation:

$$J_t(\Sigma_{t-1}) = \min_{\Sigma_t \succ 0} \operatorname{tr}(\Omega_t \Sigma_t) + \frac{\lambda}{2}\left[\beta \log\det\left(A_t \Sigma_t A_t' + W_t\right) - \log\det(\Sigma_t)\right] + \beta J_{t+1}(\Sigma_t) \tag{19}$$

subject to (16) for $t \geq 1$ and (17) for $t = 0$.

It is straightforward to verify that

$$\mathcal{V}_t\left(\Sigma_{t-1}\right) = J_t\left(\Sigma_{t-1}\right) + \frac{\lambda}{2} \log\det\left(A_{t-1}\Sigma_{t-1}A_{t-1}' + W_{t-1}\right) \tag{20}$$

for $t \geq 1$ and

$$\mathcal{V}_0\left(\Sigma_{-1}\right) = J_0\left(\Sigma_{-1}\right) + \frac{\lambda}{2} \log\det\left(\Sigma_{-1}\right). \tag{21}$$

Moreover, the optimal solution $\{\Sigma_t\}_{t=0}^{T}$ for (18) and (19) also gives the optimal solution to Problem 3 by the dynamic programming principle.

In Lemma 2 of Appendix A we show that

$$\beta \log\det\left(A_t\Sigma_tA_t' + W_t\right) - \log\det\left(\Sigma_t\right)$$

is strictly convex in Σ_t when $\beta \in (0, 1]$. Thus we can show that the problem in (19) is a convex program. However, the software CVX cannot recognize whether the difference of two concave functions is convex by its ruleset. We need to transform this problem into a DCP form. To achieve this goal, the following proposition derives a dynamic semidefinite program representation.

Proposition 2 *Suppose that $W_t \succ 0$ and $\Omega_t \succeq 0$ for $t = 0, 1, ..., T$. Then the value function $J_t\left(\Sigma_{t-1}\right)$ is strictly convex in Σ_{t-1} and satisfies the dynamic semidefinite program for $t = 0, 1, ..., T-1$:*

$$\begin{aligned} J_t\left(\Sigma_{t-1}\right) &= \min_{\Pi_t \succ 0, \Sigma_t \succ 0} \operatorname{tr}\left(\Omega_t\Sigma_t\right) - \frac{\lambda}{2}(1-\beta)\log\det\left(\Sigma_t\right) \\ &\quad + \frac{\lambda\beta}{2}\left(\log\det W_t - \log\det \Pi_t\right) + \beta J_{t+1}\left(\Sigma_t\right) \end{aligned} \tag{22}$$

subject to (16) and

$$\begin{bmatrix} \Sigma_t - \Pi_t & \Sigma_t A_t' \\ A_t\Sigma_t & W_t + A_t\Sigma_tA_t' \end{bmatrix} \succeq 0,$$

where $J_T\left(\Sigma_{T-1}\right)$ satisfies (18) and is also strictly convex. For $t = 0$, (16) is replaced by (17).

Since $J_t\left(\Sigma_{t-1}\right)$ is strictly convex for $t = 0, 1, ..., T$ and since the log-determinant function is strictly concave, the objective function in (22) as the sum of four convex functions is convex in Σ_t and Π_t. Since the constraints are linear matrix inequalities, the dynamic programming problem in Proposition 2 is a DCP. We can then apply the software CVX to derive numerical solutions efficiently. Notice that $\mathcal{V}_t\left(\Sigma_{t-1}\right)$ also satisfies a dynamic programming equation. But we do not solve it directly because $\mathcal{V}_t\left(\Sigma_{t-1}\right)$ may not be convex as it is equal to the sum of a convex function $J_t\left(\Sigma_{t-1}\right)$ and a concave function by (20).

The assumption of $W_t \succ 0$ ensures that $\log\det W_t$ is well defined. This assumption can be restrictive in economic applications. It 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 can eliminate this constraint by substituting out I_t. Another way is to introduce a depreciation or capital quality shock often used in the literature. See Section 5.3 for the details. Alternatively, we allow $W_t \succeq 0$ and present a result similar to Proposition 2 in Appendix C. 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 5.

After obtaining the solutions for $\{F_t, \Sigma_t, C_t, V_t\}$, we use the system of equations (1), (2), (6), (11), and (12) to generate impulse responses and simulations of the model.

3.2 Infinite-horizon Case

In the infinite-horizon case, all exogenous matrices A_t, B_t, Q_t, R_t, S_t, and W_t are time invariant. We can derive the solution for the infinite-horizon case by taking the limit of the finite-horizon solution as $T \to \infty$. For numerical implementation, we can apply the method of value function iteration. We present a formal analysis in Appendix D, where Proposition 9 establishes a convergence result. Here we sketch the key idea.

Under some stability conditions in the standard control theory, P_t and F_t converge to P and F given in Section 2.1 as $T \to \infty$. By (15), Ω_t converges to

$$\Omega \equiv F'(R + \beta B'PB)F \succeq 0. \tag{23}$$

Moreover, the value functions $J_t(\Sigma_{t-1})$ and $\mathcal{V}_t(\Sigma_{t-1})$ also converge to some time-invariant functions $J(\Sigma_{t-1})$ and $\mathcal{V}(\Sigma_{t-1})$ for any fixed $t \geq 1$ as $T \to \infty$. Let the optimal policy function for problem (19) be $\Sigma_t = h_t(\Sigma_{t-1})$ for a finite T. As $T \to \infty$, h_t converges to a time-invariant function h for any fixed $t \geq 1$. Since the initial no-forgetting constraint (17) is different from (16) for $t \geq 1$, the initial policy function h_0 is different from h.

In the infinite-horizon case, as $t \to \infty$, $\Sigma_t = h(\Sigma_{t-1})$ may converge to a steady state Σ. We can then recover the steady-state SNR Φ using the no-forgetting constraint

$$\Phi = \Sigma^{-1} - \left(A\Sigma A' + W\right)^{-1} \succeq 0, \tag{24}$$

and recover the steady-state information structure (C, V) using $\Phi = C'V^{-1}C$. The signal s_t takes the form $s_t = Cx_t + v_t$, where v_t is a Gaussian white noise with covariance matrix V.

3.3 Golden Rule of Information Structure

The procedure of solving the dynamics of Σ_t and the limiting steady state in the previous subsection is complicated. To simplify the steady-state solution, suppose that x_0 is drawn from the prior Gaussian distribution with covariance matrix $\Sigma_{0|-1} = \Sigma_{-1} = A\Sigma A' + W$, where Σ is the endogenous steady-state posterior covariance matrix. Then we have $\Sigma_t = \Sigma$ for all $t \geq 0$ by the Kalman filter. In this subsection we present a simple method to solve for Σ.

In the steady state, it follows from Lemma 1 that, under the optimal control policy,

$$\begin{aligned} &\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \left(x_t' Q x_t + u_t' R u_t + 2 x_t' S u_t\right)\right] \\ =\ & \bar{x}_0' P \bar{x}_0 + \operatorname{tr}\left(A' P A \Sigma\right) + \frac{1}{1-\beta} \operatorname{tr}(WP) + \frac{1}{1-\beta} \operatorname{tr}(\Omega\Sigma), \end{aligned}$$

where we have used the fact that

$$\mathbb{E}\left[x_0' P x_0\right] = \bar{x}_0' P \bar{x}_0 + \operatorname{tr}\left(P \Sigma_{-1}\right) = \bar{x}_0' P \bar{x}_0 + \operatorname{tr}\left(P\left(A\Sigma A' + W\right)\right). \tag{25}$$

The steady-state mutual information is given by

$$I\left(x_t; s_t | s^{t-1}\right) = \frac{1}{2} \log\det\left(A\Sigma A' + W\right) - \frac{1}{2} \log\det(\Sigma) \ \text{ for } t \geq 0.$$

The steady-state no-forgetting constraint becomes

$$\Sigma \preceq A\Sigma A' + W. \tag{26}$$

Now we consider the following static problem that determines Σ.

Problem 4 *(Golden rule of information structure for Problem 3)*

$$\min_{\Sigma \succ 0} \quad (1-\beta) \operatorname{tr}\left(A' P A \Sigma\right) + \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma)\right] \tag{27}$$

subject to (26).

To understand this problem, we draw an analogy to the optimal growth model with the resource constraint $C_t + K_{t+1} = f(K_t)$. The steady-state optimal capital stock satisfies the first-order condition $\beta f'(K) = 1$. But the golden rule of capital that maximizes the steady-state discounted utility (or consumption) satisfies $f'(K) = 1$. These two levels of the capital stock are generally different. Similarly, the optimal information structure for Problem 4 maximizes the steady-state discounted utility minus the discounted information cost multiplied by $1 - \beta$,[11] or equivalently, minimizes the steady-state welfare loss including the information cost. We call this solution the

[11] Notice that we have ignored the term $\bar{x}_0' P \bar{x}_0$ because they are independent of Σ.

golden rule of information structure to differentiate it from the limiting steady state studied in Section 3.2.

Sims (2003) simplifies the formulation of the steady-state problem with information-flow constraints and much of the literature follows his formulation. In particular, he studies the following static problem:

$$\min_{\Sigma \succ 0} \operatorname{tr}(\Omega\Sigma)$$

subject to (26) and

$$\log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma) \leq 2\kappa.$$

The steady-state RI problem with discounted information costs can be similarly formulated by removing the first term in (27).

The pitfall of this formulation is that maximizing the steady-state expected utility under limited information is not equivalent to minimizing the steady-state expected welfare loss for general control problems.[12] This is because the expected welfare loss nets out the initial value $\mathbb{E}\left[x_0' P x_0\right]$, but this value in the steady state is endogenous as the prior distribution for x_0 is drawn from the Gaussian distribution with endogenous covariance matrix $A\Sigma A' + W$ (see (25)). This pitfall does not arise in Problem 3 because in that problem x_0 is drawn from an exogenously given prior distribution. It also does not arise in tracking problems studied in Section 4.

As in Proposition 2, we transform Problem 4 into a DCP using a semidefinite program representation.

Proposition 3 *Suppose that $W \succ 0$ and $\Omega \succeq 0$. Then the golden-rule solution Σ to Problem 4 is the solution to the following semidefinite program:*

$$\min_{\Pi \succ 0, \Sigma \succ 0} \; (1-\beta)\operatorname{tr}\left(A'PA\Sigma\right) + \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\log\det W - \log\det \Pi\right] \tag{28}$$

subject to (26) and

$$\begin{bmatrix} \Sigma - \Pi & \Sigma A' \\ A\Sigma & A\Sigma A' + W \end{bmatrix} \succeq 0. \tag{29}$$

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 golden-rule RI problem admits a generalized reverse water-filling solution described below.

We first introduce some notations. Let

$$\bar{\Omega} = \Omega + (1-\beta)\, A'PA.$$

[12] We are extremely grateful to Chris Sims for raising this issue to us.

Let $W^{\frac{1}{2}} \succ 0$ denote the positive definite square root of W. Then the positive semidefinite matrix $W^{\frac{1}{2}}\bar{\Omega}W^{\frac{1}{2}}$ admits an eigendecomposition $W^{\frac{1}{2}}\bar{\Omega}W^{\frac{1}{2}} = U\Omega_d U'$, 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 $W^{\frac{1}{2}}\bar{\Omega}W^{\frac{1}{2}}$.

Proposition 4 *Suppose that $\Omega \succeq 0$, $W \succ 0$, and $A = \rho I$ with $|\rho| < 1$ in Problem 4. Then the golden-rule posterior covariance matrix for x_t is given by*

$$\Sigma = W^{\frac{1}{2}}U\widehat{\Sigma}U'W^{\frac{1}{2}}, \tag{30}$$

where $\widehat{\Sigma} \equiv diag\left(\widehat{\Sigma}_i\right)_{i=1}^{n_x}$ with

$$\widehat{\Sigma}_i = \min\left(\frac{1}{1-\rho^2}, \widehat{\Sigma}_i^*\right), \quad \widehat{\Sigma}_i^* = \frac{1}{2\rho^2}\left(\sqrt{1+\frac{2\rho^2\lambda}{d_i}} - 1\right). \tag{31}$$

To understand this proposition, we consider two special cases. First, in the IID case with $\rho = 0$ or $A = 0$, the solution reduces to that for the static problem with $T = 0$. The static case corresponds to $\Omega = \Omega_0$ and $W = \Sigma_{-1}$, and the optimal posterior covariance matrix is given by (30) with $\widehat{\Sigma}_i = \min\left(1, \lambda/\left(2d_i\right)\right)$. This static solution generalizes the standard reverse water-filling solution analyzed by Cover and Thomas (2006) for the problem in which Ω is an identity matrix and W is diagonal.[13] For that problem, we have $\bar{\Omega} = \Omega = I$, $W = \text{diag}\left(w_i^2\right)_{i=1}^{n_x}$. Then the optimal posterior covariance matrix is diagonal with the ith diagonal entry given by $\Sigma_i = \min\left(w_i^2, \lambda/2\right)$. This means that any prior variance higher than $\lambda/2$ is reduced to $\lambda/2$ ex post. The decision maker does not pay attention to the components of prior variances lower than $\lambda/2$ so that the corresponding posterior variances remain the same.

Second, consider the dynamic univariate case with $n_x = 1$, $\bar{\Omega} = 1$, and $W = w^2$. Then

$$\Sigma = w^2 \min\left(\frac{1}{1-\rho^2}, \frac{1}{2\rho^2}\left(\sqrt{1+\frac{2\rho^2\lambda}{w^2}} - 1\right)\right).$$

If $0 < \lambda < 2w^2/\left(1-\rho^2\right)^2$, then the posterior variance Σ is reduced from the stationary prior variance $w^2/\left(1-\rho^2\right)$ to a smaller variance. But if $\lambda \geq 2w^2/\left(1-\rho^2\right)^2$, no information is collected and $\Sigma = w^2/\left(1-\rho^2\right)$.

