Rational Inattention in Macroeconomics: A Survey

Luo, Yulei and Young, Eric

The University of Hong Kong, University of Virginia

2013

Rational Inattention in Macroeconomics: A Survey*

Yulei Luo[†]
University of Hong Kong

Eric R. Young[‡]
University of Virginia

September 2, 2013

Abstract

In this paper we survey recent works on rational inattention (RI) in macroeconomics within the dynamic linear-quadratic-Gaussian (LQG) setting. We first discuss how RI affects consumption smoothness and sensitivity, precautionary savings, asset pricing, portfolio choice, and aggregate fluctuations in the univariate case. We then discuss the applications of RI to macroeconomic models of permanent income and price-setting in the multivariate case. Finally, we briefly discuss how RI can be applied to non-LQG settings.

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

Keywords: Rational Inattention, Consumption, Asset Pricing, Precautionary Savings, Portfolio Choice, Price Setting.

---

*We thank Chris Sims for many helpful discussions on rational inattention. Luo thanks the General Research Fund (GRF) in Hong Kong (#HKU749711) for financial support. 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

The real world is complicated, with in principle hundreds of relevant variables that individuals should track in order to make economic decisions. Obviously, few (if any) households actually pay attention to that many variables. It is straightforward to assume agents ignore certain variables when they make decisions, but doing so abandons standard economic assumptions that agents will make choices that are optimal for themselves – that is, behavior is deliberate. Rational inattention (henceforth RI) to variables is the assumption that individuals have limited attention but are allowed to decide what they care most about, and therefore pay the most attention to.

In this paper we survey the applications of rational inattention in macroeconomics. RI was first proposed and introduced into economics by Sims (2003). In a previous paper (Sims 1998) he made the point that stickiness appears pervasive in the data, and therefore models that rely on some forms of adjustment costs (random opportunities to change prices or wages, for example, or costs of altering portfolios) must have adjustment costs everywhere; the resulting proliferation of essentially free parameters (the costs are estimated to match the stickiness and cannot generally be evaluated in a model-independent manner) makes the underlying models hard to test in a disciplined way and blurs the line between identifying assumptions and overidentifying restrictions (Kydland and Prescott 1996). In contrast, Sims (2003) proposed a single mechanism that generates stickiness pervasively – agents have limited information about the state of the world and "learn" slowly because they cannot process information in unlimited amounts.

The key innovation relative to standard noisy rational expectations models (like Kydland and Prescott 1982 or Cooley and Hansen 1997) is that RI hypothesis permits agents to design the distribution of noise terms by focusing limited attention on certain variables at the expense of others. Under RI, agents respond to changes in the true underlying state slowly because it takes time for them to learn exactly what the new state is – they cannot learn without error because the information flow required to perfectly describe the state is larger than their "Shannon channel" permits. Thus, the distribution of the RI-induced noise is an outcome of optimal choice will adapt to changing circumstances in the economy.[1]

In general, the RI model is intractable, and obviously intractable models are useless.[2] There

[1] Gabaix (2013) focuses on an alternative mechanism, sparsity, in which agents choose to ignore variables completely that do not affect their payoffs too much. However, his mechanism does not explain why agents don't pay attention to all variables, which RI does.

[2] We are reminded of the joke about a man searching for his car keys under the lamppost, even though he dropped them out in the dark alley. One cannot get anywhere writing down models that cannot be solved; while searching

are a number of obstacles that arise. First, the choice of the agents in an RI problem is the joint distribution of choices and states, which can easily be high-dimensional unless the agent finds it optimal to choose the distribution from a class that can be described with a small number of parameters (such as the Gaussian or normal distribution, which is completely described by only two numbers). Unfortunately, Matejka and Sims (2011) and Saint-Paul (2011) show that generically the RI problem will generate discrete distributions that have gaps and holes, making them impossible to parametrize with a low-dimensional vector; related numerical results can be found in Lewis (2009), Batchuluun, Luo, and Young (2009), and Tutino (2012). Second, the information flow constraint means that the posterior (the distribution over states after the signal is received) cannot be too far from the prior, so the state is high-dimensional as well. As a result, the researcher rapidly encounters the "curse of dimensionality" and only short horizons or small models are feasible. In addition, the law of motion that ties the prior to the posterior is not transparent, so that it becomes difficult to describe what the agent is "paying attention to", which makes explanation of the results challenging.

Fortunately, there is a case that is (nearly) analytically tractable. As shown in Sims (2003, 2010), RI models are easiest to solve and handle when random variables are all jointly normal. Specifically, if random shocks are all Gaussian, the payoff function is quadratic, all constraints are linear, and the prior distribution is Gaussian (LQG), the *ex post* distribution of the true state upon receiving the observations and the distribution of the noise will also be Gaussian; as noted already, describing a Gaussian distribution only requires knowledge to two numbers (the mean and the variance) so the dimensionality of the problem is dramatically reduced. Furthermore, the variance converges rapidly to a constant in the steady state, reducing the problem to finding the evolution of the mean and the constant value for the variance and then running the signals through the Kalman filter to link priors and posteriors. Contributions to the RI-LQG literature include Adam (2005), Kasa (2006), Luo (2008), Luo and Young (2010a, 2010b, 2013), Maćkowiak and Wiederholt (2009, 2013), Melosi (2009), Reis (2011), Paciello and Wiederholt (2012), Luo, Nie, and Young (2012), and Kim, Ko, and Yun (2012).[3]

under the lamppost might not find the keys immediately, the hope is that eventually we learn how to make the lamp bright enough to illuminate the keys.

[3]Other papers assume *ex post* Gaussian distributions and Gaussian noise but adopt exponential or constant-absolute-risk-aversion (CARA) preferences, including Peng (2004), Peng and Xiong (2005), Mondria (2010), and Van Nieuwerburgh and Veldkamp (2009, 2010). Since both the optimality of *ex post* Gaussianity and the standard Kalman filter are based on the LQG specification, the applications of these results in the RI models with CARA

In this paper we first present a general discrete-time dynamic LQG setting in which RI is introduced. We then present how RI can be applied to a variety of macroeconomic models. Particurly, we divide the RI-LQG models into two categories: (i) the univariate case and (ii) the multivariate case. In the first case, the information-processing constraint (i.e., the reduction in uncertainty about the state variable is bounded by a finite channel capacity) can uniquely determine the conditional variance of the state variable; consequently, the variance of the RI-induced noise and the Kalman gain can also be determined. In the first application, we consider the permanent income model with RI proposed and solved in Luo (2008). As shown in Luo (2008), using the method of state-space-reduction and the original multivariate model can be reduced to a univariate model and can thus be solved explicitly. Using the closed-form solution, we discuss the implications of RI for consumption dynamics and asset pricing.[4] Specifically, we discuss how RI can be a potential explanation for the excess smoothness and excess sensitivity puzzles in the consumption literature. We also argue that the RI model with fixed information-processing cost does a better job at replicating the different consumption behavior in emerging and developed small open economies studied in Aguiar and Gopinath (2007) and Luo, Nie, and Young (2012).[5] In another application, we discuss how RI increases long-run consumpion risk and reduces the optimal share invested in the risky asset; consequently, the RI model can generate more realist joint behavior of aggregate consumption and the equity return. In a robust (RB) or risk-sensitve (RS) permanent income model, we show that the interaction of RI and RB (or RS) can increase the total amount of uncertainty facing the consumer and thus leads to more precautionary savings. As shown in Luo and Young (2010a), under RB or RS, the certainty equivalence principle no longer holds and the resulting decision rules depend explicitly on the variance of the shocks, producing precautionary savings. Fortunately, the value functions are still quadratic functions of the states, leading again to the optimality of Gaussian noise in the RI model. Thus, we can preserve the tractability of the LQ permanent income model.

We then move on to study the multivariate case. In this case given channel capacity the conditional variance-covariance matrix can be obtained by solving a semidefinite programming

preferences are only approximately valid. There are no estimates of the approximation error in the literature.

[4] As argued in Sims (2010), RI can be modelled by assuming either fixed channel capacity or fixed marginal cost of information processing. In this paper we also discuss how the two ways lead to different conclusions about the effects of RI.

[5] Aguiar and Gopinath (2007) document that the relative volatility of consumption growth to income growth in emerging countries is significantly greater than that in developed countries.

problem in which the inattentive agent minimizes the expected loss due to information-processing constraints. We discuss the connection between this problem and the celebrated "reverse water-filling problem" from information theory. The basic intuition is that agents effectively have a budget of attention that they "pour" into bins representing each variable, with deeper bins representing variables that are more volatile or less important for payoffs; the unfilled part of the bin is the residual noise left in the observation of that particular variable. The solution to the problem is the variance-covariance matrix of the noise vector, from which we can recover the gain of the Kalman filter and the posterior distribution. Unlike the univariate case, preferences and budget constraints matter for the attention allocation problem, since as noted the size of the bins depend on how much particular variables affect payoffs (either directly or indirectly through the constraints). Unfortunately, analytical solutions to the semidefinite programming problem are not available, unlike the standard water-filling problem, because the variables do not generally enter the payoff function symmetrically. Numerical solutions are presented to describe how agents trade off attention.

Lastly, we study the implications of RI for price stickiness, the original motivation of Sims (1998). Following Maćkowiak and Wiederholt (2009), we study the price-setting decision of a firm that has information-flow constraints. Consistent with some VAR evidence, we find that firm prices respond immediately and strongly to individual-level productivity shocks, less quickly and completely to aggregate productivity shocks, and very slowly and incompletely to monetary policy shocks (movements in the nominal interest rate). The intuition is straightforward – individual productivity shocks are the dominant source of movements in profits, so the firm pays a lot of attention to them so that changes are quickly observed and incorporated into prices. Aggregate productivity shocks are less volatile and less important, so their movements are not tracked as closely, and monetary policy is almost completely ignored.

These results are not merely theoretical curiosities. Paciello and Wiederholt (2012) show that they fundamentally change the nature of optimal monetary policy. In the literature there has appeared the "divine coincidence" in which there appears to be no tradeoff between the two primary goals of the Federal Reserve (stabilizing output and inflation) – it turns out that stabilizing prices has the effect of stabilizing output. In models where information is exogenously restricted (Woodford 2001), this coincidence disappears – the Fed must make a decision about how much output volatility it is willing to tolerate to limit inflation volatility. In addition, there is no gain to commitment – future monetary policy promises do not matter – so the models are incapable of thinking about problems of "forward guidance". In contrast, when attention is limited but endogenous, the divine coincidence reappears and commitment is valuable because the Fed will be tempted to exploit the predetermined attention decision to move variables that agents are not tracking. If RI is a good description of the world, then it will have practical consequences for the science of monetary policy.

The paper is organized as follows. First, we present the basics of rate distortion theory, which forms the underlying technology of rational inattention. We then discuss the tracking problem of RI explicitly. Next we specialize this problem to the linear-quadratic-Gaussian environment that provides the most tractability, describing it again in generic terms. Then we consider a sequence of examples in the LQG framework that highlight various questions in macroeconomics that RI has been used to study.

2 Rate Distortion and Channel Capacity

Describing an arbitrary real number requires an infinite amount of information (think about how many digits you need to describe π) and so cannot be done in practice without some error. Rate distortion theory is about the connection between the amount of information permitted and the resulting error in representation – how well can one do if one is restricted to some prescribed finite information flow? To give a concrete example, suppose one has a random variable $X \sim N\left(0, \sigma^2\right)$ and wants to represent it with some function $Y\left(X\right)$ that requires only a finite amount of information (1 bit). Errors are punished using a squared deviation, so that the goal is to choose $Y\left(X\right)$ to minimize

$$W = E\left[\left(X - Y\left(X\right)\right)^2\right].$$

We must choose the weight and the location of the nodes, subject to the constraint that it can result in only 1 bit being transmitted. The solution is

$$Y\left(x\right) = \begin{cases} \sigma\sqrt{\frac{2}{\pi}} & \text{if } x \geq 0 \\ -\sigma\sqrt{\frac{2}{\pi}} & \text{if } x < 0; \end{cases}$$

that is, each representation equals the conditional mean of its region, with equal mass attached to each point. With more than 1 bit the problem becomes one of choosing the regions and the masses.

More generally, we can define the problem as choosing some function $Y(X)$ to minimize the expected sum of squared errors:

$$W = E\left[(X - Y(X))^T (X - Y(X))\right]$$

subject to a limit on how different the underlying density of X and the conditional density of Y given X can be, giving rise to the notion of an information rate distortion function.

Definition 1 *The information rate distortion function $R(D)$ for a source random variable X with distribution $f(x)$ and distortion measure $d(x, y)$ is defined as*

$$R(D) = \min_{p(y|x)} \{I(X; Y)\}$$

such that

$$\int_{x,y} f(x) p(y|x) d(x, y) \leq D.$$

The problem also requires that $p(y|x)$ be a distribution (nonnegative, integrate to 1) and that the joint distribution $P(x, y) = f(x) p(y|x)$ is also a distribution and satisfies the distortion constraint. The relevant result for a vector of independent Gaussian sources is now given.

Theorem 2 *Let $X_i \sim N\left(0, \sigma_i^2\right)$ for $i \in \{1, 2, ..., m\}$ be independent and let the distortion function be $d(x, y) = \sum_{i=1}^{m} (x_i - y_i)^2$. Then the information rate distortion function is*

$$R(D) = \frac{1}{2} \sum_{i=1}^{m} \log\left(\frac{\sigma_i^2}{D_i}\right)$$

where

$$D_i = \begin{cases} \lambda & \text{if } \lambda < \sigma_i^2 \\ \sigma_i^2 & \text{if } \lambda \geq \sigma_i^2 \end{cases}$$

and where λ is a constant chosen such that $\sum_{i=1}^{m} D_i = D$.

λ is related to the Lagrange multiplier on the information flow constraint and thus represents the marginal increase in information flow needed if the maximal distortion is to be decreased. This problem is often called the 'reverse water-filling problem' (see Figure **??**); the "depth" a particular variable is the variance σ_i^2, the amount of attention allocated to each variable is related to the difference between λ and σ_i^2, denoted $\widehat{\sigma}_i^2$. Essentially, one can interpret the solution as "pouring" in attention (variance reduction) to the channel with the largest current capacity until the finite attention is exhausted. The solution is

$$Y \sim N\left(0, \begin{bmatrix} \widehat{\sigma}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \widehat{\sigma}_m^2 \end{bmatrix}\right),$$

where $\widehat{\sigma}_i^2 = \max\left\{\sigma_i^2 - D_i, 0\right\}$ and $D_i = \min\left\{\lambda, \sigma_i^2\right\}$. If $\sigma_i^2 > \lambda$, $D_i = \lambda$ and $\widehat{\sigma}_i^2 = \sigma_i^2 - \lambda$; that is, variables with prior variances higher than λ are allocated attention sufficient to bring their posterior variance down to λ. In contrast, if $\sigma_i^2 < \lambda$, $D_i = \sigma_i^2$ and $\widehat{\sigma}_i^2 = 0$, so that this variable is ignored completely (see x_4 in Figure 1).[6] The geometric interpretation of the solution method is to start λ at the bottom of the graph and raise it until the shaded area equals the permitted amount of information flow.[7]

If the variables are not independent, we simply apply the Spectral Decomposition Theorem to the variance-covariance matrix Σ so that $\Sigma = P\Omega P^{-1}$ and solve the problem using the diagonal elements of Ω (which are the eigenvalues of Σ). Similarly, if we have a Gaussian stochastic process one can decompose it by frequencies (using a Fourier transfrom) and apply the problem to each frequency (which are uncorrelated).

