Signal Extraction and Rational Inattention*

Yulei Luo† — University of Hong Kong

Eric R. Young‡ — University of Virginia

April 15, 2013

Abstract

In this paper we examine the implications of two theories of informational frictions, signal extraction (SE) and rational inattention (RI), for optimal decisions and economic dynamics within the linear-quadratic-Gaussian (LQG) setting. We first show that if the variance of the noise and channel capacity (or marginal information cost) are fixed exogenously in the SE and RI problems, respectively, the two environments lead to different policy and welfare implications. We also find that if the signal-to-noise ratio and capacity in the SE and RI problems are fixed, respectively, the two theories generate the same policy implications in the univariate case, but different policy implications in the multivariate case.

JEL Classification Numbers: *C61, D81, E21.*

Keywords: Signal Extraction, Rational Inattention, Exogenous and Endogenous Noises.

* We thank Tom Sargent and Chris Sims for helpful discussions. We are also grateful for useful suggestions and comments from Shin-ichi Fukuda, Ken Kasa, Yang Lu, Jun Nie, Yi-Chang Tsai, Jenny Xu, Tao Zhu, and seminar and conference participants at the Hong Kong University of Science and Technology, the University of Tokyo, and Fudan Conference on Economic Dynamics. Luo thanks the General Research Fund (GRF) in Hong Kong and the HKU seed funding program for basic research for financial support. Young thanks the Bankard Fund for Political Economy at the University of Virginia for financial support. Part of this work was conducted while Luo was visiting the UC Davis Economics Department and the Kansas City Fed, whose hospitality is greatly appreciated. All errors are the responsibility of the authors.

† School of Economics and Finance, The University of Hong Kong, Hong Kong. E-mail: `yluo@econ.hku.hk`.

‡ Department of Economics, University of Virginia, Charlottesville, VA 22904. E-mail: `ey2d@virginia.edu`.

1 Introduction

Muth (1960) applied classical filtering methods to solve for a stochastic process for permanent income for which Friedman (1956)'s adaptive expectations hypothesis would be an optimal estimator of permanent income. Specifically, Muth (1960) solved a single-agent dynamic signal extraction (SE) problem, in which an economic agent was modeled as facing exogenous signals and noises which had to be disentangled, and showed that the exponentially weighted average of past observations of a random walk plus noise process is optimal in the sense that it minimizes the mean squared estimation error. Townsend (1983) and Sargent (1991) extended the single-agent signal extraction problem by studying multiple-agent settings in which agents extract signals from endogenous variables that are affected by other agents' signal extraction problems. Recently, there have been some papers examining the effects of heterogeneous information on economic dynamics within signal extraction settings. For example, Morris and Shin (2002) examined the welfare effects of asymmetric information in the presence of strategic complementarity; Wang (2004) examined how imperfect observations on labor income on the estimation risk and precautionary savings; and Angeletos and La'O (2009) studied how dispersed information about the underlying aggregate productivity shock contributes significant noise in the business cycle and helps explain cyclical variations in observed Solow residuals and labor wedges in the RBC setting. The key assumption in these signal extraction settings is that the stochastic properties of noises are given exogenously.

Sims (2003) first introduced rational inattention (RI) into economics within the linear-quadratic Gaussian (LQG) setting and argued that it is a plausible method for introducing sluggishness, randomness, and delay into economic models.[1] In his formulation agents have finite Shannon channel capacity, limiting their ability to process signals about the true state of the world. As a result, an impulse to the economy induces only gradual responses by individuals, as their limited capacity requires many periods to discover just how much the state has moved; one key change relative to the RE case is that consumption has a hump-shaped impulse response to income shocks.[2]

[1] For the applications of RI within the (approximate) LQG setting, see (among others) Adam (2005), Kasa (2006), Luo and Young (2010), Maćkowiak and Wiederholt (2009), Melosi (2009), Paciello and Wiederholt (2011), and Kim, Ko, and Yun (2012).

[2] See Sims (2003) and Luo (2008).

Luo (2008) used this model to explore anomalies in the consumption literature, particularly the well-known "excess sensitivity" and "excess smoothness" puzzles, employing an LQG version of the standard permanent income model (as in Hall 1978 and Flavin 1981). In that model RI is equivalent to confronting the household with a noisy signal about the value of permanent income but permitting the agents to choose the distribution of the noise terms, subject to their limited capacity. The key feature of the LQG-RI model is that the RI-induced noise is optimal and generated endogenously due to individuals' finite information-processing capacity.[3]

The main objective of this paper is to compare the two theories. Specifically, we explore the effects of SE and RI for economic dynamics, policy, and welfare within the linear-quadratic-Gaussian setting. We first study a univariate case for which the models can be solved in closed-form. The first result we find is that if the variance of the noise itself is fixed, we can use a policy experiment to distinguish SE from RI as they lead to different dynamic behavior, policy, and welfare implications. Specifically, we assume that the variance of the exogenous shock is scaled up due to a change in policy. In the SE problem with exogenous noises, an increase in the variance of the exogenous shock will lead to a different solution for the conditional variance and Kalman gain; consequently, the change in policy will eventually lead to a change in the model's dynamic behavior and the agent's welfare. In contrast, in the RI problem, if channel capacity is fixed, a change in the variance of the exogenous shock will lead to the same change in the conditional variance of the state and the variance of the noise, but has no effect on the Kalman gain. That is, inattentive agents with fixed capacity will behave as if they face noise shocks whose nature changes systematically as the dynamic properties of the economy change with policy. Furthermore, we show that once we assume that the marginal cost of information is fixed, capacity will be elastic with respect to a change in policy; consequently, the Kalman gain in this case will adjust with respect to the policy change and the change in policy has different quantitative impacts on the model dynamics under this case and the SE case.

The welfare losses of agents due to imperfect information also depend on the value of the Kalman gain. Therefore, SE and RI can lead to different policy and welfare implications in the

[3] Fully non-LQG versions of the RI problem are solved and discussed in Sims (2005, 2006), Lewis (2006), Tutino (2009), and Metajka and Sims (2010). The main feature of the non-LQG RI models is that they have either very short horizons or extremely simple setups due to numerical obstacles. We comment in the conclusion about the difficulty of extending our results to non-LQG environments.

LQ setting. We also find that in the univariate case, if the ratio of the variance of the exogenous shock to that of the noise (i.e., the signal-to-noise ratio, SNR) is fixed, the SE and RI problems are observationally equivalent in the sense that they lead to the same dynamics of the model economy when the ratio of the conditional variance to that of the noise in the SE problem is equal to $1 - 1/\exp(2\kappa)$ in the RI problem in which κ is the exogenously given channel capacity. After considering correlated shocks and noise, we find that our results remain unchanged.

We then move on to study the multivariate case in which the state vector includes multiple elements. In this case given channel capacity the conditional variance-covariance matrix can be obtained by solving a semidefinite programming problem in which the inattentive agent minimizes the expected welfare losses due to information-processing constraints. After computing the optimal steady state conditional variance-covariance matrix, we can recover the variance-covariance matrix of the noise vector and then determine the Kalman gain. In this case, we show that SE and RI will lead to different dynamic behavior and deliver different policy and welfare implication after the government implements a policy that changes the variance of the exogenous shock even if the signal-to-noise ratio is fixed. However, when modeling the multivariate SE problem, it is difficult to specify the process of the vector of noises *ex ante* without prior knowledge about the states. Ad hoc assumptions on the nature of the noise might be inconsistent the underlying efficiency conditions (equalization of the marginal utility of additional capacity across variables).[4] Therefore, RI provides a useful and microfounded way to specify the stochastic properties of the noises by solving the agent's constrained optimization problem. It is worth noting that in the multivariate RI problem, the agent's preference, budget constraint, and information-processing constraints jointly determine the values of the conditional variance of the state, the variance of the noise, and the Kalman gain, whereas in the multivariate SE problem given the variance of the noise, the propagation equation updating the conditional variance based on the budget constraint is used to determined the conditional variance and then the Kalman gain.

The remainder of the paper is organized as follows. Section 2 examines optimal decisions and economic dynamics in an LQG setting with signal extraction. Section 3 presents the RI version of the model and compares different implications of RI and SE on the dynamic behavior, policy and welfare within the LQG setting. Section 4 presents applications to permanent income models.

[4] See Melosi (2009) for an empirical investigation of this issue.

Section 5 concludes.

2 Signal Extraction in a LQG Model

2.1 Full-information Rational Expectations LQG Model

Consider the following linear-quadratic-Gaussian (LQG) model:

$$v(s_0) = \max_{\{c_t, s_{t+1}\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t \left(s_t^T Q s_t + c_t^T R c_t + 2 c_t^T W s_t \right) \right], \tag{1}$$

subject to

$$s_{t+1} = A s_t + B c_t + \varepsilon_{t+1}, \tag{2}$$

with s_0 known and given, where β is the discount factor, s_t is a $(n \times 1)$ state vector, c_t is a $(k \times 1)$ control vector, ε_{t+1} is an iid $(n \times 1)$ vector of Gaussian random variables with mean 0 and covariance matrix Ω, and $E_t[\cdot]$ denotes the mathematical expectation of a random variable given information processed at t. We assume that Q, R, and W are such that the objective function is jointly concave in s_t and c_t, and the usual conditions required for the optimal policy to exist are satisfied.

When the agent can fully observe the state s_t, the model is a standard linear-quadratic regulator problem. Solving the corresponding Bellman equation

$$s_t^T P s_t = \max_{c_t} \left\{ s_t^T Q s_t + c_t^T R c_t + 2 c_t^T W s_t + \beta E_t \left[\left(s_t^T A^T + c_t^T B^T + \varepsilon_{t+1}^T \right) P \left(A s_t + B c_t + \varepsilon_{t+1} \right) \right] \right\},$$

yields the decision rule

$$c_t^* = -F s_t, \tag{3}$$

and the Riccati equation is

$$P = Q + F^T R F - 2 F^T W + \beta \left(A^T - F^T B^T \right) P \left(A - BF \right), \tag{4}$$

where

$$F = \left(R + \beta B^T P B \right)^{-1} \left(W + \beta B^T P A \right). \tag{5}$$

Iterating on the matrix Riccati equation (4) uniquely determines P, since the equation defines a contraction mapping. Using P, we can determine F in the optimal policy (5).

2.2 Signal Extraction with Exogenous Noises

Following the signal extraction literature (e.g., Muth 1960; Lucas 1972, 1973; Morris and Shin 2002), we now assume that the agent cannot observe the true state s_t perfectly and only observes the noisy signal $s_t^* = s_t + \xi_t$ when making decisions. Here ξ_t is a $(n \times 1)$ vector of noises. The agent then estimates the state using a standard Kalman filtering equation. In the standard signal extraction problem, the stochastic property of the noise ξ_t is given *exogenously.* Specifically, assume that ξ_t is an iid Gaussian innovation with mean 0 and variance-covariance matrix Λ.[5] We point out here that the agent may not have perfect information even about the endogenous part of the state vector s_t.

