Multivariate LQG Control under Rational Inattention in Continuous Time*

Jianjun Miao†

February 2019

Abstract

I propose a multivariate linear-quadratic-Gaussian control framework with rational inattention in continuous time. I propose a three-step solution procedure and numerical methods. The critical step is to transform the problem into a rate distortion problem and derive a semidefinite programming representation. I provide generalized reverse water-filling solutions for some special cases and characterize the optimal signal dimension. I apply my approach to study a consumption/saving problem and illustrate two pitfalls in the literature.

Keywords: Rational Inattention, Optimal Control, Semidefinite Programming, Consumption, Information Theory

JEL Classifications: C6, D8, E2.

*I thank Yulei Luo for very helpful discussions. I also thank Jieran Wu for helping me solve some numerical examples.

†Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215. Email: miaoj@bu.edu

1 Introduction

Economics is the study of how people choose to allocate scarce resources to maximize their objectives. Attention is a scarce resource that is useful for people to process information. Even though many economic data are publicly available, people are boundedly rational and only pay limited attention to some of them. Sims (1998, 2003) formalizes this idea as a problem of rational inattention (RI). He proposes a framework in which the decision maker solves an optimization problem subject to an information-flow constraint. Since Sims's seminal contributions, this literatures has grown rapidly as surveyed by Sims (2011) and Maćkowiak, Matějka, and Wiederholt (2018).

Despite the fact that Sims (1998) first introduces the idea of RI in a continuous-time setup, little progress has been made in this direction. Most of the research on RI has focused on either static or discrete-time models. Only a handful of papers discussed later study control problems with RI in continuous time and are limited to univariate models only. How to formulate and solve multivariate control problems with RI in continuous time is an open question. I attempt to resolve this question in the present paper, which makes three contributions to the literature. First, I propose a multivariate linear-quadratic-Gaussian (LQG) control framework with RI in continuous time. The LQG control framework has a long tradition in mathematics, engineering, and economics, and has wide applications especially in macroeconomics. The continuous-time setup is technically more convenient to derive analytical solutions. I formulate the LQG control problem under RI 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 Brownian motion noise. The signal dimension and the linear transformation are endogenously chosen subject to the information-flow constraint.

The second contribution of my paper is to provide three sets of novel characterization results. First, I derive a generalized reverse water-filling solution for the case in which the state transition matrix is diagonal with equal diagonal elements. Cover and Thomas (2006) present the classic reverse water-filling solution for the case of independent Gaussian shocks and unweighted mean-square errors in a static setting. Miao, Wu, and Young (2019) generalize this solution to a dynamic discrete-time setting with correlated Gaussian shocks and weighted mean-square errors. In this paper I generalize this solution to a continuous-time setting. Second, I characterize the signal dimension based on the reverse water-filling solution. I show that the signal dimension cannot exceed the rank of the weighting matrix derived from the control problem. This means that the signal dimension cannot exceed the minimum of the state dimension and the control dimension. As the distortion increases or the information-flow rate decreases, the signal dimension decreases if the positive eigenvalues of the weighted innovation covariance matrix are not identical. Third, I study a pure tracking problem in which the optimal control under full information is a linear combination of exogenous states. I show that the optimal signal is always one dimensional if the states follow correlated Ornstein-Uhlenbeck processes with the same persistence parameter. Moreover the signal can be normalized as the sum of the optimal control under full information and a Brownian motion noise, similar to the discrete-time model studied by Miao, Wu, and Young (2019).

The last contribution of my paper is to propose an efficient numerical method to solve the multivariate RI model in continuous time using semidefinite programming (Vandenberghe and Boyd (1998)). I propose a three-step solution procedure. The first step is to use the separation principle to solve the optimal control problem under partial information taken the information structure as given. The second step is to transform the problem of solving the optimal information structure into a pure tracking problem under RI, which is also called the distortion rate problem in the engineering literature. The last step is to transform this problem into an inverse rate distortion problem, which admits a semidefinite programming representation. This representation can be numerically solved using the publicly available semidefinite programming solver such as SDPT3 (Toh, Todd, and Tutuncu (1999) and Tutuncu, Toh, and Todd (2003)).[1] This solver can handle optimization problems up to 100 dimensions accurately, robustly, and efficiently.

I illustrate my approach using a consumption/saving problem as an example, in which there are two persistent income shocks and one transitory income shocks. Although there are three state variables in this model, I follow Luo (2008) to reduce it to the one with a single state variable – total wealth. I then derive an analytical solution. I use my solution procedure with three state variables to derive the numerical solution and compare with the analytical solution. I find they are almost the same subject to small numerical errors. I also find that the optimal signal is one dimensional and can be normalized as total wealth plus a Brownian motion noise.

I now discuss the related literature. As mentioned earlier, most of the literature on RI studies either static or discrete-time models.[2] Because of the difficulty of multivariate control models with RI, most papers analyze the univariate case or make strong assumptions on the signal structure. For example, Peng (2005), Peng and Xiong (2006), Maćkowiak and Wiederholt (2009), and Van Niewerburgh and Veldkamp (2010), and Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) impose the signal independence assumption or some restriction on the signal form. An undesirable implication of the signal independence assumption is that initially independent states remain ex post independent. Some ad hoc restrictions on the signal form may be inconsistent with optimality ex post. Sims (2003, 2011) proposes a two-step solution procedure in discrete time, but he does not discuss how to solve RI problems in continuous time. As discussed by Miao, Wu, and Young (2019), his procedure is flawed and inefficient in discrete time.

[1] This solver can be implemented in CVX (Grant and Boyd (2008)), a Matlab package that simplifies the construction of the problem. This package also supports other semidefinite programming solvers and solves general disciplined convex programs.

[2] Matějka and McKay (2015) and Caplin and Leahy (2018) study static RI models with discrete choices.

My paper is related to that by Miao, Wu, and Young (2019) who study multivariate RI problems in discrete time by removing the signal independence assumption or other strong assumptions on the signal structure. While Miao, Wu, and Young (2019) study both finite and infinite-horizon cases, here I only focus on the infinite-horizon stationary case. Although the solution procedure is similar, the continuous-time setup studied in this paper poses new technical challenges. First, one cannot simply formulate the signal form as the state plus noise in level in continuous time as noted by Sims (1998) and Moscarini (2004). This is because the implied information rate is infinity. One solution is to assume that the decision maker samples the process at discrete intervals (Moscarini (2004)). Instead I follow the standard literature on filtering in continuous time (Liptser and Shiryaev (2001)) and assume that the signal vector satisfies a stochastic differential equation. Second, the mutual information in continuous time takes a very different form from that in discrete time. In continuous time the asymptotic mutual information per unit of time is equal to one half of the mean-square error in estimating the unobserved state vector. In discrete time the asymptotic mutual information per unit of time is equal to the reduction in the conditional entropy per period. Thus the analysis of the rate distortion problem and the derivations of the generalized reverse water-filling solutions are different. Third, the semidefinite programming representations and the derivations in continuous time and in discrete time are very different. Thus numerical algorithms are also different.

Moscarini (2004), Kasa (2006), Luo (2018), and Hébert and Woodford (2018) also study RI problems in continuous time. Moscarini (2004) analyzes a pure tracking problem given discrete sampling. Kasa (2006) and Luo (2018) study univariate control problems. They also compare the implications of RI with those of robustness proposed by Hansen and Sargent (2007). Hébert and Woodford (2018) study a continuous-time RI model as the limit of a discrete-time sequential evidence accumulation problem. Unlike in my model, the actions in their model are discrete instead of continuous. Moreover they also study general information cost functions other than the mutual information.

2 Model

In this section I present the model setup, the standard solution under full information, and the first step of my solution procedure for the model with RI.

2.1 Setup

Consider an infinite-horizon stationary LQG model with RI in continuous time. Let the n_x dimensional state vector x_t follow the linear dynamics

$$dx_t = Ax_t dt + Bu_t dt + GdW_{xt}, \tag{1}$$

where u_t is an n_u dimensional control vector and (W_{xt}) is an n_x dimensional standard Brownian motion. The matrices A, B, and G are conformable constant matrices. Assume that the innovation covariance matrix GG' is positive definite, denoted by $GG' \succ 0$.[3] The state vector x_t may contain both exogenous states and endogenous states such as capital. The matrix A is called a state transition matrix.

