

Consumption Dynamics under Information Processing Constraints*

Yulei Luo
School of Economics and Finance
The University of Hong Kong
Pokfulam Road
Hong Kong
Email: yluo@alumni.princeton.edu
Tel: (609)2403475

This version: May 2006

First version: May 2005

Abstract

This paper studies how “Rational Inattention” - a kind of information processing constraints proposed by Sims (2003)- affects consumption and saving behavior in a permanent income hypothesis (PIH) model. It is shown that incorporating RI into the otherwise standard PIH model can substantially affect the intertemporal allocation of consumption, which makes the model better explain the data in some important aspects. Specifically, I propose a tractable approach to solve the multivariate state PIH model with RI explicitly and then show that incorporating RI into the model can help resolve the excess sensitivity puzzle and the excess smoothness puzzle in the literature. Furthermore, I find that the utility costs due to RI are very trivial, which can rationalize the key assumption that consumers do not have strong incentive to devote high channel capacity in observing and processing economic information. Finally, I compare the RI hypothesis with the habit formation hypothesis and the rule-of-thumb hypothesis.

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

Keywords: *Rational Inattention, Stochastic Optimal Control, Consumption Decisions, Welfare Effects.*

*I am very grateful to Christopher A. Sims for his excellent advice and support. I also thank Per Krusell, Jonathan Parker, Ricardo Reis, Lars Svensson, Charles Whiteman, Wei Xiong, 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.

1. Introduction

Canonical macroeconomic theory assumes that agents have the ability to process as much information as they need instantaneously and thus have perfect information about the state(s). However, in reality ordinary people only have finite information-processing capacity, consequently, they cannot digest all relevant information and do not have perfect observation when making economic decisions.¹ In his pioneering work, Sims (2003) proposes a tractable Linear Quadratic Gaussian (LQG)² framework, in which agents make optimal decision under information processing constraints³ in the sense that they only have limited capacity to process information. One of the main findings in his paper is that in the RI model information imperfection emerges endogenously because agents only have finite (Shannon) capacity to process information and thus cannot observe the state perfectly. Therefore, the nature of the noise due to RI is determined endogenously and/or optimally when agents need to allocate their limited information capacity across various sources. This endogeneity of noise distinguishes the RI model from other LQG models with imperfect state observations that are widely studied in stochastic control theory (e.g., Whittle 1982, 1996).

One of the most important contributions to modern intertemporal consumption-saving theory is the idea that consumption is determined by consumers' permanent income that is defined as the expected value of lifetime resources. The typical specification of this concept is a special case, where utility functions are linear quadratic, labor income is stochastic, the interest rate is constant, there are no borrowing constraints, consumers have infinite horizons and perfect information, and expectations are formed rationally. This model is referred to as the Permanent Income Hypothesis (PIH).⁴ Although the standard PIH model has considerable intuitive appeal, a host of influential papers have revealed two robust discrepancies between the model's predictions and the empirical evidence. First, as argued in Campbell and Deaton (1989) and Deaton (1992), the PIH predicts that consumption growth should be more volatile than income growth if income growth has positive serial correlation (as the US data suggest it does), yet in the data aggregate consumption growth is much smoother than aggregate income growth. This is the *excess smoothness* puzzle: if income is nonstationary, the observed consumption growth is much too smooth, relative to the consumption path predicted by the PIH. Second, the PIH predicts that consumption growth is orthogonal to lagged or predictable income growth, yet in the data consumption growth and lagged income growth are correlated, that is, the growth of consumption can be predicted by expected or lagged changes in income (For example, Flavin 1982, Campbell and Mankiw 1989, and Deaton 1992). It is called the *excess sensitivity* puzzle. Thus aggregate consumption growth exhibits two puzzles: it is both "excessively smooth" to unanticipated income growth, and "excessively sensitive" to

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

²Note that here "Linear" means that the flow budget constraint is linear, "Quadratic" means that the objective function is quadratic, and "Gaussian" means that the exogenous force is Gaussian.

³It is labelled "Rational Inattention" in Sims (2003). In this paper, I use "information processing constraints" and "Rational Inattention" (henceforth "RI") interchangeably.

⁴Hall (1978) solves a PIH model with these assumptions and shows that consumption process is a martingale, that is, changes in consumption are unpredictable over time.

anticipated changes in income.

To match the data better, numerous economic arguments have been advanced to explain the phenomena of excess sensitivity and excess smoothness. A partial listing of these explanations includes: time-varying interest rate (Christiano 1987, Hall 1988), the buffer stock model (Carroll 1997, Ludvigson and Michaelides 2001), the rule-of-thumb consumers (Campbell and Mankiw 1989, 1990), habit formation in consumption (Deaton 1992), incomplete information (Pischke 1995), and inattentiveness and cost of planning (Reis 2003).

In section 6 of Sims (2003), he solves a simple PIH model with i.i.d. income process analytically as an example to show that with RI consumption reacts to i.i.d. shocks gradually and with delay. In this paper, I will apply the RI idea in a full-fledged PIH model with some more realistic income processes including both the trend-stationary process and the difference-stationary process⁵ and explore the implications of RI for both individual and aggregate consumption and savings behavior. Moreover, I find that RI can help resolve the excess smoothness puzzle and the excess sensitivity puzzle. The analytical solution here has the following major advantages. First, it gives us clear economic insight about the nature of the RI model and it also greatly facilitates comparative statics analysis. Second, it can be used to calibrate the structure parameter governing the average degree of inattention in the model economy. Third, it is particularly useful to understand how introducing RI help us resolve the two consumption puzzle. Finally, it can be used to compare with other PIH models with friction (e.g., the standard internal additive habit-formation model, the Campbell-Mankiw's rule-of-thumb model, and the incomplete-information model).