Under the LQG assumption, the certainty equivalence principle holds when the agent cannot observe s_t perfectly, so the decision rule under imperfect information can be written as

$$c_t^* = -F\widehat{s}_t, \tag{6}$$

where $\widehat{s}_t = E[s_t|\mathcal{I}_t]$ is the perceived state and $\mathcal{I}_t = \left\{s_t^*, s_{t-1}^*, \cdots, s_0^*\right\}$ is the information set including perceived signals until time t.

Furthermore, we assume that in the steady state, the true state follows a normal distribution after observing the noisy signals $s_t|\mathcal{I}_t \sim N(E[s_t|\mathcal{I}_t], \Sigma_t)$, where $\Sigma_t = E_t\left[(s_t - \widehat{s}_t)(s_t - \widehat{s}_t)^T\right]$ is the conditional variance-covariance matrix, and the Kalman filtering equation governs the behavior of $\widehat{s}_t$

$$\widehat{s}_{t+1} = (1 - \theta_t)(A\widehat{s}_t + Bc_t) + \theta_t s_{t+1}^*, \tag{7}$$

where θ_t is the Kalman gain to be determined.[6] Following the standard procedure in the Kalman filter literature, we have the updating equation for Σ_t,

$$\Sigma_{t+1} = (I - \theta_t) A\Sigma_t A^T (I - \theta_t)^T + (I - \theta_t)\Omega(I - \theta_t)^T + \theta_t\Lambda\theta_t^T, \tag{8}$$

and the optimal Kalman gain

$$\theta_t = \left(\Omega + A\Sigma_t A^T\right)\left(\Omega + A\Sigma_t A^T + \Lambda\right)^{-1}. \tag{9}$$

[5] Our quadratic objective function encompasses the standard tracking objective of minimizing the squared difference of the control from the target.

[6] Muth (1960) showed that the exponentially weighted average of past observations of a random walk plus a noise process is optimal in the sense that it minimizes the mean squared forecast error.

If iterations on Σ and θ using (8) and (9) converge, in the steady state we have

$$\Sigma = (I - \theta)\left(A\Sigma A^T + \Omega\right)(I - \theta)^T + \theta\Lambda\theta^T \tag{10}$$

and

$$\theta = \left(\Omega + A\Sigma A^T\right)\left(\Omega + A\Sigma A^T + \Lambda\right)^{-1}. \tag{11}$$

(Baxter, Graham, and Wright 2010 discussed the convergence of the Ricatti equation for Kalman filtering problems with endogenous variables.) Starting from the initial condition Σ_0, we can compute the steady state (θ, Σ) by iterating on (10) and (11). After computing (θ, Σ), we can obtain a complete characterization of the dynamic system. The key assumption in the SE problem is that the variance-covariance matrix of the noise, Λ, is given. Given this Λ, (10) and (11) jointly determine the steady state (θ, Σ).

It is straightforward to show that we have the following alternative equations for computing the Kalman gain and the conditional variance-covariance matrix, (θ, Σ):

$$\Sigma_{t+1} = \Psi_t - \Psi_t\left(\Psi_t + \Lambda_t\right)^{-1}\Psi_t, \tag{12}$$

and

$$\theta_t = \Sigma_t\Lambda_t^{-1}, \tag{13}$$

where $\Psi_t = A\Sigma_t A^T + \Omega$ is the conditional variance of the state *prior to* observing the new signal at $t + 1$. In the steady state, (12) and (13) reduce to

$$\Lambda^{-1} = \Sigma^{-1} - \Psi^{-1}. \tag{14}$$

and

$$\theta = \Sigma\Lambda^{-1}, \tag{15}$$

respectively. After obtaining (15), (42), (6), and (7) completely characterize the model's dynamic behavior.

3 Rational Inattention in the LQG Model

Following Sims (2003), we introduce rational inattention (RI) into the LQG model proposed in Section 2.1 by assuming agents face information-processing constraints and have only finite

Shannon channel capacity to observe the state of the world. Specifically, we use the concept of entropy from information theory to characterize the uncertainty about a random variable; the reduction in entropy is thus a natural measure of information flow. Formally, entropy is defined as the expectation of the negative of the (natural) log of the density function, $-E\left[\ln\left(f\left(X\right)\right)\right]$. For example, the entropy of a discrete distribution with equal weight on two points is simply $E\left[\ln_2\left(f\left(X\right)\right)\right]=-0.5\ln\left(0.5\right)-0.5\ln\left(0.5\right)=0.69$, and the unit of information contained in this distribution is 0.69 "nats".[7] In this case, an agent can remove all uncertainty about X if the capacity devoted to monitoring X is $\kappa=0.69$ nats.

With finite capacity $\kappa\in\left(0,\infty\right)$, a variable s following a continuous distribution cannot be observed without error and thus the information set at time $t+1$, $\mathcal{I}_{t+1}$, is generated by the entire history of noisy signals $\left\{s_j^*\right\}_{j=0}^{t+1}$. Following the literature, we assume the noisy signal takes the additive form $s_{t+1}^*=s_{t+1}+\xi_{t+1}$, where ξ_{t+1} is the endogenous noise caused by finite capacity. We further assume that ξ_{t+1} is an iid idiosyncratic shock and is independent of the fundamental shock. Note that the reason that the RI-induced noise is idiosyncratic is that the endogenous noise arises from the consumer's own internal information-processing constraint. Agents with finite capacity will choose a new signal $s_{t+1}^*\in\ \mathcal{I}_{t+1}=\left\{s_1^*,s_2^*,\cdots,s_{t+1}^*\right\}$ that reduces the uncertainty of the state variable s_{t+1} as much as possible. Formally, this idea can be described by the information constraint

$$\mathcal{H}\left(s_{t+1}|\mathcal{I}_t\right)-\mathcal{H}\left(s_{t+1}|\mathcal{I}_{t+1}\right)\leq\kappa, \tag{16}$$

where κ is the investor's information channel capacity, $\mathcal{H}\left(s_{t+1}|\ \mathcal{I}_t\right)$ denotes the entropy of the state *prior to* observing the new signal at $t+1$, and $\mathcal{H}\left(s_{t+1}|\ \mathcal{I}_{t+1}\right)$ is the entropy *after* observing the new signal. κ imposes an upper bound on the amount of information – that is, the change in the entropy – that can be transmitted in any given period. We assume that the noise ξ_{t+1} is Gaussian.[8] Finally, following the literature, we suppose that the prior distribution of s_{t+1} is Gaussian.

Under the LQG setting, as shown in Sims (2003, 2006), the true state under RI also follows a normal distribution $s_t|\mathcal{I}_t\sim N\left(E\left[s_t|\mathcal{I}_t\right],\Sigma_t\right)$, where $\Sigma_t=E_t\left[\left(s_t-\widehat{s}_t\right)\left(s_t-\widehat{s}_t\right)^T\right]$. In addition,

[7] For alternative bases for the logarithm, the unit of information differs; with log base 2 the unit of information is the 'bit' and with base 10 it is a 'dit' or a 'hartley.'

[8] As shown in Sims (2003), within the LQG setting Gaussian noise is optimal.

in the steady state the agent observes an additive noisy signal: $s_t^* = s_t + \xi_t$. Note that in the RI problem we also have the usual formula for updating the conditional variance-covariance matrix of a Gaussian distribution Σ_t:

$$\Sigma_{t+1} = \Psi_t - \Psi_t \left(\Psi_t + \Lambda_t\right)^{-1} \Psi_t. \tag{17}$$

If iterations on Σ converge (which depends on both A and Σ), (17) reduces to $\Sigma = \Psi - \Psi \left(\Psi + \Lambda\right)^{-1} \Psi$, which can be solved for

$$\Lambda^{-1} = \Sigma^{-1} - \Psi^{-1}. \tag{18}$$

Using these expressions, the Kalman gain K can be rewritten as

$$\theta = \Sigma\Lambda^{-1}. \tag{19}$$

3.1 The Univariate Case

The key difference between SE and RI is that under RI the agent faces the information-processing constraint

$$-\ln\left(\left|\Sigma_{t+1}\right|\right) + \ln\left(\left|A^T \Sigma_t A + \Omega\right|\right) \leq 2\kappa. \tag{20}$$

Since more information about the state is better in single-agent models, this constraint will be binding.[9] Considering the univariate state case $n = 1$, (20) fully determines the value of the steady state conditional variance Σ:

$$\Sigma = \frac{\Omega}{\exp\left(2\kappa\right) - A^2}, \tag{21}$$

which means that Σ is determined by the variance of the exogenous shock (Ω) and the exogenously given capacity (κ).[10] Given this Σ, we can use (18) to recover the variance of the endogenous noise (Λ):

$$\Lambda = \left(\Sigma^{-1} - \Psi^{-1}\right)^{-1}, \tag{22}$$

[9] By "better" we mean that conditional on draws by nature for the true state, the expected utility of the agent increases if information about that state is improved.

[10] Note that here we need to impose the restriction $\exp\left(2\kappa\right) - A^2 > 0$. If this condition fails, the state is not *stabilizable* and the unconditional variance diverges. Obviously we cannot directly impose conditions on A, since it is chosen by the agents in the model; it is also clear that we can, for a given model, often easily find sufficient conditions that guarantee the restriction holds.

where $\Psi = A^2\Sigma + \Omega$, and use (19) to find the Kalman gain (θ):

$$\theta = \Sigma\Lambda^{-1} = 1 - \Sigma\Psi^{-1}. \tag{23}$$

Substituting (21) and (22) into (23), we have

$$\theta = 1 - \frac{1}{\exp(2\kappa)}. \tag{24}$$

Note that (22) and (23) also hold in the SE problem. To compare the RI and SE problems in the univariate case, we first make the following assumption.

Assumption 1. Assume that Λ is fixed exogenously in the SE case.

Under Assumption 1, it is clear that in the SE problem given Λ and Ω, we can compute Σ by solving the nonlinear equation (14). After obtaining Σ, we can use (15) to determine the Kalman gain θ; thus, in this sense SE and RI have the same implications.