Unfortunately, in economics not all distortions are equal – it is not simply the variance that matters, but rather a weighted average that depends on how each variable contributes to payoffs:

$$W = E\left[(X - Y(X))^T Z (X - Y(X))\right]$$

is the appropriate objective function, where Z is a positive semidefinite symmetric matrix (needed to guarantee the minimum exists). Unless Z happens to be diagonal the water-filling solution, while still providing the intuition for the attention allocation problem, no longer supplies a constructive method for finding that solution. We will discuss the numerical problems that multivariate RI problems pose later in this survey.

[6] One can also show that any information rate distortion $R > R(D)$ is achievable by some representation (which is why we do not need to define $R(D)$ in terms of infima); this result is nontrivial, but is of little interest since we are not trying to discover the way agents "code" their observations, only what they can achieve by doing so.

[7] Gabaix (2013) introduces the notion of "sparsity" in decision-making, where agents choose to completely ignore changes in random variables that do not affect their payoffs by at least some fixed amount. RI could in principle provide a deeper foundation for his assumption, but the mechanics of his implementation are very different.

3 The Canonical Rational Inattention Tracking Problem

Suppose a decision maker has the goal of maximizing some function $u(|y-x|)$ (such as minus a quadratic loss) subject to information-processing constraints. The problem solved by the decision maker is

$$\max_{f,\mu_x}\left\{\int u(|y-x|)\,f(x,y)\,\mu_x(dx)\,\mu_y(dy)\right\} \tag{1}$$

subject to

$$\int \log(f(x,y))\,f(x,y)\,\mu_x(dx)\,\mu_y(dy)+\int \log\left(\int f(x,y')\,\mu_y(dy')\right)f(x,y)\,\mu_x(dx)\,\mu_y(dy)\leq\kappa, \tag{2}$$

$$\int f(x,y)\,\mu_x(dx)=g(y) \tag{3}$$

for given density $g(y)$, and

$$f(x,y)\geq 0 \tag{4}$$

$$\int f(x,y)\,\mu_x(dx)\,\mu_y(dy)=1. \tag{5}$$

The first constraint says that the information flow generated by an observation of y is less than κ, the second says that the marginal of the joint distribution $f(x,y)$ at any y should be the exogenous distribution of y, and the last two require that the joint density be nonnegative and integrate to 1. Rather than analyze this problem in detail (which is the goal of Matejka and Sims 2010), we simply note their key result and move on to a specialized environment.

Theorem 3 *Suppose that* $\exp(u(|y-x|))$ *is analytic and positive on the real line,* u *is non-increasing and non-constant, and* g *has bounded support. Then the density of* x *(denoted* $p(x)$*) has positive value on a finite set of points in the support of* g*.*

This result is generic – information-constrained problems with finite support will have discrete solutions – and implies that unless one knows the form of the optimal $f(x,y)$ one should expect discreteness in any approximation solution found numerically. This discreteness is not "trivial" in general, as the distributions can have "holes" (see Batchuluun, Luo, and Young 2009 for an example), and thus poses a difficult numerical problem as the number of states gets large and the problem hits many nonnegativity constraints.

In the rest of this survey we will use quadratic approximations to agent objectives and linear constraints, exploiting the following result from Sims (2003):

Theorem 4 *Suppose u is quadratic and y is Gaussian. Then $f(x, y)$ is Gaussian.*

This result substantially simplifies the numerical solutions, since Gaussian distributions have only two parameters and the IPC can be reduced to the difference of variances between the prior and the posterior (plus a technical condition that the difference be positive semidefinite, which is not necessary in the fully-nonlinear problem). Supposing that

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^{T} & \Sigma_{yy} \end{bmatrix}\right)$$

the IPC becomes

$$-\frac{1}{2}\log\left|I - \Sigma_{yy}^{-1}\Sigma_{xy}^{T}\Sigma_{xx}^{-1}\Sigma_{xy}\right| \leq \kappa.$$

Exploiting the Toeplitz structure of the variance-covariance matrices Σ_{ii} Sims (2003) notes that this constraint can be written in the frequency domain as

$$-\frac{1}{2}\int_{-\pi}^{\pi}\log\left(1 - \frac{\left|S_{xy}(\omega)\right|^{2}}{\left|S_{x}(\omega)\right|\left|S_{y}(\omega)\right|}\right)d\omega \leq \kappa,$$

where the term in parentheses is one minus the coherence between X and Y at frequency ω.[8] The interpretation of this representation of the constraint is that there should be as much comovement between the "true state" and the "estimated state" as possible, given the limit of changes in entropy. Using this formulation it is possible then to show the variance of the noise relative to the signal must increase as the frequency gets shorter, so that agents with rational inattention will not respond to high frequency variations in the target.

4 Modeling Rational Inattention within the LQG Setting

In this section we will rewrite the tracking problem as a standard linear-quadratic regulator problem and use that problem to illustrate general solutions.

[8] A Toeplitz matrix has the structure

$$\begin{bmatrix} d & e & f & g \\ c & d & e & f \\ b & c & d & e \\ a & b & c & d \end{bmatrix},$$

where all diagonals are even. The variance-covariance matrix of a sample from an autoregressive process has this form. Coherence measures the comovement of two random processes at a given frequency.

4.1 Full-information Rational Expectations LQG Model

Consider the following standard full-information rational expectations dynamic linear-quadratic-Gaussian (LQG) model:

$$v(s_0) = \max_{\{c_t, s_{t+1}\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t \left(s_t^T Q s_t + c_t^T R c_t + 2 c_t^T W s_t \right) \right], \tag{6}$$

subject to

$$s_{t+1} = A s_t + B c_t + \varepsilon_{t+1}, \tag{7}$$

with s_0 known and given, where $\beta < 1$ 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.[9]

When the agent can fully observe the state s_t, the model is a standard linear-quadratic regulator problem. Guess that the value function at t is $v(s_t) = s_t^T P s_t$. 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

$$c_t^* = -F s_t, \tag{8}$$

and the Riccati equation is

$$P = Q + F^T R F - 2 F^T W + \beta \left(A^T - F^T B^T \right) P (A - BF), \tag{9}$$

where

$$F = \left(R + \beta B^T P B \right)^{-1} \left(W + \beta B^T P A \right). \tag{10}$$

Iterating on the matrix Riccati equation (9) uniquely determines P, since the equation defines a contraction mapping from the space of negative semidefinite matrices back to itself:

$$T(P) = Q + F(P)^T R F(P) - 2 F(P)^T W + \beta \left(A^T - F(P)^T B^T \right) P (A - B F(P)).$$

Using P, we can determine F in the optimal policy (10).

[9] These conditions are known as detectability and stabilizability in the control literature. They amount to assumptions that states can be observed with sufficient accuracy and respond sufficiently to controls to avoid violating transversality conditions. Under full information detectability is irrelevant.

4.2 Introducing Rational Inattention

With finite capacity $\kappa \in (0, \infty)$, a random variable $\{s_t\}$ following a continuous distribution cannot be observed without error and thus the information set at time $t+1$, denoted $\mathcal{I}_{t+1}$, 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 shocks hitting the economy. 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 about the variable s_{t+1} as much as possible. Formally, this idea can be described by the information constraint

$$\mathcal{H}\left(s_{t+1}|\mathcal{I}_t\right) - \mathcal{H}\left(s_{t+1}|\mathcal{I}_{t+1}\right) \leq \kappa, \tag{11}$$

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 flow – that is, the change in the entropy – that can be transmitted in any given period. Finally, following the literature, we suppose that the prior distribution of s_{t+1} is Gaussian.

In the LQG environment, as has been 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, given that the noisy signal takes the additive form $s_{t+1}^* = s_{t+1} + \xi_{t+1}$, the noise ξ_{t+1} will also be Gaussian. In this case, (11) reduces to

$$\ln\left(|\Psi_t|\right) - \ln\left(|\Sigma_{t+1}|\right) \leq 2\kappa, \tag{12}$$

where $\Psi_t = A^T\Sigma_t A + \Omega$ and Σ_{t+1} are the conditional variance-covariance matrices prior to and after observing the new signal, respectively. Since more information about the state is better in single-agent models, this constraint will be binding.[10]. We will exploit the formula for updating the conditional variance-covariance matrix of a Gaussian distribution Σ in the steady state:

$$\Lambda^{-1} = \Sigma^{-1} - \Psi^{-1}. \tag{13}$$

[10] 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.

The evolution of the estimated state $\widehat{s}_t$ is governed by the Kalman filtering equation

$$\widehat{s}_{t+1} = (1-\theta)\left(A\widehat{s}_t + Bc_t\right) + \theta s^*_{t+1}, \tag{14}$$

where $\theta = \Sigma\Lambda^{-1}$. Combining (7) with (14) yields

$$\widehat{s}_{t+1} = A\widehat{s}_t + Bc_t + \eta_{t+1}, \tag{15}$$

where

$$\eta_{t+1} = \theta A\left(s_t - \widehat{s}_t\right) + \theta\left(\zeta_{t+1} + \xi_{t+1}\right)$$

with $\eta_{t+1}|\mathcal{I}_t \sim N\left(0, \Lambda_\eta\right)$ and $\Lambda_\eta = \Psi - \Sigma$.

In the RI literature, the following two-stage procedure is adopted to solve the RI problem:

1. In the first stage, we take the steady state conditional variance of the state vector (Σ) as given. Given Σ, all the other matrices, Ψ, Λ, Λ_η, and θ, can be uniquely determined, and we obtain the constrol by solving the following standard dynamic program in which the state is $\widehat{s}_t$:

$$v\left(\widehat{s}_t\right) = \max_{\{c_t\}} E_t\left[s_t^T Q s_t + c_t^T R c_t + 2c_t^T W s_t + \beta v\left(\widehat{s}_{t+1}\right)\right], \tag{16}$$

subject to (15). We exploit the certainty equivalence property of LQG models here to separate forecasting from control.

2. In the second stage, given that the value functions under full information and imperfect information are $v\left(s_t\right) = s_t^T P s_t$ and $\widehat{v}\left(\widehat{s}_t\right) = \widehat{s}_t^T \widehat{P}\widehat{s}_t$, respectively, we can compute the optimal Σ by minimizing the expected welfare loss due to RI subject to the information-processing constraint, (11) and an additional "no subsidization" constraint

$$A^T\Sigma A + \Omega \succeq \Sigma. \tag{17}$$

The "no subsidization" constraint is to be read as "the matrix defined as the difference between the posterior and the prior element-by-element must be positive semidefinite." This constraint is termed a "no subsidization" condition, as it implies that an agent cannot reduce one variable's entropy by more than κ by increasing the entropy of another; one can also interpret it as a "no forgetting" condition, since it can be viewed as prohibiting the agent from forgetting information about one variable in order to learn more about another.[11] It is worth noting that this constraint

[11] Fans of the TV show Married with Children might recall an episode where Kelly Bundy is learning facts for a trivia contest – eventually her brain fills up so that every new fact pushes an old one out. The "no forgetting" condition means that Kelly has unlimited capacity to store existing knowledge; the constraint is on the new information that she can learn.

takes the form of restrictions on eigenvalues and thus is not linear in the multivariate case while it is automatically satisfied in the univariate case. We will survey applications of the univariate case in the next section, and then survey applications in the multivariate case.

Before doing so, we think it useful to describe briefly the essence of decision-making by an agent facing an information-processing constraint. The agent is choosing a joint distribution of future states s_{t+1} and controls c_t and informing "nature" to give him an outcome randomly according to this distribution once conditioned on the true (but unknown to the agent) state. That outcome serves as the signal s^*_{t+1} which permits the agent to update her beliefs, since there is information in the signal whenever the agent chooses the distribution to be informative given that signal.

5 The Univariate Case

5.1 Fixed Channel Capacity or Fixed Marginal Cost of Information-Processing

Fixed Channel Capacity. In the univariate state case in which $n = 1$, (12) fully determines the value of the steady state conditional variance Σ:

$$\Sigma = \frac{\Omega}{\exp(2\kappa) - A^2}, \tag{18}$$

which means that Σ is entirely determined by the variance of the exogenous shock (Ω) and the exogenously given capacity (κ).[12] Given this Σ, we can use (13) to recover the variance of the endogenous noise (Λ) as

$$\Lambda = \left(\Sigma^{-1} - \Psi^{-1}\right)^{-1}, \tag{19}$$

where $\Psi = A^2\Sigma + \Omega$, and also to compute the Kalman gain θ:

$$\theta = \Sigma\Lambda^{-1} = 1 - \Sigma\Psi^{-1}, \tag{20}$$

which reduces to $\theta = 1 - \frac{1}{\exp(2\kappa)}$ using (18) and (19).

Given that κ is *fixed*, (18), (19), and (20) 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. However, if an

[12] Note that here we need to impose the restriction $\exp(2\kappa) - A^2 > 0$. If this condition fails, the state is not *stabilizable* and the unconditional variance diverges.

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.

Fixed Marginal Cost of Information-Processing. 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 (12) is constant.[13] 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 $\sum_{t=0}^{\infty} \beta^t \left(c_t - c_t^*\right)^2$, 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 (12), and we impose the restriction that $\beta A = 1$ for simplicity. Solving this problem yields the optimal steady state conditional variance:

$$\Sigma = \frac{-\left(\Omega H - \lambda A\right) + \sqrt{\left(\Omega H - \lambda A\right)^2 + 4\lambda \Omega A^2}}{2HA^2} > 0. \tag{21}$$

It is straightforward to show that as λ goes to 0, $\Sigma = 0$; and as λ goes to ∞, $\Sigma = \infty$. Comparing (21) with (18), 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:

$$\kappa = \frac{1}{2} \ln \left(A^2 + \frac{2HA^2}{-\left[H - A\left(\lambda/\Omega\right)\right] + \sqrt{\left[H - A\left(\lambda/\Omega\right)\right]^2 + 4A^2\left(\lambda/\Omega\right)}} \right). \tag{22}$$

In this case, the Kalman gain is

$$\theta = 1 - \left\{ A^2 + \frac{2HA^2}{-\left[H - A\left(\lambda/\Omega\right)\right] + \sqrt{\left[H - A\left(\lambda/\Omega\right)\right]^2 + 4A^2\left(\lambda/\Omega\right)}} \right\}^{-1}. \tag{23}$$

It is obvious that κ converges to its lower limit $\underline{\kappa} = \ln\left(A\right)$ as λ goes to ∞; and it converges to ∞ as λ goes to 0.[14] In other words, using the RI modeling strategy, the agent is allowed to adjust

[13] Formally, the assumption is that κ is a choice variable and the utility cost function is $\mu\kappa$ for some constant μ. Quadratic cost functions of the form $\mu\kappa^2/2$ could also be accomodated, where the marginal cost of attention is increasing in the amount of attention already allocated.