The main contributions of this paper are as follows. First, both individual and aggregate consumption respond with delay and gradually to various income innovations and changes in consumption can be characterized by $MA(\infty)$ processes. Second, aggregate consumption is sensitive to lagged income shocks and is smooth relative to permanent income even if current income follows a difference-stationary process, that is, RI can be a potential explanation for the two twin consumption puzzles. Third, combining incomplete-aggregate-information proposed by Pischke (95) with RI can further improve model's prediction about aggregate consumption dynamics. Fourth, the utility costs due to limited channel capacity may be negligible under plausible assumptions. Hence, it would be reasonable to assume that individuals only use low channel capacity in processing information since increasing capacity only bring them very tiny welfare improvement. Finally, we compare RI with the habit formation (HF) hypothesis and the rule-of-thumb (ROT) consumer hypothesis and find that under some restrictions on the parameters RI and HF can generate similar aggregate consumption dynamics, even if they generate different individual consumption dynamics. In addition, I show that RI can help the ROT model better explain the observed consumption dynamics.

The main mechanism behind these results is that in the RI model measurement error in observing the state is unavoidable due to information-processing constraints. Hence, with

⁵Sims (2003) provides a numerical approach to solve the multivariate state PIH model because in that case consumers need to allocate their limited capacity efficiently across various states. Here I will show that in the multivariate state case what matters for optimal consumption decision may be just a linear combination of all elements in the state vector instead of the state vector itself, and thus the multivariate state case could be reduced to the univariate state case that can be solved explicitly.

limited Shannon capacity, consumers cannot respond instantly and without error to the changes in wealth. Consequently, the responses of consumption with respect to the shocks to wealth are smooth and delayed, and changes in consumption can be predicted by past income shocks. Furthermore, the responses to income shocks will rise gradually and eventually reach a higher flat asymptote as compared with the case without information processing constraints, because the initially undetected income shocks accumulate interest before consumption reacts fully to them. Finally, endogenous noises due to RI are totally idiosyncratic because they are generated from individuals' own channels, and thus they can be cancelled out after aggregating over all individuals.

Recently, there have been some papers that incorporate information frictions into a variety of theoretical models, and that explore how imperfect information acquiring and processing affects the optimal decision rules of consumers, firms, and investors, as well as its implications for equilibrium outcomes. For example, Woodford (2001), Ball et. al. (2003), and Adam (2004) analyze the effects of imperfect common knowledge on monetary policy and inflation dynamics; Peng and Xiong (2001) discusses how information capacity constraints affect the dynamics of asset return volatility; Moscarini (2004) derives optimal time-dependent adjustment rules from the information constraints; Luo and Young (2006) examines the effects of RI on the amplification and propagation of aggregate shocks; Maćkowiak and Wiedeholt (2005) investigates how the optimal allocation and the dynamics of prices depend on the firms' environment. A number of recent papers have also explored the potential of inattentiveness from another aspects. For example, Mankiw and Reis (2002) examines the effects of inattentiveness of firms on the dynamics of output and inflation; and Reis (2003) derives the optimal decision rules for inattentive consumers and then discussed the implications of inattentiveness for individual and aggregate consumption behaviors⁶.

This paper is organized as follows. Section 2 presents a LQG PIH model with RI and discusses some implications of RI for the dynamics of both individual and aggregate consumption. Section 3 examines the interaction between RI and aggregate incomplete information. Section 4 discusses welfare implications of RI. Section 5 compares the RI model with the habit formation model and the rule-of-thumb model. Section 6 concludes with a summary of main findings and some extensions.

2. The PIH Model with Rational Inattention

The model discussed in this paper is a simple LQG PIH model with RI. The model economy is populated by a continuum of consumers (measure 1), and all individuals have identical income processes while each agent faces different realizations of this process. Following the literature, I first consider optimal consumption and saving decisions of a typical consumer with information processing constraints and then aggregate over all individuals to obtain the

⁶'RI' modeled in Sims (2003) and this paper is based on Shannon channel capacity, while Reis models 'inattentiveness' by assuming and justifying the existence of decision costs that induces agents to only infrequently update their decisions. As shown below, although the two assumptions are based on distinct mechanisms, they can generate similar aggregate consumption dynamics.

dynamics of aggregate consumption. To save notations, I use the subscription “ a ” to indicate aggregate components and variables. Specifically, suppose that the consumer maximizes his utility function subject to both the usual flow budget constraint and the information processing constraints that will be specified later. The decision problem of the typical consumer can be characterized by the following optimization problem

$$\widehat{J}(\widehat{S}_t) = \max_{\{c_t, \mathcal{D}_t\}} E_0 \sum_{t=0}^{\infty} \beta^t u(C_t)$$

subject to:

$$W_t = R(W_{t-1} - C_{t-1}) + Y_t, \quad (2.1)$$

$$S_{t+1} | \mathcal{I}_{t+1} \sim \mathcal{D}_{t+1} \quad (2.2)$$

$$S_t | \mathcal{I}_t \sim \mathcal{D}_t \quad (2.3)$$

$$\text{Given } S_0 | \mathcal{I}_0 \sim N(\widehat{S}_0, \Sigma_0). \quad (2.4)$$

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} (They will be precisely specified subsequently.), be less than channel capacity, where S_t is the actual state variable that can be W_t if income process is i.i.d., $[W_t, Y_t]$ for more general income process, or a linear combination of W_t and Y_t , \widehat{S}_t is the *perceived* state variable,⁷ \mathcal{I}_t is the information available at time t , $u(C_t) = C_t - \frac{1}{2}C_t^2$ is the utility function, (2.1) is the flow budget constraint, C_t is individual consumption, W_t is individual wealth, β is the discount factor, R is the constant gross interest rate at which the consumer can borrow and lend freely (following the literature, assume that $\beta R = 1$), Y_t is individual income process that consists of several components (this process will be specified later), and the innovations to income are Gaussian. The expectation is formed under the assumption that $\{C_t\}_0^\infty$ are chosen under the information processing constraints. $\widehat{J}(\widehat{S}_t)$ is the value function for the individual with RI, and $J(S_t)$ is the value function derived from the standard full information PIH model.⁸ Finally, define the loss function at t due to information processing constraints as the difference between these two value functions, that is, $\Delta J = J(S_t) - \widehat{J}(\widehat{S}_t)$.⁹

2.1. Information Processing Constraints

The consumer faces the information processing constraints in the sense that he only uses finite channel capacity in observing the relevant state that includes both individual and aggregate components. Following Sims (2003), I also use the concept of *entropy* from information theory to characterize the rate of information flow and then use the reduction in entropy as a measure for information. Here entropy is defined as a measure of the uncertainty about a random variable.¹⁰ With the finite capacity, the consumer will choose a signal that reduces

⁷It is also called “estimated state” or “information state” in the literature.

