

Consumption Dynamics under Information Processing

Constraints*

Yulei Luo[†]

The University of Hong Kong

*I am very grateful to Christopher A. Sims for his advice and support. I also thank Narayana Kocherlekota, two anonymous referees, Per Krusell, Jonathan Parker, Ricardo Reis, Charles Whiteman, Wei Xiong, Chi-Wa Yuen and seminar participants at Princeton University, University of Hong Kong, the 2005 Midwest Macro Meeting at Iowa, the 2005 SED at Budapest, and the 2005 ESWC at London for helpful comments.

[†]School of Economics and Finance, The University of Hong Kong, Hong Kong. E-mail: yluo@econ.hku.hk; Tel: (+852) 2859-1042; fax: (852) 2548-1152.

Abstract

This paper studies how “rational inattention (RI)” – a type of information processing constraint proposed by Sims (2003) – affects the joint dynamics of consumption and income in a permanent income model with general income processes. Specifically, I propose an analytical approach to solve the multivariate permanent income model with RI and examine its implications for optimal consumption, saving, and welfare. It is shown that RI can affect the relative volatility of consumption and provide an endogenous propagation mechanism that disentangles the short-run and long-run responses of consumption to exogenous income shocks. I also explore how aggregation reduces the impact of the RI-induced endogenous noise on consumption and thus increases the smoothness of aggregate consumption. Finally, I compare RI with four alternative hypotheses (habit formation, signal extraction, robustness, and inattentiveness) by examining their implications for the joint behavior of consumption and income.

JEL Classification Numbers: C61, D81, E21.

Keywords: Rational Inattention, Consumption and Savings Behavior, Short-run and long-run Impacts, Aggregation.

1. Introduction

Canonical macroeconomic theory assumes that agents have the ability to process as much information as they need instantaneously and completely, and thus have perfect information about the state(s). However, in reality ordinary people only have finite information-processing capacity; consequently, they cannot observe and process all available information about the state(s) perfectly when making economic decisions.¹ Sims (2003) proposed a new theory, rational inattention (RI),² to capture this fact by assuming that information about the state flows through a finite channel before it reaches the agents.³ In his analysis, Sims explicitly solved an RI version of the permanent income hypothesis (PIH) model by restricting attention to the unrealistic case of i.i.d. income shocks. In this paper, however, I consider a much larger class of income processes. First, I show that within the linear quadratic Gaussian (LQG) framework, the multivariate permanent income model with general income processes can be reduced to a univariate model with i.i.d. innovations to permanent income that can be solved in closed-form solution. I then derive an analytical representation for optimal consumption when income is the sum of a unit root process and an MA (k) process. In addition, I also work through the implications of this analytical representation for some specific income processes, and show in some numerical examples that the welfare losses for even relatively tight informational constraints are trivial.

I go on to discuss the aggregate implications of RI for individual income processes of this kind. I show that, relative to the standard rational expectation (RE) PIH model, aggregate consumption growth under RI could exhibit both “excessively smooth” to *unanticipated* income growth and “excessively sensitive” to *anticipated* changes in income. These characterizations are qualitatively consistent with the observed excess smoothness and excess sensitivity of aggregate consumption in the U.S. data. Hence, the RI hypothesis could be a potential explanation for the two empirical puzzles that cannot be explained by the standard PIH model: the *excess smoothness* puzzle and the *excess sensitivity* puzzle.⁴ In a recent paper,

¹As shown in Shannon (1948), measuring a real-valued stochastic process without error implies an infinite amount of information processing capacity.

²In this paper, I use “information processing constraints” and “Rational Inattention” (henceforth “RI”) interchangeably.

³It may be fraught with some endogenous errors which affect optimal decisions; consequently, agents cannot observe the state perfectly and cannot adjust their optimal plans instantly and completely with respect to exogenous shocks.

⁴Numerous economic arguments have been advanced to explain these two puzzles. A partial listing of these explanations includes: time-varying interest rate (Hall 1988), the buffer stock model (Ludvigson and Michaelides 2001), the rule-of-thumb consumers (Campbell and Mankiw 1989), habit formation (Deaton 1992), and incomplete aggregate information (Pischke

Reis (2003) analyzed optimal consumption decisions of agents facing costs of planning in a continuous-time model with the exponential utility and showed that inattentiveness due to costly planning could be an alternative explanation for the two puzzles. The main difference between this paper and Reis' paper is that this paper follows the line of Sims (2003) and uses the concept of Shannon channel to characterize inattention, whereas Reis modelled 'inattentiveness' by assuming the existence of decision costs that induces agents to only infrequently update their decisions. The two models have distinct mechanisms of individuals' slow adjustment in consumption but could result in similar aggregate consumption dynamics under some restrictions.⁵

Finally, I also compare RI with three alternative models (habit formation, signal extraction, and robustness) within the same PIH framework. Specifically, I explicitly derive their implications for the mapping from exogenous income processes into a joint process for consumption and income, and then discuss how these models differ in the empirical content summarized by the mapping at both individual and aggregate levels.

The main contribution of this paper is that I propose a tractable analytical approach to solve an RI version of the PIH model with general income processes. The explicit solution has the following advantages: First, it gives us clear economic insight about the nature of RI and it also greatly facilitates comparative statics analysis. Second, the welfare loss due to RI can be easily calculated. Finally, the solution to a single-agent case is in exact aggregation form and thus can be aggregated to obtain the joint dynamics of aggregate consumption and income. After solving the RI model with some specific income processes, I show that RI provides an endogenous propagation mechanism to disentangle the short-run responses from the long-run responses of consumption to various income shocks. The main mechanism behind these results is that with finite Shannon capacity, consumers cannot respond instantly and without errors to the changes in wealth; consequently, consumption reacts to the wealth shocks gradually, which means the responses to the shocks can be propagated via the slow adjustment in consumption. I also explicitly derive an expression for the relative volatility of individual consumption to income and show

1995).

⁵See Section 4 for a detailed comparison.

that the volatility of individual consumption increases with the degree of RI.

Recently, some papers have incorporated information frictions into a variety of theoretical models and explored how information imperfection affects optimal decisions of consumers, firms, and investors, as well as their implications for equilibrium and monetary policy. For example, Woodford (2001), Ball et. al. (2003), and Adam (2005) analyzed the effects of imperfect common knowledge on monetary policy and inflation dynamics; Moscarini (2004) derived optimal time-dependent adjustment rules from information constraints; Luo and Young (2006) examined the effects of RI on the amplification and propagation mechanisms of aggregate shocks; Maćkowiak and Wiedeholt (2005) investigated how the optimal allocation of attention affects the dynamics of prices.

The remainder of the paper is organized as follows. Section 2 presents an RI version of the PIH model with general income processes and discusses some implications of RI for individuals' optimal consumption decisions. Section 3 discusses the implications of RI for aggregate consumption. Section 4 compares the RI model with four alternative models. Section 5 concludes.

2. The PIH Model with Rational Inattention

2.1. The Model Set-up

Consider a simple permanent income model with a continuum of consumers facing information-processing constraints. Specifically, in this section, I focus on optimizing behavior of the typical consumer with information-processing constraints. In the next section, I will examine the implications of RI for aggregate consumption after aggregating across all consumers in the economy. Before setting up the PIH model with RI, I first present a standard rational expectation (RE) version of the PIH model with general income processes and then show that it can be reduced to a univariate PIH model with i.i.d. innovations to permanent income into which RI can be easily incorporated. This approach to reduce the state space simplifies the model greatly because multivariate versions of the RI model require a separate “no subsidization” constraint that is highly nonlinear, which makes the RI model very difficult to solve, even

within the LQG framework.⁶

A standard RE version of the PIH model can be formulated as follows

$$\max_{\{c_t\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t) \right] \quad (2.1)$$

subject to

$$w_{t+1} = R(w_t - c_t) + y_{t+1}, \quad (2.2)$$

where $u(c_t) = -\frac{1}{2}(c_t - \bar{c})^2$ is the utility function, \bar{c} is the bliss point, Equation (2.2) is the flow budget constraint, c_t is consumption, w_t is wealth, y_t is a general income process with Gaussian white innovations, β is the discount factor, R is the constant gross interest rate at which the consumer can borrow and lend freely, and $\beta R = 1$ (an assumption typically imposed in the literature),⁷ This specification follows that in Hall (1978) and Flavin (1982), and implies that optimal consumption is determined by permanent income, that is,

$$c_t = \frac{R-1}{R} s_t, \quad (2.3)$$

where

$$s_t = w_t + \sum_{j=1}^{\infty} R^{-j} E_t [y_{t+j}] \quad (2.4)$$

is the expected present value of lifetime resources, which consists of financial wealth plus human wealth.

If s_t is defined as a new state variable, we can reformulate the above PIH model as follows⁸

$$v(s_0) = \max_{\{c_t\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t) \right] \quad (2.5)$$

subject to

$$s_{t+1} = R(s_t - c_t) + \zeta_{t+1}, \quad (2.6)$$

⁶This constraint implies that the uncertainty regarding one state variable cannot be increased in order to reduce the uncertainty regarding another. In other words, information flow cannot be kept low by “forgetting” some existing information, trading it off for more information about other state variables. See Sims (2003) for a detail discussion.

⁷In Section (2.1), I also briefly discuss the effects of RI for consumption when $\beta R \neq 1$.

⁸See Appendix 5.1 for the derivation.

where

$$\zeta_{t+1} = \sum_{j=t+1}^{\infty} R^{-j+(t+1)} (E_{t+1} - E_t) [y_j] \quad (2.7)$$

is the $(t + 1)$ -innovation to permanent income and $v(s_0)$ is the consumer's value function under RE. In other words, by defining a new state variable s_t , the original multivariate optimization problem can be reduced to a univariate problem in which RI can be easily incorporated. This univariate RE PIH model leads to the well-known random walk result of Hall (1978):

$$\Delta c_{t+1} = \frac{R-1}{R} \zeta_{t+1}, \quad (2.8)$$

which means that the change in consumption depends neither on the past history of labor income nor on anticipated changes in labor income.