[14] We require here that $H \neq 0$; that is, the state must be *detectable*. If the state is not detectable there is no point in allocating attention to monitoring it.

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 (22) 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 a policy experiment: Ω is scaled up due to a change in policy. If we adopt the assumption that λ is *fixed*, (21) means that there is a less 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 of modeling RI may lead to different policy implications.[15]

5.2 Extension to Correlated Shocks and Noises

In the above analysis, we assumed that the exogenous fundamental shock and noise are uncorrelated. Luo and Young (2013) discuss how *correlated* shocks and noises affect the implications of SE and RI for the model's dynamic behavior. It is easy to think of situations where the shocks affect the measurement process; for example, in the US the Bureau of Economic Activity measures GDP and requires costly labor allocation, so that a shock that reduces productivity could simultaneously reduce true GDP and result in less labor allocated to its measurement, generating larger noise. Other examples can be found in physical systems that are less relevant for economics.

Luo and Young (2013) consider a case in which the fundamental shock (ε) and the noise shock (ξ) are correlated:

$$\begin{aligned} \operatorname{corr}\left(\varepsilon_{t+1}, \xi_{t+1}\right) &= \rho, \\ \operatorname{cov}\left(\varepsilon_{t+1}, \xi_{t+1}\right) &= \Gamma = \rho\sqrt{\Omega}\sqrt{\Lambda}, \end{aligned}$$

where ρ is the correlation coefficient between ε_{t+1} and ξ_{t+1}, $\Omega = \operatorname{var}\left(\varepsilon_{t+1}\right)$ and $\Lambda = \operatorname{var}\left(\xi_{t+1}\right)$. In this case, the Kalman gain is

$$\theta = \left(\Psi + \Gamma\right)\left(\Psi + \Lambda + 2\Gamma\right)^{-1}, \tag{24}$$

[15] Note that these two different ways to model RI are very similar to the *constraint* and *multiplier* preferences adopted by Hansen and Sargent (2007) to model aversion to model misspecification. They also established the observational equivalence between the two preferences within the LQG setting. Luo and Young (2010a) extend this equivalence to RI settings.

and the updating formula for the conditional variance is

$$\Sigma = \Psi - (\Psi + \Gamma)^2 (\Psi + \Lambda + 2\Gamma)^{-1}, \tag{25}$$

where $\Psi = \Omega + A^2\Sigma$. Just like the case without the correlation, given Λ and Γ, (24) and (25) jointly determine the steady state (θ, Σ).

Under RI, 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, (25) 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

$$\Lambda = \left(\frac{-\rho\Sigma\sqrt{\Omega} + \sqrt{\rho^2\Sigma^2\Omega - \Sigma\Psi\left[\rho^2\Omega - (\Psi - \Sigma)\right]}}{\rho^2\Omega - (\Psi - \Sigma)}\right)^2. \tag{26}$$

It is clear from (26) 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.

In contrast, consider the RI problem with a fixed information-processing cost (λ). From (21) and (26), it is clear that in the presence of correlated noise ($\rho > 0$), there is a smaller change in Σ when there is a change in Ω because $\frac{\partial \ln \Sigma}{\partial \ln \Omega} < 1$, and consequently, the change in Ω will also change θ because Λ depends on the interactions between Ω and Σ.

5.3 Applications

We now specialize our generic LQG RI problem to study the dynamics of consumption and the pricing of assets. We also introduce a related informational distortion that has proven useful at matching these facts: the idea of model uncertainty. Although we will discuss it in more detail below, we find it useful to present a brief discussion of the relationship between the two frictions. With RI, the agent is uncertain of the current value of the state, but is certain about the distribution of future states (conditional on each possible value of the current state). With model uncertainty the agent is certain of the current state but is uncertain about the distribution of future states.

5.3.1 Consumption Dynamics under RI

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.[16] In Luo (2008), households solve the following dynamic consumption-savings problem

$$v(s_0) = \max_{\{c_t\}} \left\{ E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t) \right] \right\} \tag{27}$$

subject to

$$s_{t+1} = Rs_t - c_t + \zeta_{t+1}, \tag{28}$$

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

$$s_t = w_t + \frac{1}{R} \sum_{j=0}^{\infty} R^{-j} E_t \left[y_{t+j} \right] \tag{29}$$

is permanent income (the expected present value of lifetime resources), consisting of financial wealth (w_t) plus human wealth (i.e., the expected discounted present value of current and future labor income, y),

$$\zeta_{t+1} = \frac{1}{R} \sum_{j=t+1}^{\infty} \left(\frac{1}{R} \right)^{j-(t+1)} \left(E_{t+1} - E_t \right) \left[y_j \right]; \tag{30}$$

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.[17] We assume y follows an AR(1) process with persistence coefficient $\rho \in [0,1]$, $y_{t+1} = \rho y_t + \varepsilon_{t+1}$, where $\varepsilon_{t+1} \sim N\left(0, \omega^2\right)$, $s_t = w_t + y_t / \left(R - \rho\right)$ and $\zeta_{t+1} = \varepsilon_{t+1} / \left(R - \rho\right)$.[18] 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:

$$c_t = \left(R - 1\right) s_t. \tag{31}$$

[16] 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.

[17] Note that in this case, the flow budget constraint is $w_{t+1} = Rw_t - c_t + y_t$.

[18] We can accomodate any income process that implies permanent income is finite; that is, nonstationary processes are permitted provided they do not grow too quickly on average.

As shown in Section 4.2, under RI, the consumer cannot observe the true state perfectly and he needs to estimate the state using the Kalman filtering equation

$$\widehat{s}_{t+1} = R\widehat{s}_t - c_t + \eta_{t+1}, \tag{32}$$

where $\widehat{s}_t = E_t[s_t]$ is the perceived state,

$$\eta_{t+1} = \theta R(s_t - \widehat{s}_t) + \theta(\zeta_{t+1} + \xi_{t+1}) \tag{33}$$

is the innovation to the mean of the distribution of perceived permanent income,

$$s_t - \widehat{s}_t = \frac{(1-\theta)\zeta_t}{1-(1-\theta)R\cdot L} - \frac{\theta\xi_t}{1-(1-\theta)R\cdot L}, \tag{34}$$

$E_t[\eta_{t+1}] = 0$ because the expectation is conditional on the perceived signals and inattentive agents cannot perceive the lagged shocks perfectly, $\theta = \Sigma\Lambda^{-1}$ is the Kalman gain, and given $s_0 \sim N(\widehat{s}_0, \Sigma)$.[19] Given (33), the variance of the innovation to the perceived state is $\omega_\eta^2 = \mathrm{var}(\eta_{t+1}) = \frac{\theta}{1-(1-\theta)R^2}\omega_\zeta^2 > \omega_\zeta^2$, which means that ω_η^2 reflects two sources of uncertainty facing the consumer: (i) fundamental uncertainty, ω_ζ^2 and (ii) induced uncertainty, i.e., state uncertainty due to RI, $\left[\frac{\theta}{1-(1-\theta)R^2} - 1\right]\omega_\zeta^2$. Therefore, as κ decreases, the relative importance of induced uncertainty to fundamental uncertainty becomes more and more important.

As shown in Luo (2008), the certainty equivalence principle holds in this RI-PI model and the consumption function is:

$$c_t = (R-1)\widehat{s}_t, \tag{35}$$

Combining (28), (35), with (32) yields the following expression for the change in consumption:

$$\Delta c_t = (R-1)\left[\frac{\theta\zeta_t}{1-(1-\theta)R\cdot L} + \theta\left(\xi_t - \frac{\theta R\xi_{t-1}}{1-(1-\theta)R\cdot L}\right)\right], \tag{36}$$

where L is the lag operator. We require $(1-\theta)R^2 < 1$, which is the stabilizability requirement for this model (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 (36), 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. All these moments depend on the Kalman gain θ.

[19] For the filter to converge to the steady state we need $\kappa > \ln(R) \approx R - 1$.

It is worth noting that RI has different implications for consumption dynamics in a representative agent model or a model with a continuum of identical agents. The key reason is the presence of the RI-induced endogenous noise, ξ. Specifically, if we regard the typical consumer discussed above as the representative agent, (36) can be used to examine the implications of RI on aggregate consumption. In this case, it is obvious that RI helps predict that aggregate consumption is sensitive to lagged income information.[20] Furthermore, the relative volatility of consumption growth to income growth can be written as

$$\mu \equiv \frac{\mathrm{sd}\left(\Delta c_t\right)}{\mathrm{sd}\left(\Delta y_t\right)} = \sqrt{\frac{\theta}{1-\left(1-\theta\right)R^2}} \geq 1, \tag{37}$$

when $\rho = 1$ for $\theta \leq 1$. Note that when the income process is a random walk, μ is greater than 1 when the representative consumer is inattentive, i.e., $\theta < 1$. In other words, RI increases the relative volatility of consumption growth, which exacerbates the excess smoothness puzzle documented in the consumption literature.[21] The intuition behind this result is as follows. From (36), the behavior of consumption is determined by two channels: (i) the slow propagation channel (the $1-(1-\theta)R\cdot L$ term) and (ii) the noise channel (the presence of ξ_t). In the representative agent model, the noise channel dominates the slow propagation channel and thus the volatility of consumption growth is higher under RI.

In the model with a continuum of identical consumers, aggregating across all individual consumers facing the same aggregate income process using (36) yields the expression of the change in aggregate consumption:

$$\Delta c_t = (R-1)\left[\frac{\theta\zeta_t}{1-(1-\theta)R\cdot L} + \theta\left(\overline{\xi}_t - \frac{\theta R\overline{\xi}_{t-1}}{1-(1-\theta)R\cdot L}\right)\right], \tag{38}$$

where i denotes a particular individual, $E^i\left[\cdot\right]$ is the population average, and $\overline{\xi}_t = E^i\left[\xi_t\right]$ is the common noise.[22] This expression shows that even if every consumer only faces the common shock

[20] Furthermore, as documented in Reis (2006), the impulse response of aggregate consumption to aggregate income takes a hump-shaped form, which means that aggregate consumption reacts to income shocks gradually.

[21] In the US data aggregate consumption growth is much smoother than income and is sensitive to past information. (In the US data, the smoothness ratio is around 0.56.) These two anomalies have been termed the excess smoothness and excess sensitivity puzzles in the literature.

[22] For simplicity, here we use the same notation c for aggregate consumption. A purely technical point that we ignore here is that it is not obvious that one *can* construct a continuum indepedent random variables in such a way that their joint distribution is constant over time; Judd (1985), Uhlig (1996), and Sun (2006) propose various methods for dealing with this problem.

ζ, the RI economy still has heterogeneity since each consumer faces the idiosyncratic noise induced by finite channel capacity. As argued in Sims (2003), although the randomness in an individual's response to aggregate shocks will be idiosyncratic because it arises from the individual's own information-processing constraint, there is likely a significant common component. Therefore, the common term of the idiosyncratic error, $\overline{\xi}_t$, lies between 0 and the part of the idiosyncratic error, ξ_t, caused by the common shock to permanent income, ζ_t. Formally, assume that ξ_t consists of two independent noises: $\xi_t = \overline{\xi}_t + \xi_t^i$, where $\overline{\xi}_t = E^i[\xi_t]$ and ξ_t^i are the common and idiosyncratic components of the error generated by ζ_t, respectively. A single parameter, $\widetilde{\lambda} = \mathrm{var}\left(\overline{\xi}_t\right) / \mathrm{var}\left(\xi_t\right) \in [0,1]$, can be used to measure the common source of coded information on the aggregate component (or the relative importance of $\overline{\xi}_t$ vs. ξ_t). The relative volatility in terms of θ and λ can be written as

$$\mu = \sqrt{\frac{\theta^2}{1-[(1-\theta)R]^2} + \widetilde{\lambda}^2 \left\{ \frac{\theta}{1-(1-\theta)R^2} - \frac{\theta^2}{1-[(1-\theta)R]^2} \right\}}. \tag{39}$$

The assumption of $\widetilde{\lambda} = 0$ is equivalent to assuming that individuals do not interact with each other directly or indirectly via conversation, imitation, newspapers, or other media. In this case, the excess smoothness ratio $\mu = \sqrt{\frac{\theta^2}{1-[(1-\theta)R]^2}} < 1$.

In contrast, the assumption of $\widetilde{\lambda} = 1$ (the representative agent case discussed above) means that peoples' needs for information coding are exactly the same and they completely rely on the common source of coded information (everyone watches CNN). In this case, $\mu = \sqrt{\frac{\theta}{1-(1-\theta)R^2}}$. Hence, given the value of θ, $\mu \in \left[\sqrt{\frac{\theta^2}{1-[(1-\theta)R]^2}}, \sqrt{\frac{\theta}{1-(1-\theta)R^2}}\right]$. For example, if $\theta = 50\%$ and $\lambda = 0.1$, the model predicts that $\mu = 0.58$, which is close to its empirical counterpart in U.S. data (around 0.56). The intuition is that in the model with a continuum of consumers, the slow propagation channel dominates the noise channel when $\widetilde{\lambda}$ is sufficiently low. Hence, RI provides an alternative explanation to the two consumption excesses in this RI model with a continuum of individuals.

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

$$\Sigma = \frac{-\left[\Omega\left(R-1\right) - \lambda R\right] + \sqrt{\left[\Omega\left(R-1\right) - \lambda R\right]^2 + 4\lambda\Omega R^2}}{2\left(R-1\right)R^2}. \tag{40}$$

Comparing (40) with $\Sigma = \Omega / \left(\exp\left(2\kappa\right) - R^2\right)$ obtained in the fixed capacity case, 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:

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

In this case, the Kalman gain is

$$\theta = 1-\left\{R^2 + \frac{2(R-1)R^2}{-\left[(R-1)-R\left(\lambda/\omega_\zeta^2\right)\right]+\sqrt{\left[(R-1)-R\left(\lambda/\omega_\zeta^2\right)\right]^2+4R^2\left(\lambda/\omega_\zeta^2\right)}}\right\}^{-1}. \quad (41)$$

After implementing these policies, ω_ζ^2 is scaled down to $0.5\omega_\zeta^2$ (i.e., the economy switches to a more stable environment) 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 (41). Figure (2) plots the different implications of fixed capacity and fixed information cost for consumption dynamics after implementing the stabilization policy: consumption growth reacts more slowly to the income shock when λ is fixed. The intuition behind this result is simple. In the fixed λ case some capacity will be reallocated to other sources because a reduction in macroeconomic uncertainty leads to smaller welfare losses due to RI.