Proposition 4 generalizes the preceding two special cases. In particular, d_i is the ith eigenvalue of the weighted innovation 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 innovation variances. High eigenvalues d_i are scaled down by the factor $\widehat{\Sigma}_i$ for sufficiently small information costs.

[13] See Fulton (2018) and Kőszegi and Matějka (2019) for similar results.

What kind of signal structure can generate the optimal covariance matrix Σ for Problem 4? Let the signal be $s_t = Cx_t + v_t$, where v_t is a Gaussian white noise with covariance matrix V. Using equation (24), we can recover Φ, C, and V. Then by the steady-state version of the Kalman filter, we have

$$\widehat{x}_t = \widehat{x}_{t|t-1} + \left(A\Sigma A' + W\right) C' \left[C\left(A\Sigma A' + W\right) C' + V\right]^{-1} \left(s_t - C\widehat{x}_{t|t-1}\right), \tag{32}$$

$$\widehat{x}_{t|t-1} = \left(A - BF\right) \widehat{x}_{t-1},\ \widehat{x}_{0|-1} = \bar{x}_0. \tag{33}$$

The posterior covariance matrix Σ_t of x_t will stay at Σ for all $t \geq 0$ by (13) and (14), whenever x_0 is drawn from the prior Gaussian distribution with covariance matrix $A\Sigma A' + W$. The following result characterizes the signal structure.

Proposition 5 *Suppose that $\Omega \succeq 0$, $W \succ 0$, and $A = \rho I$ with $|\rho| < 1$ in Problem 4. Then the golden-rule information structure (C, V) satisfies*

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

where $\widehat{\Sigma}_i^*$ is given in Proposition 4. The signal dimension is equal to the number of d_i such that $\lambda < 2d_i/\left(1-\rho^2\right)^2$. The signal dimension (weakly) decreases as λ increases if positive eigenvalues $d_i > 0$ are not identical.

This proposition shows that the signal dimension decreases with the information cost. The maximal dimension does not exceed the rank of the matrix $\bar{\Omega} = \Omega + (1-\beta) A'PA$, which does not exceed the minimum of the state dimension and the control dimension. For the general case, we are unable to derive analytical results, but Proposition 3 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 steady-state Kalman filter equations (32) and (33) 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 $s_t = x_t + \xi_t$. Proposition 5 shows that this signal vector may be suboptimal for multivariate problems. Moreover, any prior assumption on the signal form and dimension can lead to a suboptimal solution.

4 Tracking Problems

Consider the following tracking problem similar to that in Sims (2011). Suppose that the state vector x_t and the target y_t have a state space representation:

$$x_{t+1} = Ax_t + \eta_{t+1}, \quad y_t = Gx_t,$$

where G is a conformable matrix, x_0 is Gaussian with mean $\bar{x}_0$ and covariance matrix $\Sigma_{-1} \succ 0$, and η_{t+1} is a Gaussian white noise with covariance matrix W. The decision maker does not observe

x_t and wants to keep an action z_t close to y_t with a quadratic loss, given his or her observation of histories of signals s^t. The signal s_t satisfies (2) with $T = \infty$. The decision maker selects an optimal information structure before choosing z_t by paying an information cost of λ per nat.

Let Σ_t denote the posterior covariance matrix of x_t given information s^t. We formulate the tracking problem with discounted information costs as follows:

Problem 5 *(Tracking problem with discounted information costs)*

$$\min_{\{z_t\},\{\Sigma_t\}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t \left[(y_t - z_t)' (y_t - z_t) + \lambda I \left(x_t; s_t | s^{t-1} \right) \right]$$

subject to (17),

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

$$I \left(x_t; s_t | s^{t-1} \right) = \frac{1}{2} \log \det \left(A \Sigma_{t-1} A' + W \right) - \frac{1}{2} \log \det \left(\Sigma_t \right),$$

$$\Sigma_t \preceq A \Sigma_{t-1} A' + W, \tag{34}$$

for $t \geq 1$.

As is well known, it is optimal to set $z_t = G\mathbb{E} \left[x_t | s^t \right]$. Thus $\mathbb{E} \left[(y_t - z_t)' (y_t - z_t) \right] = \operatorname{tr} \left(G' G \Sigma_t \right)$ and this problem becomes an infinite-horizon version of Problem 3 with $\Omega = G'G$. The analysis in Section 3 applies. In particular, the golden rule of information structure solves Problem 4, where the first term in (27) is removed (or $\beta = 1$). The reason is that we do not need to consider the initial utility value $\mathbb{E} \left[x_0' P x_0 \right]$ for the tracking problem as the objective is already a loss function. Propositions 4 and 5 still apply.

To understand the distinction between the steady state as the limit point of the optimal sequence $\{\Sigma_t\}$ and the golden rule of information structure studied in Sections 3.2 and 3.3, we consider a simple univariate example with $n_x = 1$, $A = \rho$, $W = w^2$, and $G = 1$, similar to that of Sims (2011). Then the no-forgetting constraint (34) becomes the usual scalar inequality constraint. When (34) does not bind in the long run, the first-order condition gives

$$1 - \frac{\lambda}{2\Sigma_t} + \frac{\lambda}{2} \frac{\beta \rho^2}{\rho^2 \Sigma_t + w^2} = 0. \tag{35}$$

When λ is sufficiently small, this is indeed the case and the positive root Σ_t of equation (35) gives the steady-state optimal posterior variance.

By contrast, the golden rule of information structure satisfies the following first-order condition when (34) does not bind:

$$1 - \frac{\lambda}{2\Sigma} + \frac{\lambda}{2} \frac{\rho^2}{\rho^2 \Sigma + w^2} = 0. \tag{36}$$

Again this happens when λ is sufficiently small and the positive root Σ of equation (36) gives the golden rule of variance. Clearly the preceding two solutions are different for $\beta \in (0, 1)$.

It is interesting to consider the special case in which G is an n_x-dimensional row vector. Then the rank of $\Omega = G'G$ is one. We are able to derive an analytical result for the golden-rule solution when all states have the same persistence parameter ρ, but innovations are arbitrarily correlated.[14]

Proposition 6 *Consider Problem 4 where the first term $(1-\beta)\operatorname{tr}(A'PA\Sigma)$ in (27) is removed. Let G be an n_x-dimensional row vector. Suppose that $\Omega = G'G$, $W \succ 0$, and $A = \rho I$ ($|\rho| < 1$). If $\lambda \geq 2\left\|W^{1/2}G'\right\|^2 / \left(1-\rho^2\right)^2$, then no information is processed and the optimal posterior covariance matrix is given by $\Sigma = W/\left(1-\rho^2\right)$. If $0 < \lambda < 2\left\|W^{1/2}G'\right\|^2 / \left(1-\rho^2\right)^2$, then the golden-rule signal is one dimensional and can be normalized as*[15]

$$s_t = y_t + \left\|W^{1/2}G'\right\| v_t. \tag{37}$$

The variance V of v_t satisfies

$$V^{-1} = \frac{2\left\|W^{1/2}G'\right\|^2}{\lambda}\left[1 - \left(1-\rho^2\right)\widehat{\Sigma}_1^*\right] > 0,$$

and the golden-rule posterior covariance matrix Σ for x_t is given by

$$\Sigma = \frac{W}{1-\rho^2} - \frac{W\Omega W}{\left\|W^{1/2}G'\right\|^2}\left[\left(1-\rho^2\right)^{-1} - \widehat{\Sigma}_1^*\right],$$

where

$$\widehat{\Sigma}_1^* = \frac{1}{2\rho^2}\left(\sqrt{1 + \frac{2\rho^2\lambda}{\left\|W^{1/2}G'\right\|^2}} - 1\right).$$

Using numerical examples we can easily verify that Proposition 6 holds. For the general case, we are unable to derive analytical results. We can verify numerically that the signal dimension is still one even if the exogenous states have different persistence. But the signal does not take the form as in (37).

To see this, we numerically solve an example taken from Sims (2011) using the methods developed in Section 3. This example can be interpreted as a single firm's price setting problem adapted from Maćkowiak and Wiederholt (2009).[16] Let x_t represent exogenous shocks, y_t the full information profit-maximizing price, and z_t the optimal price under RI. We use the same parameter values as in Sims (2011): $\beta = 0.9$,

$$A = \begin{bmatrix} 0.95 & 0 \\ 0 & 0.4 \end{bmatrix}, \quad W = \begin{bmatrix} 0.0975 & 0 \\ 0 & 0.85 \end{bmatrix}, \quad G = [1, 1].$$

We first set $\lambda = 2$.[17] It takes about 3 seconds for a PC with Intel Core i7-7700 CPU and 16GB memory to compute the golden-rule posterior covariance matrix using CVX:

$$\Sigma = \begin{bmatrix} 0.3376 & -0.1779 \\ -0.1779 & 0.7996 \end{bmatrix}.$$

[14]When $\rho = 0$, Proposition 6 reduces to the IID case, which is also the static case studied by Fulton (2018).

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

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

[17]The parameter λ in our paper corresponds to 2λ in Sims (2011).

We also find that the eigenvalues of $A\Sigma A' + W - \Sigma$ are -8.53×10^{-6} and 0.2529. Thus the no-forgetting constraint (26) binds in the sense that $A\Sigma A' + W - \Sigma$ is singular and hence the golden-rule signal is one dimensional with an error less than 10^{-5}. The signal takes the form $s_t = [1, 0.6836]\, x_t + v_t$, where v_t is a Gaussian white noise with variance 1.1901.

We next solve for the steady state as the limit point of the sequence of optimal posterior covariance matrices. As is well known, the convergence for the method of value function iteration is sensitive to the discount factor β. For $\beta = 0.9$, it takes about 18 minutes for the same PC to get convergence of J_t and Σ_t with an error less than 10^{-4}.[18] Given the initial prior $\Sigma_{-1} = W$, it takes about 18 periods for the optimal posterior covariance matrix to converge to the steady state:

$$\Sigma = \begin{bmatrix} 0.3590 & -0.1769 \\ -0.1769 & 0.7945 \end{bmatrix}.$$

Again we find that the no-forgetting constraint (26) binds in the steady state (the eigenvalues of $A\Sigma A' + W - \Sigma$ are 3.53×10^{-5} and 0.2551) and the steady-state optimal signal is one dimensional with an error less than 10^{-4}.

When $\lambda = 0.2$ and $\beta = 0.9$, the steady-state optimal posterior covariance matrix and the golden-rule posterior covariance matrix are respectively given by

$$\Sigma = \begin{bmatrix} 0.3199 & -0.3041 \\ -0.3041 & 0.3861 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0.3181 & -0.3040 \\ -0.3040 & 0.3875 \end{bmatrix}.$$

The no-forgetting constraint (26) binds for both solutions with an error less than 10^{-4} (the two eigenvalues of $A\Sigma A' + W - \Sigma$ are -4.58×10^{-5} and 0.6020, and 1.94×10^{-5} and 0.6010, respectively) and the implied signals are one dimensional.

In summary, the golden-rule solution and the limiting steady-state solution are generally different, but close together especially for β close to 1. Both solutions imply a similar one-dimensional signal structure. Sims (2011) shows that the solutions for $\lambda = 2$ and $\lambda = 0.2$ are respectively given by

$$\Sigma = \begin{bmatrix} 0.373 & -0.174 \\ -0.174 & 0.774 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 0.318 & -0.300 \\ -0.300 & 0.380 \end{bmatrix}.$$

The implied eigenvalues of $A\Sigma A' + W - \Sigma$ are 4.5×10^{-3} and 0.2665 for $\lambda = 2$, and are 2.3×10^{-3} and 0.6050 for $\lambda = 0.2$. While Sims's solutions are close to ours, they may lead to a different conclusion regarding signal dimension and impulse response functions when errors are specified to be less than 10^{-3}.

Because the golden-rule solution is much easier to compute numerically and because much of the literature just analyzes this type of solutions following Sims (2003), we will also focus on it in the next section for a better comparison with the literature.

[18] The convergence is much faster for small values of β. For example, the program converges in about 2 minutes for $\beta = 0.4$.

5 Applications

In this section we study three applications to illustrate our results. We analyze a pure tracking problem in an equilibrium setting 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.

5.1 Equilibrium Sticky Prices

We extend the pricing problem in Section 4 to an equilibrium setting as in Maćkowiak and Wiederholt (2009). Here we present the key equilibrium conditions directly and refer the reader to their paper for detailed derivations and interpretations.

Consider an economy with a continuum of firms indexed by $j \in [0,1]$. Firm j sells good j and sets its prices to maximize the present discounted value of profits. The full-information profit-maximizing price is given by

$$p_{jt}^{*} = (1-\alpha_2)\, p_t + \alpha_2 q_t + \alpha_3 z_{jt}, \tag{38}$$

where p_t is the aggregate price level, q_t is nominal aggregate demand, and z_{jt} represents an idiosyncratic shock. The parameter $\alpha_2 \in (0,1]$ describes the degree of strategic complementarity. Suppose that z_{jt} and q_t follow exogenous AR(1) processes

$$\begin{aligned} z_{jt} &= \rho_i z_{j,t-1} + \epsilon_{jt}, \quad 0<\rho_i<1, \\ q_t &= \rho_a q_{t-1} + \epsilon_{at}, \quad 0<\rho_a<1, \end{aligned}$$

where ϵ_{jt} and ϵ_{at} are independent Gaussian white noise processes with variances σ_i^2 and σ_a^2. Assume that z_{jt} is also independent across firms $j \in [0,1]$ such that $\int \epsilon_{jt} dj = 0$.