Following Sims (2003), RI can be incorporated into the above univariate PIH framework as follows. Suppose that the typical consumer maximizing his lifetime utility function is subject to both usual budget constraints and information processing constraints that will be specified later in this paper. The typical consumer's optimizing problem under RI can be written as

$$\hat{v}(\hat{s}_0) = \max_{\{C_t, \mathcal{D}_t\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t) \right] \quad (2.9)$$

subject to

$$\begin{aligned} s_{t+1} &= R(s_t - c_t) + \zeta_{t+1}, \\ s_{t+1} | \mathcal{I}_{t+1} &\sim \mathcal{D}_{t+1}, \quad s_t | \mathcal{I}_t \sim \mathcal{D}_t, \end{aligned}$$

$s_0 | \mathcal{I}_0 \sim N(\hat{s}_0, \sigma_0^2)$ is given, and the requirement that the rate of information flow at $t + 1$ implicit in the specification of the distributions, \mathcal{D}_t and \mathcal{D}_{t+1} , be less than channel capacity.⁹ s_t is the actual state variable, $\hat{s}_t = E[s_t | \mathcal{I}_t]$ is the *perceived* state variable, \mathcal{I}_t is the information available at time t , and $\hat{v}(\hat{s}_0)$

⁹They will be precisely specified subsequently.

is the consumer's value function under RI. The expectation is formed under the assumption that $\{c_t\}_{t=0}^{\infty}$ are chosen under the information processing constraints.

2.2. Information Processing Constraints

The consumer faces information processing constraints in the sense that he only uses finite channel capacity to observe the relevant state. As in Sims (2003), I also use the concept of *entropy* from information theory to measure the uncertainty about a random variable and then use the reduction in entropy as a measure for information flow.¹⁰ With finite capacity, the consumer will choose a signal that reduces the uncertainty of the state variable. Formally, this idea can be described by the following information processing constraint:

$$\mathcal{H}(s_{t+1}|\mathcal{I}_t) - \mathcal{H}(s_{t+1}|\mathcal{I}_{t+1}) \leq \kappa, \quad (2.10)$$

where κ is the consumer's channel capacity that imposes an upper bound on the amount of information that can be transmitted via the channel,¹¹ $\mathcal{H}(s_{t+1}|\mathcal{I}_t)$ denotes the entropy of the state prior to observing the new signal at $t + 1$, and $\mathcal{H}(s_{t+1}|\mathcal{I}_{t+1})$ the entropy after observing the new signal.

Given the LQG specification, Sims (2003, 2005) showed that \mathcal{D}_t is a normal distribution with $N(\widehat{s}_t, \sigma_t^2)$, where $\widehat{s}_t = E_t[s_t]$ and $\sigma_t^2 = \text{var}_t[s_t]$ are the conditional mean and variance of the state variable, s_t , respectively. Note that in models with imperfect observations, the optimal decisions are determined by the *perceived* state, rather than the *actual* state. A difficulty arises because the perceived state is defined as the distribution of the state variable conditional on the information set available at time t , \mathcal{I}_t , which leads to a substantial increase in dimensionality when we use dynamic programming to characterize the optimization problem. Fortunately, given the LQG specification and the Gaussian distribution of s_t , the first two moments, \widehat{s}_t and σ_t^2 , are enough to characterize the perceived state. In addition, as I will show, the problem would be even simpler because σ_t^2 converges to a constant in the steady state.

¹⁰See Shannon (1948) and Cover and Thomas (1991) for details.

¹¹It may be regarded as a simple technology parameter. If the base for logarithms is 2, the unit used to measure information flow is called 'bits', and if we use the natural logarithm e , the unit is called 'nats'. Hence, 1 nat is equal to $\log_2 e = 1.433$ bits.

The constraint, (2.10), can thus be rewritten as

$$\log |\psi_t^2| - \log |\sigma_{t+1}^2| \leq 2\kappa, \quad (2.11)$$

where $\sigma_{t+1}^2 = \text{var}_{t+1} [s_{t+1}]$ and $\psi_t^2 = \text{var}_t [s_{t+1}] = R^2\sigma_t^2 + \text{var}_t [\zeta_{t+1}]$ are the posterior and prior variance of the state variable, s_{t+1} , respectively.¹² As shown in Sims (2003), in any univariate case, this information constraint completes the characterization of the optimization problem with RI and the model can be solved explicitly. Furthermore, with a finite capacity κ the optimizing consumer will choose a signal that reduces the conditional variance of s_{t+1} by a limited amount. Hence, (2.11) must be binding for the optimizing consumer:

$$\log |\psi_t^2| - \log |\sigma_{t+1}^2| = 2\kappa. \quad (2.12)$$

It is straightforward to show that in the univariate case Equation (2.12) has a steady state $\bar{\sigma}^2$. In the steady state, $\sigma_t^2 = \bar{\sigma}^2 = \frac{(\frac{R}{R-1})^2 \omega_\zeta^2}{\exp(2\kappa) - R^2}$ where $\omega_\zeta^2 = \text{var}_t [\zeta_{t+1}]$, and the consumer behaves as if he is observing a noisy signal, $s_{t+1}^* = s_{t+1} + \xi_{t+1}$, where ξ_{t+1} is the i.i.d. endogenous noise and its variance, $\lambda_t^2 = \text{var}_t [\xi_{t+1}]$, is determined by the usual updating formula of the variance of a Gaussian distribution:

$$\sigma_{t+1}^2 = \psi_t^2 - \psi_t^2(\psi_t^2 + \lambda_t^2)^{-1}\psi_t^2, \quad (2.13)$$

which means that in the steady state, $\lambda_t^2 = \bar{\lambda}^2 = \text{var}_t [\xi_{t+1}] = \frac{[(\frac{R}{R-1})^2 \omega_\zeta^2 + R^2 \bar{\sigma}^2] \bar{\sigma}^2}{(\frac{R}{R-1})^2 \omega_\zeta^2 + (R^2 - 1) \bar{\sigma}^2}$. For simplicity, I assume that initially the model economy has been at the steady state, that is, $s_0 | \mathcal{I}_0 \sim N(\hat{s}_0, \bar{\sigma}^2)$.

2.3. A Closed-form Solution of Consumption for General Income Processes

In this subsection, I close the model economy by specifying the income processes. Given the specified income processes, I will derive explicit representations for both optimal consumption and consumption growth in terms of income innovations and endogenous noises. I then explore some implications of RI for the joint dynamics of consumption and income and the smoothness of consumption.

¹²Note that here I use the fact that the entropy of a Gaussian random variable is equal to half of its logarithm variance plus some constant term.

Following Hall and Mashkin (1982), Deaton (1992), Ludvigson and Michaelides (2001), and others, I suppose that income y_t can be expressed as the sum of permanent and transitory components:

$$y_{t+1} = y_{t+1}^p + y_{t+1}^t, \quad (2.14)$$

where the superscripts p and t denote permanent and transitory, respectively. Each of these components follows its own stochastic process; y_t^p follows a random walk:

$$y_{t+1}^p = y_t^p + \varepsilon_{t+1}, \quad (2.15)$$

and the transitory component follows a stationary MA process:

$$y_{t+1}^t = \bar{y} + \sum_{j=0}^M \phi_j \eta_{t+1-j}, \quad (2.16)$$

with $\phi_0 = 1$. Note that when income follows a general stationary ARMA process, it can be converted to its MA representation. Hence, when $M = \infty$, the specification (2.16) can be used to characterize arbitrary stationary ARMA processes. In addition, it is assumed that the consumer can separately identify the two components so that consumption can respond to each of them. All innovations are assumed to have zero mean, be uncorrelated over time and with each other, and the variance of ε_t and η_t are ω_ε^2 and ω_η^2 , respectively.

Substituting the specified income processes, (2.14), (2.15), and (2.16), into (2.4) and (2.7) yields

$$s_t = w_t + \frac{1}{R-1} y_t^p + \sum_{i=0}^{M-1} \left(\sum_{j=i+1}^M R^{-(j-i)} \phi_j \right) \eta_{t-i} + \frac{1}{R-1} \bar{y}, \quad (2.17)$$

$$\zeta_t = \frac{R}{R-1} \varepsilon_t + \sum_{j=0}^M (R^{-j} \phi_j) \eta_t, \quad (2.18)$$

where $\omega_\zeta^2 = \left(\frac{R}{R-1} \right)^2 \omega_\varepsilon^2 + \left(\sum_{j=0}^M R^{-2j} \phi_j^2 \right) \omega_\eta^2$.

Formally, the consumption function and the value function under RI can be obtained by solving the

following stochastic Bellman equation

$$\widehat{v}(\widehat{s}_t) = \max_{c_t} \left\{ -\frac{1}{2} (c_t - \bar{c})^2 + \beta E_t [\widehat{v}(\widehat{s}_{t+1})] \right\} \quad (2.19)$$

subject to the usual budget constraint, (2.6), and the Kalman filter equation,

$$\widehat{s}_{t+1} = (1 - \theta) R (\widehat{s}_t - c_t) + \theta (s_{t+1} + \xi_{t+1}), \quad (2.20)$$

where $s_0 | \mathcal{I}_0 \sim N(\widehat{s}_0, \bar{\sigma}^2)$, $\widehat{s}_t = E_t[s_t]$ is the perceived state variable, θ is the optimal weight on the observation,¹³ and ξ_{t+1} is the endogenous noise. The following proposition summarizes the main results of the above PIH model with RI.

Proposition 1. *Given finite channel capacity κ , the value function is*

$$\widehat{v}(\widehat{s}_t) = -\frac{R-1}{2R} \widehat{s}_t^2 + \bar{c} \widehat{s}_t - \frac{R}{2(R-1)} \bar{c}^2 - \frac{\text{var}_t[\varpi_{t+1}]}{2R}, \quad (2.21)$$

where $\varpi_{t+1} = \theta \left[\left(\frac{\zeta_{t+1}}{1 - (1-\theta)R \cdot L} \right) + \left(\xi_{t+1} - \frac{\theta R \xi_t}{1 - (1-\theta)R \cdot L} \right) \right]$, the consumption function is

$$c_t^* = \frac{R-1}{R} \widehat{s}_t, \quad (2.22)$$

and the evolution of \widehat{s}_t is characterized by the following Kalman filter equation:

$$\widehat{s}_{t+1} = (1 - \theta) \widehat{s}_t + \theta (s_{t+1} + \xi_{t+1}), \quad (2.23)$$

where $\theta = 1 - 1/\exp(2\kappa)$ and the variance of ξ_{t+1} , $\text{var}[\xi_{t+1}]$, is $\frac{1-\theta}{\theta[1-R^2(1-\theta)]} \omega_\zeta^2$.