Suppose that the decision maker does not observe the state x_t perfectly, but observes an n_y dimensional noisy signal y_t about x_t satisfying

$$dy_t = Cx_t dt + dW_{yt},\ y_0 = 0, \tag{2}$$

where C is a conformable constant matrix and (W_{yt}) is an n_y dimensional standard Brownian motion independent of (W_{xt}). Assume that x_0 is a Gaussian random variable with mean m_0 and covariance matrix Σ_0 and is independent of (W_{xt}) and (W_{yt}). Throughout the paper I focus on the stationary case in which x_0 is drawn from the long-run conditional stationary distribution. The decision maker's information set at date t is given by the σ-algebra $\mathcal{F}_t^y$ generated by $y^t \equiv \{y_s : 0 \le s \le t\}$. The control process (u_t) is adapted to the filtration $\{\mathcal{F}_t^y\}$.

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

$$\lim_{T\to\infty} \sup \frac{1}{T} I\left(x^T; y^T\right) \le \kappa, \tag{3}$$

where $\kappa > 0$ denotes the information-flow rate or channel capacity and $I\left(x^T; y^T\right)$ denotes the (Shannon) mutual information between the processes $x^T = \{x_t : 0 \le t \le T\}$ and $y^T = \{y_t : 0 \le t \le T\}$. The mutual information $I\left(x^T; y^T\right)$ measures total uncertainty reduction from time 0 to time T after observing the signal y^T. The expression on the left-hand side of inequality in (3) measures the average uncertainty reduction per unit of time over an infinite horizon.

The decision maker can process information by choosing the information structure represented by the linear transformation C. The choice of C also means that the signal dimension n_y is endogenously chosen. One may assume that C has full row rank, meaning that each component of the signal vector is not redundant. In principle the decision maker can choose both C and the noise variance as shown in Miao, Wu, and Young (2019) in the discrete time case. For example one may assume that the signal vector satisfies

$$dy_t = Cx_t dt + \Xi dW_{yt},\ y_0 = 0, \tag{4}$$

where Ξ is a conformable constant matrix. If Ξ is invertible, this signal vector gives the same information as that in (2) when both C and Ξ can be endogenously chosen. I will illustrate this point in Sections 3 and 4. Thus I focus on (2) only.

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

The decision maker first chooses the information structure C and then selects a control $\{u_t : t \geq 0\}$ adapted to $\{\mathcal{F}_t^y\}$ to maximize a quadratic function subject to the constraints described above. I formulate his decision problem as follows:

Problem 1 *(stationary LQG problem under RI with fixed capacity)*

$$\max_{\{u_t\},C} -\mathbb{E}\left[\int_0^\infty \beta e^{-\beta t}\left(x_t'Qx_t + u_t'Ru_t + 2x_t'Su_t\right)dt\right]$$

subject to (1), (2), and (3), where the expectation is taken with respect to the stationary distribution.

One may use the Lagrange multiplier to eliminate the information-flow constraint (3) and study the following relaxed problem:

Problem 2 *(stationary LQG problem under RI with fixed information cost)*

$$\max_{\{u_t\},C} -\mathbb{E}\left[\int_0^\infty \beta e^{-\beta t}\left(x_t'Qx_t + u_t'Ru_t + 2x_t'Su_t\right)dt\right] - \lambda \lim_{T\to\infty}\sup \frac{1}{T} I\left(x^T; y^T\right)$$

subject to (1) and (2), where the expectation is taken with respect to the stationary distribution. Here $\lambda > 0$ is interpreted as the shadow cost of information.

In the engineering literature one often studies the long-run average cost criterion instead of the discounted objective:

$$\lim_{T\to\infty}\sup \frac{1}{T}\int_0^T \left(x_t'Qx_t + u_t'Ru_t + 2x_t'Su_t\right)dt.$$

One can analyze this problem as the limiting case of the infinite-horizon discounted cost problem as the discount rate β vanishes. I will not study this problem in this paper. Instead I will focus on Problem 1 only. Problem 2 can be similarly analyzed and will be omitted.

2.2 Full Information Solution

Before analyzing Problem 1, I 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 textbook by Anderson and Moore (1989). Suppose that $R \succ 0$ and

$$\begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \succeq 0.$$

Then the value function given any initial value $x_0 = x$ takes the form

$$V^{FI}(x) = -\beta x'Px - tr\left(GG'P\right), \tag{5}$$

where $tr(\cdot)$ denotes the trace operator and $P \succeq 0$ satisfies the algebraic Riccati equation

$$\beta P = Q - (PB + S)R^{-1}\left(B'P + S'\right) + \left(A'P + PA\right). \tag{6}$$

Taking the expectation with respect to the initial distribution of x_0 gives

$$V^{FI} \equiv \mathbb{E}\left[V^{FI}\left(x_0\right)\right] = -\beta\mathbb{E}\left[x_0'Px_0\right] - tr\left(GG'P\right). \tag{7}$$

The optimal control is given by

$$u_t = -Fx_t, \tag{8}$$

where

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

Following Anderson and Moore (1989), assume that the pair $\left(A - BR^{-1}S' - \beta I/2, B\right)$ is stabilizable and the pair $\left(A - BR^{-1}S' - \beta I/2, L\right)$ is detectable for $LL' = Q - SR^{-1}S'$, where I is a conformable identity matrix. Then there exists a positive semidefinite solution P to (6) and this solution is the limit of the differential Riccati equation derived from a finite-horizon problem. To ensure the stability of the Kalman-Bucy filter discussed in the next subsection, I also assume that the pair (A, G) is controllable. Notice that all the preceding conditions are sufficient, but not necessary. Other sufficient conditions are also available in the literature.

2.3 Control under Partial Information

I now turn to the decision problem under RI. I use a three-step procedure similar to that of Miao, Wu, and Young (2019), which is also related to the method of Sims (2003). In the first step I take the information structure as given and solve the optimal control problem under partial information. In this case the separation principle holds. To apply this principle, I first derive the long-run stationary Kalman-Bucy filter

$$dm_t = Am_tdt + Bu_tdt + K\left(dy_t - Cm_tdt\right), \tag{9}$$

$$0 = A\Sigma + \Sigma A' - KK' + GG', \tag{10}$$

where I define $m_t \equiv \mathbb{E}\left[x_t|\mathcal{F}_t^y\right]$, $\Sigma \equiv \mathbb{E}\left[\left(x_t - m_t\right)\left(x_t - m_t\right)'\right]$, and the Kalman gain $K = \Sigma C'$. If the pair (A, G) is controllable and the pair (A, C) is detectable, then the algebraic Riccati equation (10) admits a unique positive definite solution Σ (Anderson and Moore (1989))

Following Liptser and Shiryaev (2001), I rewrite the objective function as

$$\begin{aligned} & -\mathbb{E}\left[\int_0^\infty \beta e^{-\beta t}\left(x_t'Qx_t + u_t'Ru_t + 2x_t'Su_t\right)dt\right] \\ = \quad & -tr\left(Q\Sigma\right) - \mathbb{E}\left[\int_0^\infty \beta e^{-\beta t}\left(m_t'Qm_t + u_t'Ru_t + 2m_t'Su_t\right)dt\right]. \end{aligned}$$

I then consider the auxiliary problem

$$J\left(m\right) = \max_{\{u_t\}} -E\left[\int_0^\infty \beta e^{-\beta t}\left(m_t'Qm_t + u_t'Ru_t + 2m_t'Su_t\right)dt\right] \tag{11}$$

subject to

$$dm_t = Am_t dt + Bu_t dt + Kd\overline{W}_{yt},\ m_0 = m,$$

where $\overline{W}_{yt}$ is the innovation Brownian motion relative to the filtration $\{\mathcal{F}_t^y\}$ and satisfies

$$d\overline{W}_{yt} = dy_t - Cm_t dt.$$

The problem in (11) can be viewed as a standard LQG control problem under full information $\{\mathcal{F}_t^y\}$. It follows from (5) that the optimal control is given by $u_t = -Fm_t$ and the value function is given by

$$J(m) = -\beta m' P m - tr\left(KK'P\right). \tag{12}$$

Thus the optimal value in Problem 1, denoted by $V(m_0)$, satisfies

$$V(m_0) = -\beta m_0' P m_0 - tr\left(KK'P\right) - tr\left(Q\Sigma\right). \tag{13}$$

Proposition 1 *Suppose that the assumptions in Section 2.1 are satisfied. Let $\Sigma_0 = \Sigma$ and C be given. Then the optimal value under partial information $V(m_0)$ satisfies*

$$V^{FI} - V(m_0) = \mathbb{E}\left[(x_t - m_t)' \Omega (x_t - m_t)\right],$$

where

$$\Omega \equiv (PB + S) R^{-1} \left(B'P + S'\right) \succeq 0. \tag{14}$$

The intuition behind this result is as follows. Since the information-flow constraint cannot help in the optimization, I know that $V^{FI} > V(m_0)$. The difference is equal to the weighted mean-square error with the weight given by Ω. In the next section I describe how I solve the optimal information structure in Problem 1 under RI.