Comparison to Other Models of Sluggish Adjustment Here we briefly compare the RI model to two other formulations that imply sluggish adjustment in consumption. The first, denoted habit formation (HF), postulates that lagged consumption negatively affects the marginal utility of current consumption, either additively

$$u'(c_t, c_{t-1}) = u'(c_t - gc_{t-1})$$

for some$g > 0$, or multiplicatively

$$u'(c_t, c_{t-1}) = u'\left(\frac{c_t}{c_{t-1}^\gamma}\right)$$

for some $\gamma > 0$.[23] In both cases, since c_{t-1} is fixed and marginal utility declines with c_t, the consumer dislikes large changes in consumption. The second formulation is called inattentiveness,

[23] The lagged consumption terms can enter in more complicated ways, including with more lags (as in Otrok 2001) or in a nonlinear fashion (as in Campbell and Cochrane 1999).

and presumes that agents infrequently update the information set they use to make decisions, so that consumption will be nonresponsive to events that occur in between adjustment periods.

Luo (2008) compared the distinct implications of RI and HF for consumption dynamics, and showed that the effects of HF and RI on aggregate consumption could be similar or the same because aggregating across all individuals would weaken or even eliminate the effect of the endogenous noise on consumption growth, i.e., $\widetilde{\lambda}$ is sufficiently low or zero. HF implies that slow adjustment in consumption is optimal because consumers are assumed to prefer to smooth not only consumption but also consumption growth, whereas RI predicts that slow adjustment in consumption is optimal because capacity constraints make consumers take more time to acquire and process information. In other words, RI provides a non-psychological explanation for slow adjustment in consumption that is caused by information-processing constraints rather than a direct assumption of the structure of preferences. In addition, it is worth noting that HF affects consumption decisions both with and without uncertainty, whereas RI affects consumption through its interaction with the fundamental uncertainty and it has no impact on consumption in the absence of uncertainty.

In Reis (2006), a model of costly planning and inattentiveness, during the intervals of inattentiveness consumption dynamics are determined by the standard deterministic consumer's optimization problem, whereas at the adjustment dates, consumption is determined by the standard stochastic consumer problem. Reis (2006) showed that aggregate consumption growth exhibits slow adjustment because "news" diffuses across all individuals slowly. Hence, in the inattentiveness economy, individuals adjust consumption infrequently but completely once they choose to adjust, and all the sluggishness in aggregate consumption comes from aggregating across all individuals. In contrast, individuals under RI adjust their optimal consumption plans frequently but incompletely, and the sluggishness of aggregate consumption comes from individuals' incomplete consumption adjustments.

5.3.2 Implications for Consumption Volatility in Emerging and Developed Countries

Aguiar and Gopinath (2007) consider two groups of small economies (emerging and developed economies), and find that consumption is more volatile than income at business cycle frequencies for emerging markets, as compared to a ratio of less than one for developed markets. They show that a small open economy (SOE) RBC model driven primarily by shocks to trend growth can explain well this regularity about the relative volatility of consumption to income in emerging markets. However, by using a long time series data over $1900 - 2005$, García-Cicco, Pancrazi, and Uribe (2010) estimate an RBC model driven by the same shocks considered in Aguiar and Gopinath (2007), and find that the model does a poor job accounting for the observed business cycle fluctuations in Argentina and Mexico along a number of dimensions, including the relative volatility of consumption to income. Luo and Young (2013) show that the RI model with a fixed information-processing cost can explain the observed difference in consumption volatility in emerging and developed countries.

The PIH model presented in Section 5.3.1 can be regarded as a SOE model in which the constant interest rate is given exogenously and there are a continuum of consumers in the model economy. Using (36) and assuming that all idiosyncratic errors are canceled out after aggregation, the relative volatility of consumption growth to income growth can be written as

$$\mu \equiv \frac{\mathrm{sd}\left(\Delta c_t\right)}{\mathrm{sd}\left(\Delta y_t\right)} = \left(\frac{R-1}{R-\rho}\sqrt{\frac{1+\rho}{2}}\right)\sqrt{\frac{\theta^2}{1-\left(\left(1-\theta\right)R\right)^2}}, \tag{42}$$

where $\mathrm{sd}\left(\cdot\right)$ denotes standard deviation. It is straightforward to show that the relative consumption volatility is decreasing with the degree of imperfect state observations because $\partial\mu/\partial\theta < 0$. As shown in Luo, Nie, and Young (2012), if there is no imperfect-state-observation assumption (i.e., $\theta = 1$), the model cannot generate the empirical relative consumption volatility. For example, if $R = 1.04$, the full information model predicts that μ in emerging and developed economies would be 0.28 and 0.24, respectively. In contrast, in the data, the corresponding μ values are 1.35 and 0.98, respectively.[24] In the RI model with fixed capacity, θ is uniquely determined by fixed capacity κ and thus has no impact on the cross-country comparison if emerging and developed counties have the same average amount of channel capacity. In contrast, if we adopt the fixed information-processing cost assumption, (41) and (42) can have the potential to generate the observed difference in consumption volatility in emerging and developed countries because θ is an increasing function of income uncertainty and income uncertainty in emerging countries is much higher than that in developed counties ($\mathrm{sd}\left(\Delta y\right)/\mu(y)$ is 3.82 in emerging countries, while it is 2.07 in developed countries.) Intuitively, in developed counties consumers pay less attention to macroeconomic conditions because the fundamental uncertainty is low; consequently, the aggregate consumption process in these countries is more stable relative to the income process.

[24] See Table 1 in Luo, Nie, and Young (2012) for the estimated income processes in both emerging and developed countries.

5.3.3 Equilibrium Asset Pricing Implications

The PIH model presented in Section 5.3.1 is usually regarded as a partial equilibrium model where the return to saving R is taken as fixed. However, as noted in Hansen (1987) and Cochrane (2005), it does not have to be interpreted as a partial equilibrium result — it can be viewed instead as a general equilibrium model with a linear production technology and an exogenous income process. Given the expression of optimal consumption in terms of the state variables derived from the PIH model with imperfect-state-observation, we can price assets by treating the process of aggregate consumption that solves the model as though it were an endowment process. In this setup, equilibrium prices are shadow prices that leave the agent content with that endowment process.

In the model setting specified in Section 5.3.1, w can be regarded as capital. R can be regarded as the return on the linear technology and is not yet the interest rate (the equilibrium rate of return on one-period claims to consumption). As proposed in Cochrane (2005) and used in Luo and Young (2010b), after finding optimal consumption as in (35), we can price one-period claims using this equilibrium consumption stream. Denoting the risk free rate by R^f, we have the following Euler equation:

$$\frac{1}{R^f} \equiv E_t\left[\beta\frac{u'(c_{t+1})}{u'(c_t)}\right] = \beta E_t\left[\frac{\bar{c}-c_{t+1}}{\bar{c}-c_t}\right] = \beta = \frac{1}{R},$$

where $E_t[\cdot]$ is the consumer's expectation operator conditional on his processed information at time t. We can now use the basic pricing equation, $p = E[mx]$,[25] to compute the price of the stream of aggregate consumption (treated as the stream of endowments) as[26]

$$\begin{aligned} p_t &= E_t\left[\sum_{j=1}^{\infty}(m_{t,t+j}c_{t+j})\right] \\ &= \underbrace{\frac{1}{R-1}c_t}_{p_t^{rn}} - \underbrace{\frac{1}{\bar{c}-c_t}\Xi}_{p_t^{ra}}, \end{aligned} \tag{43}$$

where $m_{t,t+j} \equiv \beta^j\frac{u'(c_{t+j})}{u'(c_t)}$ is the stochastic discount factor, and $\Xi \equiv \sum_{j=1}^{\infty}\left(\beta^j \operatorname{var}_t(c_{t+j})\right) = \frac{(2-\theta)R}{1-R^2(1-\theta)}\omega_\zeta^2$. Denoting the risk-neutral component by p_t^{rn} and the risk-adjusted component by p_t^{ra}, we have

$$p_t^{rn} = \frac{1}{R-1}c_t \tag{44}$$

[25] Note that we know $E[mx]$ after solving the PIH model given the state variables and can use them to determine the asset price p.

[26] For the details of the derivation, see Luo and Young (2010b).

and

$$p_t^{ra} = \frac{1}{\bar{c} - c_t} \frac{(2-\theta) R}{1 - R^2 (1-\theta)} \omega_\zeta^2. \tag{45}$$

(43) yield the following implications. The first term in (43) is the risk-neutral component denoted by p_t^{rn}. This term can be regarded as the value of a perpetuity paying c_t. The second term is the risk-adjusted component, p_t^{ra}; it lowers the asset price relative to the risk-neutral level because $c_t \leq \bar{c}$ and it is decreasing with the degree of attention (θ).

From (43) and (45), it is clear that the Kalman gain (θ) also plays a key role in determining the general equilibrium asset prices under RI. For example, if the economy switches to a more stable environment due to the stabilization policy (i.e, ω_ζ^2 is scaled down to $0.5\omega_\zeta^2$), the fixed λ assumption predicts that the Kalman filter gain, $\theta = \Sigma\Lambda^{-1}$, is reduced. Given that $\lambda/\omega_\zeta^2 = 0.000135$ and $\theta = 0.79$, we can easily calculate that $\theta = 0.68$ using (41).

5.3.4 RI-induced Precautionary Savings

It is well known that the linear-quadratic model has some undesirable features. For example, it satisfies the certainty equivalence property, ruling out any response of saving to uncertainty (that is, precautionary behavior). Given that the important component of RI is the introduction of endogenous uncertainty into the household problem, it is not particularly desirable to use a model in which this uncertainty cannot manifest itself in decision rules. Fully nonlinear versions of the RI problem are solved in Sims (2006), Lewis (2009), Batchuluun, Luo, and Young (2009), and Tutino (2012); these papers show that the precautionary aspect of RI is important when channel capacity is small. But the models solved in those papers have either very short horizons or extremely simple setups due to numerical obstacles – the state of the world is the distribution of true states and this distribution is not well-behaved. It is important to find a class of models that can produce precautionary behavior while maintaining tractability in the RI setup, if the properties of finite channel capacity are going to be thoroughly explored. Luo and Young (2010a) and Luo, Nie, and Young (2012) examine how RI affects precautionary savings within the risk-sensitive or robust LQG settings. In a risk-sensitive model, agents effectively compute expectations through a distorted lens, increasing their effective risk aversion by overweighting negative outcomes. The resulting decision rules depend explicitly on the variance of the shocks, producing precautionary savings, but the value functions are still quadratic functions of the states, leading again to the optimality of Gaussian noise in the RI model. Thus, we can preserve the tractability of the LQ PIH model without being forced to accept certainty equivalence. In the robust control model of Hansen and Sargent (1995), agents are concerned about the possibility that their model is misspecified in a manner that is difficult to detect statistically; as a result, they choose their decisions as if the subjective distribution over shocks was chosen by a malevolent nature in order to minimize their expected utility (that is, the solution to a robust decision-maker's problem is the equilibrium of a max-min game between the decision-maker and nature). Hansen and Sargent (2007) present an observational equivalence result for RS and Robust models for consumption and savings decisions: any consumption path that could be generated by a model featuring risk sensitivity can also be generated by a model with robustness. Thus, introducing RI into the Robust model is again straightforward, since the model retains the optimality of Gaussian noise.

Following Hansen and Sargent (2007), Luo, Nie, and Young (2012) use the multiplier preference structure to introduce RB into the RI model proposed in Section 5.3.1:

$$\widehat{v}\left(\widehat{s}_t\right) = \max_{c_t}\left\{-\frac{1}{2}\left(c_t - \overline{c}\right)^2 + \beta \min_{m_{t+1}} E_t\left[m_{t+1}\widehat{v}\left(\widehat{s}_{t+1}\right) + \vartheta_0 m_{t+1}\ln m_{t+1}\right]\right\}, \tag{46}$$

subject to (32). Here m_{t+1} is the likelihood ratio, $E_t\left[m_{t+1}\ln m_{t+1}\right]$ is defined as the relative entropy of the distribution of the distorted model with respect to that of the approximating model, and $\vartheta_0 > 0$ is the shadow price of capacity that can reduce the distance between the two distributions, i.e., the Lagrange multiplier on the constraint:

$$E_t\left[m_{t+1}\ln m_{t+1}\right] \leq \widetilde{\eta},$$

where $\widetilde{\eta} \geq 0$ defines an entropy ball of the distribution of the distorted model with respect to that of the approximating model. Minimizing (46) with respect to m_{t+1} yields

$$m_{t+1} = \frac{\exp\left(-v\left(\widehat{s}_{t+1}\right)/\vartheta_0\right)}{E_t\left[\exp\left(-v\left(\widehat{s}_{t+1}\right)/\vartheta_0\right)\right]}, \tag{47}$$

and it is straightforward to show that substituting m_{t+1} into (46) yields the following Bellman equation:

$$\widehat{v}\left(\widehat{s}_t\right) = \max_{c_t}\left\{-\frac{1}{2}\left(c_t - \overline{c}\right)^2 + \beta\mathcal{R}_t\left[\widehat{v}\left(\widehat{s}_{t+1}\right)\right]\right\}, \tag{48}$$

where

$$\mathcal{R}_t\left[\widehat{v}\left(\widehat{s}_{t+1}\right)\right] = -\vartheta_0 \log E_t\left[\exp\left(-v\left(\widehat{s}_{t+1}\right)/\vartheta_0\right)\right],$$

subject to (32). (48) is a standard risk-sensitive (RS) dynamic programming problem and can be easily solved using the standard procedure.[27] The following proposition summarizes the solution to the RB-RI model when $\beta R = 1$.

[27] Risk-sensitivity was first introduced into the LQG framework by Jacobson (1973) and extended by Whittle (1981). Hansen and Sargent (1995) introduced discounting into the RS specification and show that the resulting decision rules are time-invariant.

Proposition 5 *Given ϑ_0 and θ, the consumption function under RB and RI is*

$$c_t = \frac{R-1}{1-\Pi}\widehat{s}_t - \frac{\Pi\overline{c}}{1-\Pi}, \tag{49}$$

where $\widehat{s}_t$ is governed by

$$\widehat{s}_{t+1} = \rho_s \widehat{s}_t + \frac{\Pi\overline{c}}{1-\Pi} + \eta_{t+1}, \tag{50}$$

where $\rho_s = \frac{1-R\Pi}{1-\Pi} \in (0,1)$,

$$\Pi = \frac{R\omega_\eta^2}{\vartheta_0} \in (0,1), \tag{51}$$

η_{t+1} is defined in (33),

$$\omega_\eta^2 = \frac{\theta}{1-(1-\theta)R^2}\omega_\zeta^2,$$

and $\theta = 1 - 1/\exp(2\kappa)$. The value function is

$$v(\widehat{s}_t) = \Omega\left(\widehat{s}_t - \frac{\overline{c}}{R-1}\right)^2 + \rho, \tag{52}$$

where $\Omega = -\frac{R(R-1)}{2(1-\Pi)}$ and $\rho = \frac{\vartheta_0}{2(R-1)}\ln\left(1-\frac{(R-1)\Pi}{1-\Pi}\right)$.