Furthermore, the change in consumption can be written as:

$$\Delta c_t^* = \theta \frac{R-1}{R} \left[\left(\frac{\zeta_t}{1 - (1-\theta)R \cdot L} \right) + \left(\xi_t - \frac{\theta R \xi_{t-1}}{1 - (1-\theta)R \cdot L} \right) \right], \quad (2.24)$$

¹³Hence, θ measures how much uncertainty can be removed upon the receipt of a new signal, and is independent of the patience of the consumer and the variance of the income shocks.

where L is the lag operator.

Proof. (See Appendix 5.2 for the proof.) ■

Note that the weight on the observation, θ , is an increasing function of channel capacity, κ , and can be used to measure how much uncertainty can be removed after observing the new signal. It increases to 100% as κ gets infinitely-large, so the standard RE PIH model is a special case of the RI PIH model in which $\kappa = +\infty$. Equations (2.6), (2.14), (2.22), and (2.23) constitute a dynamic system that can characterize the RI economy completely. In the next subsection, I will use several particular income processes to examine how the interaction of RI and the stochastic properties of income processes affects the responses of consumption with respect to the disturbances.

Equation (2.24) shows that the growth of consumption includes an MA(∞) process in terms of income shocks with decreasing coefficients, which means that consumption adjusts slowly and gradually to income shocks, with reactions that build up over time. Furthermore, consumption also reacts to its own shocks, that is, the endogenous noises, instantaneously. It thus implies that consumption should be sensitive to both current and past income shocks, as well as endogenous noises. Note that when $\theta = 100\%$, i.e., $\kappa = \infty$, the above expression reduces to the standard RE solution, $\Delta c_t = \frac{R-1}{R}\zeta_t$, which means that the change in consumption is not sensitive to past income shocks and jumps immediately in response to income shocks.

Equation (2.24) also makes it clear that the short-run (SR) and long-run (LR) impacts of the income disturbances on consumption can be disentangled by RI. Formally, the short-run and long-run responses of consumption to the permanent and transitory shocks are:

$$\begin{aligned} \text{SR}(\varepsilon_t) &= \frac{\partial(\Delta c_t^*)}{\partial \varepsilon_t} = \theta \text{ and } \text{LR}(\varepsilon_t) = \frac{\partial(\sum_{s=0}^{\infty} \Delta c_{t+s}^*)}{\partial \varepsilon_t} = \frac{\theta}{1 - (1 - \theta)R}; \\ \text{SR}(\eta_t) &= \frac{\partial(\Delta c_t^*)}{\partial \eta_t} = \theta \sum_{j=0}^M R^{-j} \phi_j \text{ and } \text{LR}(\eta_t) = \frac{\partial(\sum_{s=0}^{\infty} \Delta c_{t+s}^*)}{\partial \eta_t} = \frac{\theta}{1 - (1 - \theta)R} \sum_{j=0}^M R^{-j} \phi_j, \end{aligned}$$

respectively, which means that the degree of inattention (low κ or low θ) reduces the short-run effects of consumption to the permanent and transitory income shocks, while it increases the long-run effects of the same shocks. The interest rate R also increases the long-run effect of income shocks; the higher the

interest rate, the larger the long-run effect.¹⁴ Note that when $\kappa = \infty$ (i.e., $\theta = 100\%$), the two effects are the same. RI thus provides an endogenous mechanism to propagate the exogenous shocks.

If we regard this typical consumer as the representative agent, we can examine the implications of RI on aggregate consumption directly from Equation (2.24). The standard RE PIH model predicts that aggregate consumption is not sensitive to lagged income information and is more volatile than aggregate income if income is a random walk or a difference-stationary process, which does not fit the US aggregate data well.¹⁵ Given Equation (2.24), it is straightforward to calculate the smoothness ratio (i.e., the relative volatility of consumption to income) under RI:¹⁶

$$\mu = \frac{\text{sd}[\Delta c_t^*]}{\text{sd}[\varepsilon_t]} = \sqrt{\frac{\theta}{1 - (1 - \theta)R^2}} \geq 1, \quad (2.25)$$

where “sd” denotes standard deviation and θ is assumed to be greater than $1 - \frac{1}{R^2}$ to guarantee feasible solutions. Because $\frac{\partial \mu}{\partial \theta}$ is less than 0, RI increases the relative volatility of consumption in this single-agent case. The intuition is that imperfectly observing the state variable leads to a reduced ability to smooth consumption, similar to the results in Luo and Young (2006). In Section 3, I will discuss the aggregation issue and its implications for the smoothness of aggregate consumption.

As in Campbell (1987), we may also easily examine the predictions of the PIH model with RI for the behavior of saving. Saving d_t under RI is defined as $(\frac{R-1}{R}w_t + \frac{1}{R}y_t) - c_t^*$ and can be rewritten as¹⁷

$$d_t = -\sum_{j=1}^{\infty} R^{-j} E_t[\Delta y_{t+j}] + \frac{R-1}{R} \left[\frac{(1-\theta)\zeta_t}{1 - (1-\theta)R \cdot L} - \frac{\theta\xi_t}{1 - (1-\theta)R \cdot L} \right], \quad (2.26)$$

in which the first part of Equation (2.26) is the amount of saving due to expected income falls, as predicted by the standard RE PIH model, whereas the second part is the amount of saving caused by RI. Specifically, introducing RI generates a stochastic saving premium that consists of two MA(∞) processes. For example, when positive income shocks arrive at the beginning of period t , the consumer cannot adjust

¹⁴Note that since $\beta R = 1$, an increase in R also implies a corresponding reduction in β .

¹⁵They are called the excess sensitivity puzzle and the excess smoothness puzzle, respectively. See Deaton (1992) for a recent review on these issues.

¹⁶See Appendix 5.3 for the derivation. For simplicity, here I ignore the transitory component y^t .

¹⁷We can easily derive this expression by combining (2.3), (2.22), and the fact that $c_t^* = c_t - (c_t - c_t^*)$.

his optimal consumption plan instantly and completely due to the information constraints; consequently, the share of undetected income shocks becomes saving in that period. Hence, the RI saving premium is larger for a higher degree of inattention (lower θ), *ceteris paribus*. It is also obvious that a positive increase in the error tends to reduce the RI saving premium, and both lagged income shocks and noises still have effects on the current savings.

It is worth noting that when $\beta R > 1$ (that is, the net interest rate, $r = R - 1$, is greater than the discount rate, $\delta = (1 - \beta)/\beta$), RI has larger impacts on consumption. Specifically, when $\beta R > 1$, consumption growth under RI can be written as:¹⁸

$$\Delta c_t^* = \theta \left(1 - \frac{1}{\beta R^2}\right) \left\{ \begin{array}{l} \sum_{j=0}^{\infty} \left[\sum_{k=0, k \leq j} \left(\rho_1^{j-k} \rho_2^k \right) - \sum_{k=0, k \leq j-1} \left(\rho_1^{j-1-k} \rho_2^k \right) \right] \zeta_{t-j} \\ + \sum_{j=0}^{\infty} \left[\sum_{k=0, k \leq j} \left(\rho_1^{j-k} \rho_2^k \right) - \sum_{k=0, k \leq j-1} \left(\rho_1^{j-1-k} \rho_2^k \right) \right] \xi_{t-j} \\ - R \sum_{j=0}^{\infty} \left[\sum_{k=0, k \leq j} \left(\rho_1^{j-k} \rho_2^k \right) - \sum_{k=0, k \leq j-1} \left(\rho_1^{j-1-k} \rho_2^k \right) \right] \xi_{t-1-j} \end{array} \right\}, \quad (2.27)$$

where $\rho_1 = \frac{1}{\beta R} \in (0, 1)$ and $\rho_2 = (1 - \theta)R \in (0, 1)$. Figure 1 compares the responses of consumption to the permanent income shock (ε_t) in the two cases. It clearly shows that qualitatively, RI has similar impacts on consumption in the two cases;¹⁹ but quantitatively, RI has larger impacts on consumption when $\beta R > 1$, because the deviation of the asymptote from that for the standard RE case is larger in this case. The reason is that with higher interest rates relative to impatience (for example, increasing R and holding β constant), an income shock that initially undetected would accumulate more interest before consumption responds fully.

2.4. Applications to Some Particular Income Processes

In the above subsection, I solved an RI version of the PIH model with general income processes. In this subsection, I will work through some particular income processes to examine how RI affects optimal consumption decisions and the impulse responses of consumption with respect to the innovations to income.

¹⁸The derivation is available from the author.

¹⁹That is, consumption reacts gradually to income shocks, with steady adjustments to the corresponding asymptote.

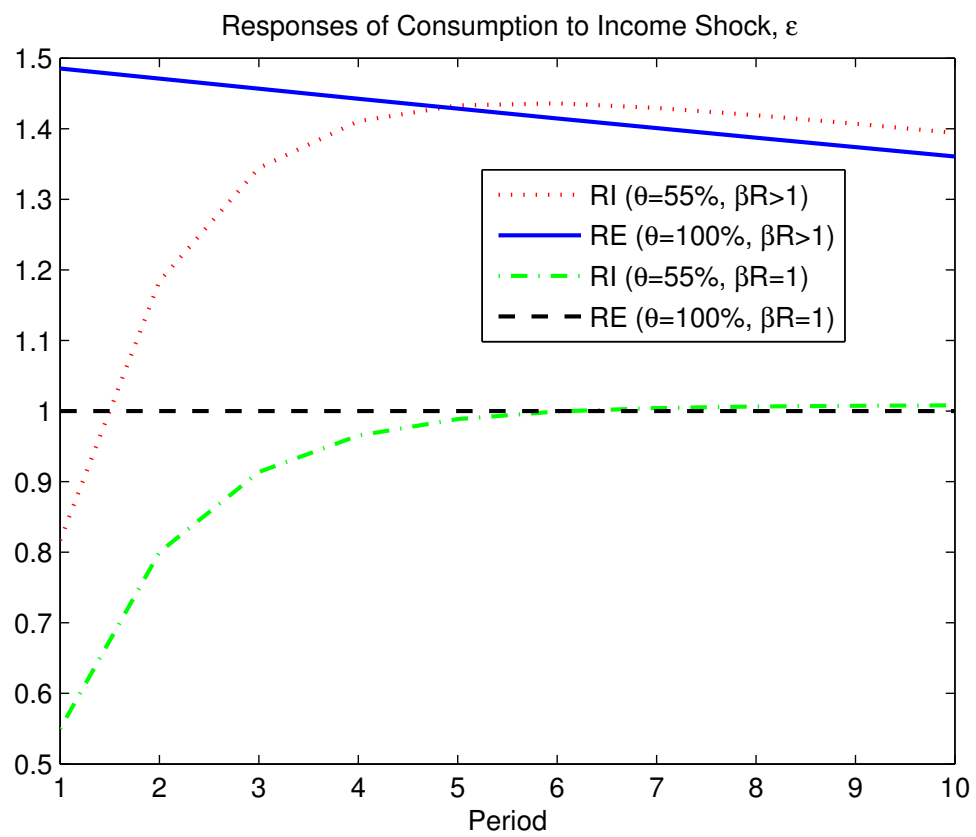


Figure 1

Given the specified income processes, (2.14), (2.15), and (2.16), consider the following three particular cases:²⁰