3 Optimal Information Structure

In this section I present the remaining two steps of my solution procedure and provide some characterization results for some special cases.

3.1 Rate Distortion Function

Sims (2003) proposes to solve the optimal information structure by minimizing $V^{FI} - V(m_0)$. In other words, the optimal information structure is to bring expected utility from the current date onward as close as possible to the expected utility value under full information. This leads to the second step of his solution method, which is also the same as my second step.

Problem 3 *(Distortion rate problem)*

$$D(\kappa) \equiv \inf_{C \in \mathcal{C}} \quad \mathbb{E}\left[(x_t - m_t)' \Omega (x_t - m_t)\right]$$

subject to (2), (3), (9), and (10), where $\mathcal{C} \subset \mathbb{R}^{n_y \times n_x}$ *is the set of matrices* C *such that* (A, C) *is a detectable pair.*

According to the information theory in the engineering literature, the objective in the problem above is the distortion between the source process x_t and its estimate m_t measured by the weighted mean-square error. The function $D(\kappa)$ is called the distortion rate function which shows the minimal distortion given that the information flow measured by the mutual information is limited by the capacity κ. Notice that the restriction of the domain of optimization $\mathcal{C}$ is to ensure that there is a unique positive definite solution to (10). Sims (2003, 2011) advocates to solve Problem 3 using first-order conditions in a discrete time setup. Following the engineering literature on information theory, I propose to solve a closely related inverse problem, which leads to the last step of my solution procedure.

Problem 4 *(Rate distortion problem)*

$$\kappa(D) \equiv \inf_{C \in \mathcal{C}} \quad \lim_{T \to \infty} \sup \frac{1}{T} I\left(x^T; y^T\right) \tag{15}$$

subject to (2), (9), (10), and

$$\mathbb{E}\left[(x_t - m_t)' \Omega (x_t - m_t)\right] \leq D. \tag{16}$$

Here $D > 0$ denotes an upper bound of the distortion. The problem above means that the decision maker chooses the information structure to minimize the information-flow rate given the distortion is limited by the bound D. The function $\kappa(D)$ is called the rate distortion function in information theory. As in Cover and Thomas (2006), one can show that both $D(\kappa)$ and $\kappa(D)$ are convex and decreasing functions and $\kappa(D)$ is the inverse function of $D(\kappa)$. Thus the rate distortion problem and the distortion rate problem are equivalent. The rate distortion problem is more convenient to analyze than the distortion rate problem because the information-flow constraint (3) is more complicated than the distortion constraint (16). Another advantage of studying the rate distortion problem is that one can always find a solution for any $D > 0$. But the rate distortion problem may not have a solution for an arbitrary capacity $\kappa > 0$. I will illustrate this point in Sections 3.2 and 4. Thus I will focus on the rate distortion problem throughout this paper.

I now explicitly compute the mutual information in (15) and the distortion in (16). I first compute the weighted mean-square error

$$\mathbb{E}\left[(x_t - m_t)' \Omega (x_t - m_t)\right] = tr(\Omega\Sigma).$$

Next it follows from Duncan (1970) and Liptser and Shiryaev (2001) that

$$I\left(x^T; y^T\right) = I\left(x^T; m^T\right) = \frac{1}{2}\int_0^T \mathbb{E}\left[\left\|C\left(x_t - m_t\right)\right\|^2\right] dt,$$

where $\|\cdot\|$ denotes the Euclidean norm. Given that $\Sigma \equiv \mathbb{E}\left[\left(x_t - m_t\right)\left(x_t - m_t\right)'\right]$ in the stationary case, I derive

$$\lim_{T\to\infty} \sup \frac{1}{T} I\left(x^T; y^T\right) = \frac{1}{2} tr\left(\Sigma C'C\right). \tag{17}$$

The matrix $C'C$ is often called a signal-to-noise ratio in the engineering literature. The intuition for (17) can be best gained from the univariate case. Then the asymptotic mutual information contained in (y_t) about (x_t) per unit of time increases with the conditional variance of x_t and with the signal-to-noise ratio. The higher the conditional variance of the state or the signal-to-noise ratio, the more informative the signal is.

I then transform Problem 4 into a more transparent form.

Lemma 1 *The rate distortion problem in Problem 4 is equivalent to the following problem:*

$$\kappa(D) = \inf_{\Sigma \succ 0} \; tr(A) + \frac{1}{2} tr\left(\Sigma^{-1} G G'\right) \tag{18}$$

subject to

$$tr(\Omega\Sigma) \leq D, \tag{19}$$

$$A\Sigma + \Sigma A' + GG' \succeq 0. \tag{20}$$

In the problem above I have eliminated the choice variable C, which can be recovered through equation (10) with $K = \Sigma C'$ once Σ is obtained. Constraint (20) corresponds to the no-forgetting constraint discussed by Sims (2003) in the discrete-time setup. It comes from equation (10) and ensures that C can be recovered. Notice that the objective in (18) is an inverse function of Σ, while the objective in discrete time is a log-determinant function (Sims (2003) and Miao, Wu, and Young (2019)). Thus one has to use a different method to analyze the problem in Lemma 1.

In the appendix I derive the following semidefinite programming representation:

Proposition 2 *Suppose that the assumptions in Section 2.1 are satisfied. Then the rate distortion problem has the semidefinite programming representation:*

$$\kappa(D) = \min_{\Pi \succeq 0, \Sigma \succ 0} \; tr(A) + \frac{1}{2} tr(\Pi)$$

subject to (19), (20), and

$$\begin{bmatrix} \Pi & G' \\ G & \Sigma \end{bmatrix} \succeq 0. \tag{21}$$

This problem admits an optimal solution $\Pi \succeq 0$ *and* $\Sigma \succ 0$.

Constraint (19) is a linear inequality. Constraints (20) and (21) are linear matrix inequalities. The problem in the proposition is convex and can be numerically solved efficiently using the publicly available package CVX. Once the conditional covariance matrix Σ is determined, the optimal information structure C is pinned down by the following result.

Proposition 3 *Let $\Sigma \succ 0$ be the solution obtained in Proposition 2. Then any matrix C that satisfies*

$$C'C = \Sigma^{-1}\left(A\Sigma + \Sigma A' + GG'\right)\Sigma^{-1} \tag{22}$$

gives the optimal information structure for the rate distortion problem. One solution is $C = MU_1'$, where M is a diagonal matrix with the square roots of all positive eigenvalues of the matrix $\Sigma^{-1}(A\Sigma + \Sigma A' + GG')\Sigma^{-1}$ on the diagonal and U_1 is a matrix with all corresponding eigenvectors on the columns.

Propositions 2 and 3 provide an efficient numerical method to solve the optimal information structure in Problem 1. In particular, given any feasible information rate $\kappa > 0$. I use Proposition 2 to find a unique distortion D such that $\kappa(D) = \kappa$ and the corresponding conditional covariance matrix Σ. Proposition 3 delivers the optimal information structure C, which can be numerically solved using the singular value decomposition and is not unique. Even though C is not unique, it follows from the Kalman-Bucy filter in (9) and (10) that the impulse response functions for the fundamental shock W_{xt} do not change. But the impulse responses to the information-processing error W_{yt} will be affected. For example, if C is a solution, then $-C$ is also a solution so that the impulse responses to the same shock W_{yt} have an opposite sign for these two solutions. Notice that the non-uniqueness of C does not affect the conditional covariance matrix Σ of the state x_t.

3.2 Generalized Reverse Water-filling Solution

In this subsection I derive some analytical results for some special cases. For convenience I treat Ω as exogenous and directly impose assumptions on it whenever necessary. Let $GG' = \Psi \succ 0$ and let $\Psi^{\frac{1}{2}}$ denote the square root of the positive definite matrix Ψ. Suppose that the positive semidefinite matrix $\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}}$ has an eigendecomposition $\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}} = U\Omega_d U'$, where U is an orthonormal matrix, $\Omega_d = diag\,(d_i)_{i=1}^{n_x}$, and $d_i \geq 0$ denotes an eigenvalue of $\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}}$ for $i = 1, 2..., n_x$. I then have the following result.