We now discuss how to use a policy experiment to distinguish RI from SE. Suppose that the variance of the exogenous shock, Ω, is scaled up due to a change in policy. In the SE problem with fixed Λ, Equations (22) and (23) imply that an increase in Ω will generally lead to a different solution for Σ and θ; consequently, the change in policy will lead to a change in the model's dynamics. Since Σ is a nonlinear function of Ω, the effect of changes in Ω on Σ could be complicated. In the next section, we will explore this relationship using some numerical examples in a permanent income model. In contrast, in the RI problem, if κ is *fixed*, (21), (22), and (23) imply that a change in Ω will lead to the same change in Σ, Ψ, and Λ, but has no impact on θ. In other words, agents with fixed capacity will behave as if facing noise whose nature changes systematically as the dynamic properties of the economy change, i.e., the change in policy does not change the model's dynamics.

Assumption 2. Assume that the signal-to-noise ratio (SNR)*, $\Omega\Lambda^{-1}$, is fixed exogenously in the SE case.*

Note that Equation (22) can be rewritten as

$$\Omega\Lambda^{-1} = \Omega\Sigma^{-1} - \left[A^2\left(\Omega\Sigma^{-1}\right)^{-1} + 1\right]^{-1}. \tag{25}$$

Under Assumption 2, since the SNR is fixed, 25) can be used to solve for $\Omega\Sigma^{-1}$. Given the SNR and $\Omega\Sigma^{-1}$, we can compute

$$\Sigma\Lambda^{-1} = \left(\Sigma\Omega^{-1}\right)\left(\Omega\Lambda^{-1}\right). \tag{26}$$

Consider the same case in which Ω is scaled up, Assumption 2 means that the exogenous noise should also be scaled up such that $\Omega\Lambda^{-1}$ is fixed at the same level; consequently, (25) leads to the same solution for $\Omega\Sigma^{-1}$ and (26) leads to the same $\Sigma\Lambda^{-1}$. The following proposition summarizes the main conclusion in this case:

Proposition 1 *Under Assumption 2 (i.e., the* SNR *is fixed), the SE and RI problems are observationally equivalent in the sense that they lead to the same dynamics if* κ *is fixed and* $\Sigma\Lambda^{-1}$ *in the SE problem is equal to* $1 - 1/\exp(2\kappa)$ *in the RI problem.*

Proof. *The proof is straightforward by comparing (24) and (26).* ■

In the above analysis, for simplicity we assume that κ remains unchanged when Ω is affected by the government policy. However, if an increase in Ω leads to higher marginal welfare losses due to imperfect observations, some capacity may be reallocated from other sources to reduce the welfare losses due to low capacity. In this case, θ will change accordingly as it is completely determined by capacity κ; consequently, the dynamic behavior of the model will also change in response to the change in Ω. We will further explore this issue in the next subsection and the permanent income model examined in Section 4.

3.1.1 Alternative Way to Model Limited Information-Processing Capacity

As argued in Sims (2010), instead of using fixed finite channel capacity to model limited information-processing ability, it is also reasonable to assume that the marginal cost of information processing is constant. That is, the Lagrange multiplier on (20) is constant. In the univariate case, if the decision rule under full information is $c_t^* = Hs_t$ and the objective of the agent with finite capacity is to minimize $\sum_{t=0}^{\infty} \beta^t (c_t - c_t^*)^2$, the optimization problem reduces to

$$\min_{\Sigma_t} \sum_{t=0}^{\infty} \beta^t \left[H^2\Sigma_t + \lambda \ln \left(\frac{A^2\Sigma_{t-1} + \Omega}{\Sigma_t} \right) \right],$$

where Σ_t is the conditional variance at t, λ is the Lagrange multiplier corresponding to (20), and we impose the restriction that $\beta A = 1$ for simplicity. Solving this problem yields the optimal conditional variance:

$$\Sigma_t = \Sigma = \frac{-(\Omega H - \lambda A) + \sqrt{(\Omega H - \lambda A)^2 + 4\lambda\Omega A^2}}{2HA^2} > 0. \tag{27}$$

It is straightforward to show that as λ goes to 0, $\Sigma = 0$; and as λ goes to ∞, $\Sigma = \infty$. Comparing (27) with (21), it is clear that the two modeling strategies are *observationally equivalent* in the sense that they lead to the same conditional variance if the following equality holds:

$$\kappa = \frac{1}{2}\ln\left(R^2 + \frac{2HA^2}{-\left[H - A(\lambda/\Omega)\right] + \sqrt{\left[H - A(\lambda/\Omega)\right]^2 + 4A^2(\lambda/\Omega)}}\right). \quad (28)$$

It is obvious that as κ converges to its lower limit $\underline{\kappa} = \ln(R)$ as λ goes to ∞; and it converges to ∞ as λ goes to 0.[11] In other words, using this RI modeling strategy, the agent is allowed to adjust the optimal level of capacity in such a way that the marginal cost of information-processing for the problem at hand remains constant. Note that this result is consistent with the concept of 'elastic' capacity proposed in Kahneman (1973).

Furthermore, it is clear from (28) that if the cost of information processing (λ) is fixed, an increase in fundamental uncertainty (Ω) will lead to higher capacity (κ) devoted to monitoring the evolution of the state. We now consider the same policy experiment discussed above: Ω is scaled up due to a change in policy. If we adopt the assumption that λ is *fixed*, (27) means that there is no change in Σ because $\frac{\partial \ln \Sigma}{\partial \ln \Omega} < 1$. (Note that in the fixed κ case, $\frac{\partial \ln \Sigma}{\partial \ln \Omega} = 1$.) Consequently, a change in Ω will change θ and the model's dynamics if the inattentive agent is facing fixed marginal cost of information. Therefore, different ways to model RI may lead to different policy implications.[12]

3.1.2 Extension to Correlated Shocks and Noises

In the above analysis, we assumed that the exogenous fundamental shock and noise are uncorrelated. We now discuss how *correlated* shocks and noises affect the implications of SE and RI for the model's dynamic behavior. In real systems, we may observe correlated shocks and noises. For example, if the system is an airplane and winds are buffeting the plane, the random gusts of wind affect both the process (the airplane dynamics) and the measurement (the sensed wind speed) if people use an anemometer to measure wind speed as an input to the Kalman filter.

[11] We require here that $H \neq 0$; that is, the state must be *detectable*.

[12] Note that these two different ways to model RI is very similar to the *constraint* and *multiplier* preferences adopted by Hansen and Sargent (2007) to model robustness. They also established the observational equivalence between the two preferences within the LQG setting.

It is straightforward to introduce correlated shocks and noises into the SE problem. Specifically, we consider the case in which the process shock (ε) and the noise (ξ) are correlated as follows:

$$\operatorname{corr}(\varepsilon_{t+1}, \xi_{t+1}) = \rho,$$
$$\operatorname{cov}(\varepsilon_{t+1}, \xi_{t+1}) = \Gamma = \rho\sqrt{\Omega}\sqrt{\Lambda},$$

where ρ is the correlation coefficient between ε_{t+1} and ξ_{t+1}, $\Omega = \operatorname{var}[\varepsilon_{t+1}]$ and $\Lambda = \operatorname{var}[\xi_{t+1}]$. Under SE, Λ is given exogenously and the correlation just introduces another exogenous stochastic dimension on the noise. As shown in Simon (2006), in this case the optimal Kalman gain can be written as

$$\theta = (\Psi + \Gamma)(\Psi + \Lambda + 2\Gamma)^{-1}, \tag{29}$$

and the updating formula for the conditional variance is:

$$\Sigma = \Psi - (\Psi + \Gamma)^2 (\Psi + \Lambda + 2\Gamma)^{-1}, \tag{30}$$

where $\Psi = \Omega + A^2\Sigma$. Just like the case without the correlation, given Λ and Γ, (29) and (30) jointly determine the steady state (θ, Σ).

In the RI problem, the correlation generalizes the assumption in Sims (2003) on the uncorrelated RI-induced noise. The presence of correlation between shocks and noises does not affect the conditional variance Σ since $\Sigma = \frac{\Omega}{A^2 - \exp(2\kappa)}$. In the steady state, (30) can be rewritten as the following quadratic equation in terms of $\sqrt{\Lambda}$: $\left[\rho^2\Omega - (\Psi - \Sigma)\right]\Lambda + 2\rho\Sigma\sqrt{\Omega}\sqrt{\Lambda} + \Sigma\Psi = 0$, which can be solved for

$$\sqrt{\Lambda} = \frac{-\rho\Sigma\sqrt{\Omega} + \sqrt{\rho^2\Sigma^2\Omega - \Sigma\Psi\left[\rho^2\Omega - (\Psi - \Sigma)\right]}}{\rho^2\Omega - (\Psi - \Sigma)}. \tag{31}$$

It is clear from (31) that if κ is fixed, the change in Ω will lead to the same change in Σ, Ψ, and Λ, but has no effect on the Kalman gain $\theta = \Sigma\Lambda^{-1}$. That is, the presence of correlated noise does not change the dynamic behavior of the model.

3.2 The Multivariate Case

In the multivariate RI problem, it is much more difficult to determine the steady state conditional variance-covariance matrix Σ because it cannot be computed analytically. Here we follow Sims (2003) and calculate the expected welfare loss due to imperfect observations under RI. Specifically, we assume that the value functions under full information and imperfect information can be written as

$$v(s_t) = s_t^T P s_t \text{ and } \widehat{v}(\widehat{s}_t) = \widehat{s}_t^T \widehat{P} \widehat{s}_t,$$

respectively.[13] We can compute the optimal Σ by minimizing the expected welfare loss due to RI,

$$E_t\left[v(s_t) - \widehat{v}(\widehat{s}_t)\right], \tag{32}$$

subject to information-processing constraints. Note that to solve this problem numerically, we need to use a two-stage procedure.[14] First, under the linear-quadratic-Gaussian assumption, the certainty equivalence principle applies and the decision rule under imperfect information,

$$c_t^* = -F\widehat{s}_t, \tag{33}$$

is independent of Σ or Λ. We then use this decision rule to determine $\widehat{v}(\widehat{s}_t)$ which depends on Σ and Λ. Applying the welfare criterion proposed in (32), we can solve for optimal steady state Σ and Λ.

Solving the problem posed in (32) is equivalent to solving the semidefinite programming problem

$$\max_{\Sigma}\left\{\operatorname{trace}(-Z\Sigma)\right\} \tag{34}$$

subject to

$$-\log(|\Sigma|) + \log\left(\left|A^T\Sigma A + \Omega\right|\right) \leq 2\kappa, \tag{35}$$

$$A^T\Sigma A + \Omega \succeq \Sigma, \tag{36}$$

where $Z = F^T RF - 2F^T W + \beta\left(F^T B^T PBF + F^T B^T PA + A^T PBF\right)$ (see online appendix 1.1 for the derivation). If the positive-definiteness constraint on $A^T\Sigma A + \Omega - \Sigma$, (36), does not bind, the first-order condition for Σ can be written as follows:

$$Z = \lambda\left[A\left(A\Sigma A^T + \Omega\right)^{-1} A^T - \Sigma^{-1}\right], \tag{37}$$

which can be reduced to:

$$\Sigma^{-1} = \left(G\Sigma G^T + G_0\right)^{-1} - \frac{Z}{\lambda}, \tag{38}$$

[13] See also Maćkowiak and Wiederholt (2009).