Case 1. Suppose that y_{t+1}^t follows an AR(1) process: $y_{t+1}^t = \bar{y} + \rho(y_t^t - \bar{y}) + \eta_{t+1}$ with $\rho \in (0, 1)$, which implies that

$$y_{t+1}^t = \bar{y} + \sum_{j=0}^{\infty} \rho^j \eta_{t+1-j}. \quad (2.28)$$

In this case, the new state variable $s_t = w_t + \frac{1}{R-1}y_t^p + \frac{\rho}{R-\rho}y_t^t$ and the innovation to permanent $\zeta_t = \frac{R}{R-1}\varepsilon_t + \frac{R}{R-\rho}\eta_t$.²¹ Thus, the short-run and long-run effects of transitory income shocks on consumption are

$$\text{SR}(\eta_t) = \theta \frac{R-1}{R-\rho} \text{ and } \text{LR}(\eta_t) = \frac{\theta}{1-(1-\theta)R} \frac{R-1}{R-\rho},$$

which means that the persistence in the income process ($\rho > 0$) increases both the short-run and long-run responses to the disturbances, but it does not change the propagation mechanism induced by RI.²² Figure 2a shows the responses of consumption to η_{t+1} with $\rho = 0.3$ and $\theta = 55\%$, 75% , and 100% (that is, $\kappa = 0.6$ bits, 1 bit, ∞ bits, respectively). Apparently, the responses of consumption are slow and gradually reach a flat asymptote as θ increases to 100% .²³ The figure also illustrates that consumption reacts to the endogenous noise, ξ_t , more quickly. Furthermore, given $\theta = 63\%$, Figure 2b plots the impulse responses to income innovations for different values of ρ . It shows that the higher the persistence coefficient ρ , the larger the impacts of the income shocks on consumption. It is worth noting that 1 bit of information transmitted is a very low number and is well below the total information-processing ability of human beings.²⁴ However, it is not implausible for ordinary consumers because they also face many other competing demands on capacity. Furthermore, low channel capacity devoted to processing available information can be rationalized because the welfare improvement from increasing capacity is tiny.²⁵

²⁰Note that if there is no permanent component and $y_{t+1}^t = \bar{y} + \eta_{t+1}$ (i.e., $M = 0$), the model reduces to the model with i.i.d. income process discussed in Sims (2003).

²¹For simplicity, here I ignore the constant term, $\frac{1}{R-1}\bar{y}$, in s_t .

²²Note that when $\rho = 0$, it reduces to the i.i.d. case in Sims (2003).

²³This flat asymptote is above the horizontal line representing the responses in the PIH model without RI. The intuition behind this result is that consumption initially does not react fully to income shocks and thus the share of undetected income shocks goes to savings and accumulates interest before the consumer fully digests the shocks.

²⁴Landauer at Bell estimated that human beings remembered about 2 bits per second under all the experimental conditions (visual, verbal, musical, or whatever).

²⁵In the next subsection, I show that the welfare losses due to RI are trivial for plausible parameter values.

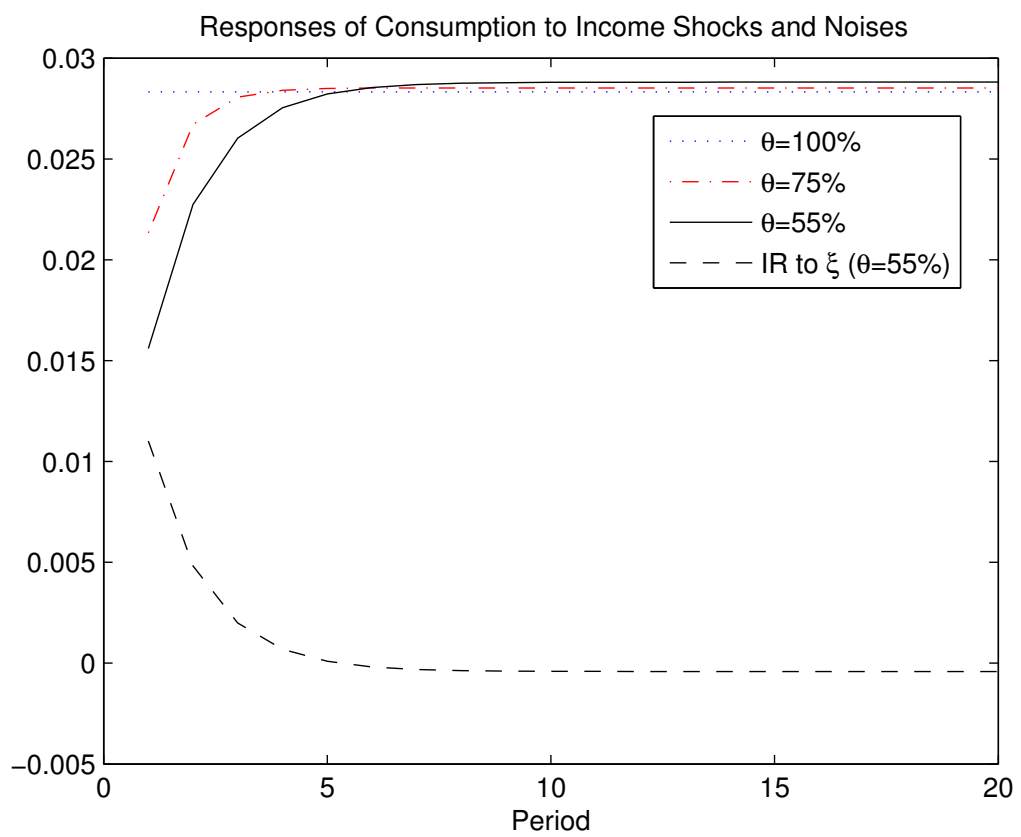


Figure 2a

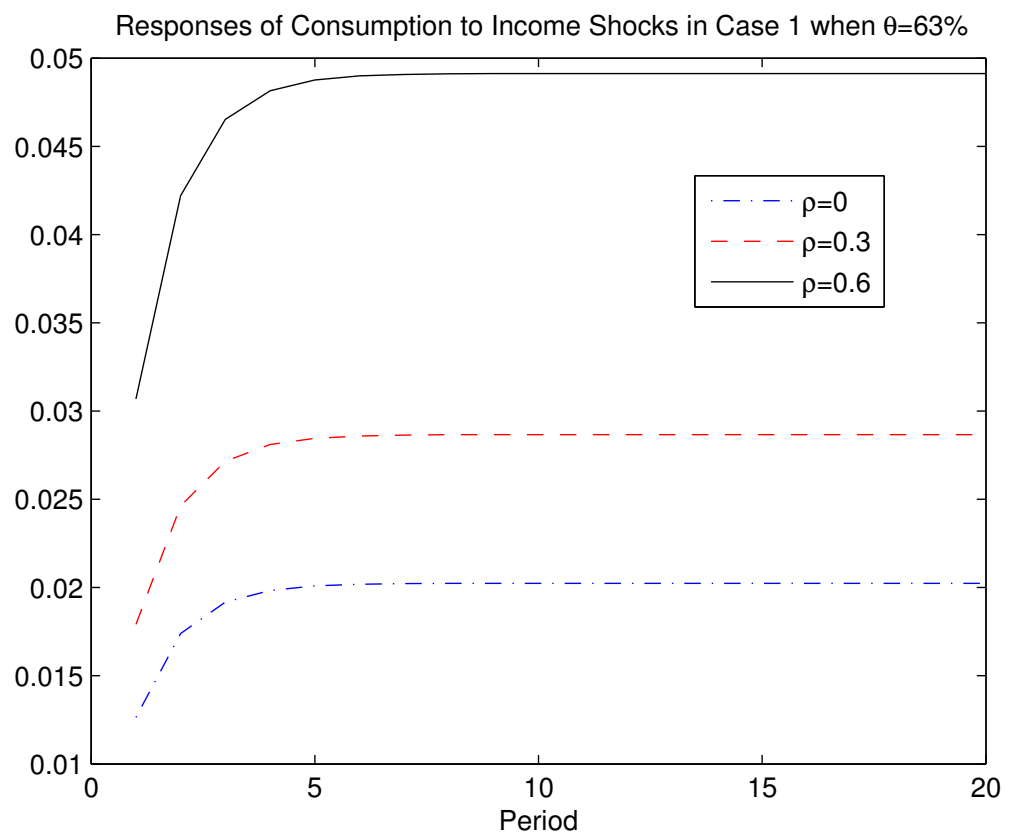


Figure 2b

Case 2. Assume that the transitory component follows an MA(2) process:²⁶

$$y_{t+1}^t = \bar{y} + \eta_{t+1} + \phi_1 \eta_t + \phi_2 \eta_{t-1}. \quad (2.29)$$

In this case, $s_t = w_t + \frac{1}{R-1} y_t^p + \left(\frac{\phi_1}{R} + \frac{\phi_2}{R^2} \right) \eta_t + \frac{\phi_2}{R} \eta_{t-1}$ and $\zeta_t = \frac{R}{R-1} \varepsilon_t + \left(1 + \frac{\phi_1}{R} + \frac{\phi_2}{R^2} \right) \eta_t$. Thus, the short-run and long-run effects of transitory income shocks on consumption can be written as

$$\text{SR}(\eta_t) = \theta H \left(1 + \frac{\phi_1}{R} + \frac{\phi_2}{R^2} \right) \text{ and } \text{LR}(\eta_t) = \frac{\theta}{1 - (1-\theta)R} H \left(1 + \frac{\phi_1}{R} + \frac{\phi_2}{R^2} \right),$$

where $H = \frac{R-1}{R}$. As in Case 1, the persistence in the income process ($\phi_1, \phi_2 > 0$) also increases both the short-run and long-run responses to the disturbances proportionally, but it does not change the propagation mechanism of the model. Given Equation (2.24), it is apparent that with finite capacity, consumption reacts to income shocks gradually and the higher the weights on the lagged innovation, ϕ_1 and ϕ_2 , the larger the impact of the income shocks on consumption.