Proposition 4 *Suppose that $\Omega \succeq 0, \Psi \succ 0$, and $A = -\rho I$, with $\rho > 0$. If $0 < D < \frac{tr(\Omega_d)}{2\rho}$, then the solution to the rate distortion problem in Lemma 1 is given by $\Sigma = \Psi^{\frac{1}{2}}U\widehat{\Sigma}U'\Psi^{\frac{1}{2}}$, where $\widehat{\Sigma} = diag\left(\widehat{\Sigma}_i\right)_{i=1}^{n_x}$ with*

$$\widehat{\Sigma}_i = \min\left\{\frac{1}{\sqrt{\alpha d_i}}, \frac{1}{2\rho}\right\},$$

and $\alpha > 0$ is the Lagrange multiplier such that $tr\left(\Omega_d \widehat{\Sigma}\right) = D$. The optimal information structure C satisfies

$$C'C = \Psi^{-\frac{1}{2}} U \ \widehat{\Sigma}^{-1} diag\left(\max\left\{1 - \frac{2\rho}{\sqrt{\alpha d_i}}, 0\right\}\right)_{i=1}^{n_x} \ \widehat{\Sigma}^{-1} U' \Psi^{-\frac{1}{2}}. \tag{23}$$

The rate distortion function is given by

$$\kappa(D) = \frac{1}{2} tr\left(\widehat{\Sigma}^{-1}\right) - \rho n_x.$$

If $D \geq \frac{tr(\Omega_d)}{2\rho}$, then $\widehat{\Sigma}_i = \frac{1}{2\rho}$ for all i and $\Sigma = \frac{\Psi}{2\rho}$.

When $A = -\rho I$ with $\rho > 0$, the state vector x_t follows an Ornstein–Uhlenbeck process conditional on a control u_t. Its stationary distribution is Gaussian with covariance matrix $\frac{\Psi}{2\rho}$. Proposition 4 establishes that if the distortion D exceeds a threshold, the decision maker does not process any information so that the conditional covariance matrix Σ is the same as the prior covariance matrix $\frac{\Psi}{2\rho}$. If the distortion D is below the threshold, the decision maker acquires information to reduce uncertainty. To understand how uncertainty is reduced, I consider two special cases. First, in the scalar case I explicitly derive the following result:

Corollary 1 *Consider the scalar case with $n_x = 1$, $\Omega = 1$, $\Psi = \sigma^2$, and $A = -\rho$. If $\rho > 0$, then*

$$\Sigma = \min\left\{D, \ \frac{\sigma^2}{2\rho}\right\}, \ \kappa(D) = \max\left\{\frac{\sigma^2}{2D} - \rho, 0\right\}.$$

If $\rho < 0$, then $\Sigma = D$ and

$$\kappa(D) = \frac{\sigma^2}{2D} - \rho > -\rho.$$

In both cases an optimal information structure is given by

$$C = \frac{1}{\Sigma}\sqrt{\sigma^2 - 2\rho\Sigma}.$$

In the scalar case described in Corollary 1, the conditional variance Σ is bounded above by the distortion D. The no-forgetting constraint implies that $\sigma^2 \geq 2\rho\Sigma$. If $\rho > 0$ and $0 < D < \frac{\sigma^2}{2\rho}$, the uncertainty is reduced from the prior stationary variance $\frac{\sigma^2}{2\rho}$ to the conditional variance D. But if $\rho < 0$ then the no-forgetting constraint does not bind and the conditional variance is equal to D. Moreover the rate $\kappa(D)$ must be larger than $-\rho > 0$.

[Insert Figure 1.]

Figure 1 illustrates the rate distortion functions for the cases of $\rho > 0$ and $\rho < 0$. If the capacity $\kappa > 0$ is specified as a primitive as in the information-flow constraint, then Corollary 1 shows that

$$\Sigma = \frac{\sigma^2}{2(\kappa + \rho)} \tag{24}$$

for $\rho > 0$. The solution above still holds for $\rho < 0$, but I need $\kappa > -\rho$. If one specifies $\kappa \in (0, -\rho]$, then the distortion rate problem has no solution. The intuition is as follows. The scalar state process x_t is nonstationary for $\rho < 0$. If the capacity is too low, it is impossible for the decision maker to acquire sufficient information to make the estimated state process stationary. That is, the Kalman-Bucy filter (10) cannot hold. Solving the rate distortion problem allows one to calibrate κ in the sensible region. I will revisit this point in Section 4.

Next I consider the multivariate case with independent shocks and unweighted mean-square error objective.

Corollary 2 *Suppose that $\Omega = I$, $\Psi = diag\left(\sigma_i^2\right)_{i=1}^{n_x} \succ 0$, and $A = -\rho I$, with $\rho > 0$. If $0 < D < \sum_{i=1}^{n_x} \frac{\sigma_i^2}{2\rho}$, then $\Sigma = diag\left(\Sigma_i\right)_{i=1}^{n_x}$ and*

$$\Sigma_i = \min\left\{\frac{\sigma_i}{\sqrt{\alpha}}, \frac{\sigma_i^2}{2\rho}\right\},$$

where $\alpha > 0$ is such that $tr\left(\Sigma\right) = D$. The rate distortion function is given by

$$\kappa\left(D\right) = \frac{1}{2}\sum_{i=1}^{n_x} \max\left\{\sigma_i\sqrt{\alpha} - 2\rho, 0\right\}.$$

The optimal information structure satisfies

$$C'C = diag\left(\frac{1}{\Sigma_i^2}\max\left\{1 - \frac{2\rho}{\sqrt{\alpha d_i}}, 0\right\}\right)_{i=1}^{n_x}. \tag{25}$$

If $D \geq \sum_{i=1}^{n_x} \frac{\sigma_i^2}{2\rho}$, then $\Sigma = \frac{\Psi}{2\rho}$.

Corollary 2 corresponds to the reverse water-filling solution in the static case discussed by Cover and Thomas (2006) and in the dynamic discrete time case studied by Miao, Wu, and Young (2019). Corollary 2 shows that if the states are independent and the objective is unweighted, then the states remain independent ex post. The conditional variances of the states are reduced according to a decreasing order. For the state with prior innovation volatility σ_i higher than $2\rho/\sqrt{\alpha}$, the uncertainty is reduced to $\Sigma_i = \sigma_i/\sqrt{\alpha} < \sigma_i^2/\left(2\rho\right)$. For any state with prior innovation volatility lower than that level, the state variance remains unchanged ex post. Intuitively the decision maker pays attention only to states with sufficiently high innovation variances and acquires information to reduce the uncertainty about those states.

For the general case described in Proposition 4, a similar reverse water-filling solution applies. The decision maker pays attention to the high eigenvalues of the weighted prior covariance matrix. The matrix $\widehat{\Sigma}$ can be interpreted as a scaling factor. The decision maker acquires information to reduce uncertainty for sufficiently high eigenvalues. Moreover even though the states are ex ante uncorrelated (Ψ is diagonal), the states can be correlated ex post when Ω is not a diagonal matrix.

The result below characterizes the signal structure for Proposition 4.

Proposition 5 *For the rate distortion problem in Lemma 1, suppose that* $\Omega \succeq 0, \Psi \succ 0$, *and* $A = -\rho I$, *with* $\rho > 0$. *Let* $m = rank\,(\Omega)$ *and* $0 < d_1 \leq d_2 \leq ... \leq d_m$. *If*

$$0 < D < \frac{\sum_{i=1}^{m} \sqrt{d_1 d_i}}{2\rho},$$

then the signal dimension is equal to m. *The signal dimension decreases with* D *when the positive eigenvalues of* $\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}}$ *are not identical.*

This proposition shows that the maximal signal dimension is equal to the rank of the matrix Ω when the state transition matrix satisfies $A = -\rho I$ with $\rho > 0$. This maximum is attained when the distortion is sufficiently low. Intuitively, to achieve a sufficiently low distortion, the decision maker has to acquire as much as information as possible subject to the information-flow constraint. As the distortion increases to a sufficiently high level, the decision acquires less information by reducing the signal dimension as long as different sources of state uncertainty are not identical.

3.3 Tracking Problem

In this subsection I consider an important special case without endogenous state. The exogenous state follows the dynamics

$$dx_t = Ax_t dt + GdW_{xt}. \tag{26}$$

Suppose that the optimal control under full information is given by $u_t^* = a'x_t$, where a is an n_x dimensional column vector. The ith component of a measures the importance of the ith state for the optimal choice u_t^*.

The decision maker solves a pure tracking problem under RI

$$\min_{u_t, C} \mathbb{E}\left[(u_t - u_t^*)^2\right]$$

subject to (2), (3), and (26). The solution is $u_t = \mathbb{E}\left[u_t^* | \mathcal{F}_t^y\right] = a'm_t$. Thus the problem can be transformed into the rate distortion problem described in Lemma 1 with $\Omega = aa'$. In this case $rank\,(\Omega) = 1$.