Each firm j does not observe q_t and z_{jt}. It acquires an optimal signal vector s_{jt} about a vector x_{jt} of unobserved states subject to discounted entropy information costs. To fit in the framework of Section 4, assume that the vector of states x_{jt} and the target p_{jt}^{*} have a state space representation. We will specify the state vector x_{jt} later.

Firm j sets price p_{jt} to track p_{jt}^{*} subject to entropy information costs. For simplicity, we focus on the long-run Golden-rule solution to the following problem:

$$\max_{p_{jt},\Sigma_j} \ \alpha_1 \mathbb{E}\left[\left(p_{jt}-p_{jt}^{*}\right)^2\right] + \lambda I\left(x_{jt}; s_{jt}|s_j^{t-1}\right), \tag{39}$$

subject to a no-forgetting constraint, where $\alpha_1 > 0$, Σ_j is the posterior covariance matrix, $s_{jt} = C_j x_{jt} + v_{jt}$, and v_{jt} is a Gaussian white noise with covariance matrix V_j. Then the optimal price under RI is given by $p_{jt} = \mathbb{E}\left[p_{jt}^{*}|s_j^t\right]$. Assume that v_{jt} is independent of all other shocks, and is independent across firms $j \in [0,1]$ such that $\int v_{jt}dj = 0$. The model is closed by the equilibrium condition:

$$p_t = \int_0^1 p_{jt}dj. \tag{40}$$

In the analysis below, we normalize $\alpha_1 = 1$.

5.1.1 No Strategic Complementarity

When there is no strategic complementarity ($\alpha_2 = 1$), we have $p^*_{jt} = q_t + \alpha_3 z_{jt}$. Then there is no equilibrium price feedback to individual pricing decisions. After defining the state vector as $x_{jt} = (z_{jt}, q_t)'$, we obtain the state space representation: $p^*_{jt} = Gx_{it}$, $G = (\alpha_3, 1)$,

$$x_{jt} = Ax_{j,t-1} + \begin{bmatrix} \epsilon_{jt} \\ \epsilon_{at} \end{bmatrix}, \ A = \begin{bmatrix} \rho_i & 0 \\ 0 & \rho_a \end{bmatrix}, \ W = \begin{bmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_a^2 \end{bmatrix}.$$

The problem (39) becomes a single firm's pricing problem under RI studied in Section 4.

Firm j's optimal price under RI is given by

$$p_{jt} = \mathbb{E}\left[p^*_{jt}|s^t_j\right] = G\mathbb{E}\left[x_{jt}|s^t_i\right] = G\widehat{x}_{jt}, \tag{41}$$

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

$$\widehat{x}_{jt} = \left(I - K_jC_j\right) A\widehat{x}_{j,t-1} + K_j s_{jt}, \tag{42}$$

for $t \geq 0$, with $\widehat{x}_{j,-1} = 0$. Here $\{C_j, V_j\}$ is derived from Proposition 1 and

$$K_j \equiv \left(A\Sigma A' + W\right) C_j' \left[C_j \left(A\Sigma A' + W\right) C_j' + V_j\right]^{-1}$$

is the Kalman gain and Σ is the solution to (39). Notice that Σ is the same for all firms j, but (C_j, V_j) may not be the same across firms because this pair is not uniquely determined. We focus on symmetric equilibrium in which all (C_j, V_j) are the same across firms j, so we remove the subscript j.

Equations (41) and (42) show that individual price responses p_{jt} to shocks through s_{jt} are determined by two effects: (i) the learning effect reflected by the term GK, and (ii) the attention allocation effect reflected by the optimal choice of information structure C, V, Σ, and K.

In Appendix E we show that the equilibrium aggregate price satisfies

$$p_t = \int_0^1 p_{jt}dt = G\int_0^1 \widehat{x}_{jt}dj = G\left[I - (I-KC)AL\right]^{-1} KC(I-AL)^{-1}\left[0,1\right]'\epsilon_{at},$$

where L represents the lag operator. Since we can verify that $KC = I - \Sigma\left(A\Sigma A' + W\right)^{-1}$, p_t is determined by G, A, and Σ, and is independent of C and V.

When $\rho_i = \rho_a$, Proposition 6 applies. Equation (37) shows that the optimal signal can be normalized as the profit-maximizing price plus a noise (i.e., $s_{jt} = p^*_{jt} + v_{jt}$). This signal form implies that the impulse responses of individual prices to the idiosyncratic shock z_{jt} are larger than to the aggregate shock q_t if and only if it carries a larger weight α_3 as shown in equations (41) and (42). The individual price responses are the same when $\alpha_3 = 1$. 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. They show that the price is more responsive to the shock with a higher variance even when $\rho_i = \rho_a$ and $\alpha_3 = 1$.

When $\rho_i \neq \rho_a$, based on numerical solutions for a wide range of parameter values, we find that the optimal signal is still one dimensional, but it does not take the normalized form of the profit-maximizing price plus a noise. Instead of presenting a detailed comparative statics analysis here, we turn to the more interesting case with strategic complementarity.

5.1.2 Strategic Complementarity

When there is strategic complementarity, i.e., $\alpha_2 \in (0,1)$, there is equilibrium price feedback in (38). The equilibrium solution becomes more involved due to higher-order beliefs. We present the technical details in Appendix E.

We focus on the equilibrium in which the aggregate price p_t follows a causal stationary process, which has an MA(∞) representation. We approximate such an equilibrium by a stationary ARMA(r,m) process $p_t = \Psi(L)\,\epsilon_{at}$ for a large enough $r \geq m+1$,[19] where L represents the lag operator and

$$\Psi(z) \equiv \frac{b_0 + b_1 z + b_2 z^2 + \cdots + b_m z^m}{1 - a_1 z - a_2 z^2 - \cdots - a_r z^r}. \tag{43}$$

Here z is a complex number in the unit circle. All coefficients in the rational function Ψ and the order (r and m) are endogenous with $a_r \neq 0$ and $b_m \neq 0$. Notice that the equilibrium aggregate price p_t contains only aggregate innovations ϵ_{at}, because idiosyncratic innovations ϵ_{jt} wash out in the aggregate.

We adopt the following state space representation (Hamilton (1994)):

$$x_{jt} = \begin{bmatrix} \rho_i & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \rho_a & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & a_1 & a_2 & \cdots & \cdots & a_{r-1} & a_r \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} x_{j,t-1} + \begin{bmatrix} \epsilon_{jt} \\ \epsilon_{at} \\ \epsilon_{at} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \tag{44}$$

$$p^*_{jt} = Gx_{jt}, \quad G = [\alpha_3, \ \ \alpha_2, \ \ (1-\alpha_2)D], \ D = \begin{bmatrix} b_0 & b_1 & \cdots & b_{r-2} & b_{r-1} \end{bmatrix}, \tag{45}$$

[19] This assumption ensures the state transition matrix A constructed in equation (44) is invertible.

where the state vector $x_{jt}' = [z_{jt}, q_t, \xi_t']$ consists of the exogenous states z_{jt}, q_t, and an endogenous state (column) vector ξ_t such that $p_t = D\xi_t$. Moreover, we set $b_{m+1} = b_{m+2} = ... = b_{r-1} = 0$. Let the $(r+2) \times 1$ noise vector be $\eta_{jt} \equiv [\epsilon_{jt}, \epsilon_{at}, \epsilon_{at}, 0, ..., 0]'$. Then η_{jt} is a Gaussian white noise and its covariance matrix W is singular. Let A denote the transition matrix in equation (44). We can check that A is invertible.

Because W is singular, we cannot apply Proposition 3 to solve for the golden-rule solution to the RI problem. But we can apply Proposition 8 in Appendix C as A is invertible. After obtaining the optimal information structure, we aggregate individual optimal prices using (40) and (41). We then obtain a fixed point problem for $(a_1, a_2, ..., a_r, b_0, b_1, ..., b_m)$. In Appendix E, we describe the algorithm to solve this fixed point problem and determine the endogenous r and m. Once obtaining these coefficients, we can determine the equilibrium aggregate price function and individual pricing rules.

We set baseline parameter values as follows: $\lambda = 0.002$, $\rho_i = \rho_a = 0.95$, $\sigma_i = 10\%$, $\sigma_a = 1\%$, $\alpha_1 = \alpha_3 = 1$, and $\alpha_2 = 0.15$. For these parameter values we find that an ARMA(2,1) process is a good approximation of the equilibrium aggregate price p_t.[20] Then the state vector x_{jt} is $r+2=4$ dimensional. We also find that the optimal signal vector s_{jt} is three dimensional and the no-forgetting constraint binds. Thus the signal vector violates the signal independence assumption in the literature (Maćkowiak and Wiederholt (2009)). Moreover, Proposition 6 does not apply in that the optimal signal cannot be normalized as the profit-maximizing price plus a noise. Importantly, the optimal signal form implies that the aggregate and idiosyncratic shocks (q_t and z_{jt}) are confounded. We will show below that this feature has interesting economic implications.

Now we consider the impact of the information cost λ on the impulse responses of the aggregate equilibrium price to a one-standard-deviation shock to the nominal aggregate demand, shown in the left panel of Figure 1. Under full information, the aggregate price moves one-to-one with the nominal aggregate demand shock so that real output does not change. Under rational inattention, the responses are dampened and delayed. The higher the information cost λ, the less responsive the aggregate price is.

[Insert Figure 1 Here.]

The right panel of Figure 1 shows the impact of the degree of strategic complementarity α_2. The case with $\alpha_2 = 1$ corresponds to the solution without strategic complementarity studied earlier. As in Maćkowiak and Wiederholt (2009), when the profit-maximizing price is less sensitive to real aggregate demand (i.e., when α_2 is lower), the response of the price level to a nominal demand shock is more dampened and delayed. The reason is that the price feedback effects are stronger.

[20]It takes about 24 seconds for a PC with Intel Core i7-7700 CPU and 16GB memory to find an equilibrium. The order of the ARMA process of the aggregate price may be higher for a high innovation variance and a high information cost.

Next we study the impact of innovation volatilities presented in Figure 2. Under the signal independence assumption, Maćkowiak and Wiederholt (2009) find when the innovation variance of a shock increases, firms shift attention toward that shock, and away from the other shock. By contrast, Figure 2 shows that when the innovation variance of a shock increases, the individual price responses to both aggregate and idiosyncratic shocks rise. Thus there is a spillover effect similar to that in Mondria (2010). The intuition is that the optimal signal structure implies that aggregate and idiosyncratic shocks are confounded. The effect of an increase in the innovation variance of one shock is transmitted to the other shock due to the learning effect via the term GK.

[Insert Figure 2 Here.]

We finally study the impact of persistence of shocks presented in Figure 3. When we change a persistence parameter ρ_i or ρ_a, we adjust the innovation variance to hold the unconditional variance fixed as in Maćkowiak and Wiederholt (2009). We find that the impact of persistence on individual price responses is ambiguous, a result similar to Maćkowiak and Wiederholt (2009). For the baseline parameter values, Figure 3 shows that individual price responses to the aggregate shock are larger if the idiosyncratic shock is less persistent. But individual price responses to the idiosyncratic shock are not monotonic with the persistence of the idiosyncratic shock. By contrast, individual price responses to the aggregate shock are larger if the aggregate shock is more persistent, even though its unconditional variance is much smaller than that of the idiosyncratic shock. But individual price responses to the idiosyncratic shock barely changes with the persistence of the aggregate shock.

[Insert Figure 3 Here.]

5.2 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

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

where $\bar{c}$ is a bliss level of consumption, w_t is wealth, and y_t is income. For simplicity we suppose $\beta\left(1+r\right)=1$. 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 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 $\sigma_2^2 = 0.003$. Unlike Sims (2003), we focus on the golden-rule information structure with discounted information costs, instead of capacity constraints.[21]

[Insert Figure 4 Here]

For the information cost parameter $\lambda = 0.01$, we find that the optimal signal vector s_t is one dimensional and $C = [1, \ 11.7433, \ 5.8978]$ and $V = 1.4319$.[22] 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 λ increases, the linear transformation C barely changes. But the signal noise variance increases significantly. Intuitively, the signal becomes more noisy when the information cost is larger.

Figure 4 plots the impulse response functions for consumption to a one-standard-deviation shock to each of the three true income components and the 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 $1+r$

[21] We have also solved for the transition dynamics of the posterior covariance matrix and its steady state. In previous version of the paper we solved the case with capacity constraints. The impulse response functions are qualitatively similar.

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

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 λ. And the initial responses to all true income shocks are larger when λ is smaller. 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 less persistent income shock (η_{2t}) is smaller, even if its innovation variance is very large.

5.3 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

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

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

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

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 control 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. Notice that the optimal capital and investment choice is independent of transitory shocks e_t.

We now solve for the long-run golden-rule information structure using the semidefinite programming approach.[23] We set baseline parameter values as in McGrattan and Prescott (2010): $\alpha = 0.26$, $\theta = 0.076$, $\delta_1 = 0.126$, $\delta_2 = 0.05$, $\tau = 0.35$, 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 $k_2 = 0.639$.

[Insert Figure 5 here.]