Case 3. Suppose that income process follows a difference-stationary process, that is,

$$y_{t+1} - y_t = \rho(y_t - y_{t-1}) + \eta_{t+1}, \quad (2.30)$$

where $\rho \in (0, 1)$.²⁷ In this case, $s_t = w_t + \frac{R+(R-1)\rho}{(R-1)(R-\rho)} y_t - \frac{\rho R}{(R-1)(R-\rho)} y_{t-1} - 1$ and $\zeta_t = \frac{R^2}{(R-1)(R-\rho)} \eta_t$. The short-run and long-run effects of income shocks are

$$\text{SR}(\eta_t) = \theta \frac{R}{R-\rho} \text{ and } \text{LR}(\eta_t) = \frac{\theta}{1 - (1-\theta)R} \frac{R}{R-\rho}. \quad (2.31)$$

Note that in this case the amplification coefficient $\frac{R^2}{(R-1)(R-\rho)}$ can be rewritten as $\frac{R}{R-1} \frac{R}{R-\rho}$, which means that the effects of RI on consumption measured by the short-run and long-run impacts can be further amplified compared to the above unit root or AR(1) cases.

²⁶Hall and Mashkin (1982) used the PSID data to estimate the joint dynamics of consumption and income and found that transitory income can be characterized quite well by an MA(2) process with positive coefficients, 0.294 and 0.114.

²⁷For simplicity, here I ignore the transitory component in the income process and focus on this non-stationary component.

2.5. Welfare Implications of Information Processing Constraints

We may examine the welfare effects of income fluctuations under RI by calculating how much utility the consumer will lose if the actual consumption path under RI deviates from the first-best consumption path under RE. Following the literature, the value functions are used to measure individuals' intertemporal welfare. Because information constraints cannot help in individuals' optimization, the average welfare difference between the two economies should be greater than 0. Specifically, I will measure the loss by computing the reduction of welfare in a steady state measured by a money metric.

Following Cochrane (1989), Pischke (1995), and others, I use a money metric to measure the welfare cost of deviating from the first-best RE solution. The levels of the expected lifetime utility, $v(s_t)$ and $\widehat{v}(\widehat{s}_t)$, can be easily derived by using different decision rules, and then we get the expected welfare losses as $E[\Delta v] = E[v(s_t) - \widehat{v}(\widehat{s}_t)]$. Dividing $E[\Delta v]$ by the marginal utility of a dollar at time t and converting it to dollars per quarter gives:

$$\text{\$ Loss/quarter} = \frac{R-1}{R} \frac{E[\Delta v]}{u'(\bar{y})} = \frac{1}{2} \left(\frac{R-1}{R} \right)^2 \frac{\gamma}{\bar{y}} [(R+1)\bar{\sigma}^2 + (\text{var}[\widehat{s}_t] - \text{var}[s_t])], \quad (2.32)$$

where \bar{y} is the mean income, and γ is the local coefficient of relative risk aversion and equals to $\frac{\bar{y}}{c-\bar{y}}$ for the utility function $u(\bar{y})$. Using the quarterly data 1954-1990 provided in Pischke (pp. 830, 1995), I set the parameters as follows: $\bar{y} = \$6,929$, $\omega_\zeta = \$2,470$, and $R = 1.01$. Using Expression (2.32), Table 1 reports utility costs for several values of the local CRRA γ and the optimal weight on observations θ (Note that it increases to 1 as κ gets infinitely-large).²⁸

It is clear from Table 1 that the welfare losses due to RI are trivial. For example, for $\gamma = 4$ and $\theta = 50\%$ (that is, 50% of the uncertainty is removed upon the receipt of a new signal; another way to measure this value is to note that it implies that the consumer processes $\kappa = 0.5$ bits of the information about the state variable per quarter.),²⁹ the loss only amounts to 18 cents per quarter. This result is similar to the findings

²⁸ Allowing $\beta R > 1$ does increase the welfare loss due to RI, but the welfare effect is still not significant. Table 2 reports the welfare losses due to RI in this case. For example, when $\theta = 50\%$, $\beta = 0.99$, and $R = 1.02$, the welfare loss due to RI is \$0.18 that is still well below \$1.

²⁹ While the value of θ or κ is low, it is not impossible because some consumers may have little knowledge about their income and wealth levels.

by Pischke (1995) who calculated utility losses in the no-aggregate-information model.³⁰ Furthermore, even for a high value of the CRRA, $\gamma = 10$, and a low channel capacity ($\theta = 30\%$, i.e., $\kappa = 0.26$ bits), the welfare loss is still minor: only around \$1.1 per quarter. One explanation is that optimal consumption functions under RI and RE have very similar volatilities, even when κ is small.³¹ Based on the above calculations, we can conclude that for plausible assumptions, the welfare losses due to RI are not significant.³² This provides some evidence that it is reasonable for consumers to devote low channel capacity to observing and processing information because the welfare improvement from adopting the first-best decision rule is trivial. In other words, although consumers can devote much more capacity to processing economic information and then improve their optimal consumption decisions, it is rational for them to not do so because the welfare improvement is tiny.

3. Implications for Aggregate Consumption

In Section 2, I derived the consumption function and the expression of the change in consumption for a single consumer under RI. We may regard this consumer as the representative agent and then obtain the implications of RI for aggregate consumption directly without aggregation. However, the RI theory implies that the endogenous noise generated via the individual's own internal channel is partially idiosyncratic and therefore aggregation might play an important role in affecting the dynamics of aggregate consumption.

To investigate the effects of aggregation, for simplicity, suppose that all consumers in the model economy have identical degrees of RI, and both permanent and transitory components in the income process can be decomposed into a common aggregate effect and an idiosyncratic effect:³³

$$\varepsilon_{t+1} = \varepsilon_{t+1}^a + \varepsilon_{t+1}^i \quad \text{and} \quad y_{t+1}^t = \bar{y} + \eta_{t+1}^a + \eta_{t+1}^i, \quad (3.1)$$

³⁰He found that in most cases the utility losses due to no information about aggregate shock to income are less than \$1 per quarter.

³¹Luo and Young (2006) found similar results in a stochastic growth model with RI.

³²This conclusion is also consistent with that in Cochrane (1989), Pischke (1995), and others.

³³For simplicity, here I assume that the stationary component in the income process is i.i.d.

where the superscripts a and i denote aggregate and idiosyncratic, respectively. In this case, the expression of the change in individual consumption with RI is:

$$\Delta c_t^{i,*} = \theta H \left[\left(\frac{\zeta_t^i}{1 - (1 - \theta)R \cdot L} \right) + \left(\xi_t^i - \frac{\theta R \xi_{t-1}^i}{1 - (1 - \theta)R \cdot L} \right) \right], \quad (3.2)$$

where $H = \frac{R-1}{R}$, $\zeta_t^i = \frac{R}{R-1}(\varepsilon_t^a + \varepsilon_t^i) + (\eta_t^a + \eta_t^i)$, ε_t^a , ε_t^i , η_t^a , and η_t^i follow the distributions $N(0, \omega_a^2)$, $N(0, \omega_i^2)$, $N(0, \nu_a^2)$, and $N(0, \nu_i^2)$, respectively, and ξ_t^i is the idiosyncratic error.

Aggregating across all consumers yields the expression of the change in aggregate consumption:

$$\Delta C_t^* = \theta H \left[\left(\frac{\varepsilon_t^a/H + \eta_t^a}{1 - (1 - \theta)R \cdot L} \right) + \left(E^i [\xi_t^i] - \frac{\theta R E^i [\xi_{t-1}^i]}{1 - (1 - \theta)R \cdot L} \right) \right], \quad (3.3)$$

where $E^i(\cdot)$ denotes the population average in that period. Because consumers can distinguish aggregate from idiosyncratic shocks, their contributions to the idiosyncratic error can also be separated. The idiosyncratic error caused by the idiosyncratic income shocks (ε_t^i and η_t^i) is truly idiosyncratic and thus must be cancelled out after aggregation. However, aggregating the idiosyncratic errors caused by the aggregate income shocks (ε_t^a and η_t^a) is tricky. As argued in Sims (2003), although the randomness in an individual's response to aggregate shocks should also be *idiosyncratic* because it arises from his *own* internal information-processing constraint, a considerable part of the idiosyncratic error is common across individuals. The intuition is that people's needs for coding macroeconomic information efficiently are similar, so they rely on common sources of coded information. However, the existing theory does not provide a way to determine the common part included in the idiosyncratic error. We only know that the common term of the idiosyncratic error, $E^i [\xi_t^i]$, is between 0 and the part of the idiosyncratic error, ξ_t^a , caused by the aggregate income shocks ε_t^a and η_t^a , and the variance of ξ_t^a is

$$\text{var} [\xi_t^a] = \frac{1 - \theta}{\theta [1 - (1 - \theta)R^2]} \left[\left(\frac{R}{R-1} \right)^2 \omega_a^2 + \nu_a^2 \right]. \quad (3.4)$$

Formally, I assume that ξ_t^a consists of two independent noises: $\xi_t^a = v_t^c + v_t^i$, where v_t^c and v_t^i are the common and idiosyncratic components of the error generated by the aggregate shock, respectively.