[14] Sims (2010) also applied this principle solve a tracking problem with information constraints.

where $G = \left(A^T\right)^{-1} A$ and $G_0 = \left(A^T\right)^{-1} \Omega A^{-1}$. We can then use standard methods to solve (38). When applied to a permanent income model in the next section, we first solve this equation and then check whether in fact (36) is satisfied by the optimal solution of Σ. If so, the problem is solved.[15]

After computing the optimal steady state Σ, we can then use (18) to determine the steady state Λ and (19) to determine the Kalman gain θ. Therefore, the key difference between SE and RI is that in the SE problem we need to specify the process of the noise first, whereas in the RI problem we need to first specify the value of channel capacity that determines the steady state conditional variance of the state by solving the semidefinite programming problem proposed in (34) subject to (35) and (36).[16] Theoretically, it is clear that after solving an RI problem, we can always reconstruct a SE problem using the resulting endogenous noise due to RI as the input, and the two models are observationally equivalent in this sense. However, it is difficult to specify the process of the vector of noises *ex ante* when modeling the multivariate SE problem.[17]

When modeling the multivariate RI problem we only need to set a value for channel capacity and then compute optimal conditional variance-covariance matrices of the state and the variance-covariance matrices of the noise vector by solving the constrained semidefinite minimization problem (34). Therefore, in the multivariate RI problem, the agent's preference, budget constraint, and information-processing constraints jointly determine the values of Σ, Λ, and θ, whereas in the multivariate SE problem given Λ, (18) that is used to determine Σ and θ only depends on the budget constraint. If the noise in SE is specified exogenously, it may violate the optimality conditions for RI; for example, Melosi (2009) showed that a particular estimated SE model does not equate the marginal utility of attention across states, implying that the variance-covariance matrix of the noise would not be consistent with *any* channel capacity. Of course, obtaining the marginal utility of attention requires solving the RI problem, so it will be difficult to specify *ex ante* an SE problem consistent with RI.

[15] If this constraint does not bind, in principle we can apply the logic of the reverse water-filling problem to solve for Σ. In practice, this method has proven numerically unstable.

[16] Note that the basic idea of solving the multivariate RI problem is the same as that in the univariate model and thus the key difference between SE and RI problems remains unchanged.

[17] This problem will be particularly difficult for non-LQG problems, since the distribution of the noise shocks will generally be impossible to specify analytically.

We now consider the different policy effects of RI and SE in the multivariate case. We first assume that initially the SE and RI problems have the same Kalman gain that generates the same dynamic behavior. Suppose that the variance-covariance matrix of the exogenous shock, Ω, is scaled up due to a change in policy.[18] In the SE problem with fixed Λ, Equations (22) and (23) imply that a change of Ω will lead to a different solution for Σ and θ, i.e., the change in policy will lead to a change in the model's dynamics. In contrast, in the multivariate RI problem, as shown in (34)-(36), a change in Ω will have complicated effects on Σ, Λ, and θ. In other words, in the multivariate case a change in policy will affect the model's behavior in both SE and RI problems. (Note that in the univariate case the change in policy does not change the model's dynamics.)

We next consider the effects of RI and SE under Assumption 2 (i.e., the SNR, $\Omega\Lambda^{-1}$, is fixed in the SE problem). As before, we assume that initially the SE and RI problems have the same Kalman gain. To illustrate how a change in Ω affects the Kalman gain in RI and SE problems under Assumption 2, we multiply Σ on both sides of (22):

$$\Sigma\Lambda^{-1} = I - \left[A\Sigma A^T\Sigma^{-1} + \left(\Omega\Lambda^{-1}\right)\left(\Lambda\Sigma^{-1}\right)\right]^{-1}, \tag{39}$$

where I is the identity matrix and we use the fact that $\Omega\Sigma^{-1} = \left(\Omega\Lambda^{-1}\right)\left(\Lambda\Sigma^{-1}\right)$. Under Assumption 2, the policy has the same impact on Ω and Λ to keep the SNR fixed. (39) clearly shows that if the policy changes Σ and then $A\Sigma A^T\Sigma^{-1}$, it will affect $\theta = \Sigma\Lambda^{-1}$ even under Assumption 2. Multiplying Ω on both sides of (22) gives

$$\Omega\Lambda^{-1} = \Omega\Sigma^{-1} - \left(A\Sigma A^T\Omega^{-1} + I\right)^{-1}, \tag{40}$$

which means that a change in Ω will lead to different Σ given that $\Omega\Lambda^{-1}$ is fixed. Note that in the univariate case, $A\Sigma A^T\Sigma^{-1} = A^2$, which means that the policy has no impact on θ, and the SE and RI problems cannot be distinguished by the policy under Assumption 2 that the SNR, $\Omega\Lambda^{-1}$, is fixed.

4 Applications to the Permanent Income Model

In this section we consider the effects of SE and RI for consumption dynamics and their policy and welfare implications in an otherwise standard permanent income model. As in the previous

[18] That is, all elements in the variance-covariance matrix are scaled up.

section we first consider applications to the univariate case and then discuss applications to the multivariate case.

We are not the first to examine this model. Sims (2003) examined how RI affects consumption dynamics when the agent only has limited capacity when processing information. Luo (2008) showed that the RI permanent income can be solved explicitly even if the income process is not iid, and then examines how RI can resolve the well-known excess smoothness and excess sensitivity puzzles; that model admits a reduction to a single state variable.[19]

4.1 The Univariate Case

Following Luo (2008), we have the following univariate version of the standard permanent income model (Hall 1978, Flavin 1981) in which households solve the dynamic consumption-savings problem

$$v(s_0) = \max_{\{c_t\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t) \right] \tag{41}$$

subject to

$$s_{t+1} = R s_t - c_t + \zeta_{t+1}, \tag{42}$$

where $u(c_t) = -\frac{1}{2}(c_t - \overline{c})^2$ is the period utility function, $\overline{c} > 0$ is the bliss point, c_t is consumption,

$$s_t = w_t + \frac{1}{R} \sum_{j=0}^{\infty} R^{-j} E_t \left[y_{t+j} \right] \tag{43}$$

is permanent income (the expected present value of lifetime resources), consisting of financial wealth (w_t) plus human wealth,

$$\zeta_{t+1} = \frac{1}{R} \sum_{j=t+1}^{\infty} \left(\frac{1}{R} \right)^{j-(t+1)} \left(E_{t+1} - E_t \right) \left[y_j \right]; \tag{44}$$

is the time $(t+1)$ innovation to permanent income with mean 0 and variance ω_ζ^2, w_t is cash-on-hand (or market resources), y_t is a general income process with Gaussian white noise innovations, β is the discount factor, and R is the constant gross interest rate at which the consumer can borrow and lend freely. We assume y follows an AR(1) process with persistence coefficient $\rho \in [0, 1]$,

[19] The excess smoothness puzzle states that consumption responds too little to permanent changes in income. The excess sensitivity puzzle states that current consumption responds to changes in income that were anticipated in earlier periods.

$y_{t+1} = \rho y_t + \varepsilon_{t+1}$, where $\varepsilon_{t+1} \sim N\left(0, \omega^2\right)$, $\zeta_{t+1} = \varepsilon_{t+1}/\left(R - \rho\right)$. For the rest of the paper we will restrict attention to points where $c_t < \overline{c}$, so that utility is increasing and concave; following the literature we impose the restriction that $\beta R = 1$, because it implies a stationary path for consumption. This specification follows that in Hall (1978) and Flavin (1981) and implies that optimal consumption is determined solely by permanent income:

$$c_t = (R - 1)\, s_t. \tag{45}$$

Within this LQG setting, the certainty equivalence principle holds and introducing SE or RI lead to the following new consumption function:

$$c_t = (R - 1)\, \widehat{s}_t, \tag{46}$$

where $\widehat{s}_t = E_t\left[s_t\right]$ is the perceived state and is governed by the following Kalman filtering equation

$$\widehat{s}_{t+1} = (1 - \theta)\left(R\widehat{s}_t - c_t\right) + \theta\left(s_{t+1} + \xi_{t+1}\right), \tag{47}$$

where θ is the Kalman gain, and given $s_0 \sim N\left(\widehat{s}_0, \sigma^2\right)$. As shown in Luo (2008), combining (42), (46), with (47) yields the following expression for the change in consumption:

$$\Delta c_t = (R - 1)\left[\frac{\theta\zeta_t}{1 - (1 - \theta)R \cdot L} + \theta\left(\xi_t - \frac{\theta R \xi_{t-1}}{1 - (1 - \theta)R \cdot L}\right)\right], \tag{48}$$

where L is the lag operator. We require $(1 - \theta)R^2 < 1$, the model equivalent of the stabilizability condition stated before (this condition implies $(1 - \theta)R < 1$ since $R > 1$). This MA(∞) process shows that the dynamic behavior of the model is strongly influenced by the Kalman gain θ. Using the explicit expression for consumption growth (48), we can compute the key stochastic properties of consumption process: the volatility of consumption growth, the persistence of consumption growth, and the correlation between consumption growth and income shocks.[20] All these moments depend on the Kalman gain. In other words, SE and RI lead to different consumption processes if and only if the resulting θ differs.

4.1.1 Policy Implications under SE and RI

In this univariate permanent income model, substituting $A = R$ into Equation (25),

$$\omega_\zeta^2 \Lambda^{-1} = \omega_\zeta^2 \Sigma^{-1} - \left[R^2\left(\omega_\zeta^2 \Sigma^{-1}\right)^{-1} + 1\right]^{-1}, \tag{49}$$

[20] See Luo (2008) for a discussion on the effects of RI on consumption dynamics.

where $\omega_\zeta^2\Sigma^{-1} = \left(\omega_\zeta^2\Lambda^{-1}\right)\left(\Lambda\Sigma^{-1}\right)$, and denote $\theta = \Sigma\Lambda^{-1}$ and $\pi = \omega_\zeta^2\Lambda^{-1}$, we obtain the following relation between π and θ:

$$\pi = \theta\left(\frac{1}{1-\theta} - R^2\right). \tag{50}$$

Solving for θ yields

$$\theta = \frac{-\left(1+\pi\right) + \sqrt{\left(1+\pi\right)^2 + 4R^2\left(\pi + R^2\right)}}{2R^2}, \tag{51}$$

where we omit the negative root of θ as both Σ and Λ must be positive. Figure 1 below illustrates the relationship between π and θ given $R = 1.02$ and $\pi \in [0.1, 10]$. It clearly shows that θ is an increasing function of π, the signal-to-noise ratio. Also, as $\pi \to \infty$, $\theta \to 1$.