Since this model features two control variables and four state variables, we can study the non-trivial determination of the information structure. We find that the signal dimensional decreases with the information cost λ. As shown in Proposition 5, the signal dimension does not exceed the minimum of the state dimension and the control dimension. To understand how the signal dimension changes, we display here the optimal signal structure for two values of λ with no adjustment costs: For $\lambda = 0.01$,

$$s_t = \begin{bmatrix} -0.94 z_t + 0.30 \tilde{k}_{1,t} + 0.16 \tilde{k}_{2,t} \\ -0.04 z_t + 0.39 \tilde{k}_{1,t} - 0.92 \tilde{k}_{2,t} \end{bmatrix} + v_t,$$

where the covariance matrix of v_t is $\text{diag}(0.03, 0.15)$, but for $\lambda = 0.08$,

$$s_t = -0.93 z_t + 0.23 \tilde{k}_{1,t} + 0.30 \tilde{k}_{2,t} + v_t,$$

where the variance of v_t is 0.65. 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.

[23] We have also numerically solved for the transition dynamics and the limiting steady state of the optimal posterior covariance matrix.

We now turn to the impulse responses of two types of capital investment to a positive one-standard-deviation shock to the persistent TFP component displayed in Figure 5. Each panel of the figure includes the full information case as well as at least one case with sufficiently high λ 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 cost λ rises, the investment responses under RI become dampened and delayed – investment rises less on impact and remains above the steady state longer. If λ is sufficiently large, then the signal becomes one dimensional and the responses are very small and persistent (see dashed lines).

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 cost, 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.[24] He documents evidence that investment at the sectoral level displays a hump-shaped response to aggregate shocks and a monotonic response to 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, and models with just rational inattention cannot match both types of impulse responses.

[24] 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.

6 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 many simplifying assumptions adopted in the literature such as signal independence 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 different roles of the shock persistence and the innovation variance, information spillover, and price comovement. 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 Lemma 1: Fix the information structure $\{C_t, V_t\}$. Consider the control problem:

$$\widehat{v}_t \equiv \min_{\{u_\tau\}} \mathbb{E}\left[\sum_{\tau=t}^{T} \beta^{\tau-t}\left(x_\tau' Q_\tau x_\tau + u_\tau' R_\tau u_\tau + 2x_\tau' S_\tau u_\tau\right) + \beta^{T+1} x_{T+1}' P_{T+1} x_{T+1} \middle| s^t\right]$$

subject to (1) and (2) from period t on. Claim that

$$\widehat{v}_t = \mathbb{E}\left[x_t P_t x_t | s^t\right] + \sum_{\tau=t}^{T} \beta^{\tau-t+1} \text{tr}\left(W_\tau P_{\tau+1}\right) + \sum_{\tau=t}^{T} \beta^{\tau-t} \text{tr}\left(\Omega_\tau \Sigma_\tau\right), \tag{A.1}$$

where P_t and Ω_t satisfy (5) and (15). We prove this claim using backward induction. In the last period T, we compute the objective function as

$$\begin{aligned} & \mathbb{E}\left[\left(x_T' Q_T x_T + u_T' R_T u_T + 2x_T' S_T u_T\right) + \beta x_{T+1}' P_{T+1} x_{T+1} | s^T\right] \qquad \text{(A.2)} \\ = \; & \mathbb{E}\left[\left(x_T' Q_T x_T + u_T' R_T u_T + 2x_T' S_T u_T\right) | s^T\right] \\ & + \beta \mathbb{E}\left[\left(A_T x_T + B_T u_T + \epsilon_{T+1}\right)' P_{T+1} \left(A_T x_T + B_T u_T + \epsilon_{T+1}\right) | s^T\right]. \end{aligned}$$

Rewrite the above expression as

$$\begin{aligned} & \mathbb{E}\left[\left(x_T' Q_T x_T + u_T' R_T u_T + 2x_T' S_T u_T\right) | s^T\right] \\ & + \beta \mathbb{E}\left[x_T' A_T' P_{T+1} A_T x_T + u_T' B_T' P_{T+1} B_T u_T + \epsilon_{T+1}' P_{T+1} \epsilon_{T+1} | s^T\right] \\ & + 2\beta \mathbb{E}\left[x_T' A_T' P_{T+1} B_T u_T | s^T\right] \end{aligned}$$

$$\begin{aligned} = \; & \beta \text{tr}\left(W_T P_{T+1}\right) + \mathbb{E}\left[x_T' Q_T x_T | s^T\right] + \beta \mathbb{E}\left[x_T' A_T' P_{T+1} A_T x_T | s^T\right] \\ & + \mathbb{E}\left[u_T'\left(R_T + \beta B_T' P_{T+1} B_T\right) u_T + 2x_T'\left(S_T + \beta A_T' P_{T+1} B_T\right) u_T | s^T\right] \end{aligned}$$

Taking the first-order condition gives the optimal control $u_T = -F_T \widehat{x}_T$, where F_T satisfies (7) for $t = T$. Substituting this equation back into the objective function yields

$$\widehat{v}_T = \mathbb{E}\left[x_T' P_T x_T | s^T\right] + \beta \text{tr}\left(W_T P_{T+1}\right) + \text{tr}\left(\Omega_T \Sigma_T\right),$$

where P_T satisfies (5) for $t = T$ and where we notice that x_T conditional on s^T is Gaussian with mean $\hat{x}_T$ and covariance matrix Σ_T.

Suppose that (A.1) holds for $\widehat{v}_{t+1}$ in period $t+1$. By dynamic programming, we have

$$\widehat{v}_t = \min_{u_t} \mathbb{E}\left[\left(x_t' Q_t x_t + u_t' R_t u_t + 2x_t' S_t u_t\right) + \beta \widehat{v}_{t+1} | s^t\right].$$

Rewriting the objective function by the induction hypothesis yields

$$
\begin{aligned}
& \mathbb{E}\left[\left(x_t'Q_tx_t + u_t'R_tu_t + 2x_t'S_tu_t\right) + \beta\widehat{v}_{t+1}|s^t\right] \\
= \; & \mathbb{E}\left[\left(x_t'Q_tx_t + u_t'R_tu_t + 2x_t'S_tu_t\right)|s^t\right] + \beta\mathbb{E}\left[x_{t+1}P_{t+1}x_{t+1}|s^t\right] \\
& + \sum_{\tau=t+1}^{T}\beta^{\tau-t+1}\mathrm{tr}\left(W_\tau P_{\tau+1}\right) + \sum_{\tau=t+1}^{T}\beta^{\tau-t}\mathrm{tr}\left(\Omega_\tau\Sigma_\tau\right).
\end{aligned}
$$

The expression on the second line has the same form as in (A.2). By the previous analysis, we deduce that the optimal policy is given by $u_t = -F_t\widehat{x}_t$, where F_t satisfies (7). Substituting this policy back into the preceding objective function, we find that the resulting objective function equals

$$
\begin{aligned}
& \mathbb{E}\left[x_t'P_tx_t|s^t\right] + \beta\mathrm{tr}\left(W_tP_{t+1}\right) + \mathrm{tr}\left(\Omega_t\Sigma_t\right) \\
& + \sum_{\tau=t+1}^{T}\beta^{\tau-t+1}\mathrm{tr}\left(W_\tau P_{\tau+1}\right) + \sum_{\tau=t+1}^{T}\beta^{\tau-t}\mathrm{tr}\left(\Omega_\tau\Sigma_\tau\right),
\end{aligned}
$$

where P_t satisfies (5). Thus $\widehat{v}_t$ takes the form in (A.1), completing the induction proof. Finally, letting $t = 0$ and taking unconditional expectations, we obtain the desired result. Q.E.D.

Proof of Proposition 1: For simplicity we omit the time t subscript for all variables in the proof. By the singular-value decomposition of a positive semidefinite matrix, there exists an $n_x \times n_x$ orthogonal matrix U 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 $\widehat{\Psi} = \mathrm{diag}(\varphi_1, ..., \varphi_m)$ is an $m \times m$ diagonal matrix and $\{\varphi_i\}_{i=1}^m$ are the positive eigenvalues of Φ. Clearly, $\mathrm{rank}(\Phi) = m \leq n_x$. The matrix Φ can be factored into $\Psi = \Delta'\widehat{\Psi}\Delta$, 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 2: We first prove the following lemma:

Lemma 2 *Suppose that $W_t \succ 0$, $\Sigma_t \succ 0$, and $\beta \in (0, 1]$. Then*

$$
\beta \log\det\left(A_t\Sigma_tA_t' + W_t\right) - \log\det\left(\Sigma_t\right)
$$

is a strictly convex function of Σ_t.

Proof: We can write

$$
\begin{aligned}
& \beta \log\det\left(A_t\Sigma_tA_t' + W_t\right) - \log\det\left(\Sigma_t\right) \\
= \; & \beta\left[\log\det\left(A_t\Sigma_tA_t' + W_t\right) - \log\det\left(\Sigma_t\right)\right] - (1-\beta)\log\det\left(\Sigma_t\right).
\end{aligned}
$$

Since the log-determinant function is strictly concave, it suffice to prove that the expression in the square bracket is convex in Σ_t. The matrix determinant lemma (Theorem 18.1.1 in Harville (1997)) implies that

$$\log\det\left(A_t\Sigma_tA_t'+W_t\right)-\log\det\left(\Sigma_t\right)=\log\det W_t+\log\det\left(\Sigma_t^{-1}+A_t'W_t^{-1}A_t\right). \tag{A.3}$$

By Diggavi and Cover (2001), the last expression is convex in Σ_t as desired. Notice that Sims (2003, page 678) proves the convexity by direct differentiation assuming A_t is invertible. Q.E.D. □

Now we prove that $J_t\left(\Sigma_{t-1}\right)$ is strictly convex in Σ_{t-1} for $t=0,1,...,T$ by backward induction. In the last period, it follows from (18) that $J_T\left(\Sigma_{T-1}\right)$ is strictly convex in Σ_{T-1}. Suppose that $J_{t+1}\left(\Sigma_t\right)$ is strictly convex in Σ_t for any $t\leq T-1$. Then, by Lemma 2, the objective function in (19) is strictly convex. Since the constraint set is convex, we can verify that $J_t\left(\Sigma_{t-1}\right)$ is strictly convex.

Finally we transform the dynamic programming problem (19) into a semidefinite program representation. The matrix determinant lemma (Theorem 18.1.1 in Harville (1997)) implies that the preceding expression is equal to

$$\log\det\left(A_t\Sigma_tA_t'+W_t\right)-\log\det\left(\Sigma_t\right)=\log\det W_t-\log\det\left(\Sigma_t^{-1}+A_t'W_t^{-1}A_t\right)^{-1}. \tag{A.4}$$

Due to the monotonicity of the determinant function, we have

$$-\log\det\left(\Sigma_t^{-1}+A_t'W_t^{-1}A_t\right)^{-1}=\min_{\Pi_t\succ 0}\quad -\log\det\Pi_t$$

subject to

$$\Pi_t\preceq\left(\Sigma_t^{-1}+A_t'W_t^{-1}A_t\right)^{-1}. \tag{A.5}$$

Apply the matrix inversion formula to rewrite (A.5) as

$$\Pi_t\preceq\Sigma_t-\Sigma_tA_t'\left(W_t+A_t\Sigma_tA_t'\right)^{-1}A_t\Sigma_t,$$

which is equivalent to

$$\begin{bmatrix}\Sigma_t-\Pi_t & \Sigma_tA_t'\\ A_t\Sigma_t & W_t+A_t\Sigma_tA_t'\end{bmatrix}\succeq 0, \tag{A.6}$$

by the Schur complement property. By (A.4) and the preceding derivations, we have

$$\log\det\left(A_t\Sigma_tA_t'+W_t\right)=\min_{\Pi_t\succ 0}\quad -\log\det\Pi_t+\log\det W_t+\log\det\left(\Sigma_t\right)$$

subject to (A.6). Replacing $\log\det\left(A_t\Sigma_tA_t'+W_t\right)$ in (19) with the preceding minimized value, we obtain the representation in the proposition. Q.E.D.

Proof of Proposition 3: Using a similar transformation in the proof of Proposition 2, we can derive the desired semidefinite program representation. We will not repeat the detailed derivation. Q.E.D.