Hence,

$$E^i [\xi_t^i] = v_t^c. \quad (3.5)$$

A parameter $\lambda = \frac{\text{var}[v_t^c]}{\text{var}[\xi_t^a]} \in [0, 1]$ can be used to measure the common source of coded information on the aggregate component (or the relative importance of v_t^c to ξ_t^a). The excess smoothness ratio in terms of θ and λ can be written as³⁴

$$\mu = \frac{\text{sd}[\Delta C_t^*]}{\text{sd}[\varepsilon_t^a]} = \sqrt{\frac{\theta^2}{1 - ((1 - \theta)R)^2} + \lambda^2 \left[\frac{\theta}{1 - (1 - \theta)R^2} - \frac{\theta^2}{1 - ((1 - \theta)R)^2} \right]}. \quad (3.6)$$

Note that the assumption of $\lambda = 0$ is equivalent to assume 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 $\mu = \sqrt{\frac{\theta^2}{1 - ((1 - \theta)R)^2}}$. By contrast, the assumption of $\lambda = 1$ (it is just the single-agent case discussed in Section 2.) means that people's needs for information coding are exactly the same and they completely rely on the common source of coded information. In this case, $\mu = \sqrt{\frac{\theta}{1 - (1 - \theta)R^2}}$. Hence, given the value of θ , the smoothness ratio of aggregate consumption, μ , should be between a lower bound reached when $\lambda = 0$ and an upper bound reached when $\lambda = 1$: $\mu \in \left[\sqrt{\frac{\theta^2}{1 - ((1 - \theta)R)^2}}, \sqrt{\frac{\theta}{1 - (1 - \theta)R^2}} \right]$.³⁵ Figure 3 illustrates the relationship between the smoothness ratio (μ) and the combination of (θ, λ) . The figure clearly shows that some combinations of (θ, λ) can generate the smoothness ratio that is well below 1.³⁶ 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 0.5). Hence, RI might provide an alternative explanation to the two consumption excesses in this RI model with a continuum of individuals because in the presence of RI aggregate consumption reacts to aggregate income shocks gradually and the excess smoothness ratio is less than 1 for some combinations of (θ, λ) .

³⁴See Appendix 5.5 for the derivation.

³⁵For simplicity, here I ignore the transitory aggregate component because the permanent aggregate component dominates in the US data.

³⁶In the RE version of the PIH model (i.e., $\theta = 100\%$), $\mu = 1$, that is, aggregate consumption should have the same volatility as aggregate income, the very opposite of the stylized fact that aggregate consumption is much smoother than income.

The Excess Smoothness of Aggregate Consumption

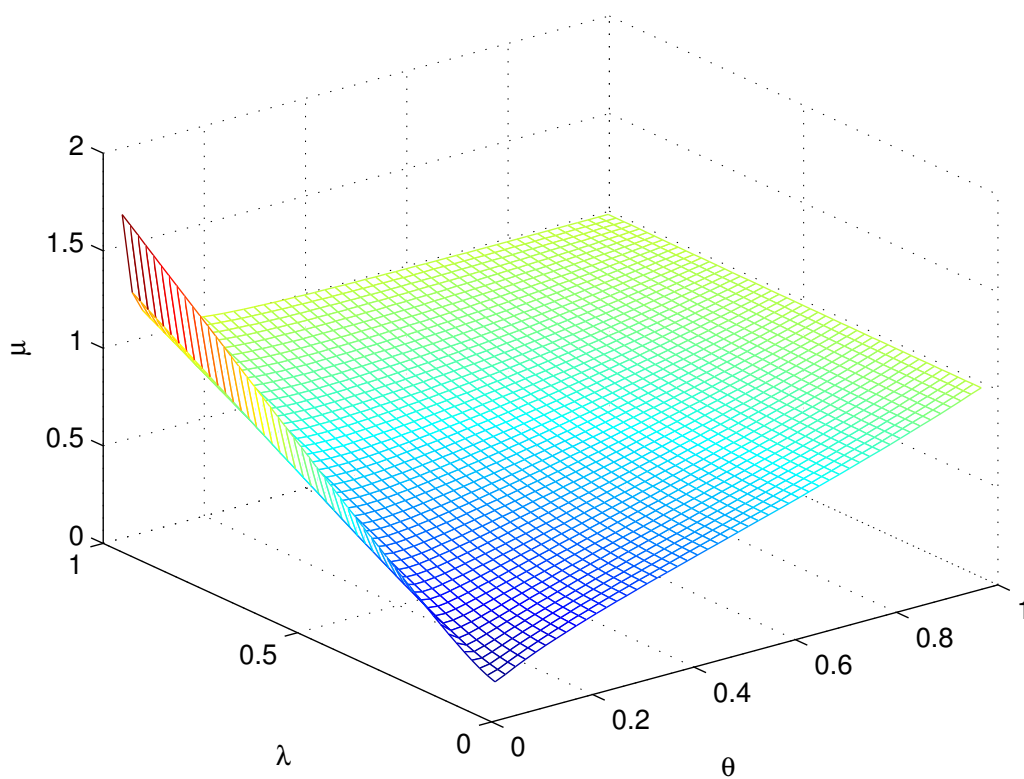


Figure 3

4. Comparison of Consumption Behavior under Alternative Hypotheses

Sims (2003) compared the RI theory with other information-related theories: the information delay theory and the signal extraction theory, and argued that they have different implications for the effects of changes in policy on the dynamic properties of the economy. In this section, I first derive the mapping from the specified income process to the joint stochastic process of consumption and income for each alternative hypothesis within the PIH framework, and then discuss how the models differ in their empirical content as summarized by the mapping.

4.1. Habit Formation

Habit formation (HF) has been modelled directly as a structure of preferences in which psychological factors make consumers prefer to gradual adjustment in consumption, thereby consumption volatility is more painful than it would be in the absence of habits. Consider a simple PIH model with internal and additive habit formation, in which higher consumption in last period creates a habit that lowers utility in this period. Specifically, the typical consumer with HF has the utility $u(\tilde{c}_t) = \tilde{c}_t - \frac{1}{2}\tilde{c}_t^2$, where $\tilde{c}_t = c_t - \gamma c_{t-1}$ and γ is the parameter governing the habit. In this model, optimizing consumers take into account the effects of their current consumption on their future reference stock. Given the same budget constraint and stochastic processes for income y_t specified in Section 2, the change in individual consumption can be written as follows

$$\Delta c_t = \left(1 - \frac{\gamma}{R}\right) \frac{\varepsilon_t + H\eta_t}{1 - \gamma \cdot L}, \quad (4.1)$$

where $H = \frac{R-1}{R}$.

Comparing Equation (4.1) with Equation (2.24), it is apparent that the two models generate different mappings from stochastic processes for y_t into the joint stochastic processes for (c_t, y_t) . Specifically, individual consumption under RI reacts not only to income shocks but also to its *own* endogenous noises induced by finite capacity. The endogenous noise can be regarded as a type of “consumption shock” or “demand shock”. Let’s consider a special case in which $1 - \theta = \frac{\gamma}{R}$. In this case, Equations (2.24) and

(4.1) show that although the two models have the same pattern of the optimal consumption response to income shocks, consumption growth under RI is more volatile than that under HF because the noise terms due to RI also contribute to the relative volatility of consumption.³⁷

However, the effects of HF and RI on aggregate consumption could be similar because aggregating across all individuals would weaken or even eliminate the impacts of the endogenous noises on consumption growth. Indeed, as discussed in Section 3, if all consumers are completely isolated (that is, there is no communication among them), and $1 - \theta = \frac{\gamma}{R}$, the two models can generate the same joint dynamics of aggregate consumption and income.

HF implies that slow adjustment in consumption is optimal because consumers are assumed to prefer to smooth not only consumption but also consumption growth, while 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. Hence, although these two models have distinct mechanisms for slow consumption adjustment, they are observationally equivalent in terms of the joint behavior of aggregate consumption and income under some restrictions. Finally, it is worth noting that HF could affect consumption decisions both in the case without uncertainty and that with uncertainty, whereas RI affects consumption through its interaction with the fundamental uncertainty and it has no impact on consumption in the absence of uncertainty.

4.2. Incomplete Information and Signal Extraction

Consider a simple signal extraction (SE) version of the PIH model presented in Section 2.1, in which consumers only observe the total income process but have no way to separate different components in the income processes specified by (2.14), (2.15), and (2.16). Hence, in first differences, $\Delta y_t = \varepsilon_t + \eta_t - \eta_{t-1}$.

³⁷Note that when the uncertainty due to fundamental income shocks reduces, the volatility of noises will be reduced accordingly and then the properties of the dynamics of individual consumption in the two models become more and more similar if $1 - \theta = \frac{\gamma}{R}$.

The best that consumers can do is to recognize that the first difference of income is an MA(1) process

$$\Delta y_t = u_t - \mu u_{t-1}, \quad (4.2)$$

where μ is determined by the relative variances of the transitory and permanent income shocks and $1 - \mu$ measures the persistent effects of the compound shock u_t on income.³⁸ Hence, the case of $\mu = 0$ corresponds only to the permanent shock, whereas $\mu = 1$ corresponds only to the transitory shock. The expression for the change in consumption can thus be written as

$$\Delta c_t = \left(1 - \frac{\mu}{R}\right) u_t = \left(1 - \frac{\mu}{R}\right) \frac{\Delta y_t}{1 - \mu \cdot L}, \quad (4.3)$$

where $\Delta y_t = \varepsilon_t + \eta_t - \eta_{t-1}$. In this model the parameter μ determines the degree of slow adjustment in consumption. Similarly, compared with Equation (2.24), Equation (4.3) shows that although SE also leads to slow adjustment in consumption, it does not generate endogenous noises.

As shown in Pischke (1995), if ε_t is the aggregate shock and η_t is the idiosyncratic shock, all the idiosyncratic components would be cancelled out after aggregating across all consumers, thereby generating an $AR(1)$ process of the change in aggregate consumption. The intuition is that aggregate consumption reacts only gradually to aggregate income because aggregate shocks are more persistent and are only partially observed. Hence, aggregation under this type of incomplete information has an effect that is very close to the effect of RI under some restrictions or the effect of HF.

It is worth emphasizing that if RI can be modelled endogenously by assuming that individuals are allowed to choose to allocate the amount of attention among different elements in the state vector, it would be optimal for them to devote low attention to monitoring the aggregate component because aggregate component is less important for individuals' optimal decisions.³⁹ Hence, RI has a potential

³⁸Note that $\mu = -\left(1 - \sqrt{1 - 4\varphi^2}\right) / 2\varphi$ where $\varphi = -\omega_\eta^2 / (\omega_\varepsilon^2 + 2\omega_\eta^2)$.

³⁹Maćkowiak and Wiederholt (2005) formally examined how attention allocation affects firms' pricing decisions and showed that firms allocate less attention in processing aggregate information. However, the problem of attention allocation would be more complicated in the multivariate PIH model than in the firm's pricing model because in the PIH model with RI the consumers need to track both the *endogenous* state variable (individual wealth) and the exogenous state variable (the persistent income process) and thus an additional nonlinear "no subsidization" constraint that complicates the model greatly need to be imposed to address how to allocate attention across these variables. See Sims (2003) for a detailed

to endogenize the ignorance of aggregate information included in individual income processes and thus could provide a micro-foundation for slow adjustment in aggregate consumption due to the exogenous incomplete information.