⁸“Full information” here means that individuals are assumed to have unlimited channel capacity and thus can observe the state perfectly.

⁹See Appendix A for a detailed derivation for these two value functions.

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

the uncertainty of the state. Formally, this idea can be described by the following information processing constraint

$$\mathcal{H}(S_{t+1}|\mathcal{I}_t) - \mathcal{H}(S_{t+1}|\mathcal{I}_{t+1}) \leq \kappa \quad (2.5)$$

where κ is the consumer's channel capacity¹¹ that imposes an upper bound on the amount of information that can be transmitted via the channel,¹² $\mathcal{H}(S_{t+1}|\mathcal{I}_t)$ 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.

So far I haven't shown that \mathcal{D}_t is a normal distribution. I propose the following procedure to deduce its property. First, conjecture that the loss function the agent used to deduce the distribution of the actual state S_t is quadratic in terms of $S_t - \hat{S}_t$. Hence, minimizing this guessed loss function subject to information constraints implies that $S_t|\mathcal{I}_t \sim N(\hat{S}_t, \Sigma_t)$.¹³ Second, using this key property, it is easy to derive the optimal consumption function and the value function with RI. Third, given the derived value functions in both the full-information case and the RI case, the loss function at t , $\Delta J = J(S_t) - \hat{J}(\hat{S}_t)$, is indeed a quadratic function in terms of $S_t - \hat{S}_t$. Hence, we can now verify the conjecture that \mathcal{D}_t is a normal distribution $N(\hat{S}_t, \Sigma_t)$.¹⁴ The constraint, (2.5), can thus be written as

$$\log |\Psi_t| - \log |\Sigma_{t+1}| \leq 2\kappa \quad (2.6)$$

where $\Sigma_{t+1} = \text{var}_{t+1}(S_{t+1})$ and $\Psi_t = \text{var}_t(S_{t+1})$ are the posterior and prior variance-covariance matrices of the state vector, respectively. Note 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. Sims (2003) shows that in the one-dimensional state case (e.g., income is i.i.d.), this information constraint completes the characterization of the optimization problem with RI and the model can be solved analytically.¹⁵ I will show later that we can reduce the multivariate state PIH case¹⁶ to the univariate state case and thus solve it explicitly. Hence, in this paper, (2.6) is the unique information processing constraint. Note that with a finite capacity κ the optimizing consumer will choose a signal that reduces the conditional variance of S_{t+1} by a maximum (limited) amount. Hence, in this case information imperfection emerges endogenously since

¹¹If 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.

¹²We can regard it as a simple technology parameter. It can be modeled exogeneously or endogeneously. Note that information channels only limit the overall amount of information flowing to agents, while the agents might also need to allocate the capacity efficiently across different elements in the state vector, that is, decide which element of the state vector to observe with what precision.

¹³See Appendix B for detailed derivations.

¹⁴We can treat the above problem as a two-step optimization procedure. First, given the observed signal about the state, we could derive the optimal decision rule in the case with imperfect information. Second, given the optimal decision rule, the individual minimizes the expected loss function by choosing the information system, that is, the property of the signal.

¹⁵However, for the multivariate state case we need another information constraint, $\Psi_t \succeq \Sigma_{t+1}$, to complete the characterization. This constraint is used to rule out the possibility that $\Psi_t - \Sigma_{t+1}$ might not be positive semi-definite because information flow cannot be kept low by forgetting some existing information, and trading this off for increased precision about other elements in the state vector.

¹⁶In this case, income process is assumed to be trend-stationary or difference-stationary.

consumers are assumed to choose the nature of signal and then receive information about the state through the information channel that is contaminated with the endogenous noise. As a result, this idiosyncratic noise could generate private information about the common shock (e.g., the business-cycle shocks) hitting the economy.

It is straightforward to show that in the univariate case the equation (2.6) has a steady state Σ . In steady state the consumer is assumed to behave as if observing a noisy measurement which is $S_{t+1}^* = S_{t+1} + \xi_{t+1}$,¹⁷ where ξ_{t+1} is the endogenous noise and its variance $\Lambda_t = \text{var}_t(\xi_{t+1})$ is determined by the usual updating formula of the variance of a Gaussian distribution based on a linear observation:

$$\Sigma_{t+1} = \Psi_t - \Psi_t(\Psi_t + \Lambda_t)^{-1}\Psi_t. \quad (2.7)$$

Note that in steady state $\Sigma = \Psi - \Psi(\Psi + \Lambda)^{-1}\Psi$, which can be solved as $\Lambda = (\Sigma^{-1} - \Psi^{-1})^{-1}$.

We can also write the updating recursion of the conditional mean \hat{S}_t , that is, the Kalman filter equation, as

$$\hat{S}_t = F(\hat{S}_{t-1}, S_t + \xi_t) \quad (2.8)$$

where the linear function F is determined by the nature of income process and the definition of the relevant state S_t , as well as the value of Λ and Σ . Once we specify them, we can easily obtain this function form.

The fact that optimal control does not affect state estimation procedures under LQ assumptions and that optimal controls are certainty equivalent versions of optimal deterministic controls under LQG assumptions is referred to as the separation principle. This means that optimization of state estimation and control can be decoupled under these assumptions. See Whittle (1982, 1996) for detailed discussions. Hence, in our RI model once we derive the full-information linear optimal consumption rule $C_t = G(S_t)$ and the optimal state estimation \hat{S}_t from (2.7) and (2.8), we can obtain the optimal consumption rule in the RI model: $C_t = G(\hat{S}_t)$. This optimal consumption rule, the flow budget constraint (2.1), and the Kalman filter equation (2.8) constitute a dynamic system that can characterize our RI economy completely.

2.2. Consumption Dynamics

In this subsection, I first derive a closed-form expression of the changes in individual consumption in terms of income innovations and noises. Given the analytical solutions, we can aggregate consumption over all individuals and use the stylized aggregate facts to calibrate the key model parameter governing the average degree of inattention in the model economy. I then discuss how incorporating RI into the standard PIH model can help resolve the excess smoothness puzzle and the excess sensitivity puzzle. In addition, I will also discuss how RI affects saving behavior.