Proposition 6 *Suppose that* $\Omega = a'a$ $(a \neq 0)$, $\Psi \succ 0$, *and* $A = -\rho I$, *with* $\rho > 0$. *Suppose that* $0 < D < \frac{1}{2\rho}\left\|\Psi^{1/2}a\right\|^2$. *Then the optimal conditional covariance matrix is*

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

The optimal signal is one dimensional and can be normalized as

$$dy_t = u_t^* dt + \Xi dW_{yt}, \tag{27}$$

where

$$\Xi = \frac{D}{\sqrt{\left\|\Psi^{1/2}a\right\|^2 - 2\rho D}}.$$

If $D \geq \frac{1}{2\rho}\left\|\Psi^{1/2}a\right\|^2$, *then* $\Sigma = \frac{\Psi}{2\rho}$ *and the decision maker does not process any information.*

An important application of this case is the monopoly pricing problem analyzed by Maćkowiak and Wiederholt (2009) in discrete time. Translated into the continuous-time setup here, each component of the state vector x_t represents a different source of exogenous shocks. The control u_t represents the optimal product price, which is a linear combination of the shocks. Then the optimal signal is one dimensional and can be normalized as the optimal price under full information plus a noise as long as all shocks have the same persistence. The shock innovations can be correlated. By the Kalman-Bucy filter in (9), given the one dimensional signal, the initial responses to the same size of the innovation shock for different sources of uncertainty are the same if all components of the vector a are the same. This result is independent of the size of the innovation variance Ψ.

The intuition is that there is an information spillover effect without the signal independence assumption. Given the one dimensional signal, a shock to a state with high innovation variance is confused with the shock to a state with low innovation variance, as in the standard signal extraction problem. Under rational inattention, the decision maker can choose the attention allocation, which is represented by the linear transformation C in the signal vector. It follows from (27) that the attention allocation is determined by the importance of the state measured by the vector a in that $C = a$. Thus, if all states are equally important, then the decision maker pays equal attention to all states independent of their innovation variances.[4] This result is different from that of Maćkowiak and Wiederholt (2009) based on the signal independence assumption.

I am unable to derive analytical resuts when different states have different persistence. I verify numerically that Proposition 6 fails in this case. I find that the signal dimension is still one dimensional, but the signal cannot be normalized to the form as in (27). In particular, for the case of two states, I find that the firm pays more attention to the more persistent state independent of the size of the innovation variance. In particular $C_1 > C_2$ when $a_1 = a_2$ and $\rho_1 < \rho_2$ for $C = (C_1, C_2)$, $a = (a_1, a_2)'$, and $A = diag(-\rho_1, -\rho_2)$.

4 Application

In this section I study a consumption/saving problem to illustrate my previous results. The discrete-time counterpart is analyzed by Sims (2003) and Luo (2008). A related continuous-time model without RI is analyzed by Wang (2004).

[4] Miao, Wu, and Young (2019) establish the same result in discrete time. They also conduct extensive numerical experiments for the case in which the different sources of shocks have different persistence. I will not study this issue here.

Suppose that the decision maker solves the following problem under full information

$$\max_{\{u_t\}} -E\left[\int_0^{\infty} \beta e^{-\beta t} \left(c_t - \bar{c}\right)^2 dt\right]$$

subject to

$$\begin{aligned} dw_t &= rw_t dt - c_t dt + z_{1t} dt + z_{2t} dt + \sigma_w dZ_t, \\ dz_{1t} &= \left(\bar{z}_1 - \rho_1 z_{1t}\right) dt + \sigma_{1z} dW_{1t}, \\ dz_{2t} &= \left(\bar{z}_2 - \rho_2 z_{2t}\right) dt + \sigma_{2z} dW_{2t}, \end{aligned}$$

where Z_t, W_{1t}, and W_{2t} are independent standard Brownian motions, w_t denotes financial wealth, and z_{1t} and z_{2t} denote two persistent income components. For simplicity I assume that $\beta = r > 0$, $\bar{c} = \bar{z}_1 = \bar{z}_2 = 0$, $\rho_1 > 0$, $\rho_2 > 0$, $\sigma_{1z} > 0$, $\sigma_{2z} > 0$, and $\sigma_w > 0$. The optimal consumption rule under full information is given by $c_t = rs_t$, where s_t denotes total wealth defined as

$$s_t = w_t + \frac{z_{1t}}{r+\rho_1} + \frac{z_{2t}}{r+\rho_2} = \left[1, \frac{1}{r+\rho_1}, \frac{1}{r+\rho_2}\right] x_t. \tag{28}$$

Here $x_t = (w_t, z_{1t}, z_{2t})'$ denotes the state vector.

Now suppose that the decision maker is boundedly rational and faces an information-flow constraint. The signal y_t satisfies (2) and the information-flow constraint is given by (3). For a numerical illustration, I set parameter values as $r = 0.02$, $\rho_1 = 0.1$, $\rho_2 = 0.5$, $\sigma_w^2 = 0.01$, $\sigma_{1z}^2 = 0.05$, and $\sigma_{2z}^2 = 0.01$. Figure 2 presents the rate distortion function, which shows that the capacity κ must be higher than a lower bound close to 0.02. I find that the signal y_t is always one dimensional. In particular, when $\kappa = 0.0904$, the conditional covariance matrix and the signal-to-noise ratio are given by

$$\Sigma = \begin{bmatrix} 9.2204 & 0.2846 & 0.0146 \\ 0.2846 & 0.1577 & -0.0006 \\ 0.0146 & -0.0006 & 0.0100 \end{bmatrix}, \quad C'C = \begin{bmatrix} 0.0072 & 0.0603 & 0.0139 \\ 0.0603 & 0.5025 & 0.1160 \\ 0.0139 & 0.1160 & 0.0268 \end{bmatrix}.$$

Using the singular value decomposition, I derive a solution for the linear transformation $C = [0.0851, 0.7089, 0.1636]$. Even though the two income shocks z_{1t} and z_{2t} are ex ante independent, they are negatively correlated ex post.

[Insert Figure 2 Here.]

To verify my numerical solution, I follow Luo (2008) to derive a closed-form solution. I use total wealth s_t as the single state variable, which follows the dynamics

$$ds_t = \left(rs_t - c_t\right) dt + \sigma dB_t, \tag{29}$$

where B_t is a standard Brownian motion satisfying

$$\sigma dB_t \equiv \frac{\sigma_{1z}}{r+\rho_1}dW_{1t} + \frac{1}{r+\rho_2}\sigma_{2z}dW_{2t} + \sigma_w dZ_t,$$

with

$$\sigma \equiv \sqrt{\left(\frac{\sigma_{1z}}{r+\rho_1}\right)^2 + \left(\frac{\sigma_{2z}}{r+\rho_2}\right)^2 + \sigma_w^2}.$$

Then the optimal consumption rule under full information given (29) is the same as that when the vector x_t is used as the state.[5] The decision maker under RI only needs to track the one-dimensional state s_t. Let the scalar signal y_t satisfy

$$dy_t = hs_t dt + dW_{yt}.$$

The optimal consumption rule under RI is given by $c_t = r\widehat{s}_t$, where $\widehat{s}_t \equiv \mathbb{E}\left[s_t|\mathcal{F}_t^y\right]$. The stationary Kalman-Bucy filter is given by

$$\begin{aligned} d\widehat{s}_t &= (r\widehat{s}_t - c_t)\,dt + \Sigma h\,(dy_t - h\widehat{s}_t dt)\,, && (30)\\ 0 &= 2r\Sigma - (\Sigma h)^2 + \sigma^2, && (31)\end{aligned}$$

where $\Sigma = \mathbb{E}\left[(s_t - \widehat{s}_t)^2\right]$. It follows from (17) that the information-flow constraint is given by

$$\frac{1}{2}\Sigma h^2 \leq \kappa.$$

The no-forgetting constraint in (20) becomes $2r\Sigma + \sigma^2 \geq 0$, which never binds. When the information-flow constraint binds, I use (31) to solve for the optimal conditional covariance

$$\Sigma = \frac{\sigma^2}{2(\kappa - r)},$$

where I impose the assumption that $\kappa > r$ (also see (24)). I then obtain

$$h = \frac{2\sqrt{\kappa(\kappa - r)}}{\sigma}.$$

The optimal signal can also be equivalently rewritten as

$$dy_t = s_t dt + h^{-1}dW_{yt}.$$

This corresponds to the state plus noise signal in the discrete-time setup of Luo (2008).