4.3. Model Misspecification and Robustness

Robust control emerged in the engineering literature in the 1970s, and was introduced into economics and further developed by Hansen, Sargent, and others. A simple version of robust optimal control considers such a question: How to make decisions when the agent does not know the probability model that generates the data? The agent with the preference for robustness considers a range of models, and makes decisions that maximize utility given the worst possible model. Following Hansen and Sargent (2005), the robust consumer's optimization problem within the PIH framework can be written as

$$v(s_t) = \max_{c_t} \min_{w_t} \left\{ -\frac{1}{2} (c_t - \bar{c})^2 + \beta [\vartheta w_t^2 + E_t[v(s_{t+1})]] \right\} \quad (4.4)$$

subject to

$$s_{t+1} = R(s_t - c_t) + (\zeta_{t+1} + w_t), \quad (4.5)$$

where w_t distorts the mean of the innovation and $\vartheta > 0$ controls how bad the error can be. Solving this problem gives the change in consumption⁴⁰

$$\Delta c_t = \frac{R-1}{R - \omega_\zeta^2 / (2\vartheta)} \left[-\frac{\omega_\zeta^2}{2\vartheta} c_{t-1} + \left(\frac{R}{R-1} \varepsilon_t + \eta_t \right) \right], \quad (4.6)$$

Equation (4.6) shows that robustness raises the sensitivity of changes in consumption to unanticipated changes in income ζ_t because $\vartheta > 0$ and then $\frac{R-1}{R - \omega_\zeta^2 / (2\vartheta)} > \frac{R-1}{R}$. This type of excess sensitivity is referred to as “making hay while the sun shines” in the literature and it does not occur in the standard RE model, the RI model, the HF model, or the SE model. In addition, consumption growth under robustness also depends on lagged information (past consumption) and the effect of past consumption disappears

discussion.

⁴⁰When ϑ approaches infinity, the robust solution reduces to the standard RE solution. The detailed derivations for the consumption function and the change in consumption under robustness are available from the author by request.

as ϑ approaches infinity. Equation (4.6) also clearly shows that the relative volatility of consumption to income increases as the degree of robustness (ϑ) increases. After summing over all consumers, the joint behavior of aggregate consumption and income under robustness has the same structure as that of individual consumption and income, which means that robustness reduces the smoothness of aggregate consumption to aggregate income. Thus, robustness itself exaggerates the standard RE PIH model's prediction for the joint dynamics of aggregate consumption and income and therefore needs to combine with other assumptions such as habit formation and adjustment costs to better explain the data in this aspect.

4.4. Costly Planning and Inattentiveness

In Reis' costly planning and inattentiveness model, during the intervals of inattentiveness, consumption dynamics is determined by the standard deterministic consumer's optimizing problem, whereas at the adjustment dates, consumption is determined by the standard stochastic consumer problem. Reis showed that aggregate consumption growth between two consecutive periods, t and $t + 1$, in the inattentiveness economy can be written as

$$\Delta C_{t+1} = \text{constant} + \Phi(0) e_{t+1} + \Phi(1) e_t + \dots + \Phi(I) e_{t+1-I}, \quad (4.7)$$

where $\Phi(s) \geq \Phi(s+1) \geq 0$ for $s = 1, 2, \dots, I$, and $\{e_t\}$ are mutually uncorrelated "news" unpredictable one period ahead. Expression (4.7) reveals that aggregate consumption 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. By 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. Conclusions

In this paper, I proposed an analytical approach to solve an RI version of the PIH model with general income processes and examined the effects of RI on optimal consumption decisions. I then worked through some specific income processes to examine the short-run and long-run responses of consumption to various income shocks. In addition, I analyzed welfare consequences due to RI and found that the welfare loss that results from deviating from the consumption path under RE is tiny. By examining the implications of RI for aggregate consumption, I showed that RI has a potential to resolve the twin consumption puzzles (the excess smoothness puzzle and the excess sensitivity puzzle). Finally, I compared the RI model with four alternative models by examining their implications for the joint dynamics of consumption and income.

Thus far, I have not considered costly attention and just set channel capacity κ exogenously. Because attention is a scarce economic resource and there are many competing demands, it is natural to ask how to determine optimal channel capacity endogenously. An interesting direction for future research would be to model endogenous inattention and examine its implications for optimal decisions. Second, since RI largely affects the optimal intertemporal allocation of consumption, it is also worthwhile examining if incorporating RI into the heterogeneous-agent stochastic general equilibrium models can help us answer some important questions in the literature. For example, can RI play a role in shaping the skewed wealth distribution in the U.S.? Can RI increase aggregate saving in general equilibrium? I leave these interesting topics to future research.

Appendix

5.1. Defining a New State Variable

Given that the consumption function is $c_t = \frac{R-1}{R} \left(w_t + \sum_{j=1}^{\infty} R^{-j} E_t [y_{t+j}] \right)$, the original budget constraint (2.2) can be rewritten as

$$\begin{aligned} & w_{t+1} + \sum_{j=1}^{\infty} R^{-j} E_{t+1} [y_{t+1+j}] \\ &= R \left(w_t + \sum_{j=1}^{\infty} R^{-j} E_t [y_{t+j}] - c_t \right) + y_{t+1} + \sum_{j=1}^{\infty} R^{-j} E_{t+1} [y_{t+1+j}] - R \sum_{j=1}^{\infty} R^{-j} E_t [y_{t+j}] \\ &= R \left(w_t + \sum_{j=1}^{\infty} R^{-j} E_t [y_{t+j}] - c_t \right) + \zeta_{t+1}, \end{aligned}$$

where the $(t+1)$ -innovation $\zeta_{t+1} = \sum_{j=t+1}^{\infty} R^{-j+(t+1)} (E_{t+1} - E_t) [y_j]$. Denote $s_t = w_t + \sum_{j=1}^{\infty} R^{-j} E_t [y_{t+j}]$, the new budget constraint and the consumption function can be rewritten as

$$s_{t+1} = R(s_t - c_t) + \zeta_{t+1} \text{ and } c_t = \frac{R-1}{R} s_t, \quad (5.1)$$

respectively.

5.2. Proof of Proposition 1

Solving for the value function and the consumption function under RI

Under the LQG specification, *the separation principle* applies, that is, state estimation and optimal control can be decoupled. See Whittle (1996) for a detailed discussion. In other words, we can just replace the actual state variable s_t with the perceived state variable \widehat{s}_t in Equation (2.3) and then obtain Equation (2.22). The following procedure includes detailed derivations of the value function and the consumption function under RI.

First, conjecture that the value function takes the following quadratic form

$$v(\widehat{s}_t) = A_0 + A_1 \widehat{s}_t + A_2 \widehat{s}_t^2, \quad (5.2)$$

where A_0, A_1 , and A_2 are constants to be determined. Substituting Equations (2.6), (2.20), and (5.2) into the Bellman equation, (2.19), yields

$$A_0 + A_1 \widehat{s}_t + A_2 \widehat{s}_t^2 = \max_{c_t} \left\{ -\frac{1}{2} (c_t - \bar{c})^2 + \beta E_t [A_0 + A_1 \widehat{s}_{t+1} + A_2 \widehat{s}_{t+1}^2] \right\}. \quad (5.3)$$

Second, performing the indicated optimization yields the following first-order condition

$$-(c_t - \bar{c}) - 2A_2 R E_t [\widehat{s}_t - c_t] - A_1 = 0,$$

which means that

$$c_t = \frac{-2RA_2}{1 - 2RA_2} \widehat{s}_t + \frac{\bar{c} - A_1}{1 - 2RA_2}. \quad (5.4)$$

Substituting Equations (2.6), (2.20), and (5.4), back into Equation (5.3) to arrive at the following equation:

$$\begin{aligned} & A_0 + A_1 \widehat{s}_t + A_2 \widehat{s}_t^2 \\ &= -\frac{1}{2} \left(\frac{-2RA_2}{1 - 2RA_2} \widehat{s}_t + \frac{\bar{c} - A_1}{1 - 2RA_2} - \bar{c} \right)^2 + \beta E_t [A_0 + A_1 \widehat{s}_{t+1} + A_2 \widehat{s}_{t+1}^2] \\ &= -\frac{1}{2} \left[\left(\frac{-2RA_2}{1 - 2RA_2} \right)^2 \widehat{s}_t^2 + 2 \left(\frac{-2RA_2}{1 - 2RA_2} \right) \left(\frac{2RA_2 \bar{c} - A_1}{1 - 2RA_2} \right) \widehat{s}_t + \left(\frac{2RA_2 \bar{c} - A_1}{1 - 2RA_2} \right)^2 \right] \\ &+ \beta A_0 + A_1 E_t \left[\frac{1}{1 - 2RA_2} \widehat{s}_t - \frac{\bar{c} - A_1}{1 - 2RA_2} \right] \\ &+ \beta A_2 \left\{ \left(\frac{R}{1 - 2RA_2} \right)^2 \widehat{s}_t^2 + \frac{2}{1 - 2RA_2} \left(-\frac{R(\bar{c} - A_1)}{1 - 2RA_2} \right) \widehat{s}_t + \text{var}[\varpi_{t+1}] + R^2 \left(\frac{\bar{c} - A_1}{1 - 2RA_2} \right)^2 \right\}, \end{aligned}$$

where $\varpi_{t+1} = \theta \left[\left(\frac{\xi_{t+1}}{1 - (1 - \theta)R \cdot L} \right) + \left(\xi_{t+1} - \frac{\theta R \xi_t}{1 - (1 - \theta)R \cdot L} \right) \right]$. Third, collecting and matching terms, the constants turn out to be

$$A_2 = -\frac{R - 1}{2R}, \quad A_1 = \bar{c}, \quad A_0 = -\frac{R}{2(R - 1)} \bar{c}^2 - \frac{\text{var}[\varpi_{t+1}]}{2R}.$$

Substituting these back into Equations (5.2) and (5.4) yields the value function, (2.21), and the consumption function, (2.22), of the text.