In the RI version of the permanent income model, we have

$$\Sigma = \frac{\Omega}{\exp\left(2\kappa\right) - R^2}, \tag{52}$$

$$\Lambda = \left(\Sigma^{-1} - \Psi^{-1}\right)^{-1}, \tag{53}$$

where $\Psi = R^2\Sigma + \Omega$. Using (52) and (53), the Kalman filter gain under RI can be written as

$$\theta = \Sigma\Lambda^{-1} = 1 - \frac{1}{\exp\left(2\kappa\right)}. \tag{54}$$

Comparing (51) with (54), it is clear that the signal to noise ratio (π) and the level of channel capacity (κ) have one-to-one correspondence. Figure 2 shows that the relationship between κ and π when the SE and RI problems are observationally equivalent in the sense that they lead to the same consumption dynamics governed by the Kalman gain θ. This result is consistent with the general conclusion we obtained using Assumption 2 in the previous section.

Using the same expression for θ, (51), we can examine how Assumption 1 can be used to distinguish SE and RI when implementing a change in government policy. Specifically, in the SE problem, we assume that before the government implements stabilization policies, the signal-to-noise ratio $\pi = \frac{\omega_\zeta^2}{\Lambda} = 2$. In this case, $\theta = 0.79$. After the government implements these policies, the variance of the shock to permanent income will be reduced from ω_ζ^2 to $0.5\omega_\zeta^2$. Since Λ is fixed under Assumption 1, $\pi = \frac{\omega_\zeta^2}{\Lambda}$ will fall from 2 to 1; consequently, $\theta = 0.74$. We now assume that the RI and SE problems are observationally equivalent in the sense that they lead to the same $\theta = 0.79$ *before* implementing the stabilization policies. After implementing these policies, ω_ζ^2 will be scaled down to $0.5\omega_\zeta^2$, and the RI theory predicts that both Σ and Λ will be scaled down to 0.5Σ and

0.5Λ, respectively.[21] Consequently, the Kalman filter gain, $\theta = \Sigma\Lambda^{-1} = 0.79$, remains unchanged. In other words, stabilization policy have different implications for consumption dynamics in the SE and RI models.

Alternatively, if we assume that the cost of information processing (λ) is fixed, the optimal conditional variance equals

$$\Sigma = \frac{-\left[\Omega\left(R-1\right)-\lambda R\right]+\sqrt{\left[\Omega\left(R-1\right)-\lambda R\right]^{2}+4\lambda\Omega R^{2}}}{2\left(R-1\right)R^{2}}. \tag{55}$$

Comparing (55) with (52), it is clear that the two modeling strategies are *observationally equivalent* in the sense that they lead to the same conditional variance if the following equality holds:

$$\kappa = \frac{1}{2}\ln\left(R^{2}+\frac{2\left(R-1\right)R^{2}}{-\left[\left(R-1\right)-R\left(\lambda/\omega_{\zeta}^{2}\right)\right]+\sqrt{\left[\left(R-1\right)-R\left(\lambda/\omega_{\zeta}^{2}\right)\right]^{2}+4R^{2}\left(\lambda/\omega_{\zeta}^{2}\right)}}\right). \tag{56}$$

In this case, the Kalman gain is

$$\theta = 1-\left\{R^{2}+\frac{2\left(R-1\right)R^{2}}{-\left[\left(R-1\right)-R\left(\lambda/\omega_{\zeta}^{2}\right)\right]+\sqrt{\left[\left(R-1\right)-R\left(\lambda/\omega_{\zeta}^{2}\right)\right]^{2}+4R^{2}\left(\lambda/\omega_{\zeta}^{2}\right)}}\right\}^{-1}. \tag{57}$$

After implementing these policies, ω_{ζ}^{2} is scaled down to $0.5\omega_{\zeta}^{2}$, and the fixed λ theory predicts that the Kalman filter gain, $\theta = \Sigma\Lambda^{-1}$, is reduced. For example, before the government implements stabilization policies, we have $\lambda/\omega_{\zeta}^{2} = 0.000135$ and $\theta = 0.79$. After the policy, we can easily calculate that $\theta = 0.68$ using (57). Figure 3 plots the different implications of SE and RI for consumption dynamics after implementing the stabilization policy: consumption growth falls more (less) under SE than RI when κ is fixed (when λ is fixed), since the Kalman gain decreases (increases).[22]

4.1.2 Welfare Effects of Imperfect Information under SE and RI

Given the restriction that $\beta R = 1$, the value function for the RI or SE models is

$$\widehat{v}\left(\widehat{s}_{0}\right) = -\frac{\left(R-1\right)R}{2}\widehat{s}_{0}^{2}+R\overline{c}\widehat{s}_{0}-\frac{1}{2}R\left(\frac{1}{R-1}\overline{c}^{2}+\omega_{\eta}^{2}\right), \tag{58}$$

[21] A proof is straightforward from Expressions (52) and (53).

[22] In principle one might be able to use policy changes to distinguish the models, although in practice identification would seem to be a serious obstacle.

where

$$\omega_\eta^2 = \text{var}\left[\eta_{t+1}\right] = \frac{\theta}{1-(1-\theta)R^2}\omega_\zeta^2 > \omega_\zeta^2, \tag{59}$$

and

$$\eta_{t+1} = \theta\left[\left(\frac{\zeta_{t+1}}{1-(1-\theta)R\cdot L}\right) + \left(\xi_{t+1} - \frac{\theta R\xi_t}{1-(1-\theta)R\cdot L}\right)\right]. \tag{60}$$

With the Kalman gain $\theta < 1$, $\omega_\eta^2 > \omega_\zeta^2$ and it then follows that ω_η^2 is decreasing in θ.

Proposition 2 $\frac{\partial\omega_\eta^2}{\partial\theta} < 0$.

Proof. By simple calculation we obtain

$$\frac{\partial\omega_\eta^2}{\partial\theta} = \frac{\left(1-R^2\right)\omega_\zeta^2}{\left[1-(1-\theta)R^2\right]^2} < 0 \tag{61}$$

because $R > 1$ and $1-(1-\theta)R^2 > 0$. ■

The value function (58), together with (59) and (61) clearly show that imperfect observations due to finite capacity lead to more uncertainty about the state, which thus increases welfare losses. More importantly, they also show that after implementing the stabilization policy discussed in the preceding subsection, SE and RI will lead to different welfare implications of imperfect observations. The reason is that the policy will lead to different Kalman gain in the SE and RI problem and thus affects the welfare losses due to imperfect information. Using the same example discussed above, when ω_ζ^2 is reduced to $0.5\omega_\zeta^2$, $\theta = 0.79$ under RI when κ is fixed, where it is equal to 0.74 under SE. That is, given the same initial conditions, the stabilization policy will reduce the welfare of agents under SE, whereas it has no effect on agents with RI. The intuition is that the stochastic property of the noise in the RI problem changes accordingly in response to the change in the policy.

Since imperfect information about the state cannot help in decision making, we can use an alternative welfare criterion, the expected welfare gap between the unconstrained value function and the constrained value function conditional on the processed information at the current period, to evaluate different welfare implications of SE and RI.[23] Note that the unconstrained value function in the full-information case can be written as

$$v\left(s_t\right) = -\frac{(R-1)R}{2}s_t^2 + R\overline{c}s_t - \frac{1}{2}R\left(\frac{1}{R-1}\overline{c}^2 + \omega_\zeta^2\right). \tag{62}$$

[23] Imperfect information cannot make a single decision-maker better off, but of course that result need not hold when agents interact.

The expected welfare gap due to imperfect information, $\Delta(\Sigma, \theta)$, can thus be written as follows:

$$\begin{aligned}\Delta(\Sigma,\theta) &= E_t\left[v(s_t) - \widehat{v}(\widehat{s}_t)\right] \\ &= -\frac{(R-1)R}{2}\operatorname{var}_t[s_t] - \frac{1}{2}R\left(\omega_\zeta^2 - \omega_\eta^2\right) \\ &= -\frac{(R-1)R}{2}\Sigma + \frac{1}{2}R\left[\frac{\theta}{1-(1-\theta)R^2} - 1\right]\omega_\zeta^2,\end{aligned} \tag{63}$$

where the expectation, $E_t[\cdot]$, is conditional on processed information at time t and Σ is the conditional variance of the state. (See online appendix 1.2 for the derivation.) From (63), we can clearly see that imperfect information affects the welfare gap via two channels:

1. The post-observation variance, i.e., the conditional variance of the state, Σ. The first term in (63) means that a larger conditional variance will decrease welfare. The intuition behind this result is that $\widehat{s}_t^2$ in the constrained value function is in the time-t information set, while s_t^2 in the constrained value function is not in the information set; consequently, $E_t\left[s_t^2\right] > \widehat{s}_t^2$. However, when R is close to 1, this term is close to 0 and thus has little effect on welfare.

2. The innovation to the level of perceived permanent income, η_{t+1}, is more volatile than that to the level of actual permanent income, ζ_{t+1}. That is, $\omega_\eta^2 > \omega_\zeta^2$. Therefore, the second term in (63) means that imperfect information will reduce welfare by increasing the volatility of the innovation to the perceived state. Note that the unconstrained value function is determined by the dynamics of the actual state s_t:

$$s_{t+1} = Rs_t - c_t + \zeta_{t+1}, \tag{64}$$

whereas the constrained value function is determined by the dynamics of $\widehat{s}_t$:

$$\widehat{s}_{t+1} = R\widehat{s}_t - c_t + \eta_{t+1}. \tag{65}$$

As will be shown later in (67), the second channel will dominate the first channel and thus imperfect information always leads to welfare losses.