I close the model economy by specifying individual labor income process. Following Deaton (1992), Pischke (1995), and others, I suppose that the process for individual income can be de-

¹⁷As shown in Appendix B, optimal use of the channel implies that the channel output signal takes this additive form.

composed into a difference stationary aggregate component with innovations that are common to all consumers and a transitory idiosyncratic component with shocks that are uncorrelated across consumers.¹⁸ Formally,

$$Y_{t+1} = Y_{a,t+1} + Y_{i,t+1} \quad (2.9)$$

where

$$Y_{a,t+1} - Y_{a,t} = \rho(Y_{a,t} - Y_{a,t-1}) + \varepsilon_{t+1} \text{ and } Y_{i,t+1} - \bar{Y} = \eta_{t+1} \quad (2.10)$$

where $Y_{a,t}$ is common component, ρ is the AR(1) coefficient of income growth, and $Y_{i,t}$ is transitory idiosyncratic component. It is assumed that the consumer can separately identify the two components, so that consumption can respond to each of them. Permanent innovations to income are consolidated into the income base and it is possible that the consumer can recognize which is which. All innovations are assumed to have zero mean. Furthermore, suppose that the all shocks are uncorrelated over time and uncorrelated with each other and the variance of ε_t and η_t are ω^2 and σ^2 , respectively.

To investigate the implications of RI for consumption dynamics, it is useful to review the main predictions from the standard full-information PIH model. The following consumption function represents the PIH theory:

$$C_t = PI_t = \frac{R-1}{R} [W_t + \sum_{j=1}^{\infty} R^{-j} E_t(Y_{t+j})] \quad (2.11)$$

where PI_t is permanent income defined as the annuity value of the consumer's net worth including the present discounted value of expected future labor income as well as non-human wealth. Hence, consumption is determined by permanent income instead of current income. Furthermore, the first difference of the above consumption function can be written as

$$\Delta C_t = \Delta PI_t = (R-1) \sum_{j=1}^{\infty} R^{-j} (E_t - E_{t-1}) Y_{t+j-1}, \quad (2.12)$$

which is equal to $\frac{R}{R-\rho} \varepsilon_t + \eta_t$ if income process follows (2.9).

Hence, the change in consumption is equal to the change in permanent income and depends only on the revision in expectations of future labor income. In other words, under the PIH, consumption process is a martingale, and the change in consumption depends neither on the past history of labor income nor on anticipated changes in labor income. This is the well-known random walk result of Hall (1978). However, a number of empirical studies have found that the correlation between consumption growth and lagged income growth is significant. Thus, (2.12) does not fit the US aggregate data well. Further, (2.12) also implied that aggregate consumption growth should be more volatile than income growth if aggregate income growth has positive serial correlation ($\rho > 0$) or aggregate income process is random walk ($\rho = 0$). Hence, this model's prediction is at odds with the joint aggregate consumption and income

¹⁸Note that individual component could also include permanent shocks. See Carroll (1997), Ludvigson and Michaelides (2001), and Storesletten et. al (2001) for details. Here I ignore these shocks of individual income because including them does not change any main results in this paper.

process in the data, in which consumption growth is much less volatile than income growth. As argued in Deaton (1992), excess smoothness and excess sensitivity are not different, but two aspects of same phenomenon. If changes in consumption are orthogonal to lagged information, then they must be equal to changes in permanent income, and they cannot be too smooth.¹⁹ Although these empirical findings have led to a large number of potential explanations, here we offer RI as an alternative potential explanation for these phenomena.

In the presence of RI, it is assumed that consumers cannot observe the state(s) perfectly since ordinary people only have a limited channel capacity when processing relevant economic information.²⁰ Consequently, noise emerges endogenously and the nature of noise is determined by the model parameters. Another difficulty emerging from imperfect observations is that the *effective* state in this type of optimal control problems is not the traditional state variable, but the so-called information state which is defined as the distribution of the state variable conditional on the information set available at time t , \mathcal{I}_t . In other words, it makes an expansion of state space to the space of distributions on the state vector S_t . This means a substantial increase in dimensionality when we use dynamic programming to characterize this optimization problem. Consequently, it makes the model with information processing constraints very difficult to solve, which is called “the curse of dimensionality” in the literature.²¹ Fortunately, the Linear-Quadratic-Gaussian assumptions simplify the dimensionality problem greatly because the conditional distribution of the state vector is Gaussian, that is., the first two moments, the conditional mean \hat{S}_t and the conditional covariance matrix Σ_t , are enough to characterize the effective state. Actually, in some cases the problem could be even simpler in that Σ_t evolves over time by the determinant rules that are independent of optimal policy or Σ_t converges to a constant matrix in the steady state.

2.2.1. Individual Consumption Behavior

For simplicity, here I assume that $\rho = 0$, that is, income process follows a random walk.²² Substituting the specified income process, (2.9), into (2.11) gives

$$C_t = \frac{R-1}{R} \left(W_t + \frac{1}{R-1} Y_{a,t} \right) + \frac{1}{R} \bar{Y} \quad (2.13)$$

Note that in this PIH model, there are two state variables: individual wealth level (W_t) and permanent component of income ($Y_{a,t}$). Sims (2003) proposes a numerical procedure to solve the multivariate state PIH model with RI, here I show that there exist an explicit solution

¹⁹See Deaton (1992) for a recent review on these issues.

²⁰Note that observing the state perfectly requires an unlimited rate of information transferred, which is at odds with reality since ordinary people only have finite channel capacity.

²¹Another set of models which also have the problem of “the curse of dimensionality” is the heterogeneous-agent model with both idiosyncratic shock and aggregate shock. In those models, the measure becomes a state variable and thus the model have infinite states. Krusell and Smith (1998) solved this problem by approximating the measure by using a small number of constituent moments and these moments can be used to predict future prices. And it is shown in their papers that using aggregate (average) capital stock as an approximation for the distribution is enough to solve the models numerically with very small errors.