Deriving the expression of the change in consumption

The change in consumption can be derived by combining Equations (2.6), (2.22), with (2.23). First, the change in the effective state $\Delta\widehat{s}_{t+1}$ can be written as

$$\begin{aligned}\Delta\widehat{s}_{t+1} &= \theta(s_{t+1} - \widehat{s}_t) + \theta\xi_{t+1} \\ &= \theta R(s_t - \widehat{s}_t) + \theta(\zeta_{t+1} + \xi_{t+1}).\end{aligned}$$

Second, $s_t - \widehat{s}_t$ can be found as follows

$$\begin{aligned}s_{t+1} - \widehat{s}_{t+1} &= R(s_t - c_t^*) + \zeta_{t+1} + \bar{y} - (1 - \theta)\widehat{s}_t - \theta(s_{t+1} + \xi_{t+1}) \\ &= (1 - \theta)R(s_t - \widehat{s}_t) + (1 - \theta)\zeta_{t+1} - \theta\xi_{t+1},\end{aligned}$$

which means that $s_t - \widehat{s}_t = \frac{(1-\theta)\zeta_t - \theta\xi_t}{1 - (1-\theta)R \cdot L}$, where L is the lag operator. Combining the expressions of $\Delta\widehat{s}_{t+1}$ and $s_t - \widehat{s}_t$ gives

$$\Delta\widehat{s}_{t+1} = \theta \left[\frac{\zeta_{t+1}}{1 - (1 - \theta)R \cdot L} + \left(\xi_{t+1} - \frac{\theta R \xi_t}{1 - (1 - \theta)R \cdot L} \right) \right], \quad (5.5)$$

from which the expression of the change in consumption can be obtained directly: $\Delta c_t^* = \frac{R-1}{R} \Delta\widehat{s}_t$.

In addition, given that $\theta = 1 - \frac{1}{\exp(2\kappa)}$ and $\bar{\sigma}^2 = \frac{\omega_\varepsilon^2}{\exp(2\kappa) - R^2} > 0$, it is straightforward to prove that $(1 - \theta)R < 1$.

5.3. The Smoothness of Individual Consumption Growth

Equation (2.24) implies that the variance of consumption growth

$$\begin{aligned}\text{var} [\Delta c_t^*] &= \theta^2 \left[\frac{1}{1 - (R(1 - \theta))^2} \omega_\varepsilon^2 + \omega_\xi^2 + \frac{(\theta R)^2}{1 - (R(1 - \theta))^2} \omega_\xi^2 \right] \\ &= \theta^2 \left[\frac{1}{1 - (R(1 - \theta))^2} + \frac{1 + R^2(2\theta - 1)}{1 - (R(1 - \theta))^2} \frac{1}{(1/(1 - \theta) - R^2)\theta} \right] \omega_\varepsilon^2 \\ &= \frac{\theta}{1 - R^2(1 - \theta)} \omega_\varepsilon^2,\end{aligned}$$

where I use the facts that $\omega_\xi^2 = \text{var} [\zeta_{t+1}] = \frac{\bar{\sigma}^2}{\theta}$, $\bar{\sigma}^2 = \frac{\omega_\varepsilon^2}{\exp(2\kappa) - R^2}$, and $\theta = 1 - \frac{1}{\exp(2\kappa)}$.

5.4. The Welfare Effects of RI

Calculating $\text{var}[s_t]$ and $\text{var}[\hat{s}_t]$

Equations (2.6) and (2.20) can be written in the following matrix form

$$A \begin{bmatrix} s_{t+1} \\ \hat{s}_{t+1} \end{bmatrix} = B \begin{bmatrix} s_t \\ \hat{s}_t \end{bmatrix} + \begin{bmatrix} \zeta_{t+1} \\ \theta \xi_{t+1} \end{bmatrix}, \quad (5.6)$$

where $A = \begin{bmatrix} 1 & 0 \\ -\theta & 1 \end{bmatrix}$ and $B = \begin{bmatrix} R & 1-R \\ 0 & 1-\theta \end{bmatrix}$. Taking unconditional variances we find that

$$A \Sigma_s A' = B \Sigma_s B' + \begin{bmatrix} \omega_\zeta^2 & 0 \\ 0 & \theta^2 \omega_\xi^2 \end{bmatrix} \quad (5.7)$$

where $\Sigma_s = \begin{bmatrix} \text{var}[s] & \text{covar}[s, \hat{s}] \\ \text{covar}[s, \hat{s}] & \text{var}[\hat{s}] \end{bmatrix}$. This expression is a standard discrete Lyapounov equation.

Given the numerical values of R , θ , and ω_ζ^2 , it is straightforward to calculate Σ_s numerically.

Calculating the welfare loss due to RI

Given the value function under RI, $\hat{v}(\hat{s}_t)$, derived in appendix 5.2 and the standard value function under RE,

$$v(s_t) = -\frac{R-1}{2R} s_t^2 + \bar{c} s_t - \frac{R}{2(R-1)} \bar{c}^2 - \frac{\text{var}[\zeta_{t+1}]}{2R}, \quad (5.8)$$

where $\zeta_{t+1} = \frac{R}{R-1} \varepsilon_{t+1} + \eta_{t+1}$, it is straightforward to show that the welfare loss due to RI can be written as

$$E[\Delta v] = E[v(s_t) - \hat{v}(\hat{s}_t)] \quad (5.9)$$

$$= -\frac{R-1}{2R} (\text{var}[s_t] - \text{var}[\hat{s}_t]) - \frac{1}{2R} (\text{var}[\zeta_{t+1}] - \text{var}[\varpi_{t+1}]) \quad (5.10)$$

$$= -\frac{R-1}{2R} (\text{var}[s_t] - \text{var}[\hat{s}_t]) + \frac{1}{2R} \frac{(1-\theta)(R^2-1)}{1-(1-\theta)R^2} \omega_\zeta^2, \quad (5.11)$$

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

5.5. The Smoothness of Aggregate Consumption Growth

Using Equations (3.3), (3.4), and (3.5), the variance of aggregate consumption growth is simply

$$\begin{aligned} \text{var} [\Delta C_t^*] &= \theta^2 \left\{ \frac{1}{1 - (R(1 - \theta))^2} \omega_a^2 + \lambda^2 \left[\omega_\xi^2 + \frac{(\theta R)^2}{1 - (R(1 - \theta))^2} \omega_\xi^2 \right] \right\} \\ &= \theta^2 \left[\frac{1}{1 - (R(1 - \theta))^2} + \lambda^2 \frac{1 + R^2 (2\theta - 1)}{1 - (R(1 - \theta))^2} \frac{1}{(1/(1 - \theta) - R^2)\theta} \right] \omega_a^2 \\ &= \theta^2 \left\{ \frac{1}{1 - (R(1 - \theta))^2} + \lambda^2 \left[\frac{1}{(1 - R^2 (1 - \theta))\theta} - \frac{1}{1 - (R(1 - \theta))^2} \right] \right\} \omega_a^2, \end{aligned}$$

which is just Equation (3.6). Hence, when $\lambda = 0$, $\text{var} [\Delta C_t^*] = \frac{\theta^2}{1 - (R(1 - \theta))^2} \omega_a^2$. When $\lambda = 1$, $\text{var} [\Delta C_t^*] = \frac{\theta}{1 - R^2(1 - \theta)} \omega_a^2$. Since $\frac{1}{(1 - R^2(1 - \theta))\theta} - \frac{1}{1 - (R(1 - \theta))^2} > 0$ when $R^2(1 - \theta) < 1$, the excess smoothness ratio μ decreases with the value of λ , holding constant all other parameters.

References

- [1] Adam, K., 2005. Optimal monetary policy in the presence of imperfect common knowledge, *Journal of Monetary Economics* 54 (2), 267-301.
- [2] Ball, L., Mankiw, G., Reis, R., 2005. Monetary policy for inattentive economies, *Journal of Monetary Economics* 52 (4), 703-725.
- [3] Campbell, J., 1987. Does saving anticipate declining labor income? An Alternative Test of the Permanent Income Hypothesis, *Econometrica* 55, 1249-1273.
- [4] Campbell, J., Deaton A., 1989. Why is consumption so smooth?, *Review of Economic Studies* 56, 357-374.
- [5] Campbell, J., Mankiw, G., 1989. Consumption, income, and the interest rates: reinterpreting the time series evidence, *NBER Macroeconomics Annual 1989*, Cambridge, MA: MIT Press, 185-215.

- [6] Cochrane, J., 1989. The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives, *American Economic Review* 79, 319-337.
- [7] Cover, T.M., Thomas, J.A., 1991. *Elements of Information Theory*. New York: John Wiley and Sons.
- [8] Deaton, A., 1992. *Understanding Consumption*. Oxford University Press, Oxford.
- [9] Flavin, M., 1981. The adjustments of consumption to changing expectations about future income, *Journal of Political Economy* 89, 974-1009.
- [10] Hall, R., 1978. Stochastic implications of the life cycle-permanent income hypothesis: Theory and Evidence, *Journal of Political Economy* 91, 249-265.
- [11] Hall, R., Mashkin, F., 1982. The sensitivity of consumption to transitory income: estimates from panel data on households, *Econometrica* 50, 461-481.
- [12] Hansen, Lars., Sargent, T., 2005. *Robustness*. Manuscript, University of Chicago and New York University.
- [13] Ludvigson, S., Michaelides, A., 2001. Does buffer stock saving explain the smoothness and excess sensitivity of consumption? *American Economic Review* 91 (3), 631-647.
- [14] Luo, Y., Young, E.R., 2006. *Rational inattention and aggregate fluctuations*, Manuscript, University of Hong Kong and University of Virginia.
- [15] Maćkowiak, B., Wiederholt, M., 2005. *Optimal sticky prices under rational inattention*, Manuscript, Humboldt University Berlin.
- [16] Moscarini, G., 2004. Limited information capacity as a source of inertia, *Journal of Economic Control and Dynamics* 28(10), 2003-2035.
- [17] Pischke, J.S., 1995. Individual income, incomplete information, and aggregate consumption, *Econometrica* 63 (4), 805-840.
- [18] Reis, R., 2006. Inattentive consumer, *Journal of Monetary Economics*, 53 (8), 1761-1800.

- [19] Shannon, C.E., 1948. A mathematical theory of communication, *The Bell System Technical Journal* 27, 623-656.
- [20] Sims, C.A., 1998. Stickiness, *Carnegie-Rochester Conference Series on Public Policy*, Vol 49(1), 317-356.
- [21] Sims, C.A., 2003. Implications of rational inattention, *Journal of Monetary Economics* 50, 3, 665-690.
- [22] Sims, C.A., 2005. Rational inattention: a research agenda," Manuscript, Princeton University.
- [23] Whittle, P., 1996. *Optimal Control: Basics and Beyond*. John Wiley & Sons.
- [24] Woodford, M., 2001. Imperfect common knowledge and the effects of monetary policy, Princeton University Press, in *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*.