Proof. See Online Appendix in Luo, Nie, and Young (2012). $\Pi < 1$ can be obtained because the second-order condition for the optimization problem is

$$\frac{R(R-1)}{2\left(1-R\omega_\eta^2/(\vartheta_0)\right)} > 0, \text{ i.e., } \Pi < 1.$$

■

It is clear from (49) and (51) determine the effects of model uncertainty due to RB and state uncertainty due to RI on the marginal propensity to consume *out of perceived permanent income* (MPC_η) and the constant precautionary saving premium. It is clear from these two expressions that Π governs how RB and RI interact and then affect the consumption function and precautionary savings. Since Π is increasing with the degrees of RB (less ϑ_0) and RI (less κ and θ), it is straightforward to show that either RB or RI leads to more constant precautionary savings and higher marginal propensity to consume, holding other factors constant and given that $\Pi < 1$.

Furthermore, in this case it is worth noting that RB and RI affect consumption and precautionary savings through distinct channels. RI affects Π via increasing the variance of the innovation to the perceived state, ω_η^2, whereas RB affects Π via changing the structure of the response of consumption to income shocks. Furthermore, if we consider the marginal propensity to consume *out of true permanent income,*

$$MPC_{\zeta} \equiv \frac{R-1}{1-R\theta/\left[\vartheta_0\left(1-(1-\theta)R^2\right)\right]\omega_{\zeta}^2}\theta, \tag{53}$$

we have the same conclusion about how induced uncertainty due to RB affects the consumption function but obtain different conclusions about how induced uncertainty due to RI affects the consumption function:

$$\frac{\partial\left(MPC_{\zeta}\right)}{\partial\vartheta_0}<0, \frac{\partial\left(MPC_{\zeta}\right)}{\partial\theta}>0.$$

Note that there is a distinction between the model proposed above and a similar one used in Luo and Young (2010a). In those other papers, agents were assumed to trust the Kalman filter they use to process information, meaning that decisions were only robust to misspecification of the income process. An implicit assumption in the two papers is that the evil agent (the minimizing agent) has the same information set as the consumer (the maximizing agent). In that model Π was independent of θ, and for the questions at hand here the resulting values were too small. By adding the additional concern for robustness developed here, we are able to strengthen the effects of robustness on decisions.

5.3.5 Long-run Consumption Risk and Portfolio Choice

Luo (2010) adopted the log-linear approximation method proposed by Campbell and Viceira (2002) and Campbell (1993) to solve an RI version of the intertemporal portfolio choice model after considering the long-term consumption risk facing investors. A major advantage of the log-linearization approach is that we can approximate the original nonlinear problem by a log linear-quadratic (LQ) framework when the coefficient of relative risk aversion is close to 1 and thus can justify the optimality of Gaussian posterior uncertainty under RI.[28]

In the standard consumption and portfolio choice model, investors choose consumption and

[28] Luo and Young (2012) extended the result in Luo (2010) to non-expected utility preferences that separate attitudes toward risk aversion across states from risk aversion over time (often called intertemporal substitution); these preferences also have an interpretation as representing the model uncertainty aversion discussed previously. Rather than extend an already long survey, we concentrate only on the expected utility case here. For completeness, we note only that the key implication is that the investor have a preference for early resolution of uncertainty; as RI pushes uncertainty revelation into the distant future, the fraction of risky assets declines and/or the premium for holding them rises.

asset holdings to maximize the intertemporal time-separable utility, defined over consumption:

$$\max_{\{C_t,\alpha_t\}_{t=0}^{\infty}} E_0\left[\sum_{t=0}^{\infty}\beta^t u(C_t)\right], \tag{54}$$

where $u(C_t) = \frac{C_t^{1-\gamma}-1}{1-\gamma}$ is the power utility function, C_t represents individual's consumption at time t, β is the discount factor, and $\gamma \geq 1$ is the coefficient of relative risk aversion. There are two tradable financial assets in the model economy: Asset e is risky, with iid one-period log (continuously compounded) return $r_{t+1}^e \sim N(\mu, \omega^2)$, while the other asset f is riskless, with constant log return given by r^f. We refer to asset e as the market portfolio of equities, and asset f as savings or checking accounts. The intertemporal budget constraint for the investor is

$$A_{t+1} = R_{t+1}^p (A_t - C_t), \tag{55}$$

where A_{t+1} is the individual's financial wealth which is defined as the value of financial assets carried over from period t at the beginning of period $t+1$, $A_t - C_t$ is savings, and R_{t+1}^p is the one-period gross return on savings given by

$$R_{t+1}^p = \alpha_t\left(R_{t+1}^e - R^f\right) + R^f, \tag{56}$$

where $R_{t+1}^e = \exp(r_{t+1}^e)$, $R^f = \exp(r^f)$, and $\alpha_t = \alpha$ is the proportion of savings invested in the risky asset.[29] As shown in Campbell (1993), the following approximate expression for the log return on wealth holds:

$$r_{t+1}^p = \alpha(r_{t+1}^e - r^f) + r^f + \frac{1}{2}\alpha(1-\alpha)\omega^2. \tag{57}$$

Following the log-linearization method proposed in Campbell (1993), Viceira (2001), and Campbell and Viceira (2002), the original intertemporal budget constraint, (55), can be written in log-linear form:

$$\Delta a_{t+1} = (1 - 1/\phi)(c_t - a_t) + \psi + r_{t+1}^p, \tag{58}$$

where $c - a = E[c_t - a_t]$ is the unconditional expectation of $c_t - a_t$, $\phi = 1 - \exp(c - a)$, $\psi = \log\phi - (1 - 1/\phi)\log(1-\phi)$, and lowercase letters denote logs.

When γ is close to 1, the original CRRA utility function can be approximately by a linear-quadratic form: $c_t + \frac{1}{2}(1-\gamma)c_t^2$. This log-LQG model can therefore fit the RI-LQG framework (at least approximately). Solving this RI problem yields:[30]

$$c_t = \log(1-\beta) + \widehat{a}_t, \tag{59}$$

[29] Given iid equity returns and power utility function, the share invested in equities, α_t, is constant over time.

[30] Here we have used the approximation result that $\phi = \beta$. Note that ϕ is independent of the degree of attention, κ, because ϕ approaches to β as γ converges to 1.

and the perceived state $\widehat{a}_t$ is characterized by the following Kalman filter equation:

$$\widehat{a}_{t+1} = (1-\theta)\left[\left(1-\frac{1}{\beta}\right)c_t + \frac{1}{\beta}\widehat{a}_t - \log\beta + \psi\right] + \theta a^*_{t+1}. \tag{60}$$

Combining them gives the evolution of $\widehat{a}_t$:

$$\widehat{a}_{t+1} = (1-\theta)\widehat{a}_t + \theta a^*_{t+1}, \tag{61}$$

where $\theta = 1 - 1/\exp(2\kappa)$ is the Kalman gain, $a^*_{t+1} = a_{t+1} + \xi_{t+1}$ is the observed signal, and ξ_{t+1} are the iid noise with $\operatorname{var}(\xi_{t+1}) = \Sigma/\theta$. Hence, combining equations (58), (59), with (61) gives the expression for individual consumption growth:[31]

$$\Delta c_{t+1} = \theta\left[\frac{\alpha u_{t+1}}{1-((1-\theta)/\beta)\cdot L} + \left(\xi_{t+1} - \frac{(\theta/\beta)\,\xi_t}{1-((1-\theta)/\beta)\cdot L}\right)\right], \tag{62}$$

where L is the lag operator and $u_{t+1} = r^e_{t+1} - \mu$.

Since consumption reacts to the shock to the equity return slowly and with delay, we need to consider the long-run risk when determining optimal asset allocation. Following Parker (2003) and Parker and Julliard (2005), we define the long-term consumption risk as the covariance of asset returns and consumption growth over the period of the return and many following periods. With finite channel capacity, we have the following equality for the risky asset e and the risk-free asset f:

$$E_t\left[R^e_{t+1}C^{-\gamma}_{t+1+S}\right] = E_t\left[R^f C^{-\gamma}_{t+1+S}\right],$$

which can be transformed to the following stationary form:

$$E_t\left[R^e_{t+1}(C_{t+1+S}/C_t)^{-\gamma}\right] = E_t\left[R^f(C_{t+1+S}/C_t)^{-\gamma}\right], \tag{63}$$

where the expectation $E_t[\cdot]$ is conditional on the entire history of the economy up to t and S is the horizon of many periods in the future over which consumption response under RI is studied. Given the expression for consumption dynamics, it is clear that S is infinitely large because consumption takes infinite periods to react to the exogenous shock. The standard equality, $E_t[R^e_{t+1}C^{-\gamma}_{t+1}] = E_t[R^f C^{-\gamma}_{t+1}]$, does not hold here because consumption reacts slowly with respect to the innovations to equity returns and thus cannot finish adjusting immediately and completely. Log-linearizing equation (63) yields

$$E_t\left[r^e_{t+1}\right] - r^f + \frac{1}{2}\omega^2 = \sum_{j=0}^{S}\operatorname{covar}_t\left[\Delta c_{t+j+1}, r^e_{t+1}\right], \tag{64}$$

[31] Note that this MA(∞) expression requires that $(1-\theta)/\beta < 1$, which is equivalent to $\kappa > \frac{1}{2}\log(1/\beta)$.

where we have used $\gamma \simeq 1$, $c_{t+1+S} - c_t = \sum_{j=0}^{S} \Delta c_{t+1+j}$, and Δc_{t+1+j} is given by (62). Specifically, the long-run impact of the innovation in the equity return on consumption growth can be rewritten as

$$\lim_{S\to\infty} \left(\sum_{j=0}^{S} \text{covar}_t \left[\Delta c_{t+1+j}, r_{t+1}^{e} \right] \right) = \varsigma \alpha \omega^2, \tag{65}$$

where ς is the ultimate consumption risk:

$$\varsigma = \frac{\theta}{1 - (1-\theta)/\beta} > 1 \tag{66}$$

when the restriction $1 - (1-\theta)/\beta > 0$, i.e., $\theta > 1 - \beta$ holds. The optimal share invested in the equity is thus:

$$\alpha^* = \frac{1}{\varsigma} \frac{\mu - r^f + 0.5\omega^2}{\gamma \omega^2}, \tag{67}$$

where $1/\varsigma = [1 - (1-\theta)/\beta]/\theta < 1$. Furthermore, the consumption function is

$$c_t^* = \log(1-\beta) + \widehat{a}_t, \tag{68}$$

where (61), $\psi = \log\beta - (1 - 1/\beta)\log(1-\beta)$,[32] ξ_t are the iid idiosyncratic noise with $\omega_{\xi}^2 = \text{var}[\xi_{t+1}] = \Sigma/\theta$, and $\Sigma = \frac{\alpha^{*2}\omega^2}{\exp(2\kappa) - (1/\beta)^2}$ is the steady state conditional variance. The change in individual consumption can thus be written as

$$\Delta c_{t+1}^* = \theta \left[\frac{\alpha^* u_{t+1}}{1 - ((1-\theta)/\beta) \cdot L} + \left(\xi_{t+1} - \frac{(\theta/\beta)\xi_t}{1 - ((1-\theta)/\beta) \cdot L} \right) \right]. \tag{69}$$

Proposition 6 *Consider the optimal consumption and investment decisions (67)-(69). Suppose that γ is close to 1, and $\theta > 1 - \beta$. We then obtain the following expression for aggregate consumption growth under RI:*

$$\Delta c_{t+1}^* = \frac{\theta \alpha^* u_{t+1}}{1 - ((1-\theta)/\beta) \cdot L}, \tag{70}$$

where θ is the average degree of inattention in the model economy and α^* is the average optimal share invested in the risky asset and is given by (67).[33] *This MA(∞) process implies that: (1) the standard deviation of consumption growth is*

$$\text{sd}\left[\Delta c_{t+1}^*\right] = \lambda \alpha \omega, \tag{71}$$

[32] Here we use the facts that $\beta = 1 - \exp(c-a)$ and $c - a = E[c_t - a_t]$ is the steady state value.

[33] That is, $\alpha^* = \frac{1-(1-\theta)/\beta}{\theta}\alpha$ and $\alpha = \frac{\mu - r^f + 0.5\omega^2}{\omega^2}$ is the optimal share invested in the stock market in the full-information RE model.

where the excess smoothness ratio $\lambda = (1-(1-\theta)/\beta) \ / \sqrt{1-((1-\theta)/\beta)^2}$, *(2) the correlation between consumption growth and equity return is*

$$\operatorname{corr}\left[\Delta c_{t+1}^*, r_{t+1}^e\right] = \sqrt{1-((1-\theta)/\beta)^2}, \tag{72}$$

(3) the covariance between aggregate consumption growth and the asset return is

$$\operatorname{covar}\left[\Delta c_{t+1}^*, r_{t+1}^e\right] = \theta\alpha^*\omega^2, \tag{73}$$

(4) the first-order autocorrelation of consumption growth is

$$\rho_{\Delta c(1)} = \operatorname{corr}\left[\Delta c_t^*, \Delta c_{t+1}^*\right] = (1-\theta)/\beta, \tag{74}$$

where $j \geq 1$, *and 5) the covariance between consumption growth and lagged equity returns is*

$$\operatorname{covar}\left[\Delta c_{t+1}^*, r_{t+1-j}^e\right] = \theta\alpha^*\left((1-\theta)/\beta\right)^j\omega^2, \tag{75}$$

where $j \geq 1$.

Proof. Using (70), it is straightforward to obtain the above results. ■

Expression (70) clearly shows that aggregate consumption adjusts gradually to the shock to the equity return; RI affects the amplification and propagation mechanism of the model by two channels: (1) reducing the optimal share invested in the equity return α^*, as $\alpha^* = \frac{1-(1-\theta)/\beta}{\theta}\alpha$, and (2) introducing a slow response of consumption to the equity return due to finite channel capacity, $\frac{\theta}{1-((1-\theta)/\beta)\cdot L}$. Expression (70) also implies that to guarantee the existence of solution, $1-(1-\theta)/\beta$ must be greater than 0; note that this condition also guarantees that the optimal share in the risky asset, α^*, should be always greater than 0 as $\frac{1}{\varsigma} = \frac{1-(1-\theta)/\beta}{\theta} > 0$.

Since $\beta, \theta \in (0,1)$, it is clear from (71)-(75) that RI can bring the model and the data closer along the following dimensions: (i) RI reduces the relative volatility of consumption to the equity return, (ii) it reduces the correlation and covariance between consumption and the equity return, (iii) it generates positive serial correlation in consumption growth, and (iv) it generates positive covariances between consumption growth and lagged equity returns.