²²I will solve the case with $\rho > 0$ when I calibrate the key model parameter. Here I just use this simple case to explore the implications of RI for individual and aggregate consumption behavior.

for this PIH model with RI given the ARMA(1,1) income process. (2.13) means that a linear combination of two individual state variables, W_t and $Y_{a,t}$, determines optimal consumption level. Hence, we may define this linear combination as a new state variable,

$$S_t = W_t + \frac{1}{R-1}Y_{a,t}, \quad (2.14)$$

and the consumption function becomes

$$C_t = \frac{R-1}{R}S_t + \frac{1}{R}\bar{Y} \quad (2.15)$$

and it is straightforward to demonstrate that the dynamics of this new state is consistent with the original flow budget constraint because

$$W_{t+1} + \frac{1}{R-1}Y_{a,t+1} = R[(W_t + \frac{1}{R-1}Y_{a,t}) - C_t] + \frac{R}{R-1}\varepsilon_{t+1} + \eta_{t+1} + \bar{Y},$$

which can be rewritten in terms of S_t :

$$S_{t+1} = R(S_t - C_t) + \zeta_{t+1} + \bar{Y} \quad (2.16)$$

where $\zeta_{t+1} = \frac{R}{R-1}\varepsilon_{t+1} + \eta_{t+1}$.

Introducing RI in this model yields the following consumption function since certainty equivalence holds in the LQG model

$$C_t = \frac{R-1}{R}\hat{S}_t + \frac{1}{R}\bar{Y} \quad (2.17)$$

where $\hat{S}_t = E(S_t|\mathcal{I}_t)$.

Given the model specification, the conditional distribution of S_t given information at time t , \mathcal{I}_t , is $N(\hat{S}_t, \sigma_{\hat{S}_t}^2)$. Thus, the new flow budget constraint implies that

$$E_t(S_{t+1}) = \hat{S}_t \text{ and } \text{var}_t(S_{t+1}) = R^2\sigma_{\hat{S}_t}^2 + (\frac{R}{R-1})^2\omega^2 + \sigma^2$$

Now we can characterize the nature of the endogenous noise and then the evolution of the effective state \hat{S}_t as follows. First, the first information constraint (2.6) implies

$$\kappa = \frac{1}{2}[\log((\frac{R}{R-1})^2\omega^2 + \sigma^2 + R^2\sigma_{\hat{S}_t}^2) - \log(\sigma_{\hat{S}_{t+1}}^2)]$$

and it has a steady state $\bar{\sigma}_S^2 = \frac{(\frac{R}{R-1})^2\omega^2 + \sigma^2}{\exp(2\kappa) - R^2}$. The consumer behaves as if observing a noisy measurement $S_{t+1}^* = S_{t+1} + \xi_{t+1}$ in steady state and the noise ξ_{t+1} is independent of all

previous shocks and noise with²³

$$\text{var}(\xi_{t+1}) = \frac{[(\frac{R}{R-1})^2\omega^2 + \sigma^2 + R^2\bar{\sigma}_S^2]\bar{\sigma}_S^2}{(\frac{R}{R-1})^2\omega^2 + \sigma^2 + (R^2 - 1)\bar{\sigma}_S^2}$$

Finally, we derive the Kalman filter equation that describes the evolution of the effective state \widehat{S}_t :

$$\widehat{S}_{t+1} = (1 - \theta)\widehat{S}_t + \theta(S_{t+1} + \xi_{t+1}) \quad (2.18)$$

where $\theta = \bar{\sigma}_S^2 / \text{var}(\xi_{t+1}) = 1 - 1/\exp(2\kappa)$ is the optimal weight on observation. Note that when $\theta = 1$ (i.e., channel capacity $\kappa = +\infty$), the RI model reduces to the standard full information PIH model.

Equation (2.9), (2.16), (2.17), and (2.18) form a dynamic system of the model economy. Based on these equations, we can easily obtain the impulse response functions of consumption. Figure 1 below shows the responses of consumption with respect to two components in income and one error shock, with channel capacity $\kappa = 1$ bit. The two horizontal lines accompanied with the the responses to both permanent and transitory shocks represent the levels of the flat responses of consumption in the absence of information capacity constraints. It is obvious that the impulse responses of consumption to two income shocks exhibit delay and gradually reach a flat asymptote. As pointed out in Sims (2003), this flat asymptote is above the horizontal line representing the responses in the PIH model without RI because consumption does not react fully to income shocks initially and then the share of undetected income shocks goes to savings and accumulates interest before the consumer digests the shocks fully. The property that consumption responds gradually and with delay to income shocks is an important potential for explaining not only individual and aggregate consumption behavior but also other observed business cycle and asset returns phenomena. For example, many monetary DSGE models imply that both real spending and inflation jump immediately in response to shocks, in contradiction to a host of empirical evidence showing that both price and real variables exhibit gradual and hump-shaped responses to real and monetary shocks. Hence, once RI mechanism can be introduced into these models in a reasonable way, the model will fit the data better.²⁴

²³We derive the variance of ξ_{t+1} as follows: since

$$\begin{bmatrix} S_{t+1} \\ S_{t+1}^* \end{bmatrix} \sim N \left(\begin{bmatrix} \widehat{S}_t \\ \widehat{S}_t \end{bmatrix}, \begin{bmatrix} \Omega & \Omega \\ \Omega & \Omega + \text{var}(\xi_{t+1}) \end{bmatrix} \right)$$

where $\Omega = (\frac{R}{R-1})^2\omega^2 + \sigma^2 + R^2\bar{\sigma}_S^2$, we have

$$\begin{aligned} \bar{\sigma}_S^2 &= \text{var}(S_{t+1}|S_{t+1}^*) \\ &= \Omega - \frac{\Omega^2}{\Omega + \text{var}(\xi_{t+1})}, \end{aligned}$$

from which we can easily solve for the expression of $\text{var}(\xi_{t+1})$.

²⁴In Luo and Young (2006), RI is introduced into an otherwise standard stochastic growth and it is shown that RI can be an additional propagation mechanism for aggregate shocks in the sense that main real macroeconomic variables exhibit gradual and hump-shaped responses to aggregate shocks and the autocorrelation function of output growth is positive in the first two periods.