When using x_t as the state vector, let the linear transformation be $C = [C_1, C_2, C_3]$ in the signal in (2). For a wide range of parameter values, I verify numerically that

$$\frac{C}{C_1} = \left[1, \frac{1}{r+\rho_1}, \frac{1}{r+\rho_2}\right]$$

[5] One can easily verify this claim using equations (6) and (8). Notice that $P = 0$ and $c = 0$ given $\overline{c} = 0$ give a solution. But the implied wealth explodes, violating the no-Ponzi game condition. Thus I rule out this solution.

and $h = C_1$ subject to small numerical errors. Thus I claim that the numerical solution using the semidefinite programming approach with a three-dimensional state vector is very close to the true analytical solution. A similar result is obtained by Miao, Wu, and Young (2019) in a discrete-time setup. However, I am unable to provide a formal proof.

This example illustrates two pitfalls of the existing analysis in the literature. First, if one solves the distortion rate problem as in the second step of Sims's approach, the problem may not have a solution if the capacity κ is calibrated to be too low. By contrast, solving the rate distortion problem can guide the researcher to select the domain of κ. Second, if one makes the signal independence assumption or any other ad hoc assumptions on the signal form, then the solution can be suboptimal. As Sims (2003, 2011) shows, one can solve for the conditional covariance matrix for the state independent of the signal structure, just like what I have shown in Lemma 1 and Proposition 2. But one has to recover the optimal signal structure using Proposition 3. It turns out that the prior assumption on the signal structure in the literature is often inconsistent with the optimal signal form ex post.

5 Conclusion

I have formulated a multivariate LQG control framework with rational inattention in continuous time. I have proposed a three-step solution procedure. The key step is to transform the tracking problem under RI into a rate distortion problem. I derive a semidefinite programming representation for this problem, which can be solved numerically using an efficient publicly available package CVX. I also provide generalized reverse water-filling solutions for some special cases and characterize the optimal signal form without strong prior assumptions. My solution approach can be applied in many other problems in economics and finance.

Appendix

A Proofs

Proof of Proposition 1: Since

$$\mathbb{E}\left[x_0' P x_0\right] = tr\left(\Sigma_0 P\right) + m_0' P m_0,$$

it follows from (7) and (13) that

$$\begin{aligned} V\left(m_0\right) &= -\beta\left\{\mathbb{E}\left[x_0' P x_0\right] - tr\left(\Sigma_0 P\right)\right\} - tr\left(Q\Sigma\right) - \frac{1}{2} tr\left(KK'P\right) \\ &= V^{FI} + \frac{1}{2} tr\left(GG'P\right) + \beta tr\left(\Sigma_0 P\right) - tr\left(Q\Sigma\right) - \frac{1}{2} tr\left(KK'P\right) \\ &= V^{FI} + \beta tr\left(\Sigma_0 P\right) - tr\left(\left(KK' - GG'\right) P\right) - tr\left(Q\Sigma\right). \end{aligned}$$

Using the rotation invariance property of the trace operator and equations (6) and (10), I derive

$$\begin{aligned} & tr\left(Q\Sigma\right) + tr\left(\left(KK' - GG'\right) P\right) - \beta tr\left(\Sigma P\right) \\ = \; & tr\left(Q\Sigma\right) + tr\left(\left(A\Sigma + \Sigma A'\right) P\right) - \beta tr\left(\Sigma P\right) \\ = \; & tr\left(\left(Q + A'P + PA\right)\Sigma\right) - \beta tr\left(\Sigma P\right) \\ = \; & tr\left(\Omega\Sigma\right), \end{aligned}$$

where I have defined (14). Given $\Sigma_0 = \Sigma$ and $\mathbb{E}\left[\left(x_t - m_t\right)' \Omega \left(x_t - m_t\right)\right] = tr\left(\Omega\Sigma\right)$, I obtain the desired result. Q.E.D.

Proof of Lemma 1: Using the formulas derived in the main text, I rewrite Problem 4 as

$$\kappa\left(D\right) = \inf_{C \in \mathcal{C}} \frac{1}{2} tr\left(C\Sigma C'\right)$$

subject to (19) and

$$0 = A\Sigma + \Sigma A' - \Sigma C' C \Sigma + GG'. \tag{A.1}$$

By Lemma 1 of Tanaka, Skoglund, and Ugrinovskii (2017), I replace equation (A.1) with the no-forgetting constraint (20). Using (A.1) to rewrite the objective yields

$$\begin{aligned} \frac{1}{2} tr\left(\Sigma C' C\right) &= \frac{1}{2} tr\left(\Sigma C' C \Sigma \Sigma^{-1}\right) \\ &= \frac{1}{2} tr\left(\left(A\Sigma + \Sigma A' + GG'\right)\Sigma^{-1}\right) \\ &= tr\left(A\right) + \frac{1}{2} tr\left(G'\Sigma^{-1} G\right). \end{aligned}$$

I have eliminated the choice variable C and obtained the desired result. Q.E.D.

Proof of Proposition 2: By Lemma 1 the objective function is

$$
\begin{aligned}
tr\left(A\right)+\frac{1}{2}tr\left(G'\Sigma^{-1}G\right) &= \min_{\Pi\succeq 0}\left\{tr\left(A\right)+\frac{1}{2}tr\left(\Pi\right):G'\Sigma^{-1}G\preceq\Pi\right\} \\
&= \min_{\Pi}\left\{tr\left(A\right)+\frac{1}{2}tr\left(\Pi\right):\ (21)\text{ holds}\right\},
\end{aligned}
$$

where the last equality uses the Schur complement formula. I then obtain the semidefinite programming representation in the proposition. The existence of an optimal solution $\Sigma\succ 0$ and $\Pi\succeq 0$ follows from Theorem 2 of Tanaka, Skoglund, and Ugrinovskii (2017). Q.E.D.

Proof of Proposition 3: Notice that $A\Sigma+\Sigma A'+GG'$ is a positive semidefinite matrix by the no-forgetting constraint (20). Since $\Sigma\succ 0$, I apply the singular value decomposition to derive

$$
\begin{aligned}
C'C &= \Sigma^{-1}\left(A\Sigma+\Sigma A'+GG'\right)\Sigma^{-1} \\
&= UVU'=[U_1,U_2]\begin{bmatrix} V_p & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} U_1' \\ U_2' \end{bmatrix}=U_1V_pU_1'=\left(U_1M\right)\left(U_1M\right)',
\end{aligned}
$$

where U is a unitary matrix and V_p is a diagonal matrix with all positive singular values on the diagonal. Notice that U_1 and U_2 are conformable matrices and $M=V_p^{\frac{1}{2}}$. I then obtain the desired result. Q.E.D.

Proof of Proposition 4: Recall that I define $GG'=\Psi$ and the positive semidefinite matrix $\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}}$ has an eigendecomposition $\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}}=U\Omega_dU'$, where U is an orthonormal matrix, $\Omega_d=diag\left(d_i\right)_{i=1}^{n_x}$, and $d_i\geq 0$ denotes an eigenvalue of $\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}}$ for $i=1,2...,n_x$. Define

$$
\widehat{\Sigma}=U'\Psi^{-\frac{1}{2}}\Sigma\Psi^{-\frac{1}{2}}U.
$$

I then obtain

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

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

$$
\begin{aligned}
tr\left(\Omega\Sigma\right) &= tr\left(\Psi^{-\frac{1}{2}}U\left(U'\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}}U\right)U'\Psi^{-\frac{1}{2}}\Sigma\right) \\
&= tr\left(\Omega_dU'\Psi^{-\frac{1}{2}}\Sigma\Psi^{-\frac{1}{2}}U\right)=tr\left(\Omega_d\widehat{\Sigma}\right),
\end{aligned}
$$

and

$$
tr\left(\Sigma^{-1}GG'\right)=tr\left(\Sigma^{-1}\Psi\right)=tr\left(\Psi^{-\frac{1}{2}}U\widehat{\Sigma}^{-1}U'\Psi^{-\frac{1}{2}}\Psi\right)=tr\left(\widehat{\Sigma}^{-1}\right).
$$

Given $A=-\rho I$, the no-forgetting constraint (20) becomes

$$
\Sigma\preceq\frac{GG'}{2\rho}\Leftrightarrow U'\Psi^{-\frac{1}{2}}\Sigma\Psi^{-\frac{1}{2}}U\preceq\frac{I}{2\rho}\Leftrightarrow\widehat{\Sigma}\preceq\frac{I}{2\rho}.
$$

Thus the rate distortion problem in Lemma 1 becomes

$$\kappa(D) = \min_{\widehat{\Sigma} \succ 0} \frac{1}{2} tr\left(\widehat{\Sigma}^{-1}\right) - \rho n_x$$

subject to

$$\begin{aligned} tr\left(\Omega_d \widehat{\Sigma}\right) &\leq D, & \text{(A.2)} \\ \widehat{\Sigma} &\preceq \frac{I}{2\rho}. & \text{(A.3)} \end{aligned}$$

Ignoring the no-forgetting constraint (A.3) for now and deriving the first-order condition, I obtain

$$\widehat{\Sigma}^{-2} = \alpha \Omega_d,$$

where $\alpha > 0$ is the Lagrange multiplier associated with (A.2). Since Ω_d is diagonal, it follows that $\widehat{\Sigma}$ is also diagonal with the diagonal elements:

$$\widehat{\Sigma}_i = \frac{1}{\sqrt{\alpha d_i}}, \; i = 1, ..., n_x.$$

If $d_i = 0$, I set $\widehat{\Sigma}_i = \infty$. Given (A.3), I obtain $\widehat{\Sigma}_i = \min\left\{\frac{1}{\sqrt{\alpha d_i}}, \frac{1}{2\rho}\right\}$.