It is clear that the relative volatility of consumption growth decreases with the degree of attention (θ) given β. Furthermore, it is also clear from (71) that RI reduces the relative volatility by two channels: (i) reducing the optimal share invested in the equity return $\left(\frac{1-(1-\theta)/\beta}{\theta}\right)$ and (ii) the gradual response of consumption to the equity return due to finite channel capacity

$\left(\sqrt{\frac{\theta^2}{1-((1-\theta)/\beta)^2}}\right)$.[34] Using (72), by simple calculation we obtain that $\frac{\partial \operatorname{corr}}{\partial \theta} > 0$, that is, given β, the correlation between aggregate consumption and the equity return (corr) increases with θ.

Related Papers on Portfolio Choice There are two papers that we want to mention briefly that use an alternative information friction, inattentiveness in the sense of Reis (2006), to study portfolio choice: Huang and Liu (2007) and Abel, Eberly, and Panageas (2009). In those cases, agents have to update their information sets but face costs that lead them to do so infrequently. The implications are similar, but the underlying source of the cost of adjustment is unclear – the costs of acquiring information do not seem to be large (for example, it is not particularly costly to read the Wall Street Journal one morning); RI, on the other hand, is explicit about why information flow is limited and thus can, in principle, be discplined externally.[35]

6 The Multivariate Case

We now move to presenting the multivariate RI environment in the LQG setting.

6.1 General Setting

In the multivariate RI problem, it is more difficult to determine the steady state conditional variance-covariance matrix Σ because it cannot be computed analytically. Following Sims (2003), Luo and Young (2013) showed that solving the problem posed in (16) and (77) in Section 4.2 is equivalent to solving the semidefinite programming problem

$$\max_{\Sigma} \{\operatorname{trace}(-Z\Sigma)\} \tag{76}$$

subject to

$$-\log(|\Sigma|) + \log\left(\left|A^T \Sigma A + \Omega\right|\right) \leq 2\kappa, \tag{77}$$

$$A^T \Sigma A + \Omega \succeq \Sigma, \tag{78}$$

[34] It is worth noting that (71) also implies that RI has a potential to explain the equity volatility puzzle because $\lambda < 1$; that is, the same volatility of consumption growth implies higher volatility of the equity in the presence of RI. See Campbell (2003) for a detailed discussion for this puzzle.

[35] As an example, Landauer (1986) used experimental data to estimate the amount of information a human can process; the difficult part is taking that flow and determining how much is being allocated to economic decisisons.

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 of Luo and Young 2013 for the derivation). If the positive-definiteness constraint on $A^T \Sigma A + \Omega - \Sigma$, (78), does not bind, the first-order condition for Σ can be written as:

$$\Sigma^{-1} = \left(G\Sigma G^T + G_0\right)^{-1} - \frac{Z}{\lambda}, \tag{79}$$

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 (79). When applied to a permanent income model in the next subsection, we first solve this equation and then check whether in fact (78) is satisfied by the optimal solution of Σ. If so, the problem is solved. After computing the optimal steady state Σ, we can then use (13) to determine the steady state Λ and $\theta = \Sigma\Lambda^{-1}$ to determine the Kalman gain θ.

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 (76). 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, (13) that is used to determine Σ and θ only depends on the budget constraint.[36]

6.2 The Multivariate Permanent Income Model

In this section we solve for optimal steady state Σ and Λ in a parametric multivariate RI permanent income model. 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

$$w_{t+1} = Rw_t - c_t + y_{t+1}, \tag{80}$$

[36] Since the noise in the traditional SE problem 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.

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 ($\varepsilon_{y,t}$):

$$y_t = \overline{y} + x_t + z_t + \varepsilon_{y,t},$$

$$x_t = 0.99x_{t-1} + \varepsilon_{x,t},$$

$$z_t = 0.95z_{t-1} + \varepsilon_{z,t},$$

with

$$\Omega = \mathrm{var}\begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{x,t} \\ \varepsilon_{z,t} \end{bmatrix} = 10^{-3}\begin{bmatrix} 0.9 & 0 & 0 \\ 0 & 0.009 & 0 \\ 0 & 0 & 0.27 \end{bmatrix}, \tag{81}$$

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 $u(c_t) = -\frac{1}{2}(c_t - \overline{c})^2$ and setting $\beta = 0.95$, we can compute that

$$\Sigma = 10^{-3}\begin{bmatrix} 0.1399 & -0.0737 & -0.0110 \\ -0.0737 & 0.1596 & -0.1820 \\ -0.0110 & -0.1820 & 0.5555 \end{bmatrix}, \tag{82}$$

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 (82) 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 (81). 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.[37] The intuition behind this result is that the optimizing agent devotes much less capacity to monitoring the slow-moving component, which leads to greater effects on the conditional variance term. Figure (3) 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

[37] 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).

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

$$\Sigma = 10^{-3}\begin{bmatrix} 0.0787 & -0.0419 & 0.0153 \\ -0.0419 & 0.1172 & -0.1926 \\ 0.0153 & -0.1926 & 0.5170 \end{bmatrix}. \tag{83}$$

Comparing (82) with (83), 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 the welfare losses due to RI. To see this clearly, the matrix Z is displayed here:

$$Z = 10^{-2}\begin{bmatrix} -0.0204 & -0.6732 & -0.2769 \\ -0.6732 & -22.2156 & -9.1363 \\ -0.2769 & -9.1363 & -3.7573 \end{bmatrix}. \tag{84}$$

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 for this problem (78) is not binding for any variable. It turns out that solving the problem when κ is small enough that the constraint would actually bind is very difficult – even frontier constrained optimization routines have trouble handling the problem. It is feasible in this small model to systematically explore the effects of the constraint binding by testing all combinations, but for larger models this approach would become infeasible. The ideal approach would be to extend the logic of the water-filling problem to the RI problem, but as noted above this extension does not appear to be possible. The problem is that the variables interact in two places – in the objective function and in the information flow constraint. Because the variables will be correlated, one cannot simply construct the bins independently – instead, the variables must be rotated into an orthogonal set using a Spectral Decomposition of Σ. The problem is that this rotation does not decouple the variables in the objective function, where they are combined by the Z matrix, and there does not seem to be a way to method to handle this interdependence in general. However, one special case arises where solutions are straightforward to obtain – when there is no state variable under the control of the decision-maker. We illustrate this environment next using a model of firms making pricing decisions.

6.3 Attention Allocation in a Price-Setting Model

Our last application examines the price-setting decision of a monopolistically-competitive firm that continuously observes noisy signals about two random variables: one aggregate variable and one firm-specific variable. Under the full-information assumption, the profit-maximizing price (in logs) of firm i, p_i, can be written as

$$p_i^f = p + \alpha_x x + \alpha_z z_i, \tag{85}$$

where p is the log of the aggregate price level, x is the log of aggregate output, z_i is an idiosyncratic demand shock, and α_x and α_z are coefficients that depend on structural parameters in the profit function (the superscript f in the price function indicates full information). All the variables on the RHS are assumed to be normally distributed. Since the sum of two normal variables is also normal, we can summarize the aggregate condition as $y = p + \alpha_x x$ such that

$$p_i^f = y + \alpha_z z_i, \tag{86}$$

where y and z are assumed to be Gaussian variables with mean 0 and variances σ_y^2 and σ_z^2, respectively. Under RI, the typical firm cannot observe y and z_i perfectly, so all it can observe are noisy signals y^* on the aggregate state and z^* on the idiosyncratic state:

$$y^* = y + \xi_y \text{ and } z^* = z_i + \xi_z, \tag{87}$$

where the noises due to RI, ξ_y and ξ_z are Gaussian variables with mean 0 and variances ω_y^2 and ω_z^2, respectively. They are assumed to be independent with each other and are also not correlated across firms.[38] Given the observed independent noisy signals, the optimal price can be written as

$$p_i^* = E\left[y|y^*\right] + \alpha_z E\left[z_i|z^*\right]. \tag{88}$$

Since p_i^* is different from the full-information solution p_i^f, the firm suffers profit losses from RI. The profit loss function of the firm can be written as

$$\Delta\pi = \frac{\gamma}{2}\left(p_i^f - p_i^*\right)^2, \tag{89}$$

[38] Maćowiak and Wiederholt (2009) show that independent signals are optimal.

where $\gamma > 0$. Given (88) and (89), the attention allocation problem can be reduced to

$$\min \operatorname{var}\left(p_i^f | y^*, z^*\right) = \sigma^2_{y|y^*} + \alpha_2^2 \sigma^2_{z|z^*}, \tag{90}$$

subject to the information-processing constraint (IPC):

$$\left[\frac{1}{2}\ln\left(2\pi\sigma_y^2\right) - \frac{1}{2}\ln\left(2\pi\sigma^2_{y|y^*}\right)\right] + \left[\frac{1}{2}\ln\left(2\pi\sigma_z^2\right) + \frac{1}{2}\ln\left(2\pi\sigma^2_{z|z^*}\right)\right] \leq \kappa,$$

which can be reduced to $\frac{\sigma^{-2}_{y|y^*}\sigma^{-2}_{z|z^*}}{\sigma_y^{-2}\sigma_z^{-2}} \leq \exp(2\kappa)$, where $\sigma^2_{y|y^*}$ and $\sigma^2_{z|z^*}$ are posterior conditional variances of y and z_i, respectively, and we use the facts that (1) for a quadratic objective function and Gaussian state variables, it is optimal to choose the joint density of the states and noisy signals to be also Gaussian and (2) the aggregate and idiosyncratic variables are uncorrelated. After using the IPC to substitute out $\sigma^{-2}_{z|z^*}$, the optimal solution of $\sigma^2_{y|y^*}$:

$$\frac{\sigma^{-2}_{y|y^*}}{\sigma_y^{-2}} = \exp(\kappa)\left|\frac{1}{\alpha_z}\right|\frac{\sigma_y}{\sigma_z} \in (1, \exp(2\kappa)), \tag{91}$$

which gives the ratio of posterior to prior precision of briefs about the aggregate condition under the optimal attention allocation.[39] (91) provides several important implications for the optimal attention allocation to the aggregate condition. First, greater values of $\sigma^{-2}_{y|y^*}/\sigma_y^{-2}$ mean that the firm pays more attention to the aggregate signal. Second, given α_z, $\sigma^{-2}_{y|y^*}/\sigma_y^{-2}$ is increasing with the relative importance of the prior variances of the aggregate and firm-specific conditions. Since the firm-specific shock is ten times more volatile than the aggregate shock as calibrated in Maćkowiak and Wiederholt (2009), the firm optimally pays much more attention to the firm-specific shock. Similarly, we can obtain the optimal $\frac{\sigma^{-2}_{z|z^*}}{\sigma_z^{-2}}$:

$$\frac{\sigma^{-2}_{z|z^*}}{\sigma_z^{-2}} = \exp(\kappa)\left|\alpha_z\right|\left(\frac{\sigma_y}{\sigma_z}\right)^{-1}. \tag{92}$$

Finally, using (91) and (92), we can easily recover the variances of the noises, ω_y^2 and ω_z^2,

$$\omega_y^{-2} = \sigma^{-2}_{y|y^*} - \sigma_y^{-2}, \tag{93}$$

$$\omega_z^{-2} = \sigma^{-2}_{z|z^*} - \sigma_z^{-2}. \tag{94}$$

[39] Note that this ratio must be greater than 1, as otherwise the posterior variance would be higher than the prior variance, which means that the firms forget. If this ratio is greater than $\exp(2\kappa)$, it violates the information-processing constraint.

Figures (4) and (5) display the dynamics of the price setting model. As can easily be seen, the responses to the idiosyncratic shock under RE and RI are very similar, with only a one-period gap in the initial period; the response to the noise shock lasts only for one period. As the solution we derived above shows, almost all attention is allocated to tracking the idiosyncratic state, since it is both highly volatile and very important for profits. The response to the aggregate shock is sluggish, since the firm devotes little attention to tracking it, and so it takes a long time for the RE and RI impulse responses to converge to each other. Figures (6) and (7) show the aggregate consequences of RI – while the individual price level does track the target price very closely, the aggregate price level is very smooth relative to the RE case. Thus, RI may play an important role in generating the sluggish response of prices discussed in Maćkowiak and Wiederholt (2013), where the authors document that prices respond rapidly to productivity (particularly idiosyncratic productivity) but slowly to monetary policy.[40]

Finally, in the context of this model, Paciello and Wiederholt (2012) show how RI matters for optimal policy. Under full information optimal monetary policy suffers from a standard form of time inconsistency (see Kydland and Prescott 1977 for the definition of time inconsistency and its relation to optimal government policy). Specifically, if some prices do not adjust completely, the central bank will want to announce zero inflation (in order to keep price dispersion to a minimum) but then *ex post* reduce prices using surprise inflation; the problem is that since agents know this incentive exists, the initial optimal policy cannot be implemented. RI adds a second element to this story – the government will also want to exploit the fixed attention decision by shifting variables that agents are not monitoring; again, households will anticipate this deviation and so the initially-optimal plan is not implementable.

7 Conclusions

In this paper we have surveyed the applications of rational inattention in macroeconomics. As happens with most surveys, we have left quite a bit out, including interesting applications in search and matching (Cheremukhin, Restrepo-Echavarria, and Tutino 2013), discrete choice (Matějka and McKay 2012), and planning for rare events (Maćkowiak and Wiederholt 2012). We have also left out alternative formulations of the attention problem, most notably Woodford (2012). More

[40] Other papers have built on this basic model to include capital (Menkulasi 2009), fiscal policy (Dworczak 2011), and multi-product firms (Pasten and Schoenle 2012). The multi-product firm model can actually undo the sluggishness of aggregate states to shocks, depending on the exact nature of the information flow constraint.

research is needed before we can confidently say how attention should be modeled, but we are certain that a lot of attention will be allocated to the problem.

References

[1] Abel, Andrew B., Janice C. Eberly, and Stavros Panageas (2007), "Optimal Inattention to the Stock Market with Information Costs and Transactions Costs," NBER Working Paper 15010.

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

[3] Aguiar, Mark and Gita Gopinath (2007), "Emerging Market Business Cycles: The Cycle Is the Trend," *Journal of Political Economy* 115(1), 69-102.

[4] Batchuluun, Atlantsetseg, Yulei Luo, and Eric R. Young (2009), "Portfolio Allocation with Information-Processing Constraints," manuscript.

[5] Campbell, John Y. (1993), "Intertemporal Asset Pricing without Consumption Data," *American Economic Review* 83(3), 487-512.

[6] Campbell, John Y. and John H. Cochrane (1999), "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy* 107(2), 205-251.

[7] Campbell, John Y. and Luis M. Viceira (2002), *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press.

[8] Cheremukin, Anton, Paulina Restrepo-Echavarria, and Antonella Tutino (2013), "The Assignment of Workers to Jobs with Endogenous Information Selection," manuscript.

[9] Cochrane, J. (2005), *Asset Pricing*, Princeton University Press.

[10] Cooley, Thomas F. and Gary D. Hansen (1997), "Unanticipated Money Growth and the Business Cycle Reconsidered," *Journal of Money, Credit, and Banking* 29(4), 624-648.