Proof of Proposition 4: The matrix determinant lemma implies that

$$\log\det(A\Sigma A' + W) - \log\det\Sigma = \log\det W - \log\det\left(\Sigma^{-1} + A'W^{-1}A\right)^{-1}.$$

Thus Problem 4 becomes

$$\min_{\Pi,\Sigma\succ 0} \ \mathrm{tr}\left(\bar{\Omega}\Sigma\right) + \frac{\lambda}{2}\left[\log\det W - \log\det\Pi\right] \tag{A.7}$$

subject to

$$\Pi = \left(\Sigma^{-1} + A'W^{-1}A\right)^{-1}, \tag{A.8}$$

$$A\Sigma A' + W \succeq \Sigma. \tag{A.9}$$

Recall the positive semidefinite matrix $W^{\frac{1}{2}}\bar{\Omega}W^{\frac{1}{2}}$ admits an eigendecomposition $W^{\frac{1}{2}}\bar{\Omega}W^{\frac{1}{2}} = U\Omega_d U'$. 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 we can derive that

$$\Pi = W^{\frac{1}{2}}U\widehat{\Pi}U'W^{\frac{1}{2}}, \ \Sigma = W^{\frac{1}{2}}U\widehat{\Sigma}U'W^{\frac{1}{2}}, \ \mathrm{tr}\left(\bar{\Omega}\Sigma\right) = \mathrm{tr}\left(\Omega_d\widehat{\Sigma}\right),$$

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

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

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

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

Now the problem in (A.7) is equivalent to

$$\min_{\widehat{\Pi},\widehat{\Sigma}} \ \mathrm{tr}\left(\Omega_d\widehat{\Sigma}\right) - \frac{\lambda}{2}\log\det\widehat{\Pi}$$

subject to (A.10) and (A.11). 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. As a result the optimal solution for $\widehat{\Pi}$ must be diagonal. Let $\widehat{\Pi} = diag\left(\widehat{\Pi}_i\right)_{i=1}^{n_x}$. By (A.10), $\widehat{\Sigma}$ is also diagonal and its diagonal elements are given by

$$\widehat{\Sigma}_i = \left(\widehat{\Pi}_i^{-1} - \rho^2\right)^{-1}, \; i = 1, 2, ..., n_x. \tag{A.12}$$

Thus the problem is equivalent to

$$\min_{\widehat{\Pi}_i} \; \mathrm{tr}\left(\Omega_d \widehat{\Sigma}\right) - \frac{\lambda}{2}\sum_{i=1}^{n_x} \log \widehat{\Pi}_i$$

subject to (A.12) and

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

Equivalently rewriting this problem in terms of $\widehat{\Sigma}_i$ using (A.12) yields

$$\min_{\widehat{\Sigma}_i} \; \sum_{i=1}^{n_x} d_i \widehat{\Sigma}_i + \frac{\lambda}{2}\sum_{i=1}^{n_x} \log\left(\rho^2 + \frac{1}{\widehat{\Sigma}_i}\right) \tag{A.13}$$

subject to

$$0 < \widehat{\Sigma}_i \leq \frac{1}{1-\rho^2}, \quad i = 1, ..., n_x. \tag{A.14}$$

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

$$\widehat{\Sigma}_i = \min\left(\frac{1}{1-\rho^2}, \widehat{\Sigma}_i^*\right), \tag{A.15}$$

where

$$\widehat{\Sigma}_i^* = \frac{1}{2\rho^2}\left(\sqrt{1 + \frac{2\rho^2\lambda}{d_i}} - 1\right). \tag{A.16}$$

The proof is completed. Q.E.D.

Proof of Proposition 5: The optimal signal-to-noise ratio is given by

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

where the last equality follows from (A.15) and (A.16). The dimension of the signal is determined by the rank of the inside diagonal matrix, which is determined by the number of d_i such that

$$\frac{2d_i}{\lambda}\left[1 - \left(1-\rho^2\right)\widehat{\Sigma}_i^*\right] > 0.$$

Using equation (A.16) we obtain the desired result. Q.E.D.

Proof of Proposition 6: 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 $d_1 \equiv \left\|W^{1/2}G'\right\|^2$ and an associated unit eigenvector $W^{\frac{1}{2}}G'/\left\|W^{1/2}G'\right\|$ where $\|\cdot\|$ denotes the Euclidean norm. To prove this claim we verify that

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

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

By Propositions 4 and 5, 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

$$\widehat{\Sigma}_1^* = \frac{1}{2\rho^2}\left(\sqrt{1+\frac{2\rho^2\lambda}{d_1}} - 1\right).$$

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

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

If $\lambda \geq 2d_1/\left(1-\rho^2\right)^2$, we can check that $\widehat{\Sigma}_i = 1/\left(1-\rho^2\right)$ for all i so that $\Sigma = W/\left(1-\rho^2\right)$ and no information is collected. There is only one positive element in the above inside diagonal matrix if $0 < \lambda < 2d_1/\left(1-\rho^2\right)^2$, which is

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

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

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

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

The optimal conditional covariance in the proposition follows from Proposition 4. In particular,

$$\Sigma = W^{\frac{1}{2}}U\left[\begin{array}{cc}\widehat{\Sigma}_1^* & 0 \\ 0 & \frac{1}{1-\rho^2}I\end{array}\right]U'W^{\frac{1}{2}}.$$

Partition $U = [U_1, U_2]$ conformably, where $U_1 = W^{\frac{1}{2}}G' / \left\| W^{1/2}G' \right\|$. Then we have $U_1U_1' + U_2U_2' = I$. Thus

$$\Sigma = W^{\frac{1}{2}} \left[\frac{I}{1-\rho^2} - U_1U_1' \left(\frac{1}{1-\rho^2} - \Sigma_1^* \right) \right] W^{\frac{1}{2}}.$$

Simplifying yields the expression in the proposition.

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

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

We then obtain (37). Q.E.D.

B RI Problems with Period-by-Period Capacity Constraints

In this appendix we study Problem 1 with period-by-period capacity constraints. As in the analysis of Section 2, we can show that the optimal information structure is determined by the following problem:

Problem 6 *(Optimal information structure for Problem 1)*

$$\min_{\{\Sigma_t\}_{t=0}^T} \sum_{t=0}^T \beta^t \mathrm{tr}\left(\Omega_t \Sigma_t\right)$$

subject to

$$\log\det\left(A_{t-1}\Sigma_{t-1}A_{t-1}' + W_{t-1}\right) - \log\det\left(\Sigma_t\right) \leq 2\kappa, \tag{B.1}$$

$$\log\det\left(\Sigma_{-1}\right) - \log\det\left(\Sigma_0\right) \leq 2\kappa, \tag{B.2}$$

$$\Sigma_t \preceq A_{t-1}\Sigma_{t-1}A_{t-1}' + W_{t-1}, \tag{B.3}$$

$$\Sigma_0 \preceq \Sigma_{-1}, \tag{B.4}$$

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

Since the log-determinant function is concave, the constraint set may not be convex in $\{\Sigma_t\}_{t=0}^T$. Thus the Kuhn-Tucker conditions may not be optimal. By dynamic programming, the value function satisfies the Bellman equation

$$J_t\left(\Sigma_{t-1}\right) = \min_{\Sigma_t \succ 0} \ \mathrm{tr}\left(\Omega_t \Sigma_t\right) + \beta J_{t+1}\left(\Sigma_t\right)$$

subject to (B.1) and (B.3) for $t \geq 1$. In the last period T, $J_{T+1}\left(\Sigma_T\right) \equiv 0$. In the initial period, we have

$$J_0\left(\Sigma_{-1}\right) = \min_{\Sigma_0 \succ 0} \ \mathrm{tr}\left(\Omega_0 \Sigma_0\right) + \beta J_1\left(\Sigma_0\right)$$

subject to (B.2) and (B.4). Since the log-determinant function is concave, the value function $J_t\left(\Sigma_{t-1}\right)$ may not be convex for $t = 0, 1, ..., T$. This can be easily seen for $J_T\left(\Sigma_{T-1}\right)$ in the last period using the envelope theorem. For a univariate problem with $n_x = 1$, Σ_t is a scalar and we can rewrite (B.1) and (B.2) as linear scalar constraints so that $J_t(\Sigma_{t-1})$ is convex.

Nonconvexity poses substantial difficulty when solving the above dynamic programming problem. This issue does not arise when solving for the long-run golden rule of information structure.

Problem 7 *(Golden rule of information structure for Problem 6)*

$$\min_{\Sigma \succ 0} \quad (1-\beta)\operatorname{tr}\left(A'PA\Sigma\right) + \operatorname{tr}(\Omega\Sigma) \tag{B.5}$$

subject to (26) and

$$\log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma) \leq 2\kappa.$$

By Lemma 2, $\log\det(A\Sigma A' + W) - \log\det(\Sigma)$ is a convex function of Σ. Thus the above problem is a convex program. This problem is the same as that in Sims (2003) except that there is a new term in (B.5) as discussed in Section 3.3. Notice that software CVX does not recognize that $\log\det(A\Sigma A' + W) - \log\det(\Sigma)$ is convex in Σ by its ruleset.

To apply CVX, we need to transform Problem 7 into a DCP. As in the proof of Proposition 2, we can show that $\log\det(A\Sigma A' + W) - \log\det(\Sigma) = c(\Sigma)$, where $c(\Sigma)$ is a new function defined as

$$c(\Sigma) \equiv \min_{\Pi \succ 0} \; -\log\det\Pi + \log\det W$$

subject to

$$\begin{bmatrix} \Sigma - \Pi & \Sigma A' \\ A\Sigma & W + A\Sigma A' \end{bmatrix} \succeq 0. \tag{B.6}$$

Since the objective function is convex and the constraint is a linear matrix inequality, $c(\Sigma)$ is convex in Σ and can be added to the CVX atom library. We then transform Problem 7 into the following DCP:

$$\min_{\Sigma \succ 0} \quad (1-\beta)\operatorname{tr}\left(A'PA\Sigma\right) + \operatorname{tr}(\Omega\Sigma) \tag{B.7}$$

subject to (26) and

$$c(\Sigma) \leq 2\kappa.$$

For tracking problems, the term $(1-\beta)\operatorname{tr}(A'PA\Sigma)$ does not appear in (B.7). We have used this method to numerically solve the pricing example in Section 4.

In an earlier version of our paper, we solve the following inverse problem as in the rate-distortion theory in the engineering literature:

$$R(D) \equiv \min_{\Sigma \succ 0} \quad \frac{1}{2}\log\det\left(A\Sigma A' + W\right) - \frac{1}{2}\log\det(\Sigma) \tag{B.8}$$

subject to (26) and

$$(1-\beta)\operatorname{tr}\left(A'PA\Sigma\right) + \operatorname{tr}(\Omega\Sigma) \leq D.$$

The function $R(D)$ is decreasing and convex in D. Given any capacity $\kappa > 0$, we can find D using this function and then solve the corresponding Σ. The earlier version of our paper also derives results similar to Propositions 4 and 5. We omit the details here.

C The Case of $W \succeq 0$

When $W \succeq 0$, we cannot apply the semidefinite program representation in Proposition 2 to solve the dynamic programming problem in (19) for the finite-horizon RI problem. We can impose a new assumption (the state transition matrix is invertible) and use a different representation.

Proposition 7 *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 value function $J_t(\Sigma_{t-1})$ is convex in Σ_{t-1} for $t = 0, 1, ..., T$ and satisfies the dynamic semidefinite program:*

$$\begin{aligned} J_t(\Sigma_{t-1}) &= \min_{\Psi_t \succ 0, \Sigma_t \succ 0} \operatorname{tr}(\Omega_t \Sigma_t) - \frac{\lambda}{2}(1-\beta)\log\det(\Sigma_t) \\ &\quad + \frac{\lambda\beta}{2}(2\log|\det A_t| - \log\det\Psi_t) + \beta J_{t+1}(\Sigma_t) \end{aligned} \tag{C.1}$$

subject to (16) for $t \geq 1$ and (17) for $t = 0$, and

$$\begin{bmatrix} I - \Psi_t & M_t' \\ M_t & A_t\Sigma_t A_t' + W_t \end{bmatrix} \succeq 0, \tag{C.2}$$

where $W_t = M_t M_t'$ with $M_t \succeq 0$. Moreover, $J_T(\Sigma_{T-1})$ satisfies (18).

Proof: We can apply the same proof for Proposition 2 to show that $J_t(\Sigma_{t-1})$ is convex using the Bellman equation (19). Now we derive a different semidefinite program representation. Since $W_t \succeq 0$, we have the decomposition $W_t = M_t M_t'$ with $M_t \succeq 0$. Since A_t is invertible, $A_t\Sigma_t A_t'$ is also invertible. Applying the matrix determinant lemma yields

$$\det\left(A_t\Sigma_t A_t' + W_t\right) = \det\left(I + M_t'\left(A_t\Sigma_t A_t'\right)^{-1} M_t\right)\det\left(A_t\Sigma_t A_t'\right).$$

Thus we have

$$\begin{aligned} & \log\det\left(A_t\Sigma_t A_t' + W_t\right) - \log\det(\Sigma_t) \\ = & -\log\det\left(I + M_t'\left(A_t\Sigma_t A_t'\right)^{-1} M_t\right)^{-1} + \log\det\left(A_t\Sigma_t A_t'\right) - \log\det(\Sigma_t) \\ = & -\log\det\left(I + M_t'\left(A_t\Sigma_t A_t'\right)^{-1} M_t\right)^{-1} + 2\log|\det A_t|. \end{aligned}$$

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

$$\min_{\Psi_t} \quad 2\log|\det A_t| - \log\det\Psi_t$$

subject to

$$0 \prec \Psi_t \preceq \left(I + M_t' \left(A_t \Sigma_t A_t'\right)^{-1} M_t\right)^{-1}. \tag{C.3}$$

Now use the matrix inversion lemma to get

$$\left(I + M_t' \left(A_t \Sigma_t A_t'\right)^{-1} M_t\right)^{-1} = I - M_t' \left(A_t \Sigma_t A_t' + M_t M_t'\right) M_t.$$

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

$$\begin{bmatrix} I - \Psi_t & M_t' \\ M_t & A_t \Sigma_t A_t' + W_t \end{bmatrix} \succeq 0. \tag{C.4}$$

In sum, we have shown that

$$\log\det\left(A_t \Sigma_t A_t' + W_t\right) = \min_{\Psi_t \succ 0} \quad 2\log|\det A_t| - \log\det\Psi_t + \log\det\left(\Sigma_t\right)$$

subject to (C.4). Substituting this equation into (19) yields the desired result. 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 $W \succeq 0$ and it is singular. So Proposition 2 does not apply. As long as A_p is invertible so that A is invertible, we can apply Proposition 7 to solve the model numerically.

For the infinite-horizon RI problem, we can take limits in Proposition 7 as $T \to \infty$ to obtain the infinite-horizon semidefinite program representation. We can then derive the steady-state solution for Σ_t as $t \to \infty$.

We can also modify Proposition 3 to derive the golden-rule solution for Problem 4.