To evaluate the different effects of SE and RI on the welfare losses under government stabilization policies, we divide ω_ζ^2 on both sides of (63) and obtain

$$\widetilde{\Delta}(\pi, K) \triangleq \frac{\Delta(\Sigma,\theta)}{\omega_\zeta^2} = -\frac{(R-1)R}{2}\frac{\theta}{\pi} + \frac{1}{2}R\left[\frac{\theta}{1-(1-\theta)R^2} - 1\right], \tag{66}$$

where we use the fact that $\frac{\Sigma}{\omega_\zeta^2}=\frac{\Sigma}{\Lambda}\frac{\Lambda}{\omega_\zeta^2}=\frac{\theta}{\pi}$, where $\pi=\omega_\zeta^2/\Lambda$ is defined as the signal-to-noise ratio. Using (66), we can examine how stabilization policies affect the welfare losses under SE and RI. Specifically, after implementing the stabilization policy (i.e., ω_ζ^2 is reduced to $0.5\omega_\zeta^2$), SE and RI will lead to different welfare losses via Σ and θ. Under RI, when κ is fixed and ω_ζ^2 is reduced to $0.5\omega_\zeta^2$, the conditional variance will fall from Σ to 0.5Σ but $\theta=0.79$ will remain unchanged. (66) therefore implies that $\frac{\Delta(\Sigma,\theta)}{\omega_\zeta^2}$ will remain the same after implementing the policy as RI has no impact on the signal-to-noise ratio π. In contrast, under SE, under Assumption 1 (i.e., Λ is fixed), θ will be reduced to 0.74. (66) therefore implies that $\frac{\Delta(\Sigma,\theta)}{\omega_\zeta^2}$ will be reduced after implementing the policy as π will fall to 0.5π. That is, given the same initial conditions ($\theta=0.79$), the stabilization policy will lead to smaller welfare losses in the SE model than in the RI model. Note that under RI, substituting $\Sigma=\frac{\omega_\zeta^2}{1/(1-\theta)-R^2}$, into (63), we can further simplify (63) as follows:

$$\Delta(\Sigma,K)=\frac{1}{2}\omega_\zeta^2(R-1)\frac{(1-\theta)R^2}{1-(1-\theta)R^2}, \tag{67}$$

which means that given ω_ζ^2, the welfare loss is decreasing with the Kalman gain, i.e., is increasing with the degree of inattention (see online appendix for the derivation).

Following Barro (2007), we use (58) to compute the welfare effects of changes in channel capacity and compare them with those from proportionate changes in the initial level of the perceived state ($\widehat{s}_0$). Specifically, the relative *marginal* welfare losses (rmw) due to imperfect information *at different capacity* (κ) can be written as

$$\text{rmw}=\frac{\partial\widehat{v}/\partial\kappa}{(\partial\widehat{v}/\partial\widehat{s}_0)\,\widehat{s}_0}=-\frac{1}{-(R-1)\,\widehat{s}_0^2+\widehat{c}\widehat{s}_0}\frac{\left(1-R^2\right)\exp(-2\kappa)\,\omega_\zeta^2}{\left[1-\exp(-2\kappa)\,R^2\right]^2}>0, \tag{68}$$

where $\frac{\partial\widehat{v}}{\partial\widehat{s}_0}=-(R-1)\,R\widehat{s}_0+R\overline{c}>0$, $\frac{\partial\widehat{v}}{\partial\kappa}=-\frac{1}{2}R\frac{\partial\omega_\eta^2}{\partial K}\frac{\partial K}{\partial\kappa}>0$ is evaluated for given $\widehat{s}_0$, Expression (68) gives the proportionate increase in $\widehat{s}_0$ that compensates, at the margin, for an reduction in capacity κ devoted to monitoring the state, in the sense of preserving the lifetime utility. Using Expression (68), it is straightforward to show that $\frac{\partial(\text{rmw})}{\partial\kappa}<0$. Denote by $f(\kappa)=\frac{\exp(-2\kappa)}{[1-\exp(-2\kappa)R^2]^2}$. It is clear that only the $f(\kappa)$ term is important for the effects of RI on rmw . Figure 4 illustrates how capacity κ affects $f(\kappa)$ and rmw when $R=1.01$. It clearly shows that RI can have significant effects the relative marginal welfare losses when capacity devoted to monitoring the state is low. For example, when $\kappa=0.2$ nats, $f=16.6$, whereas $f=0.45$ when $\kappa=1$ nat.

To do quantitative welfare analysis, we need to know the level of the initial level of permanent income, $\widehat{s}_0$. For simplicity we assume that $\widehat{s}_0$ is just mean permanent income. To compute $\widehat{s}_0$, denote by γ the local coefficient of relative risk aversion, which equals $\gamma = \frac{E[y]}{\overline{c}-E[y]}$ for the utility function $u(\cdot)$ evaluated at mean income $E[y]$. Here we impose $\beta = 0.9971$ such that the annual real interest rate is 2.5%. We then follow the procedure used in Hansen and Sargent (2004) and use the estimated one-factor endowment process as follows

$$y_{t+1} = 0.9992 y_t + \varepsilon_{t+1}, \tag{69}$$

and ε_{t+1} follows an iid process distributed as $N\left(0, 5.5819^2\right)$. Here we set the coefficient of variation of endowment, $\mathrm{sd}[y_t]/E[y_t]$, to be 0.1, which can be used to compute the mean income level $E[y] = 1396$ and then the value of the bliss point $\overline{c}$ that generates reasonable relative risk aversion γ. For example, when the local CRRA γ is set to 1, we have $\overline{c} = 2E[y] = 2792$. Furthermore, assume that the ratio of mean financial wealth to mean labor income, $E[w]/E[y]$, is 5.[24] Since $s_t = w_t + \frac{1}{R-0.9992} y_t$ and $\zeta_t = \frac{\varepsilon_t}{R-0.9992}$ we have $E[s] = \left(5 + \frac{1}{R-0.9992}\right) E[y]$ .Given this specification and set the values of κ and R, we can use (68) to compute the welfare effects of finite capacity quantitatively. Figure 5 illustrates the values of rmw at different capacity for given γ. We can see that rmw is decreasing with κ, i.e., the proportionate increase in $\widehat{s}_0$ that compensates for an reduction in κ in the sense of preserving the lifetime utility is increasing with the degree of inattention.[25] For example, given $\gamma = 1$, when $\kappa = 0.2$ nats, rmw $= 2.7127 \cdot 10^{-2}\%$, whereas rmw $= 9.0105 \cdot 10^{-4}\%$ when $\kappa = 1$ nat. That is, if the agent's capacity is reduced from 1 bit to 0.2 nats, the proportionate increase in $\widehat{s}_0$ that compensates for an reduction in κ in the sense of preserving the expected utility will be increased by about 30 times. Hence, if the level of $\widehat{s}_0$ is large, the agent would have strong incentive to reallocate more capacity to monitor this state if he is allowed to adjust his capacity.

After the government implements the stabilization policies that reduces the variance of the shock from ω_ζ^2 to $0.5\omega_\zeta^2$, the economy switches to a more stable environment. If we relax the assumption that κ is fixed and assume that the marginal cost of information λ is fixed, the corresponding κ will be reduced from 0.78 to 0.57 nats, i.e., some capacity will be reallocated to other sources to increase the economic efficiency because a reduction in macroeconomic uncertainty

[24] This number varies substantially for different individuals, from 2 to 20. 5 is the average wealth/income ratio in the Survey of Consumer Finances 2001.

[25] In addition, given κ, rmw is increasing with γ. That is, agents who are more risk averse require more compensation for a reduction in capacity to maintain the initial level of expected utility.

leads to smaller marginal welfare losses due to RI. In this case, the Kalman gain θ will fall from 0.79 to 0.68 accordingly; consequently, the dynamic behavior in the RI model will also change in response to the change in ω_ζ^2.

4.2 The Multivariate Case

In this section we solve for optimal steady state Σ and Λ in a parametric multivariate RI permanent income model and then illustrate the differences between RI and SE problems. This example is similar to that discussed in Sims (2003) and considers multiple income shocks with different stochastic properties. Specifically, we assume that the original budget constraint is as follows

$$w_{t+1} = Rw_t - c_t + y_{t+1}, \tag{70}$$

where w_t is the amount of cash-in-hand, and the income process y_t have two persistent components (x and z) and one transitory component ($\varepsilon_{y,t}$):

$$\begin{aligned} y_t &= \overline{y} + x_t + z_t + \varepsilon_{y,t}, \\ x_t &= 0.99 x_{t-1} + \varepsilon_{x,t}, \\ z_t &= 0.95 z_{t-1} + \varepsilon_{z,t}, \end{aligned}$$

with

$$\Omega = \mathrm{var}\begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \\ \varepsilon_{z,t} \end{bmatrix} = 10^{-3}\begin{bmatrix} 0.9 & 0 & 0 \\ 0 & 0.009 & 0 \\ 0 & 0 & 0.27 \end{bmatrix}, \tag{71}$$

where x_t is the most persistent and smooth component and $\varepsilon_{y,t}$ is the most transitory and volatile component. For the quadratic utility function $u(c_t) = -\frac{1}{2}\left(c_t - \overline{c}\right)^2$, the model economy can be characterized as the following three equations system:

$$\begin{bmatrix} w_t \\ x_t \\ z_t \end{bmatrix} = \begin{bmatrix} R & 0.99 & 0.95 \\ 0 & 0.99 & 0 \\ 0 & 0 & 0.95 \end{bmatrix}\begin{bmatrix} w_{t-1} \\ x_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}(c_{t-1} - \overline{c}) + \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \\ \varepsilon_{z,t} \end{bmatrix} + \begin{bmatrix} \overline{y} - \overline{c} \\ 0 \\ 0 \end{bmatrix},$$

where β is set to 0.95. Using the first welfare criterion (32) provided in Section 3.2, we can compute that

$$\Sigma = 10^{-3} \begin{bmatrix} 0.1399 & -0.0737 & -0.0110 \\ -0.0737 & 0.1596 & -0.1820 \\ -0.0110 & -0.1820 & 0.5555 \end{bmatrix}, \tag{72}$$

when capacity $\kappa = 2.2$ bits, which can be used to compute the variance of the noise Λ using $\Lambda^{-1} = \Sigma^{-1} - \Psi^{-1}$, and then compute the Kalman gain according to $\theta = \Sigma\Lambda^{-1}$. It is clear from (72) that due to the low capacity devoted to monitoring the state, the post-observation variances (i.e., the conditional variances) of the x and z components are both greater than the corresponding innovation variances in (71). More importantly, the conditional variance of the slow-moving x component is 18 times larger than its corresponding innovation variance, whereas that of the fast-moving z component is only 2 times larger than its innovation variance.[26] The intuition behind this result is that the optimizing agent devotes much less capacity to monitoring the slow-moving component, which leads to greater impacts on the conditional variance term. Figure 6 plots the impulse responses of consumption to the income shocks and noises. It shows that consumption reacts to the income shocks gradually and with delay, and reacts to the corresponding noises promptly. In addition, we can see that the response of consumption to the slow-moving x component is much more damped than that to the fast-moving z component. It is also worth noting that since the agent only cares about the trace of $Z\Sigma$ and the symmetric matrix Z is negative semidefinite, the agent with low capacity will choose to make the post-observations of the states be negatively correlated. This correlation conserves capacity by permitting some information about each state to be transmitted using a single nat.