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

[12] Dworcezk, Piotr (2011), "Fiscal Policy under Rational Inattention," manuscript.

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

[14] Gabaix, Xavier (2013), "Sparse Dynamic Programming and Aggregate Fluctuations," manuscript.

[15] García-Cicco, Javier, Roberto Pancrazi, and Martín Uribe (2010), “Real Business Cycles in Emerging Countries?” *American Economic Review* 100(5), 2510-2531.

[16] Hall, Robert E. (1978), “Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence,” *Journal of Political Economy* 86(6), 971-987.

[17] Hansen, Lars Peter (1987), “Calculating Asset Prices in Three Example Economies,” in Bewley, Truman F. ed., *Advances in Econometrics, Fifth World Congress*, Cambridge University Press.

[18] Hansen, Lars Peter and Thomas J. Sargent (1995), “Discounted Linear Exponential Quadratic Gaussian Control,” IEEE Transactions on Automatic Control 40, 968-971.

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

[20] Huang, Lixin and Hong Liu (2007), “Rational Inattention and Portfolio Selection,” *Journal of Finance* 62(4), 1999-2040.

[21] Jacobson, David H. (1973), “Optimal Stochastic Linear Systems with Exponential Performance Criteria and Their Relation to Deterministic Differential Games,” IEEE Transactions on Automatic Control 18, 124-131.

[22] Judd, Kenneth L. (1985), "The Law of Large Numbers with a Continuum of IID Random Variables," *Journal of Economic Theory* 35(1), 19-25.

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

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

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

[26] Kydland, Finn E. and Edward C. Prescott (1977), "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy* 85(3), 473-491.

[27] Kydland, Finn E. and Edward C. Prescott (1982), "Time to Build and Aggregate Fluctuations," *Econometrica* 50(6), 1345-1370.

[28] Kydland, Finn E. and Edward C. Prescott (1996), "The Computational Experiment: An Econometric Tool," *Journal of Economic Perspectives* 10(1), 69-85.

[29] Landauer, Thomas K. (1986), "How Much Do People Remember? Some Estimates of the Quantity of Learned Information in Long-Term Memory," *Cognitive Science* 10, 477-493.

[30] Lewis, Kurt (2009), "The Two-Period Rational Inattention Model: Accelerations and Analyses," *Computational Economics* 33(1), 79-97.

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

[32] Luo, Yulei (2010), "Rational Inattention, Long-run Consumption Risk, and Portfolio Choice," *Review of Economic Dynamics* 13(4), 843-860.

[33] Luo, Yulei, Jun Nie, and Eric R. Young (2012), "Robustness, Information-Processing Constraints, and the Current Account in Small Open Economies," *Journal of International Economics* 88(1), 104-120.

[34] Luo, Yulei and Eric R. Young (2009), "Rational Inattention and Aggregate Fluctuations," *The B.E. Journal of Macroeconomics: Vol. 9: Iss. 1 (Contributions)*, Article 14, 1-41.

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

[36] Luo, Yulei and Eric R. Young (2010b), "Asset Pricing under Information-Processing Constraints," *Economics Letters* 107(1), 26-29.

[37] Luo, Yulei and Eric R. Young (2012), "Long-Run Consumption Risk and Asset Allocation under Recursive Utility and Rational Inattention," manuscript.

[38] Luo, Yulei and Eric R. Young (2013), "Signal Extraction and Rational Inattention," forthcoming, *Economic Inquiry*.

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

[40] Maćkowiak, Bartosz and Mirko Wiederholt (2012), "Inattention to Rate Events," manuscript.

[41] Maćkowiak, Bartosz and Mirko Wiederholt (2013), "Business Cycle Dynamics under Rational Inattention," manuscript.

[42] Matějka, Filip, and Alisdair McKay (2012), "Simple Market Equilibria with Rationally Inattentive Consumers," *American Economic Review* 102(3): 24-29.

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

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

[45] Menkulasi, Jeta (2010), "Rational Inattention and Changes in Macroeconomic Volatility," manuscript, University of Maryland.

[46] Mondria, Jordi (2010), "Portfolio Choice, Attention Allocation, and Price Comovement," *Journal of Economic Theory* 145(5), 1837–1864.

[47] Otrok, Christopher (2001), "On Measuring the Welfare Cost of Business Cycles," *Journal of Monetary Economics* 47(1), 61-92.

[48] Paciello, Luigi and Mirko Wiederholt (2012), "Exogenous Information, Endogenous Information, and Optimal Monetary Policy," forthcoming, *Review of Economic Studies*.

[49] Parker, Jonathan (2003), "Consumption Risk and Expected Stock Returns," *American Economic Review* 93(2), 376-382.

[50] Parker, Jonathan and Christian Julliard (2005), "Consumption Risk and the Cross-section of Expected Returns," *Journal of Political Economy* 113(1), 185-222.

[51] Pasten, Ernesto and Raphael Schoenle (2012), "Rational Inattention, Multi-Product Firms, and the Neutrality of Money," manuscript.

[52] Peng, Lin (2004), "Learning with Information Capacity Constraints," *Journal of Financial and Quantitative Analysis* 40, 307-330.

[53] Peng, Lin and Wei Xiong (2006), "Investor Attention, Overconfidence and Category Learning," *Journal of Financial Economics* 80 (3), 563-602.

[54] Reis, Ricardo (2006), "Inattentive Consumers," *Journal of Monetary Economics* 53, 1761-800.

[55] Reis, Ricardo (2011), "When Should Policymakers Make Announcements?" manuscript.

[56] Saint-Paul, Gilles (2011), "A 'Quantized' Approach to Rational Inattention," manuscript.

[57] Sims, Christopher A. (1998), "Stickiness," *Carnegie-Rochester Conference Series on Public Policy* 49, 317-356.

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

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

[60] Sims, Christopher A. (2010), "Rational Inattention and Monetary Economics," *Handbook of Monetary Economics*.

[61] Sun, Yeneng (2006), "The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks," *Journal of Economic Theory* 126(1), 31-69.

[62] Tutino, Antonella (2012), "Rationally Inattentive Consumption Choices," forthcoming, *Review of Economic Dynamics*.

[63] Uhlig, Harald (1996), "A Law of Large Numbers for Large Economies," *Economic Theory* 8(1), 41-50.

[64] Van Nieuwerburgh, Stijn and Laura Veldkamp (2009), "Information Immobility and the Home Bias Puzzle," *Journal of Finance* 64(3), 1187-1215

[65] Van Nieuwerburgh, Stijn and Laura Veldkamp (2010), "Information Acquisition and Under-Diversification," *Review of Economic Studies* 77(2), 779-805.

[66] Veldkamp, Laura (2011), *Information Choice in Macroeconomics and Finance*, Princeton University Press.

[67] Woodford, Michael (2001), "Imperfect Common Knowledge and the Effects of Monetary Policy," in *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, Princeton University Press.

[68] Woodford, Michael (2012), "Inattentive Valuation and Reference-Dependent Choice," manuscript.

[69] Yang, Ming (2013), "Coordination with Flexible Information Acquisition," manuscript, Princeton University Economic Theory Center.