Proposition 8 *Suppose that A is invertible, $\Omega \succeq 0$, and $W \succeq 0$. Then the golden-rule solution Σ for Problem 4 is the solution to the following semidefinite program:*

$$\min_{\Psi,\Sigma\succ 0} \quad (1-\beta)\operatorname{tr}\left(A'PA\Sigma\right) + \operatorname{tr}(\Omega\Sigma) - \frac{\lambda}{2}\log\det\Psi \tag{C.5}$$

subject to (26) and

$$\begin{bmatrix} I-\Psi & M' \\ M & A\Sigma A'+W \end{bmatrix} \succeq 0, \tag{C.6}$$

where $W = MM'$ with $M \succeq 0$.

Proof: Use the method in the proof of Proposition 7 to derive

$$\log\det\left(A\Sigma A'+W\right) = \min_{\Psi\succ 0} \quad 2\log|\det A| - \log\det\Psi + \log\det(\Sigma)$$

subject to (C.6). Substituting above equation into (27) and eliminating the constant term $2\log|\det A|$, we obtain the desired result. □

Finally, we can apply Proposition 8 by removing the term $(1-\beta)\operatorname{tr}(A'PA\Sigma)$ to derive the golden rule Σ for the infinite-horizon tracking problem under RI in Section 4.

Notice that the above two propositions can be applied to solve models with ARMA(p,q) processes as shown in Section 5.1.2 as long as we can derive a state space representation.

D Infinite-Horizon Case

We study the following infinite-horizon problem with discounted information costs at time 1:

$$\min_{\{\Sigma_t\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} \beta^{t-1}\left[\operatorname{tr}(\Omega\Sigma_t) + \frac{\lambda}{2}\left(\log\det\left(A\Sigma_{t-1}A'+W\right) - \log\det\Sigma_t\right)\right] \tag{D.1}$$

subject to

$$\Sigma_t \preceq A\Sigma_{t-1}A'+W, \; t=1,2,..., \; \Sigma_0 \text{ given.} \tag{D.2}$$

Define the value function as $\mathcal{V}(\Sigma_0)$. By the dynamic programming principle (Stokey and Lucas with Prescott (1989) and Miao (2014)), it satisfies the Bellman equation

$$\mathcal{V}(\Sigma_0) = \min_{\Sigma\in\Gamma(\Sigma_0)} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\log\det\left(A\Sigma_0A'+W\right) - \log\det\Sigma\right] + \beta\mathcal{V}(\Sigma),$$

where

$$\Gamma(\Sigma_0) \equiv \left\{\Sigma \succ 0 : \Sigma \preceq A\Sigma_0A'+W\right\}. \tag{D.3}$$

To convert this problem into a DCP, we study an auxiliary problem. Define

$$J(\Sigma_0) \equiv \mathcal{V}(\Sigma_0) - \frac{\lambda}{2}\log\det\left(A\Sigma_0A'+W\right).$$

Then it satisfies the Bellman equation:

$$J(\Sigma_0) = \min_{\Sigma \in \Gamma(\Sigma_0)} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta \log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma)\right] + \beta J(\Sigma). \tag{D.4}$$

Let $\Sigma = h(\Sigma_0)$ be the associated optimal policy function. The policy function h generates a sequence of optimal covariance matrices $\{\Sigma_t\}_{t=1}^{\infty}$ through $\Sigma_t = h(\Sigma_{t-1})$, $t \geq 1$. Notice that the above problem is not a bounded discounted dynamic programming problem. We use the method of successive approximations (value function iteration) to analyze it.

Define the value function

$$f_0(\Sigma_0) \equiv \min_{\Sigma \in \Gamma(\Sigma_0)} \operatorname{tr}(\Omega\Sigma) - \frac{\lambda}{2}\log\det(\Sigma). \tag{D.5}$$

Because the constraint set in (D.3) is convex and the log-determinant function is strictly concave, the problem in (D.5) is a convex program and hence $f_0(\Sigma_0)$ is also strictly convex.

Define the Bellman operator $\mathbf{B}$ on the set of functions of positive semidefinite matrices:

$$\mathbf{B}(f)(\Sigma_0) \equiv \min_{\Sigma \in \Gamma(\Sigma_0)} \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta \log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma)\right] + \beta f(\Sigma). \tag{D.6}$$

Iterating this operator, we can construct a sequence of functions:

$$f_k(\Sigma_0) = \mathbf{B}^k(f_0)(\Sigma_0), \quad k \geq 1.$$

By induction and Lemma 2, each function $f_k(\cdot)$ is strictly convex and is obtained by solving a DCP problem. Let the corresponding optimal policy function be $\Sigma = h_k(\Sigma_0)$.

Say a sequence of matrices $\{\Sigma_t\}_{t=1}^{\infty}$ is feasible if $\Sigma_t \in \Gamma(\Sigma_{t-1})$ for each $t \geq 1$.

Proposition 9 *Suppose that $\Omega \succeq 0$. For any $\Sigma_0 \succ 0$, if there is a feasible sequence of matrices $\{\Sigma_t\}_{t=1}^{\infty}$ such that the objective in (D.1) is finite, then $f_k(\Sigma_0)$ increases monotonically to a finite limit function $J(\Sigma_0)$ as $k \to \infty$, which satisfies (D.4). Moreover, $h_k(\Sigma_0)$ converges to $h(\Sigma_0)$ pointwise on any compact set as $k \to \infty$.*

Proof: We first show that $f_1(\Sigma_0) \geq f_0(\Sigma_0)$. For any $\Sigma \in \Gamma(\Sigma_0)$, let $\Sigma^* \in \Gamma(\Sigma)$ be the optimal solution that attains the value $f_0(\Sigma)$. Then since $\Sigma^* \preceq A\Sigma A' + W$, we have

$$\log\det\left(A\Sigma A' + W\right) \geq \log\det(\Sigma^*).$$

It follows that

$$\begin{aligned}
& \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta \log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma)\right] + \beta f_0(\Sigma) \qquad \text{(D.7)}\\
= \;& \operatorname{tr}(\Omega\Sigma) + \frac{\lambda}{2}\left[\beta \log\det\left(A\Sigma A' + W\right) - \log\det(\Sigma)\right] \\
& + \beta\left[\operatorname{tr}(\Omega\Sigma^*) - \frac{\lambda}{2}\log\det(\Sigma^*)\right] \\
\geq \;& \operatorname{tr}(\Omega\Sigma) - \frac{\lambda}{2}\log\det(\Sigma) \geq f_0(\Sigma_0),
\end{aligned}$$

where we have used the fact that $\mathrm{tr}\left(\Omega\Sigma^{*}\right) \geq 0$ as $\Omega \succeq 0$ and $\Sigma^{*} \succ 0$. Minimizing the expression on the first line of (D.7) over $\Sigma \in \Gamma\left(\Sigma_0\right)$ yields $f_1\left(\Sigma_0\right) \geq f_0\left(\Sigma_0\right).$

It is easy to see that $\mathbf{B}\left(f\right) \geq \mathbf{B}\left(g\right),$ if $f \geq g$. Thus we can show that $f_{k+1}\left(\Sigma_0\right) \geq f_k\left(\Sigma_0\right)$ by induction. By assumption, for any $\Sigma_0 \succ 0$, there is a feasible sequence of matrices $\{\Sigma_t\}_{t=1}^{\infty}$ such that the objective in (D.1) is finite. Thus the increasing sequence $\{f_k\left(\Sigma_0\right)\}$ is bounded above and has a finite limit. Let the limit function be $J\left(\Sigma_0\right).$ To show J satisfies (D.4), notice that

$$f_k\left(\Sigma_0\right) = \mathbf{B}\left(f_{k-1}\right)\left(\Sigma_0\right) \leq \mathbf{B}\left(J\right)\left(\Sigma_0\right).$$

On the other hand,

$$J\left(\Sigma_0\right) \geq f_k\left(\Sigma_0\right) = \mathbf{B}\left(f_{k-1}\right)\left(\Sigma_0\right).$$

Taking limits on the above two inequalities yields $J\left(\Sigma_0\right) = \mathbf{B}\left(J\right)\left(\Sigma_0\right).$

By induction, each function $f_k\left(\Sigma_0\right)$ is strictly convex and hence the policy function h_k is unique. The limit function J is convex. Since $J = \mathbf{B}\left(J\right)$ and the objective function in (D.4) is strictly convex, J is also strictly convex. Thus the policy function h is also unique. Since f_k is continuous, $f_k\left(\Sigma_0\right)$ converges to $J\left(\Sigma_0\right)$ uniformly on any compact set. By Theorem 3.8 of Stokey and Lucas with Prescott (1989), $h_k\left(\Sigma_0\right)$ converges to $h\left(\Sigma_0\right)$ pointwise. □

Sims (2011) suggests to use the first-order conditions to solve (D.1). If the no-forgetting constraints (D.2) do not bind, it is straightforward to derive the first-order conditions. He mentioned to use dynamic programming and Cholesky decomposition when the no-forgetting constraints bind, but he does not provide a detailed formal analysis. We now present the first-order conditions for (D.1) by taking into account of binding constraints from convex analysis.

Notice that the choice variable Σ_t is in the set of $n_x \times n_x$ positive semidefinite matrices, denoted by $\mathcal{S}_{+}^{n_x}$. Define the inner product $\bullet$ for any two elements X and Y in this set as $X \bullet Y = \mathrm{tr}\left(X'Y\right).$ The dual cone of $\mathcal{S}_{+}^{n_x}$ is itself. Introduce a slack variable $Z_t \succeq 0$ such that $A\Sigma_{t-1}A' + W = Z_t + \Sigma_t$. Following Vandenberghe and Boyd (1996) and Vandenberghe, Boyd, and Wu (1998), we define the Lagrangian for (D.1) as

$$\begin{aligned} \mathcal{L} &= \max_{\{\Lambda_t\}} \min_{\{\Sigma_t\},\{Z_t\}} \sum_{t=1}^{\infty} \beta^{t-1} \left[\mathrm{tr}\left(\Omega\Sigma_t\right) + \frac{\lambda}{2}\left(\log\det\left(A\Sigma_{t-1}A' + W\right) - \log\det\Sigma_t\right)\right] \\ &\quad + \sum_{t=1}^{\infty} \beta^{t-1} \Lambda_t \bullet \left(Z_t + \Sigma_t - A\Sigma_{t-1}A' - W\right), \end{aligned}$$

where the dual variable $\Lambda_t \in \mathcal{S}_{+}^{n_x}$ is the Lagrange multiplier associated with (D.2).

We can write the Kuhn-Tucker conditions as follows:

$$0 = \Omega - \frac{\lambda}{2}\Sigma_t^{-1} + \frac{\beta\lambda}{2} A\left(A\Sigma_t A' + W\right)^{-1} A' + \Lambda_t - \beta A'\Lambda_{t+1}A, \tag{D.8}$$

$$\Lambda_t \bullet Z_t = 0, \tag{D.9}$$

$$\Lambda_t \succeq 0, \quad Z_t \succeq 0, \tag{D.10}$$

for all $t \geq 1$. Equation (D.8) is the first-order condition for Σ_t. Equation (D.9) is the complementary slackness condition. Condition (D.10) requires $\Lambda_t, Z_t \in \mathcal{S}_+^{n_x}$. The system of an infinite sequence of these conditions is difficult to solve numerically. The key difficulty is that we need to ensure positive semidefiniteness of Λ_t and Z_t.

The mathematics and engineering literature has developed rigorous theories to study semidefinite programs. Efficient and robust software to numerically solve such problems is also publicly available on the internet. Most software uses the primal-dual interior point method. Here we introduce the basic idea of this method. The key step is to add a parametrized barrier function to the objective function in (D.1):

$$-\nu \sum_{t=1}^{\infty} \beta^{t-1} \log \det Z_t,\ \nu > 0.$$

This function ensures that Z_t is invertible. Introduce a similar barrier function to (D.1) for the dual variable Λ_t. Use the Newton method to solve the first-order conditions associated with the modified (D.1) for any $\nu > 0$. Taking limits as $\nu \to 0$, we obtain the solution for (D.1). Notice that the mathematics and engineering literature typically focuses on static semidefinite programs. Our approach is to transform problem (D.1) into a dynamic semidefinite programming form. At each time we solve a semidefinite program using CVX and iterate the Bellman operator until the value function converges.

E Equilibrium Sticky Prices

In this appendix we derive the equilibrium solution for the model in Section 5.1.2 and provide a numerical algorithm to solve the equilibrium. We focus on the long-run stationary equilibrium. Suppose that the equilibrium aggregate price level can be approximated by a stationary ARMA process: $p_t = \Psi(L)\,\epsilon_{at}$, where Ψ is given by

$$\Psi(z) = \frac{b_0 + b_1 z + b_2 z^2 + \cdots + b_m z^m}{1 - a_1 z - a_2 z^2 - \cdots - a_r z^r}. \tag{E.1}$$

We will solve for an equilibrium with $r \geq m + 1$.

As discussed in Section 5.1.2, we can construct a state space representation for firm j :

$$x_{jt} = A x_{j,t-1} + \eta_{jt}, \tag{E.2}$$

$$p_{jt}^* = G x_{jt},\ s_{jt} = C_j x_{jt} + v_{jt}, \tag{E.3}$$

where A and G are given in (44) and (45), and $\eta_{jt} = [\epsilon_{jt}, \epsilon_{at}, \epsilon_{at}, 0, ..., 0]'$ and v_{jt} are independent Gaussian white noise processes with covariance matrices W and V_j. Assume that v_{jt} satisfies

$\int_0^1 v_{jt}dj = 0$. Notice that $W \succeq 0$ and $V_{jt} \succ 0$ by our construction. In particular, the $(1,1)$ entry of W is σ_i^2, the $(2,2)$, $(2,3)$, $(3,2)$, and $(3,3)$ entries are σ_a^2, and all other entries are zero. We can easily check that W is singular and A is nonsingular.