[Insert Figure 1 about here]

Alternatively, to gain some further insights about the effects of RI on the dynamics of consumption and savings and to make the model testable, it is useful to derive an analytical expression of the changes in consumption or savings in terms of income shocks. Fortunately, given our model specifications, we do have such an analytical expression. Combining equation (2.17) and (2.18) yields the following consumption evolution equation

$$C_t = (1 - \theta)C_{t-1} + [1 - (1 - \theta)](H_1S_t + H_0) + \theta H_1 \xi_t \quad (2.19)$$

where $H_1 = \frac{R-1}{R}$, $H_0 = \frac{1}{R}\bar{Y}$, and $H_1S_t + H_0$ is the permanent income. We can read from the above equation that current consumption can be expressed as a weighted average of past consumption and permanent income; the more the inattention, the more weight will be assigned to past consumption.

Substituting the specified income process, (2.9) when $\rho = 0$, into expression (2.12) yields

$$\Delta C_t = H_1 \zeta_t, \quad (2.20)$$

which means that the change in consumption is not sensitive to past income shocks and jumps immediately in response to income shocks. By contrast, the change in consumption in the model with RI can be written as

$$\Delta C_t = \theta H_1 \frac{\zeta_t}{1 - (1 - \theta)R \cdot L} + \theta H_1 \left[\xi_t + \frac{-\theta R \xi_{t-1}}{1 - (1 - \theta)R \cdot L} \right] \quad (2.21)$$

where L is the lag operator and we use the formula²⁵

$$S_t - \widehat{S}_t = \frac{(1 - \theta)\zeta_t - \theta\xi_t}{1 - (1 - \theta)R \cdot L}. \quad (2.22)$$

Note that given the expression of θ and $\bar{\sigma}_S^2 > 0$, i.e., $\exp(2\kappa) - R^2 > 0$, it is straightforward to prove that $(1 - \theta)R < 1$. This expression implies that the changes in consumption respond to all current and lagged income shocks, and it is also consistent with what we see from Figure 1. We define $I(i)$ as the coefficients attached to all income shocks ζ_{t-i} and $J(i)$ as the coefficients attached to endogenous noise ξ_{t-i} , and have

$$I(i) = \theta H_1 [(1 - \theta)R]^i \text{ if } i \geq 0; J(i) = \begin{cases} \theta H_1 & \text{if } i = 0 \\ -\theta^2 (R - 1) [(1 - \theta)R]^{i-1} & \text{if } i \geq 1 \end{cases} \quad (2.23)$$

Equation (2.21) shows that the growth of consumption is an MA(∞) process with decreasing coefficients and it implies that consumption adjust slowly and gradually to income shocks, with reactions that build up over time. Note that when $\theta = 1$, i.e., $\kappa = \infty$, the above expression reduces to (2.20). In other words, this expression of the change in individual consumption implies that individual consumption should be sensitive to both current and past income

²⁵This expression can be obtained by combining (2.16), (2.17), with (2.18).

shocks, as well as endogenous noise ξ_t . Parker (1999) provides the evidence that consumption responds to past news on after-tax income and examines the impact of Social Security tax withholding. The news in his paper is unpredictable and consumption reacts to the news with delay.

Given the optimal consumption functions with and without RI, (2.17) and (2.15), denote the RI saving premium of the consumer as

$$\begin{aligned} SP_t &= \frac{R-1}{R}(S_t - \widehat{S}_t) \\ &= \frac{R-1}{R} \left[\frac{(1-\theta)\zeta_t}{1-(1-\theta)R \cdot L} - \frac{\theta\xi_t}{1-(1-\theta)R \cdot L} \right]. \end{aligned} \tag{2.24}$$

This makes it clear that introducing RI generates a stochastic saving premium that consists of two MA(∞) processes. For example, when positive income shocks (ζ_t) arrives at the beginning of every period, the consumer cannot adjust his optimal consumption plan with respect to the innovations instantly and completely due to information constraint, and consequently, the share of undetected income shocks becomes saving premium in that period. (2.24) also confirms the intuition that the RI saving premium is larger, for (i) a higher interest rate (R), and (ii) a higher degree of inattention (lower θ), *ceteris paribus*. Furthermore, positive endogenous noises tend to reduce the RI saving premium. It is also obvious that lagged income shocks and noises still have impacts on the current RI saving premium. Finally, it is worth noting that the mean RI saving premium is zero because $E(SP_t) = 0$.

2.2.2. Implications for Excess Sensitivity and Excess Smoothness of Aggregate Consumption

The change in aggregate consumption is found by summing over all individuals. Allowing for heterogenous degrees of inattention (θ) does not affect the main results in this paper. For example, suppose that the degree of inattention follows a uniform distribution or a normal distribution with the same mean of θ , we can obtain almost the same MA expression for aggregate consumption as the one based on the assumption that every individual has identical degree of inattention. In other words, inattention dispersion among individuals does not affect the responses of aggregate consumption with respect to lagged aggregate income shocks.²⁶ Hence, for simplicity here we suppose that every consumer has the same degree of inattention. Because the idiosyncratic components of income process are mutually uncorrelated they will sum to zero in a large population. In addition, the endogenous noises due to information processing constraint may be idiosyncratic in that they arise from individuals' own internal information processing channel. Thus, they can also be cancelled out after aggregating over all consumers. Denote aggregate consumption as $C_{a,t}$, we can write the expression for the change in aggregate consumption as²⁷

²⁶The derivation of the changes in aggregate consumption in the case with heterogeneous degree of inattention is available from the author by request.

²⁷We obtain the following expression by summing (2.21) over all individuals with identical degree of RI.

$$\Delta C_{a,t} = \theta \frac{\varepsilon_t}{1 - (1 - \theta)R \cdot L}. \quad (2.25)$$

Expression (2.25) has a number of interesting implications. First, aggregate consumption reacts to aggregate income shocks gradually and with delay. This model prediction can match the empirical evidence because the data on aggregate consumption expenditure exhibit a significant delay and hump-shaped response to various income shocks.²⁸ Reis (2003) also provides a simple analysis of the response of aggregate consumption by estimating a structure vector autoregression (VAR) on aggregate consumption and income growth. Figure 2 reported in his paper shows that aggregate consumption has a delayed adjustment to the shocks. In addition, that figure also shows that the adjustment is not very delayed: most of the adjustment is finished with one year, which is consistent with Figure 1 above, in which the impulse response of consumption with respect to aggregate income shocks reaches the flat asymptote after 4-5 quarters.