If $D \geq \frac{tr(\Omega_d)}{2\rho}$, the distortion constraint (A.2) does not bind. Each $\widehat{\Sigma}_i$ attains the upper bound $\widehat{\Sigma}_i = 1/(2\rho)$ and thus $\Sigma = \Psi/(2\rho)$. If $0 < D < \frac{tr(\Omega_d)}{2\rho}$, then (A.2) binds and the Lagrange multiplier $\alpha > 0$ is such that $tr\left(\Omega_d \widehat{\Sigma}\right) = D$.

By (22) the optimal information structure satisfies

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

Simplifying yields the desired result. Q.E.D.

Proof of Corollary 1: For the case of $\rho > 0$ I apply Proposition 1 directly. In particular $\Omega_d = \sigma^2$ and $U = 1$. When $0 < D < \frac{\sigma^2}{2\rho}$, the distortion constraint binds. The Lagrange multiplier $\alpha > 0$ satisfies $\frac{\sigma}{\sqrt{\alpha}} = D$ so that $\alpha = \frac{\sigma^2}{D^2}$.

For the case of $\rho < 0$, the no-forgetting constraint (20) becomes

$$-2\rho\Sigma + \sigma^2 \geq 0,$$

which never binds. The distortion constraint always binds and hence $\Sigma = D$. The rest of the proof is trivial. Q.E.D.

Proof of Corollary 2: Under the assumption in the corollary, I have $U = I$ and $\Omega_d = \Psi$. Thus $d_i = \sigma_i^2$. The result then follows from Proposition 4 and is omitted. Q.E.D.

Proof of Proposition 5: I first solve for the Lagrange multiplier α. I claim that

$$\hat{\Sigma}_i = \frac{1}{\sqrt{\alpha d_i}} < \frac{1}{2\rho}, \ i = 1, ..., m.$$

For $i = m+1, ..., n_x$, $d_i = 0$ and $\hat{\Sigma}_i = 1/(2\rho)$. It follows from $tr\left(\Omega_d \hat{\Sigma}\right) = D$ that

$$\sqrt{\alpha} = \frac{\sum_{i=1}^m \sqrt{d_i}}{D}.$$

Given the assumption on D in the proposition, I find that the above expression for $\hat{\Sigma}_i$ is indeed the optimal solution. The inside diagonal matrix in (23) has exactly m positive diagonal elements. Thus the signal dimension is equal to $m = rank(\Omega)$.

Next suppose that the positive eigenvalues are not all identical. For simplicity let $0 < d_1 < d_2 \leq ... \leq d_m$. I claim that if D satisfies

$$\frac{\sum_{i=1}^m \sqrt{d_1 d_i}}{2\rho} \leq D < \frac{\sum_{i=1}^m \sqrt{d_2 d_i}}{2\rho} + \frac{d_1}{2\rho}, \tag{A.4}$$

then the signal dimension is $m-1$. I show that

$$\hat{\Sigma}_1 = \frac{1}{\sqrt{\alpha d_1}} \geq \frac{1}{2\rho}, \ \hat{\Sigma}_i = \frac{1}{\sqrt{\alpha d_i}} < \frac{1}{2\rho}, \ i = 2, ..., m. \tag{A.5}$$

To verify this result, I use $tr\left(\Omega_d \hat{\Sigma}\right) = D$ to compute

$$\frac{1}{\sqrt{\alpha}} = \frac{D - d_1/(2\rho)}{\sum_{i=2}^m \sqrt{d_i}}.$$

This equation and (A.4) imply that the inequalities in (A.5) are indeed satisfied and the expressions for $\hat{\Sigma}_i$, $i = 1, ..., m$, in (A.5) are indeed the solution. For $i = m+1, ..., n_x$, $d_i = 0$ and $\hat{\Sigma}_i = 1/(2\rho)$. Thus the inside diagonal matrix in (23) has exactly $m-1$ positive diagonal elements, implying that the signal dimension is equal to $m-1$. For other more general cases, the proof is similar and omitted. Q.E.D.

Proof of Proposition 6: Since $rank(\Omega) = 1$, I have $rank\left(\Psi^{\frac{1}{2}} \Omega \Psi^{\frac{1}{2}}\right) = 1$. I claim that matrix $\Psi^{\frac{1}{2}} \Omega \Psi^{\frac{1}{2}}$ has a unique positive eigenvalue $d_1 \equiv \left\|\Psi^{1/2} a\right\|^2$ and an associated unit eigenvector $\Psi^{\frac{1}{2}} a / \left\|\Psi^{1/2} a\right\|$ where $\|\cdot\|$ denotes the Euclidean norm. To prove this claim I verify that

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

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

Suppose that $0 < D < \frac{d_1}{2\rho}$. It follows from Proposition 4 that

$$\widehat{\Sigma}_1 = \min\left(\frac{1}{\sqrt{\alpha d_1}}, \frac{1}{2\rho}\right), \quad \widehat{\Sigma}_i = \frac{1}{2\rho}, \ i = 2, ..., n_x,$$

where $\alpha > 0$ is such that the distortion constraint binds $D = \frac{d_1}{\sqrt{\alpha d_1}}$. I can solve explicitly

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

The optimal information structure satisfies

$$C'C = \Psi^{-\frac{1}{2}} U \widehat{\Sigma}^{-1} \text{ diag}\left\{\max\left(1 - \frac{2\rho}{\sqrt{\alpha d_1}}, 0\right)_{i=1}^{n_x}\right\} \widehat{\Sigma}^{-1} U' \Psi^{-\frac{1}{2}}, \tag{A.6}$$

where U is a unitary matrix in the eigendecomposition $\Psi^{\frac{1}{2}}\Omega\Psi^{\frac{1}{2}} = U\Omega_d U'$. Partition $U = [U_1, U_2]$ conformably, where $U_1 = \Psi^{\frac{1}{2}}a/\left\|\Psi^{1/2}a\right\|$. There is only one positive element in the inside diagonal matrix in (A.6), which is

$$1 - \frac{2\rho}{\sqrt{\alpha d_1}} = 1 - \frac{2\rho D}{d_1} > 0.$$

The optimal information structure C corresponds to the eigenvector associated with the positive eigenvalue. It follows from (A.6) that

$$C' = \frac{1}{D}\sqrt{d_1^2 - 2\rho D d_1} \frac{a}{\left\|\Psi^{1/2}a\right\|}.$$

I normalize C as $C^* = a'$ so that the normalized optimal signal takes the form in (27).

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

$$\Sigma = \Psi^{\frac{1}{2}} U \begin{bmatrix} \frac{D}{d_1} & 0 \\ 0 & \frac{1}{2\rho}I \end{bmatrix} U' \Psi^{\frac{1}{2}}.$$

Since $U = [U_1, U_2]$ with $U_1 = \Psi^{\frac{1}{2}}a/\left\|\Psi^{1/2}a\right\|$, it follows that $U_1U_1' + U_2U_2' = I$. Thus

$$\Sigma = \Psi^{\frac{1}{2}}\left[\frac{I}{2\rho} - U_1U_1'\left(\frac{1}{2\rho} - \frac{D}{d_1}\right)\right]\Psi^{\frac{1}{2}}.$$

Simplifying yields the expression in the proposition.

If $D \geq \frac{d_1}{2\rho}$, one can check that $\widehat{\Sigma}_i = \frac{1}{2\rho}$ for all i so that $\Sigma = \frac{\Psi}{2\rho}$ and the decision maker does not process any information. Q.E.D.

References

Anderson, Brian O., and John B. Moore, 1989, *Optimal Control: Linear Quadratic Methods*, NJ Prentice-Hall Inc.

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.

Duncan, Tyrone E., 1970, On the Calculation of Mutual Information, *SIAM Journal on Applied Mathematics* 19, 215-220.