We use the results in Appendix C to solve for the golden rule of information structure under RI. We then obtain the posterior covariance matrix Σ for x_{jt} and (C_j, V_j). Notice that Σ is the same for all firms j, but (C_j, V_j) may not be the same across firms because this pair is not uniquely determined. We focus only on symmetric equilibrium in which (C_j, V_j) is the same for all j. Thus we remove the subscript j.

The optimal price under RI for firm j is given by

$$p_{jt} = \mathbb{E}\left[p_{jt}^*|s_j^t\right] = G\mathbb{E}\left[x_{jt}|s_j^t\right] = G\widehat{x}_{jt}, \tag{E.4}$$

The Kalman filter gives

$$\widehat{x}_{jt} = (I - KC)\, A\widehat{x}_{j,t-1} + Ks_{jt}, \tag{E.5}$$

where the Kalman gain is given by

$$K = \left(A\Sigma A' + W\right) C' \left[C\left(A\Sigma A' + W\right) C' + V\right]^{-1}.$$

Using the matrix inversion lemma, we can show that

$$KC = \left(A\Sigma A' + W\right) C' \left[C\left(A\Sigma A' + W\right) C' + V\right]^{-1} C = I - \Sigma\left(A\Sigma A' + W\right)^{-1}, \tag{E.6}$$

which is independent of C and V.

Assume that all eigenvalues of $(I - KC)\,A$ lie in the unit circle. Using the lag operator L, we can rewrite (E.5) as

$$\widehat{x}_{jt} = X\,(L)\, s_{jt}, \tag{E.7}$$

where

$$X(z) \equiv [I - (I - KC)Az]^{-1}\, K,$$

and z is in the unit circle on the complex space. It follows from (E.3) and (E.7) that

$$\widehat{x}_{jt} = X(L)Cx_{jt} + X(L)v_{jt}.$$

Assuming that all eigenvalues of A are in the unit circle, we can rewrite (E.2) as

$$x_{jt} = (I - AL)^{-1}\eta_{jt}.$$

It follows from the preceding two equations that

$$\widehat{x}_{jt} = X(L)C(I - AL)^{-1}\eta_{jt} + X(L)v_{jt}.$$

Aggregating across j yields

$$\int_0^1 \widehat{x}_{jt} dj = X(L)C(I - AL)^{-1} M\epsilon_a, \tag{E.8}$$

where $M \equiv [0, 1, 1, 0, ..., 0]'$ is a $(r + 2)$-dimensional vector and we have used the assumptions

$$\int_0^1 v_{jt} dj = 0, \quad \int_0^1 \epsilon_{jt} dj = 0.$$

It follows from (E.4) and (E.8) that the aggregate price level satisfies

$$p_t = \int_0^1 p_{jt} dt = G \int_0^1 \widehat{x}_{jt} dj = GX(L)C(I - AL)^{-1} M\epsilon_a.$$

Given the conjectured form of the equilibrium aggregate price $p_t = \Psi(L)\,\epsilon_{at}$, we obtain the equilibrium condition:

$$\Psi(z) = GX(z)C(I - Az)^{-1} M, \tag{E.9}$$

where

$$X(z)C = [I - (I - KC)Az]^{-1} KC,$$

is independent of (C, V) by (E.6). Equation (E.9) is a functional equation for the coefficients $(a_1, a_2, ..., a_r, b_0, b_1, ..., b_m)$. The solution determines the equilibrium pricing function Ψ.

We use the following algorithm to solve for these coefficients.[25]

Step 0. Initialize $k \geq 2$. Let $\{z_1, ..., z_N\}$ be an evenly spaced grid on $(-1, 1)$ for some integer N.

Step 1. Given a positive integer k, set $r = k$ and $m = k-1$. Initialize the polynomial coefficients $c \equiv (a_1, a_2, ..., a_r, b_0, b_1, ..., b_m)$.

Step 2. Given r, m, and c, compute the values $\{\Psi(z_i)\}_{i=1}^N$, where $\Psi(z)$ is given by (E.1).

Step 3. Derive the state space representation in (E.2) and (E.3). Use Proposition 8 to derive the golden-rule solution to the RI problem to obtain (C, V) and Σ. This step can be implemented in CVX.

Step 4. Compute the updated pricing function values

$$\Psi^+(z_i) \equiv GX(z_i)C(I - Az_i)^{-1} M, \; i = 1, 2, ...N.$$

Find the updated polynomial coefficients $c^+ \equiv \left(a_1^+, a_2^+, ..., a_{r^+}^+, b_0^+, b_1^+, ..., b_{m^+}^+\right)$ such that the implied rational function $\Psi^+(z)$ fits the set of values $\{\Psi^+(z_i)\}_{i=1}^N$. Here r^+ and m^+ are the maximal integers such that $a_{r^+}^+ \neq 0$, $b_{m^+}^+ \neq 0$, $r^+ \leq k$, and $r^+ \geq m^+ + 1$.

Step 5. Set $c := c^+$, $r := r^+$, and $m := m^+$. Repeat Steps 2-4 until

$$\frac{\sqrt{\sum_{i=1}^N \left[\Psi^+(z_i) - \Psi(z_i)\right]^2}}{\sqrt{\sum_{i=1}^N \left[\Psi(z_i)\right]^2}} < \epsilon_1.$$

[25] We have applied the toolbox, Ztran, developed by Han, Tan, and Wu (2019).

for some prespecified tolerance level $\epsilon_1 > 0$.

Step 6. If there is no convergence in Step 5, set $k := k + 1$ and go to Step 1. Otherwise, let the solution obtained in Step 5 be $\Psi^*(z)$. Find a rational function $\hat{\Psi}(z)$ for an ARMA(r, m) process that fits the values $\{\Psi^*(z_i)\}_{i=1}^N$ without the upper bound k restriction on the orders r and m. Check whether the distance between the MA(∞) representations (or the impulse response functions) for the ARMA processes implied by $\hat{\Psi}(z)$ and $\Psi^*(z)$ is within some prespecified tolerance level $\epsilon_2 > 0$. If so, then stop; otherwise, set $k := k + 1$ and go to Step 1.

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.

CVX Research, Inc., 2011, CVX: Matlab Software for Disciplined Convex Programming, version 2.0. http://cvxr.com/cvx.

Dewan, Amuj, 2018, Costly Multidimensional Information, working paper, Column University.

Diggavi, Suhas N., and Thomas M. Cover, 2011, The Worst Additive Noise under a Covariance Constraint, *IEEE Transactions on Information Theory* 47, 3072–3081.

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.

Grant, Michael C., 2004, Disciplined Convex Programming. PhD thesis, Information Systems Laboratory, Department of Electrical Engineering. Stanford University.

Grant, Michael C., and Stephen P. Boyd, 2008, Graph Implementations for Nonsmooth Convex Programs. In *Recent Advances in Learning and Control* (a tribute to M. Vidyasagar), V. Blondel, S. Boyd, and H. Kimura (eds), pp. 95-110, Lecture Notes in Control and Information Sciences, Springer.

Grant, Michael C, Stephen P. Boyd, and Yinyu Ye, 2006, Disciplined Convex Programming. In *Global Optimization: From Theory to Implementation,* L. Liberti and N. Maculan (eds), pp. 155-210, in the book series *Nonconvex Optimization and its Applications*, Springer.

Hamilton, James D., 1994, *Time Series Analysis*, Princeton University Press, Princeton NJ.

Han, Zhao, Fei Tan, and Jieran Wu, 2019, Analytic Policy Function Iteration, Working Paper, St. Louis University

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

Hébert, Benjamin, and Michael Woodford, 2018, Rational Inattention in Continuous Time, working paper, Columbia University.

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

Kőszegi, Botand and Filip Matějka, 2019, Choice Simplification: A Theory of Mental Budgeting and Naive Diversification, forthcoming in *Quarterly Journal of Economics*.

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.

Stokey, Nancy L. and Robert E. Lucas Jr., with Edward C. Prescott, 1989, *Recursive Methods in Economic Dynamics*, Harvard University 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.

Miao, Jianjun, 2019, Multivariate LQG Control under Rational Inattention in Continuous Time, working paper, Boston University.

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.

Steiner, Jakub, Colin Stewart, and Filip Matějka, 2017, Rational Inattention Dynamics: Inertia and Delay in Decision-Making, *Econometrica* 85, 521-553.

Sturm, Jos F., 1999, Using SeDuMi 1.02, A MATLAB Toolbox for Optimization over Symmetric Cones. In *Optimization Methods and Software*, 11-12: 625-633. Special issue on Interior Point Methods (CD supplement with software).

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 and Boyd, 1996, Semidefinite Programming, *SIAM Review* 38, 49-95.

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.

Zhong, Weijie, 2019, Optimal Dynamic Information Acquisition, working paper, Columbia University.

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

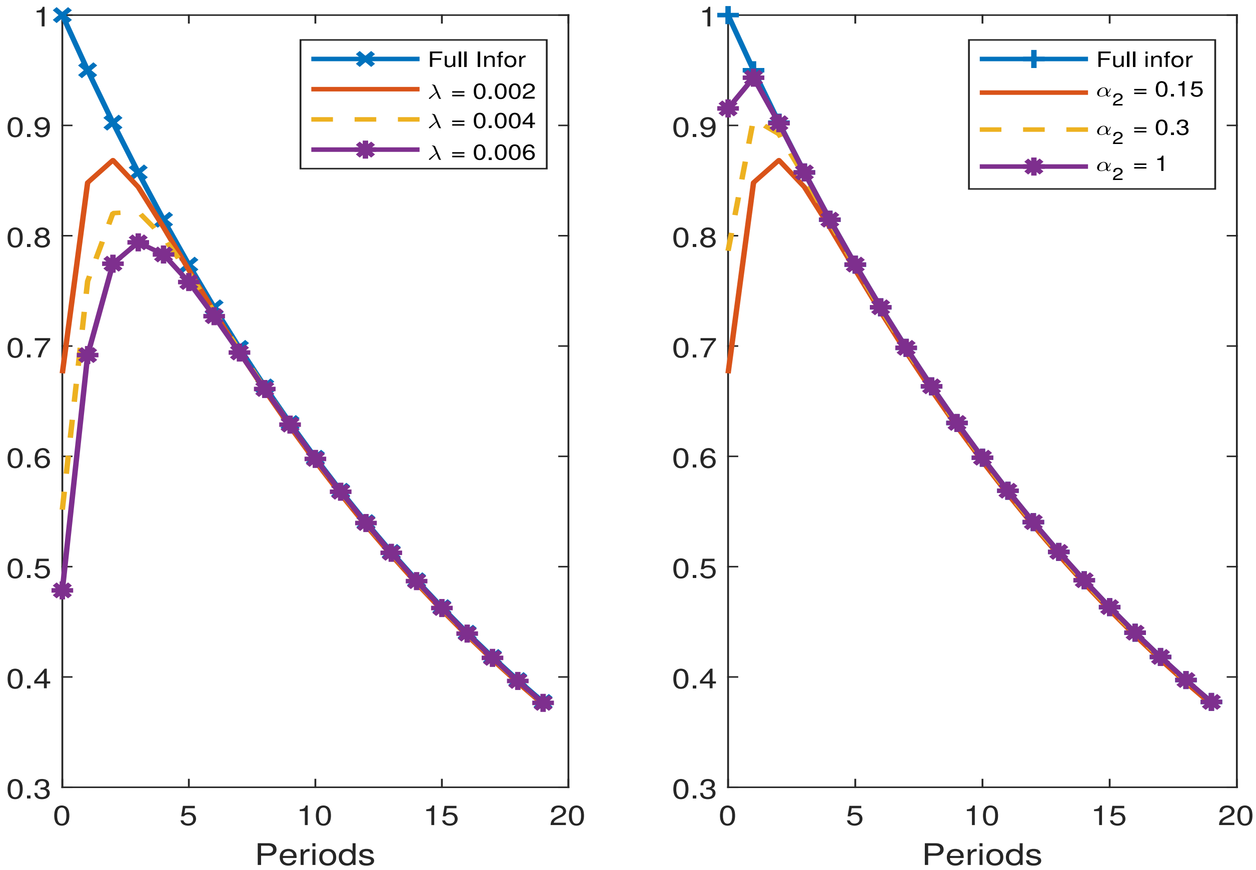

Figure 1: Impulse responses of the aggregate price to a one-standard-deviation innovation in nominal aggregate demand.