Second, unlike in the standard full-information PIH model, here aggregate consumption is not a random walk and the change in aggregate consumption reacts to lagged shocks to aggregate income. We can also rewrite (2.25) as

$$\Delta C_{a,t} = (1 - \theta)R\Delta C_{a,t-1} + \theta\varepsilon_t, \quad (2.26)$$

which means that aggregate consumption now follows an AR(1) in first differences. Hence, introducing RI can help resolve the excess sensitivity puzzle: aggregate consumption growth is excessively sensitive to lagged aggregate income growth.

Third, note that in the standard full-information PIH model, $\Delta C_{a,t} = \varepsilon_t$. Taking the standard deviation of each side yields

$$\lambda = \frac{sd(\Delta C_{a,t})}{sd(\varepsilon_t)} = 1,$$

where λ is defined as the excess smoothness ratio that is the ratio of the standard deviation of aggregate consumption growth to the standard deviation of *innovations to aggregate income* and is equal to 1 in the standard PIH model. Note that it is just the consumption response predicted by the standard full-information model. Deaton (1992) finds that the empirical counterpart of this variance ratio is actually too low based on an estimated difference stationary income process. Thus consumption growth exhibits excess smoothness. By contrast, in the presence of RI, (2.25) implies that

$$\lambda = \frac{\theta}{\sqrt{1 - ((1 - \theta)R)^2}}, \quad (2.27)$$

which implies that introducing RI can lead to excess smoothness of consumption at the aggre-

²⁸See Fuhrer (2001) for a discussion.

gate level since

$$\frac{\partial \lambda}{\partial \theta} \simeq \frac{\theta}{[1 - ((1 - \theta)R)^2]^{3/2}} > 0.$$

Furthermore, following the consumption literature, we may compare our model with other PIH models by looking at their different predictions for the shape of the normalized power spectrum of consumption growth equation (2.25). Following Gali (1991), Deaton (1992) and Reis (2003), we may also define the excess smoothness ratio as $\mu = 1/\sqrt{2\pi f_{\Delta C}(0)}$ that is just the ratio of standard derivation of the change in consumption to the standard deviation of the change in *permanent income*. Hence, the standard PIH model predicts that $\mu = 1$, while $\mu < 1$ implies that consumption growth is excessively smooth relative to income growth. Note the formula for the power spectrum is

$$h_{\Delta C}(\varpi) = \frac{1}{2\pi} [\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos(\varpi k)] \quad (2.28)$$

Given the equation of the changes in consumption (2.25), the corresponding autocovariance function is

$$\gamma_k = \sum_{i=0}^{\infty} [\omega^2 I(k+i)I(i)] \quad (2.29)$$

where $I(i) = \theta((1 - \theta)R)^i > 0$ for $i = 0, 1, 2, \dots$. Substituting the expression of γ_k in (2.28) and normalizing it by γ_0 yields

$$f_{\Delta C}(\varpi) = \frac{1}{2\pi} [1 + 2 \frac{\sum_{k=1}^{\infty} \sum_{i=0}^{\infty} [\omega^2 I(k+i)I(i)] \cos(\varpi k)}{\sum_{i=0}^{\infty} [\omega^2 I(i)^2]}]. \quad (2.30)$$

Evaluating it at $\varpi = 0$ yields

$$\mu = \sqrt{\frac{\sum_{i=0}^{\infty} I(i)^2}{[\sum_{i=0}^{\infty} I(i)]^2}} = \sqrt{\frac{1 - (1 - \theta)R}{1 + (1 - \theta)R}}, \quad (2.31)$$

which also implies that introducing RI can reduce the excess smoothness ratio and thus help resolve the excess smoothness puzzle since

$$\frac{\partial \mu}{\partial \theta} > 0.$$

Figure 2 below shows the relationship between the channel capacity κ and the excess smoothness ratio μ , and it is obvious that the excess smoothness increases with information channel capacity.

[Insert Figure 2 about here]

It is also easy to see that summing the saving premium, (2.24), over all consumers yields

the aggregate saving premium

$$SP_{a,t} = \frac{R-1}{R} \frac{(1-\theta)\zeta_t}{1-(1-\theta)R \cdot L} \quad (2.32)$$

which is also an MA(∞) in terms of aggregate income shocks and have zero mean. Note that when θ goes to 1, $SP_{a,t}$ converges to 0 as predicted by the full information case. (2.24) implies that positive aggregate shocks to income can have positive impacts on aggregate savings premium and the effects last for many periods. The intuition is that every consumer in the model economy face the same information processing constraints and adjust their optimal consumption plan with delay, hence the undetected amount of aggregate shocks goes to aggregate saving and thus increases the aggregate saving premium $SP_{a,t}$. Of course, negative aggregate shocks have opposite effects on aggregate saving premium.²⁹

Finally, it is worth emphasizing that the optimal aggregate consumption calculated in the presence of RI differs substantially from the predictions of a representative agent (RA) model. The reason is that when there is one agent in the economy, the idiosyncratic noises due to RI cannot sum to zero and these terms will affect the smoothness of aggregate consumption significantly. Specifically, since equation (2.21) still holds for the RA, it is straightforward to calculate the excess smoothness of consumption in this RA model as

$$\lambda_{RA} = \frac{sd(\Delta C_{RA,t})}{sd(\varepsilon_t)} = \sqrt{\frac{\theta^2}{1 - ((1-\theta)R)^2} + \left(1 + \frac{(\theta R)^2}{1 - ((1-\theta)R)^2}\right) \frac{\theta}{1/(1-\theta) - R^2}} \quad (2.33)$$

where there is no idiosyncratic income shocks (η_t), aggregate income follows a random walk, $\Delta Y_t = \varepsilon_t$, and following the same procedure above the steady state conditional variance of S_t ($\bar{\sigma}_S^2$) and the variance of noise ($\text{var}(\xi_t)$) are defined as

$$\bar{\sigma}_S^2 = \frac{\left(\frac{R}{R-1}\right)^2 \omega^2}{\exp(2\kappa) - R^2} \quad \text{and} \quad \text{var}(\xi_t) = \frac{\left[\left(\frac{R}{R-1}\right)^2 \omega^2 + R^2 \bar{\sigma}_S^2\right] \bar{\sigma}_S^2}{\left(\frac{R}{R-1}\right)^2 \omega^2 + (R^2 - 1) \bar{\sigma}_S^2}.$$

Note that when R is close to 1, (2.33) implies that

$$\lambda_{RA} \simeq \sqrt{\frac{\theta(2-\theta)}{1-(1-\theta)^2}} \geq \sqrt{\frac{\theta^2}{1-(1-\theta)^2}} \simeq \lambda \quad (2.34)$$

since $\theta \leq 1$. This inequality implies that aggregate consumption generated from the RA model is more volatile than that from our benchmark model in which there are a continuum of agents. The reason is that in the benchmark model the induced noises will sum to zero after aggregating while in the RI setup these noises are left. We can also see this feature from Figure 3 below. For any given level of attention, the RA model generate higher degree of smoothness in aggregate consumption.