When we relax the information-processing capacity and increase κ to 2.8 nats, the conditional covariance matrix becomes

$$\Sigma = 10^{-3} \begin{bmatrix} 0.0787 & -0.0419 & 0.0153 \\ -0.0419 & 0.1172 & -0.1926 \\ 0.0153 & -0.1926 & 0.5170 \end{bmatrix}. \tag{73}$$

[26] Alternatively, we can also see that the conditional variance of the x component is about 3 times smaller than its corresponding unconditional variance (0.4523), whereas that of the z component is about 5 times smaller than its corresponding unconditional variance (2.7692).

Comparing (72) with (73), we can see that relaxing information-processing capacity has the largest impact on the conditional variance of the endogenous state variable w: the post-observation variance of w is reduced to about half the initial value. The intuition behind this result is that the endogenous variable plays the most important role in affecting the welfare losses due to RI. To see this clearly, the matrix Z is displayed here:

$$Z = 10^{-2} \begin{bmatrix} -0.0204 & -0.6732 & -0.2769 \\ -0.6732 & -22.2156 & -9.1363 \\ -0.2769 & -9.1363 & -3.7573 \end{bmatrix}. \tag{74}$$

While w per unit has less of an effect on welfare, it is proportionally much larger than either of the other two state variables. It is also clear that as the information constraint is relaxed the agent chooses to allocate more capacity to monitoring the slow-moving component x than to monitoring the z component.

Note that in the RI problem (72) is optimal in sense that it minimizes the expected welfare losses due to finite information-processing capacity by allocating fixed capacity optimally across different elements in the state vector. In contrast, in the SE problem, Λ must be specified first and then Σ and K can be computed. However, it is difficult to specify Λ without prior knowledge about the states. Ad hoc assumptions on Λ might be inconsistent with the underlying efficiency conditions. Therefore, RI could provide a useful way to specify the stochastic properties of the noises by solving the agent's optimization problem subject to information constraints. As we have noted previously, Melosi (2009) presents an application of this idea; he notes that a particular estimated model shows that the marginal utility of information is not equated across variables and is thus inconsistent with RI (that is, inconsistent with *any* value for κ.)

5 Conclusions

In this paper we have explored the implications of two informational frictions theories, signal extraction and rational inattention, for economic behavior, policy, and welfare within the linear-quadratic-Gaussian (LQG) setting. First, we showed that if the variance of the noise itself is fixed exogenously, the two theories can be distinguished as they lead to different dynamics and welfare after implementing government policies. Second, we showed that if the signal-to-noise ratio (SNR) in the SE problem is fixed and channel capacity in the RI problem, SE and RI is observationally equivalent in the sense that they lead to the same dynamics even after implementing policies in the univariate case, whereas they generate different policy and welfare implications in the multivariate case. Furthermore, in the multivariate case we showed that under RI the agent's preference, budget constraint, and information-processing constraints jointly determine the stochastic properties of the post-observation variance and endogenous noise; hence, RI provides an efficient way to specify the nature of the Kalman gain that governs the model's dynamics.

Distinguishing between SE and RI has policy relevance beyond the simple examples we consider here. Paciello and Wiederholt (2011) study the optimal monetary policy in a model that nests both SE and RI. They find that the policy can differ across the two types of models – the key is that under RI the policymaker can affect the attention allocation between different shocks. In particular, they find that the *divine coincidence* of no tradeoff between stabilizing prices and output holds under RI, but not under SE.

Extending our results to compare environments outside the linear-quadratic-Gaussian setup will be challenging. Solving the SE problem is straightforward, although computation of the conditional expectations may be difficult depending on the distributional assumptions. The RI problem is difficult to solve, however, because the optimal joint distribution of states and controls is typically not of a known form and not easy to approximate; the optimality of discrete solutions discussed in Matejka and Sims (2010) make it difficult to characterize the distribution in terms of a small number of parameters. It seems therefore unlikely that SE and RI environments will be observationally equivalent.

6 Online Appendix (Not for Publication)

6.1 Solving for the Steady State Conditional Variance-Covariance Matrix

Assume that the value functions under full information and imperfect information can be written as

$$v(s_t) = s_t^T P s_t$$

and

$$\widehat{v}(\widehat{s}_t) = \widehat{s}_t^T \widehat{P} \widehat{s}_t$$

respectively. Note that the two value functions satisfy the following Bellman equations

$$v(s_t) = s_t^T \left(Q + F^T R F - 2F^T W\right) s_t + \beta E_t \left(v\left(s_{t+1}^*\right)\right), \tag{75}$$

$$\widehat{v}(\widehat{s}_t) = E_t \left[s_t^T Q s_t + \widehat{s}_t^T F^T R F \widehat{s}_t - 2\widehat{s}_t^T F^T W s_t\right] + \beta E_t \left[\widehat{v}(\widehat{s}_{t+1})\right], \tag{76}$$

where

$$s_{t+1}^* = A s_t - B F s_t + \varepsilon_{t+1}$$

is the value of s_{t+1} when the agent can observe s_t perfectly. The agent thus chooses the steady state conditional covariance matrix Σ to minimize the expected welfare loss due to imperfect observations:

$$E_t \left[v(s_t)\right] - \widehat{v}(\widehat{s}_t) = E_t \left[s_t^T P s_t\right] - \widehat{s}_t^T \widehat{P} \widehat{s}_t. \tag{77}$$

Substituting the two Bellman equations into this objective function gives

$$\begin{aligned}
& E_t \left[s_t^T P s_t\right] - \widehat{s}_t^T \widehat{P} \widehat{s}_t \\
& = E_t \left[s_t^T \left(Q + F^T R F - 2F^T W\right) s_t\right] + \beta E_t \left[s_{t+1}^{*T} P s_{t+1}^*\right] \\
& - E_t \left[s_t^T Q s_t + \widehat{s}_t^T F^T R F \widehat{s}_t - 2\widehat{s}_t^T F^T W s_t\right] - \beta E_t \left[\widehat{s}_{t+1}^T \widehat{P} \widehat{s}_{t+1}\right] \\
& = E_t \left[s_t^T \left(Q + F^T R F - 2F^T W\right) s_t\right] - E_t \left[s_t^T Q s_t + \widehat{s}_t^T F^T R F \widehat{s}_t - 2\widehat{s}_t^T F^T W s_t\right] \\
& + \beta E_t \left[s_{t+1}^{*T} P s_{t+1}^* - \widehat{s}_{t+1}^T \widehat{P} \widehat{s}_{t+1}\right] \\
& = E_t \left[s_t^T \left(Q + F^T R F - 2F^T W\right) s_t\right] - E_t \left[s_t^T Q s_t + \widehat{s}_t^T F^T R F \widehat{s}_t - 2\widehat{s}_t^T F^T W s_t\right] \\
& + \beta E_t \left[s_{t+1}^{*T} P s_{t+1}^* - \widehat{s}_{t+1}^T \widehat{P} \widehat{s}_{t+1}\right] \\
& = \text{trace}\left(\left(F^T R F - 2F^T W\right) \Sigma\right) + \beta E_t \left[s_{t+1}^{*T} P s_{t+1}^* - \widehat{s}_{t+1}^T \widehat{P} \widehat{s}_{t+1}\right] \\
& = \text{trace}\left(\left(F^T R F - 2F^T W\right) \Sigma\right) + \beta E_t \left[s_{t+1}^{*T} P s_{t+1}^* - s_{t+1}^T P s_{t+1} + s_{t+1}^T P s_{t+1} - \widehat{s}_{t+1}^T \widehat{P} \widehat{s}_{t+1}\right]
\end{aligned}$$

Given the LQ setting, $E_t \left[s_t^T P s_t\right] - \widehat{s}_t^T \widehat{P} \widehat{s}_t$ is a constant. Call this constant M. Then

$$M = E_t \left[s_t^T P s_t\right] - \widehat{s}_t^T \widehat{P} \widehat{s}_t = \text{trace}\left(\left(F^T R F - 2F^T W\right) \Sigma\right) + \beta E_t \left[s_{t+1}^{*T} P s_{t+1}^* - s_{t+1}^T P s_{t+1}\right] + \beta M,$$

which means that

$$(1-\beta) M = \text{trace}\left(\left(F^T R F - 2F^T W\right) \Sigma\right) + \beta E_t \left[s_{t+1}^{*T} P s_{t+1}^* - s_{t+1}^T P s_{t+1}\right].$$

Using the definitions of s_{t+1}^{*} and s_{t+1}, we obtain

$$
\begin{aligned}
(1-\beta) M &= \operatorname{trace}\left(\left(F^T R F-2 F^T W\right) \Sigma\right)+\beta E_t\left[\begin{array}{c}
\left(s_t^T-\widehat{s}_t^T\right) F^T B^T P B F\left(s_t-\widehat{s}_t\right) \\
+2\left(s_t^T-\widehat{s}_t^T\right) F^T B^T P\left(A s_t-B F \widehat{s}_t+\varepsilon_{t+1}\right)
\end{array}\right] \\
&= \operatorname{trace}\left(\left[F^T R F-2 F^T W+\beta\left(F^T B^T P B F+F^T B^T P A+A^T P B F\right)\right] \Sigma\right) \\
&= \operatorname{trace}(Z \Sigma),
\end{aligned}
$$

where Z is a constant matrix.

6.2 Deriving the Conditional Welfare Gap

Given the value functions under RE and RI in the main tex, we have

$$
\begin{aligned}
\Delta(\Sigma, K) &= E_t\left[v\left(s_t\right)-\widehat{v}\left(\widehat{s}_t\right)\right] \qquad (78)\\
&= -\frac{(R-1) R}{2} E_t\left[s_t^2\right]+R \bar{c} E_t\left[s_t\right]-\frac{1}{2} R\left(\frac{1}{R-1} \bar{c}^2+\omega_\zeta^2\right) \\
&\quad -\left\{-\frac{(R-1) R}{2} \widehat{s}_t^2+R \bar{c} \widehat{s}_t-\frac{1}{2} R\left(\frac{1}{R-1} \bar{c}^2+\omega_\eta^2\right)\right\} \\
&= -\frac{(R-1) R}{2}\left(\operatorname{var}_t\left[s_t\right]+\widehat{s}_t^2\right)-\frac{1}{2} R \omega_\zeta^2-\left[-\frac{(R-1) R}{2} \widehat{s}_t^2-\frac{1}{2} R \omega_\eta^2\right] \\
&= -\frac{(R-1) R}{2} \operatorname{var}_t\left[s_t\right]-\frac{1}{2} R\left(\omega_\zeta^2-\omega_\eta^2\right) \\
&= -\frac{(R-1) R}{2} \Sigma+\frac{1}{2} R\left[\frac{\theta}{1-(1-\theta) R^2}-1\right] \omega_\zeta^2,
\end{aligned}
$$

which is just the expression for the welfare loss in the main text. The above expression can be further simplified as follows:

$$
\begin{aligned}
\Delta(\Sigma, K) &= -\frac{(R-1) R}{2} \Sigma-\frac{1}{2} R\left(\omega_\zeta^2-\omega_\eta^2\right) \\
&= -\frac{(R-1) R}{2} \Sigma+\frac{1}{2} R\left[\frac{\theta}{1-(1-\theta) R^2}-1\right] \omega_\zeta^2 \\
&= \frac{1}{2} \omega_\zeta^2(R-1) \frac{(1-\theta) R^2}{1-(1-\theta) R^2},
\end{aligned}
$$

where we use that the fact that $\Sigma=\frac{\omega_\zeta^2}{1/(1-\theta)-R^2}$ and $\omega_\eta^2=\frac{\theta}{1-(1-\theta) R^2} \omega_\zeta^2$.