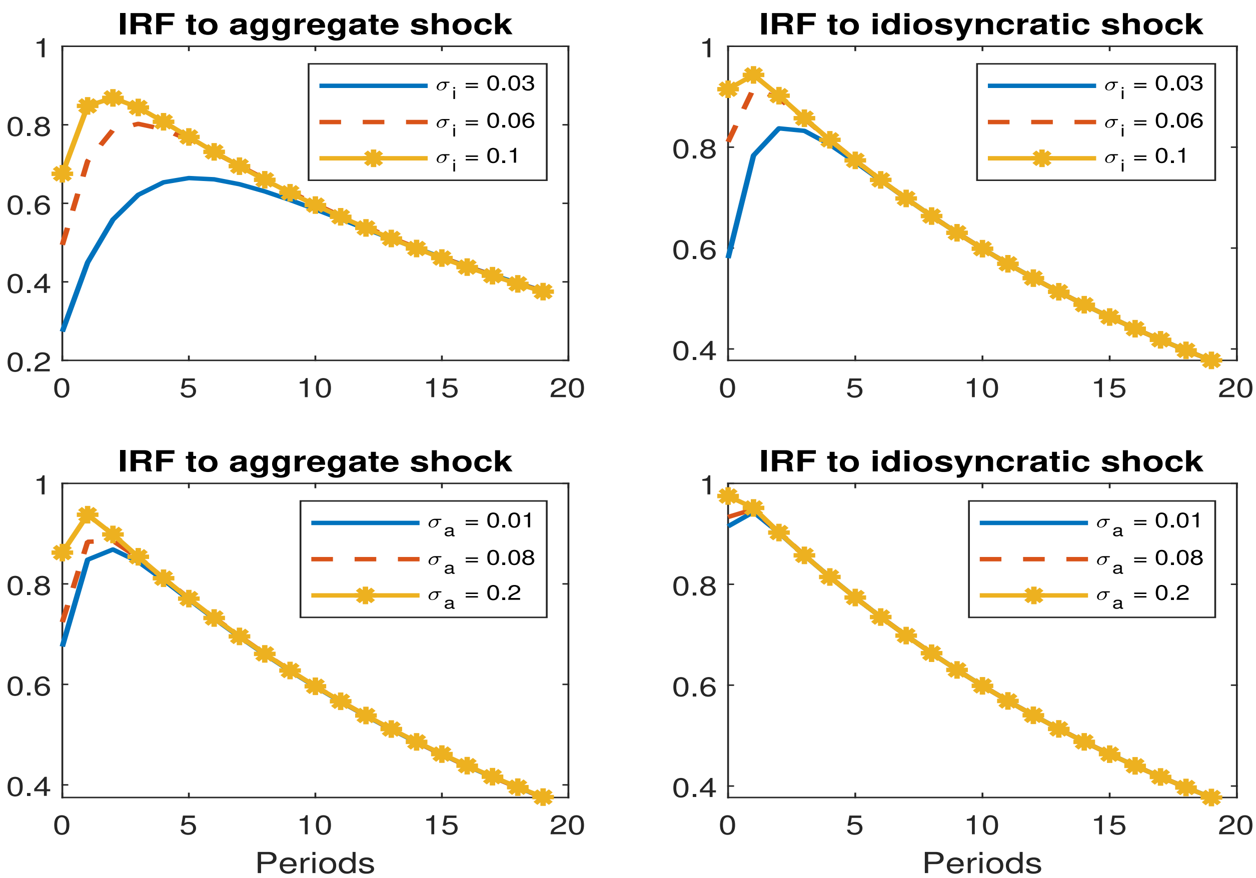

Figure 2: Impulse responses of the individual price to a one-standard-deviation innovation in nominal aggregate demand and idiosyncratic productivity for different innovation variances.

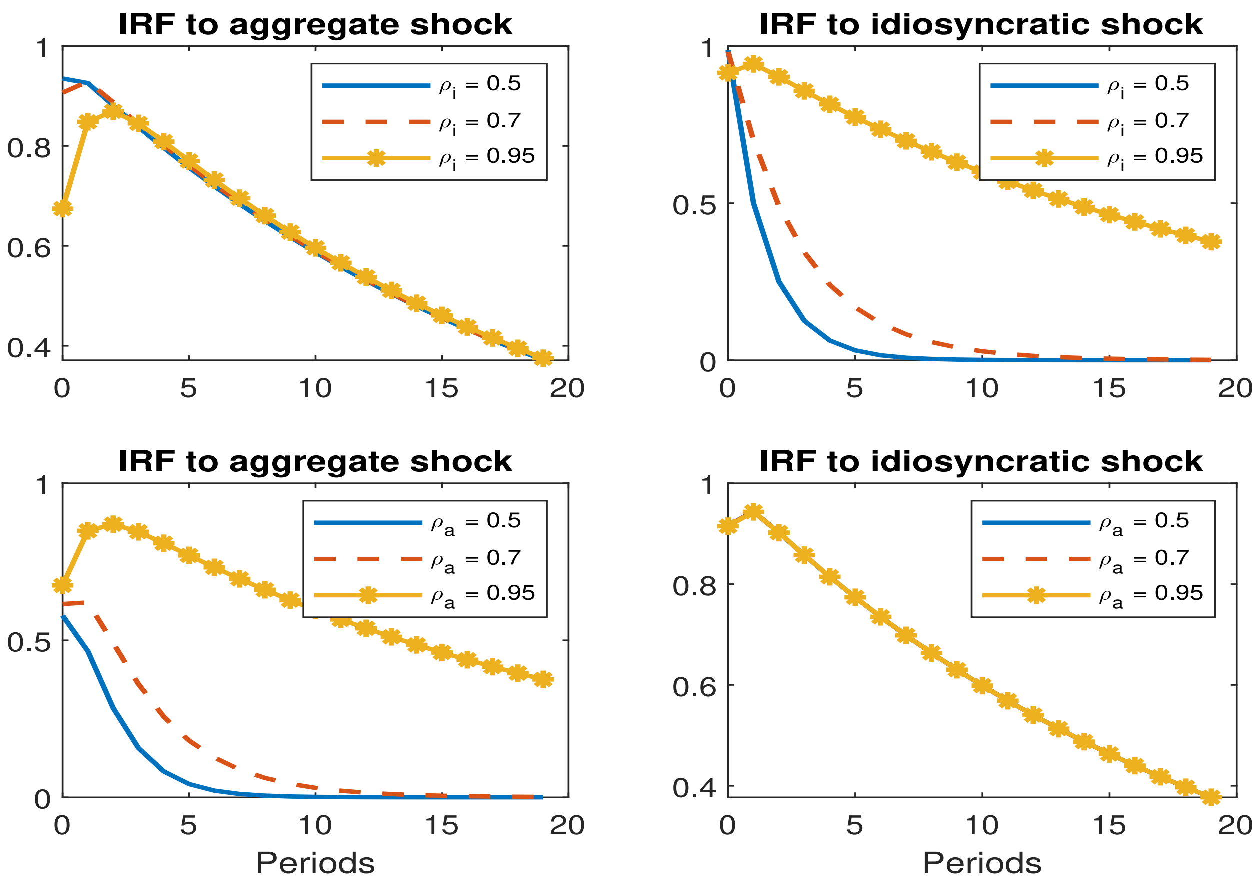

Figure 3: Impulse responses of the individual price to a one-standard-deviation innovation in nominal aggregate demand and idiosyncratic productivity for different persistence of shocks.

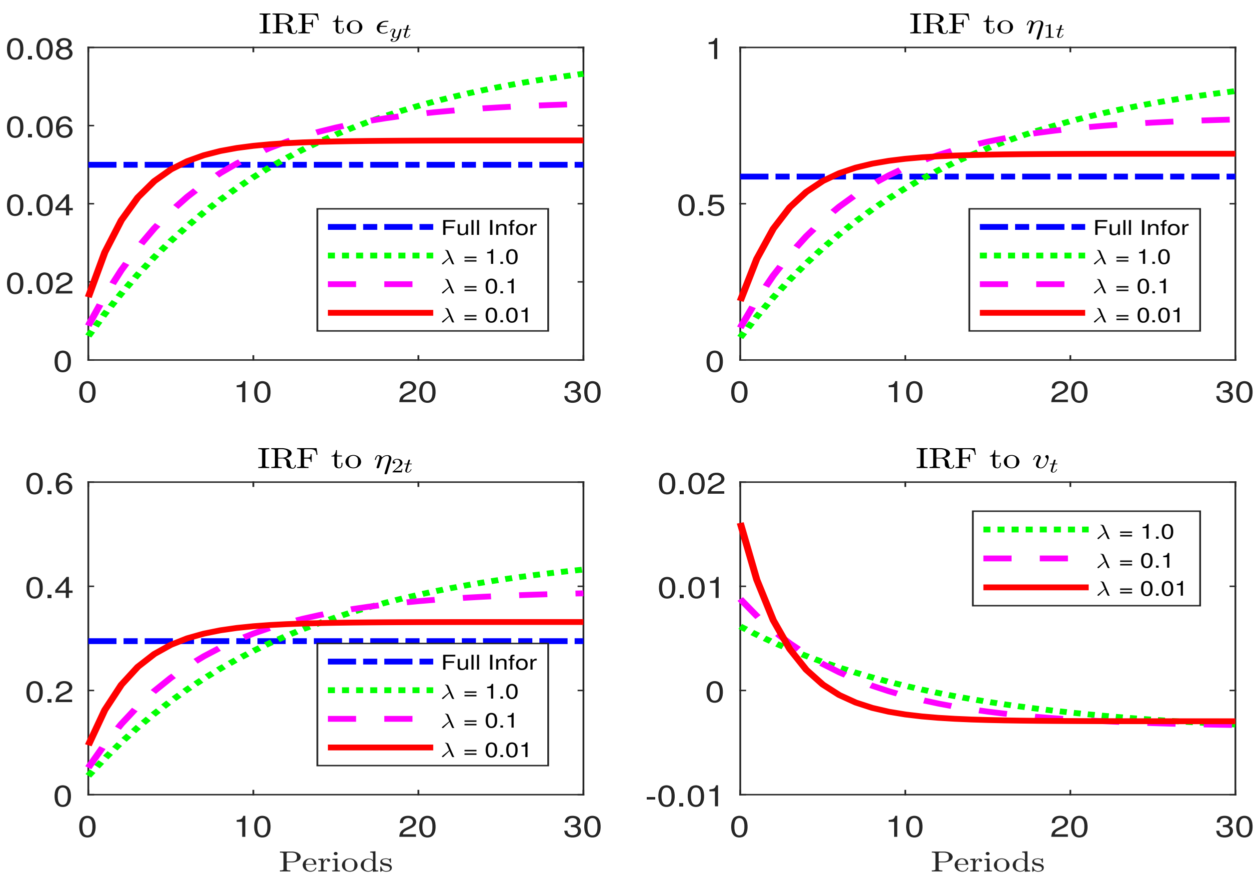

Figure 4: Impulse responses of consumption to a one-standard-deviation innovation in various shocks for different information cost parameter values.

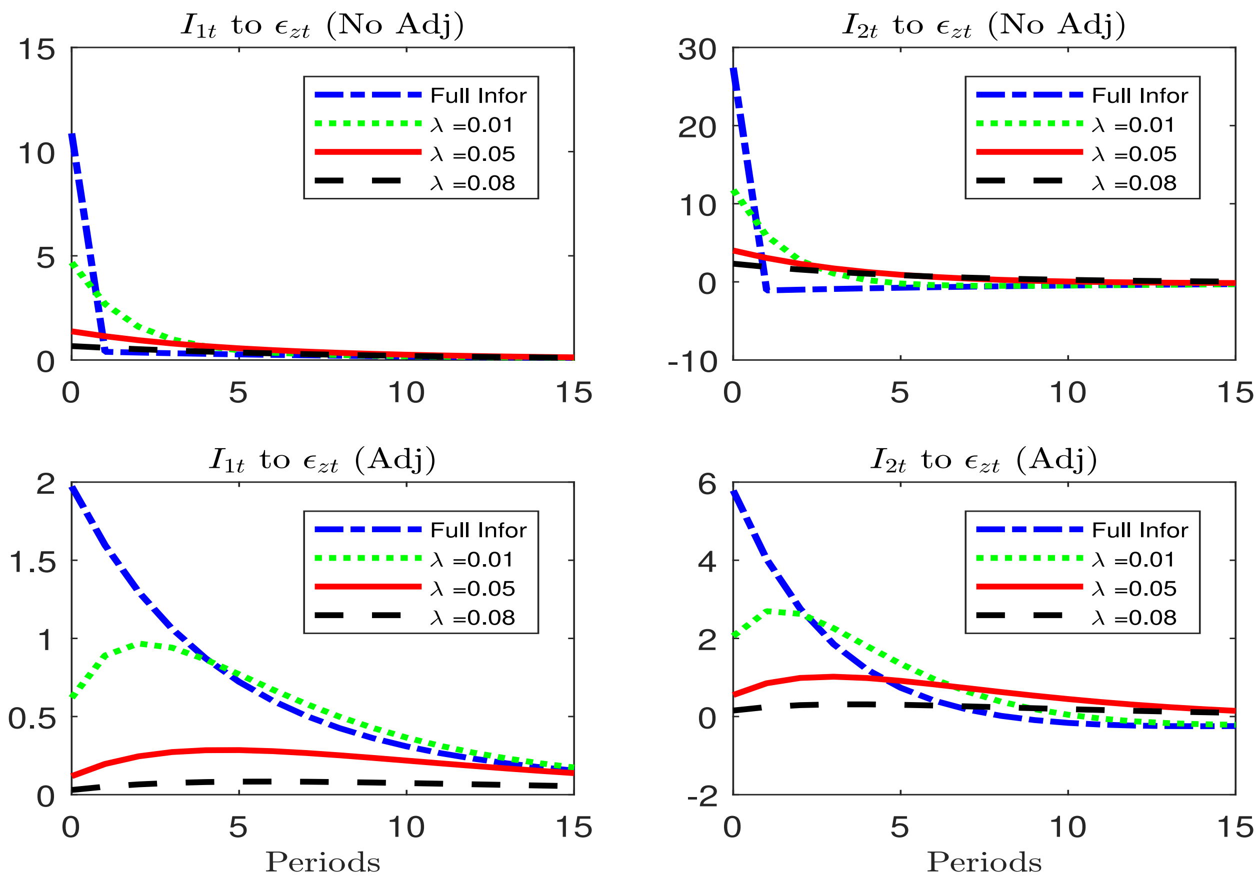

Figure 5: Impulse responses of tangible and intangible investment to a one-standard-deviation persistent TFP shock for different information costs.