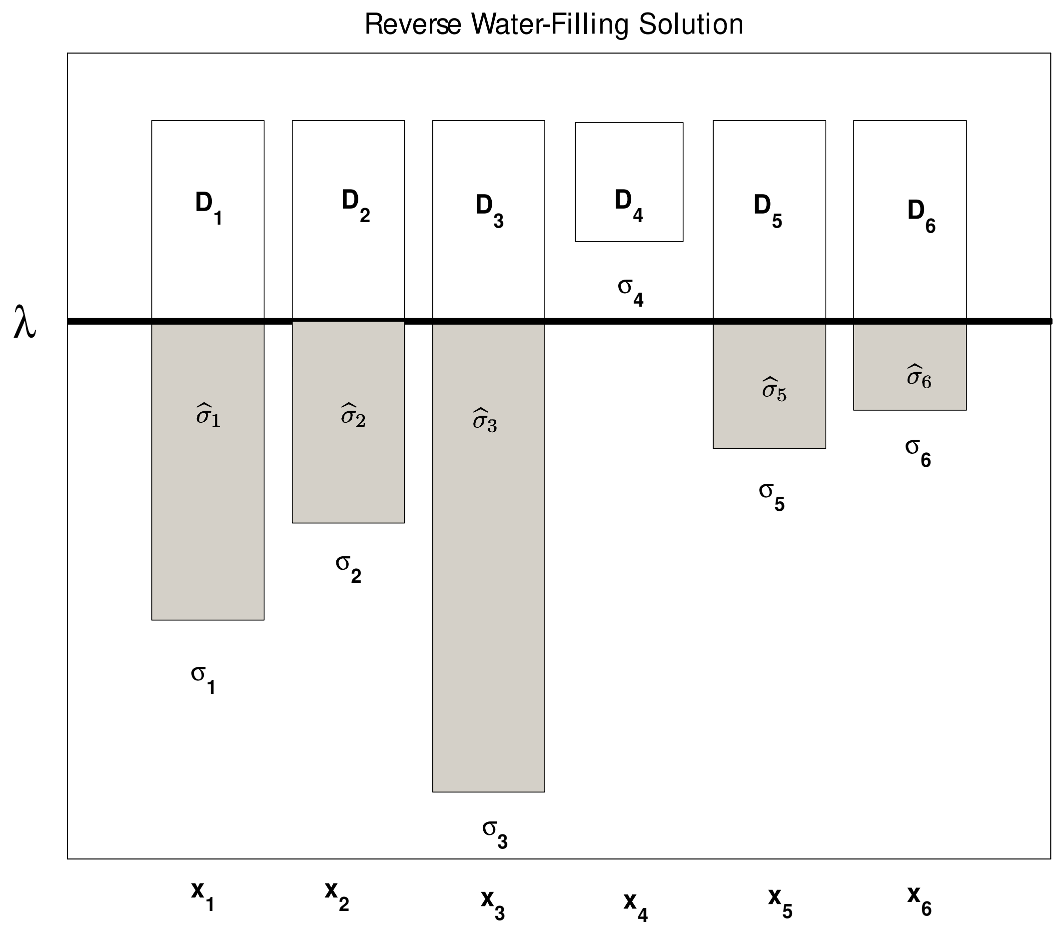

Figure 1: Reverse Water-Filling Solution

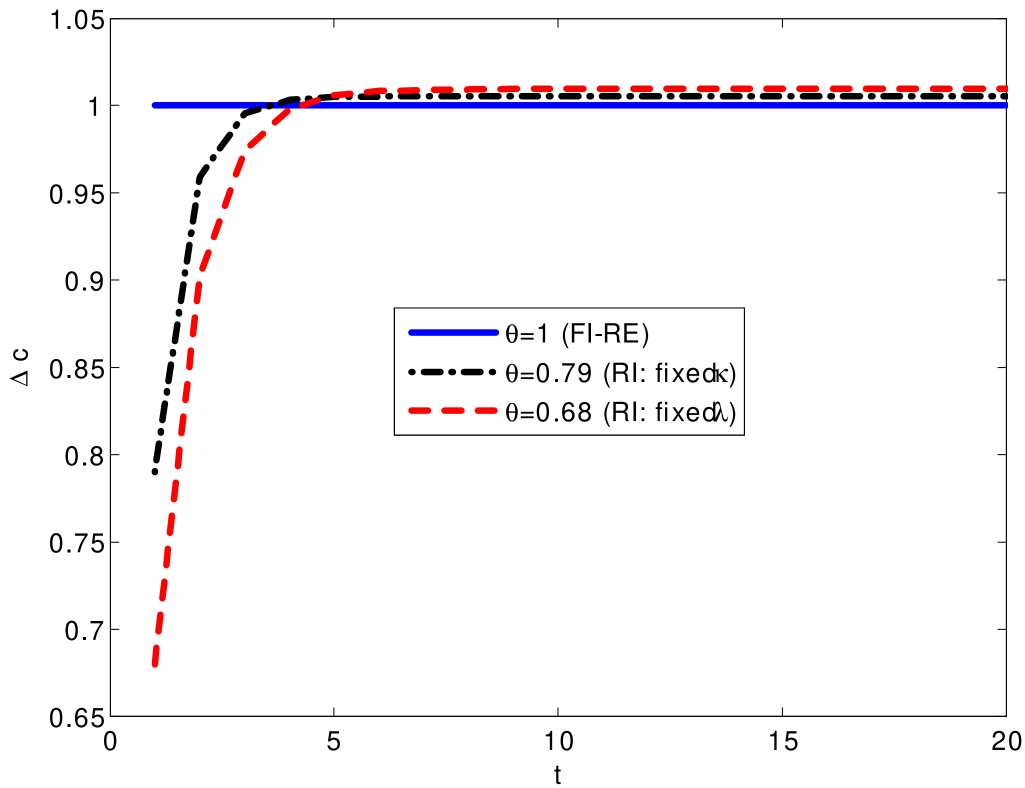

Figure 2: Impulse Response Functions

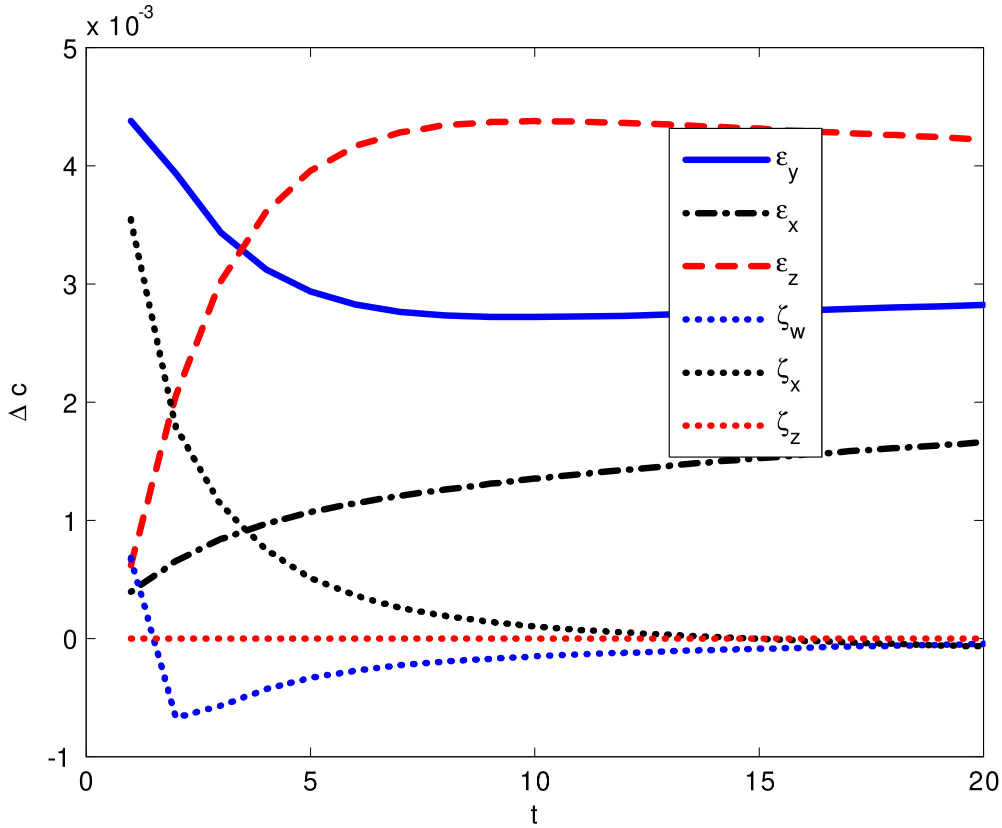

Figure 3: Impulse Response Functions

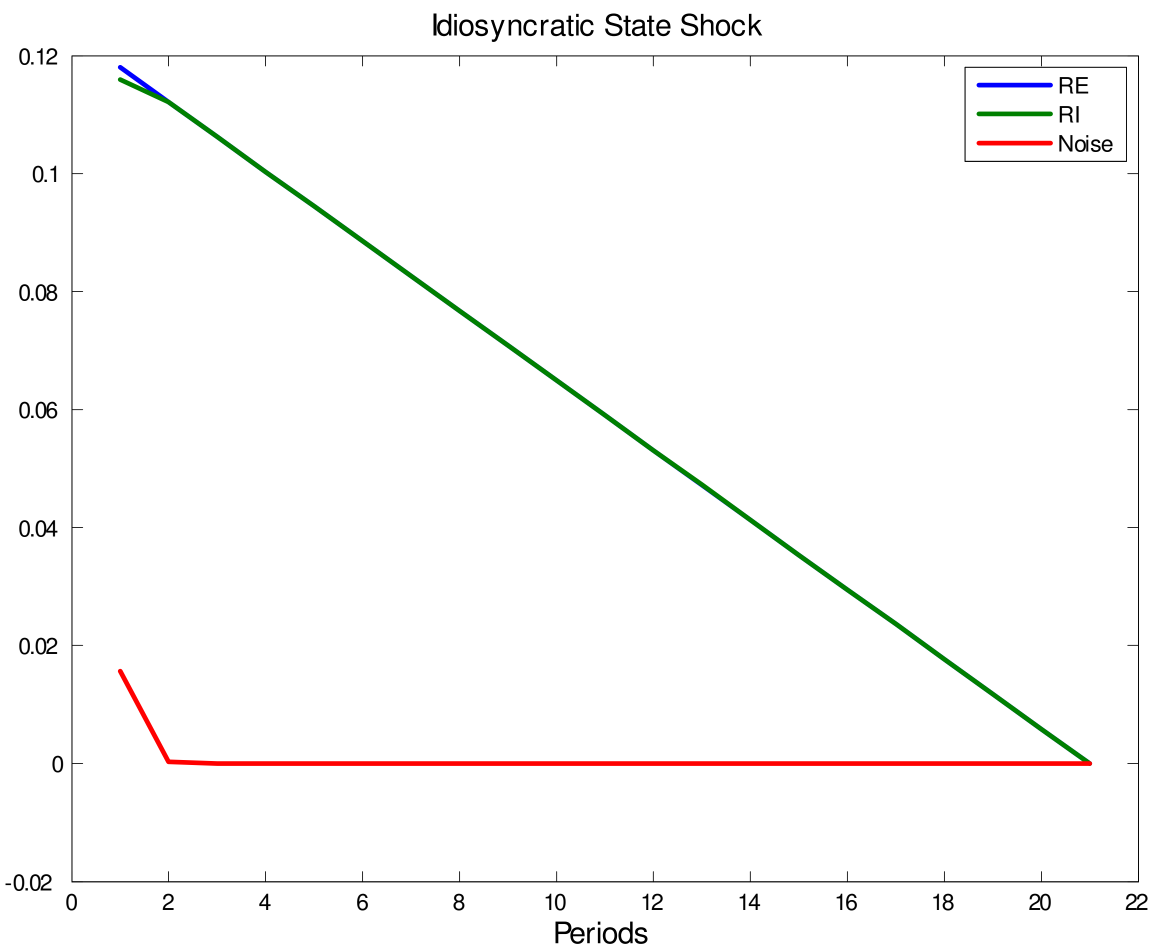

Figure 4: Impulse Response Functions

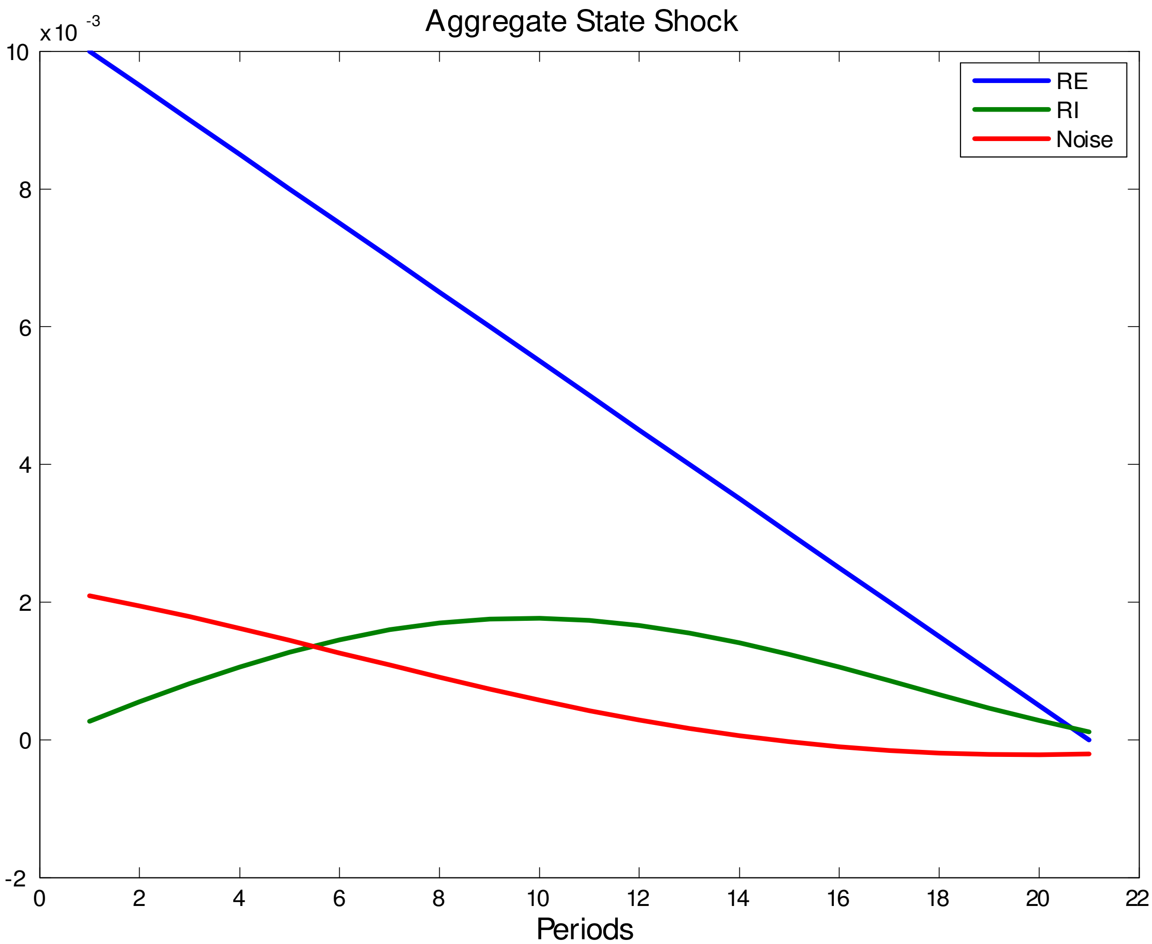

Figure 5: Impulse Response Functions

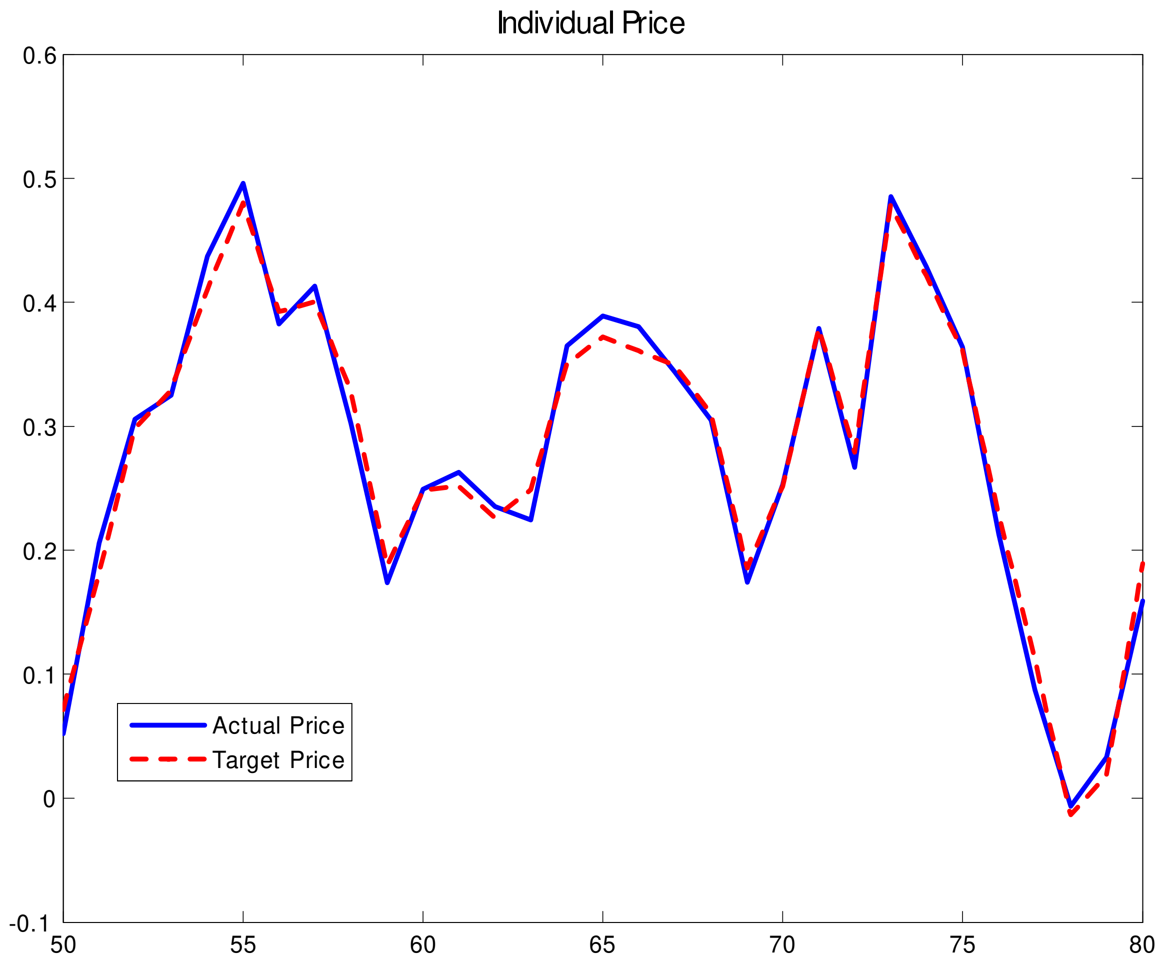

Figure 6: Simulation of Individual Prices

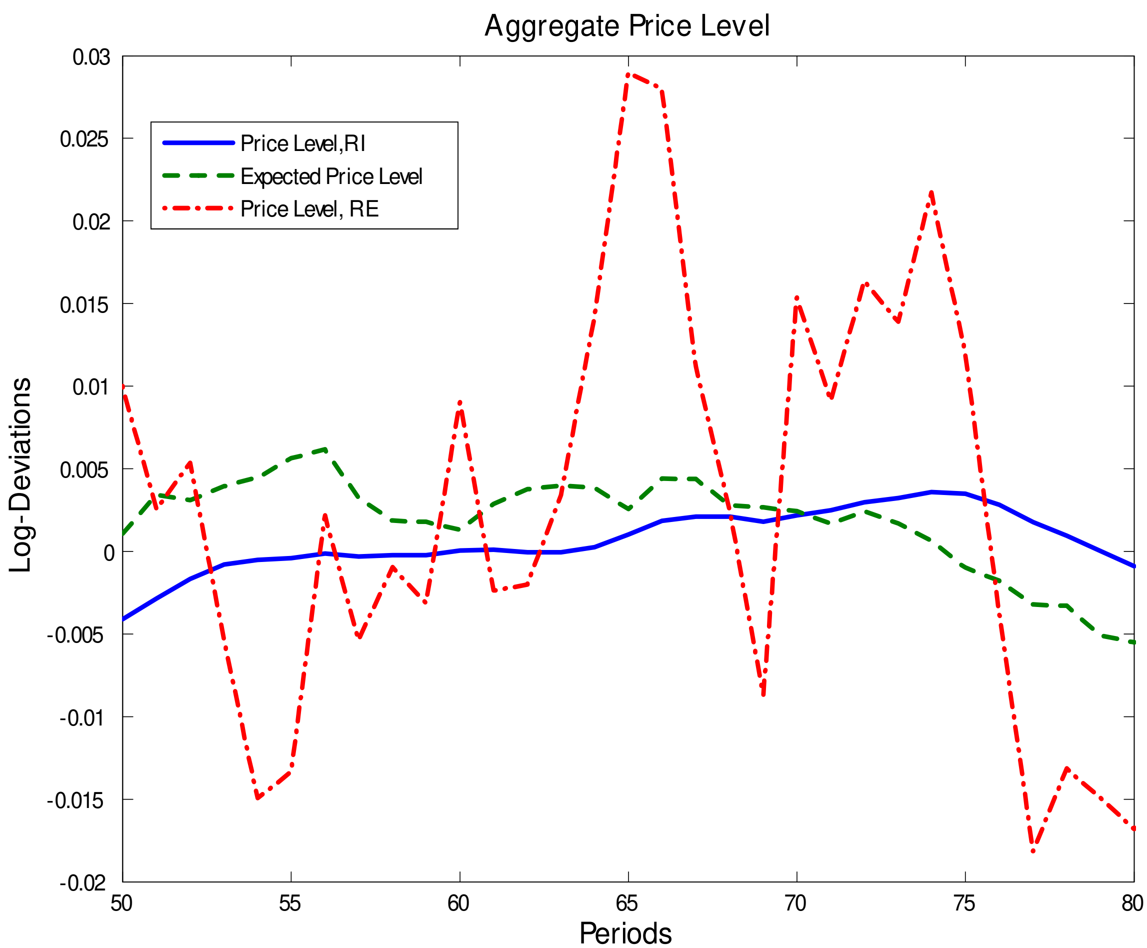

Figure 7: Simulation of Aggregate Price