References

[1] Adam, Klaus (2005), "Optimal Monetary Policy in the Presence of Imperfect Common Knowledge," *Journal of Monetary Economics* 54(2), 267-301.

[2] Angeletos, George-Marios and Jennifer La'O (2009), "Noisy Business Cycles," forthcoming, *NBER Macroeconomics Annual 2009*.

[3] Barro, Robert J. (2007), "On the Welfare Costs of Consumption Uncertainty," manuscript.

[4] Baxter, Brad, Liam Graham, and Stephen Wright (2010), "Invertible and Non-Invertible Information Sets in Dynamic Stochastic General Equilibrium," forthcoming, *Journal of Economic Dynamics and Control*.

[5] Cover, Thomas M. and Joy A. Thomas (1991), *Elements of Information Theory*, John Wiley and Sons.

[6] Flavin, Marjorie A. (1981), "The Adjustment of Consumption to Changing Expectations About Future Income," *Journal of Political Economy* 89(5), 974-1009.

[7] Hall, Robert E. (1978), "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy* 86(6), 971-87.

[8] Hansen, Lars Peter and Thomas J. Sargent (2004), "'Certainty Equivalence' and 'Model Uncertainty'," http://www.federalreserve.gov/events/conferences/mmp2004/pdf/hansensargent.pdf.

[9] Hansen, Lars Peter and Thomas J. Sargent (2007), *Robustness*, Princeton University Press.

[10] Kahneman, Daniel (1973), *Attention and Effort*, Prentice-Hall Press.

[11] Kasa, Kenneth (2006), "Robustness and Information Processing," *Review of Economic Dynamics* 9(1), 1-33.

[12] Kim, Jinsook, Eunmi Ko, and Tack Yun (2012), "The Role of Bounded Rationality in Macro-Finance Affine Term-Structure Models," manuscript.

[13] Lewis, Kurt F. (2006), "The Life-Cycle Effects of Information-Processing Constraints," manuscript.

[14] Lorenzoni, Guido (2009), "A Theory of Demand Shocks," *American Economic Review* 99(5), 2050-2084.

[15] Lucas, Robert E., Jr. (1972), "Expectations and the Neutrality of Money," *Journal of Economic Theory* 4(1), 103-124.

[16] Luo, Yulei (2008), "Consumption Dynamics under Information Processing Constraints," *Review of Economic Dynamics* 11(2), 366-385.

[17] Luo, Yulei and Eric R. Young (2010), "Risk-sensitive Consumption and Savings under Rational Inattention," *American Economic Journal: Macroeconomics* 2(4), 281-325.

[18] Maćkowiak, Bartosz and Mirko Wiederholt (2009), "Optimal Sticky Prices under Rational Inattention," *American Economic Review* 99(3), 769-803.

[19] Matejka, Filip and Christopher A. Sims (2010), "Discrete Actions in Information-Constrained Tracking Problems," manuscript.

[20] Melosi, Leonardo (2009), "A Likelihood Analysis of Models with Information Frictions," Penn Institute for Economic Research Working Paper 09-009.

[21] Morris, Stephen and Hyun Song Shin (2002), "The Social Value of Public Information," *American Economic Review* 92(5), 1521-1534.

[22] Muth, John F. (1960), "Optimal Properties of Exponentially Weighted Forecasts," *Journal of the American Statistical Association* 55(290), 299-306.

[23] Paciello, Luigi and Mirko Wiederholt (2011), "Exogenous Information, Endogenous Information, and Optimal Monetary Policy," manuscript.

[24] Sargent, Thomas J. (1991), "Equilibrium with Signal Extraction from Endogenous Variables," *Journal of Economic Dynamics and Control* 15(2), 245-273.

[25] Simon, Dan (2006), *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*, John Wiley & Sons.

[26] Sims, Christopher A. (2003), "Implications of Rational Inattention," *Journal of Monetary Economics* 50(3), 665-690.

[27] Sims, Christopher A. (2006), "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review* 96(2), 158-163.

[28] Sims, Christopher A. (2010), "Rational Inattention and Monetary Economics," forthcoming in *Handbook of Monetary Policy.*

[29] Townsend, Robert M. (1983), "Forecasting the Forecasts of Others," *Journal of Political Economy* 91(4), 546-588.

[30] Tutino, Antonella (2008), "The Rigidity of Choice: Life Cycle Savings with Information-Processing Limits." Federal Reserve Board Finance and Economics Discussion Series 2008-62.

[31] Wang, Neng (2004), "Precautionary Saving and Partially Observed Income," *Journal of Monetary Economics* 51(8), 1645-1681.

[32] Van Nieuwerburgh and Veldkamp (2006), "Learning Asymmetries in Real Business Cycles," *Journal of Monetary Economics* 53(4), 753-772.

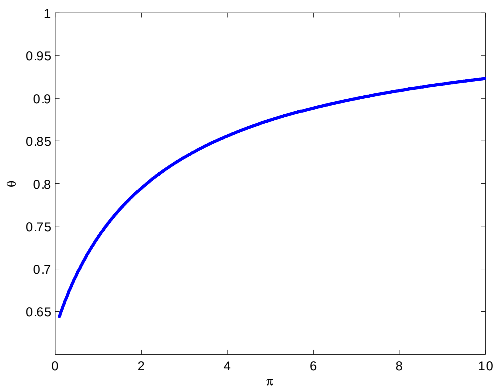

Figure 1: Relationship between π and μ

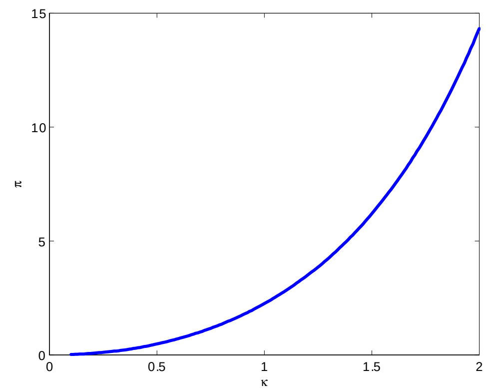

Figure 2: Relationship between κ and π

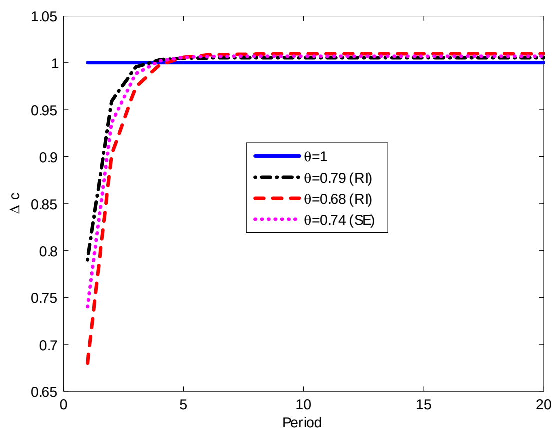

Figure 3: Consumption Dynamics under SE and RI after Policy

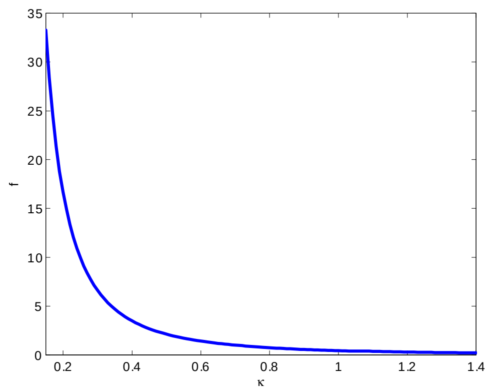

Figure 4: Effects of RI on RMW

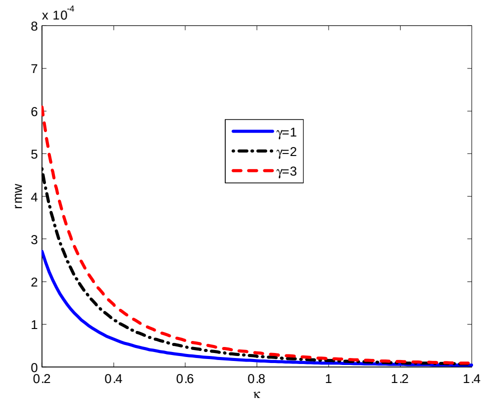

Figure 5: Effects of RI on RMW

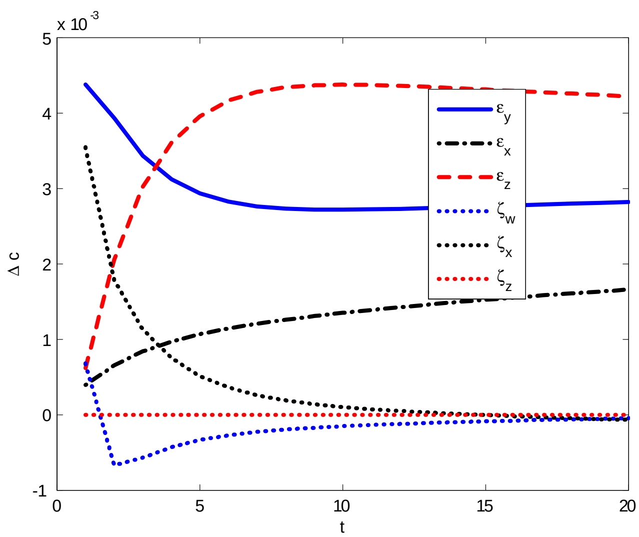

Figure 6: Responses of Consumption to Income Shocks and Noises