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

Hansen, Lars P., and Thomas Sargent, 2007, *Robustness*, Princeton, Princeton University Press.

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.

Kasa, Kenneth, 2006, Robustness and Information Processing, Review of Economic Dynamics 9, 1-33.

Liptser, Robert S. and Albert N. Shiryaev, 2001, *Statistics of Random Processes*, Vol. I & II, 2nd edition, Springer, Berlin.

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

Luo, Yulei, 2018, Robustly Strategic Consumption-Portfolio Rules with Information Frictions, forthcoming in *Management Science*.

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

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.

Miao, Jianjun, Jieran Wu, and Eric Young, 2019, Multivariate Rational Inattention, working paper, Boston University.

Moscarini, Giuseppe, 2004, Limited Information Capacity as a Source of Inertia, *Journal of Economic Dynamics and Control* 28, 2003-2035.

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.

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

Tanaka, Takashi, Mikael Skoglund, and Valeri Ugrinovskii, 2017, Optimal Sensor Design and Zero-Delay Source Coding for Continuous-Time Vector Gauss-Markov Processes, The 56th IEEE Conference on Decision and Control.

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

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

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

Vandenberghe, Lieven, and Stephen Boyd, 1998, Semidefinite Programming, *SIAM Review* 38, 49–95.

Wang, Neng, 2004, Precautionary Saving and Partially Observed Income, *Journal of Monetary Economics* 51, 1645—1681

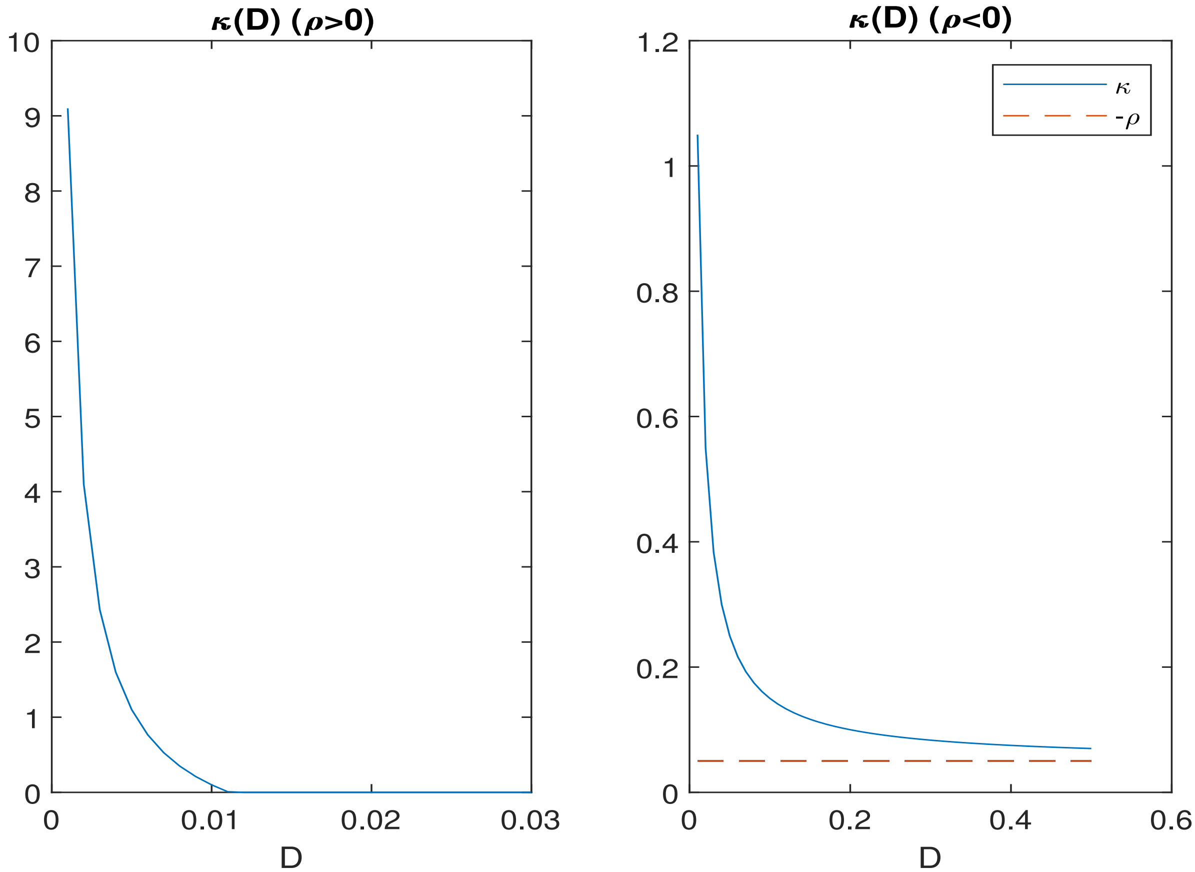

Figure 1: Rate distortion functions for the scalar case with $\rho > 0$ and $\rho < 0$.

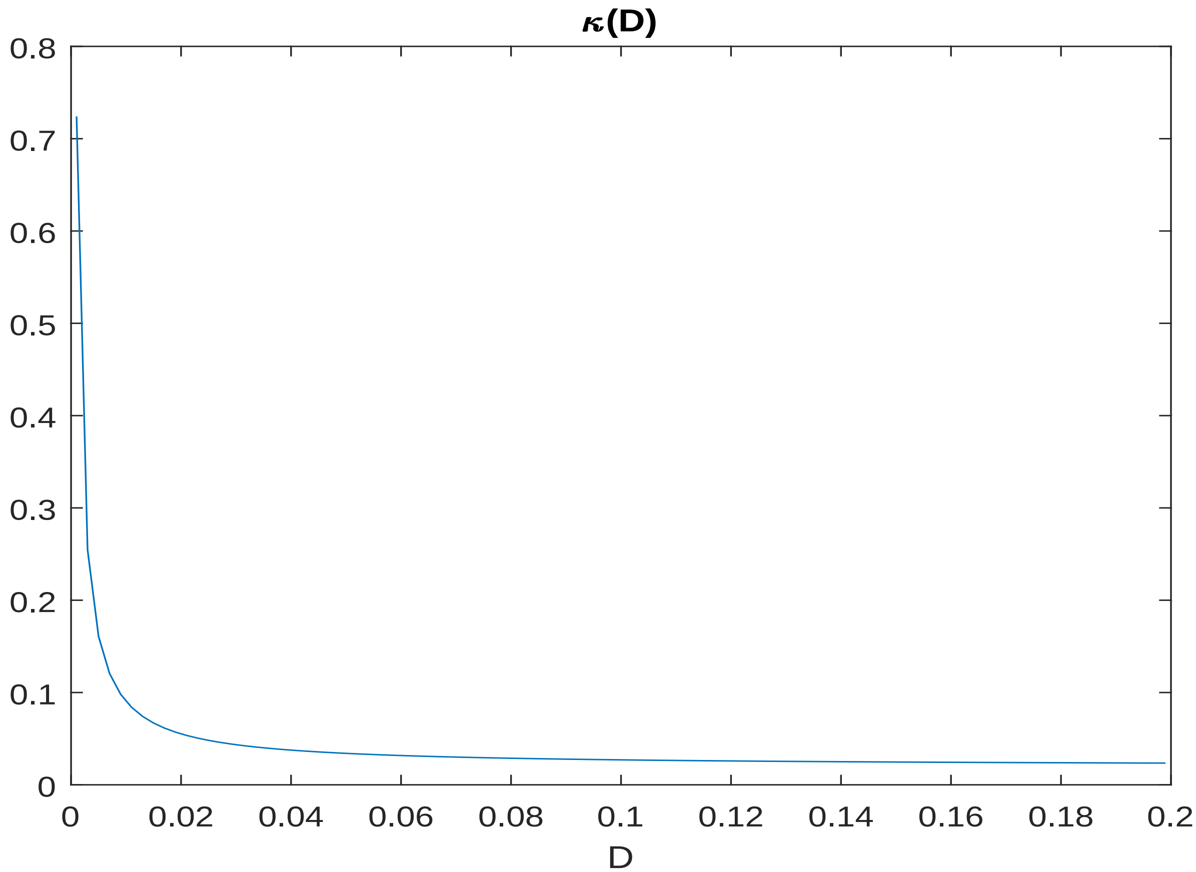

Figure 2: Rate distortion function for the consumption/saving model. Parameter values are $r = 0.02$, $\rho_1 = 0.1$, $\rho_2 = 0.5$, $\sigma_w^2 = 0.01$, $\sigma_{1z}^2 = 0.05$, and $\sigma_{2z}^2 = 0.01$.