[Insert Figure 3 about here]

²⁹Note that the mean level of $SP_{a,t}$ is zero because $E(SP_{a,t}) = 0$.

2.3. Aggregate Stylized Facts, the Model Predictions, and Calibration

In this subsection, I will first report some stylized facts about income and consumption processes in aggregate data. This has two purposes. First, we can establish some simple time series aggregate income process. Combined it with the findings of the previous subsection allows us to obtain predictions from the RI model. Second, we will also report the results on aggregate consumption and use them to calibrate the main structure parameter in the RI model.

The following table is borrowed from Pischke (Table V and VI, Pischke (1995)). The income series refers to labor income and consumption expenditure includes nondurables and services. As in Deaton (1992), Pischke (1995) also estimated AR(1) income process by OLS and found that this specification fits data well. In the table, it is clear that the regression coefficient of consumption changes on lagged income changes is around 0.11 and significant. In other words, the standard PIH model can not predict this feature in the data since consumption is martingale and does not depend on lagged income in the standard PIH model. Furthermore, the excess smoothness ratio in the table is only around 0.578 which is well below that predicted by the standard PIH model.

[Insert Table 1 about here]

We now report the predictions of the RI model and then use the reported stylized facts to calibrate the model structure parameter, κ . Following Pischke (1995), using the specification of income process ($\rho > 0$), (2.9) and (2.10), and following the same procedure as above, we can easily derive the closed-form expression of ΔC_t in the RI model.³⁰ The key step is defining a new state variable

$$S_t = W_t + \frac{R + (R - 1)\rho}{(R - 1)(R - \rho)} Y_{a,t} - \frac{\rho R}{(R - 1)(R - \rho)} Y_{a,t-1}.$$

and then the flow budget constraint can be written in terms of S_t :

$$S_{t+1} = R(S_t - C_t) + \zeta_{t+1} + \bar{Y}$$

where $\zeta_{t+1} = \frac{R^2}{(R-1)(R-\rho)} \varepsilon_{t+1} + \eta_{t+1}$. Hence, we can express the change in aggregate consumption as follows

$$\Delta C_{a,t} = \theta \frac{R}{R - \rho} \frac{\varepsilon_t}{1 - (1 - \theta)R \cdot L}, \quad (2.35)$$

³⁰Note that given this income process the full-information PIH model predicts that the change in aggregate consumption is

$$\Delta C_{a,t} = \frac{R}{R - \rho} \varepsilon_t > \varepsilon_t,$$

that is, the excess smoothness ratio should be greater than 1 and thus the implied consumption is too volatile relative to current income.

and this MA (∞) process can be rewritten as an AR(1) process

$$\Delta C_{a,t} = (1 - \theta)R\Delta C_{a,t-1} + \theta \frac{R}{R - \rho} \varepsilon_t. \quad (2.36)$$

It is easy to see that the aggregate consumption in this equation is not a random walk as the standard PIH model predicts. Consumption now follows an AR(1) in first differences. This intuition behind this is simple. Suppose that an aggregate shock hits the model economy, the consumers can not digest the current state completely due to limited channel capacity. As a result, they will change their consumption but not by as much as the shock calls for. Since the shock is persistent, in the next period their income is higher than expected and they will increase their consumption further and so on.

We can then use these aggregate facts reported in Table 1 to calibrate the average degree of inattention in the model economy (κ or θ). Specifically, two stylized facts (the regression coefficient of $\Delta C_{a,t}$ on $\Delta Y_{a,t-1}$ and the excess smoothness ratio of $\Delta C_{a,t}$) can be used to calibrate κ . I first set $\rho = 0.288$ in aggregate income process and $R = 1.01$ quarterly. Define the excess smoothness of aggregate consumption as the ratio of the standard deviation of consumption changes to the standard deviation of aggregate income innovations, we have

$$\lambda = \frac{sd(\Delta C_{a,t})}{sd(\varepsilon_t)} = \frac{R}{R - \rho} \frac{\theta}{\sqrt{1 - ((1 - \theta)R)^2}}, \quad (2.37)$$

and this ratio can be used to calibrate the value for κ . Specifically, table 1 reports that $\lambda = 0.578$, hence one standard deviation from the point estimate encompasses the range from $\lambda - 0.055$ to $\lambda + 0.055$ and thus the confidence interval of θ is around $[0.24, 0.34]$. Since $\theta = 1 - 1/\exp(2\kappa)$, κ is around $[0.2 \text{ bits}, 0.3 \text{ bits}]$. Similarly, two standard deviations from the point estimates encompasses the range from $\lambda - 2 \cdot 0.055$ to $\lambda + 2 \cdot 0.055$, the range of θ is around $[0.2, 0.36]$, that is, κ is around $[0.16 \text{ bits}, 0.32 \text{ bits}]$.

Furthermore, we may also use the facts about excess sensitivity of aggregate consumption to calibrate κ , that is, the coefficient of $\Delta C_{a,t}$ on $\Delta Y_{a,t-1}$. Suppose that the econometrician estimates the following equation:

$$\Delta C_{a,t} = \alpha + \beta \Delta Y_{a,t-1} + e_{t+1}. \quad (2.38)$$

If the data is generated by equation (2.38), the expected value of β is

$$\begin{aligned} \hat{\beta} &= \frac{\text{covar}(\Delta C_{a,t}, \Delta Y_{a,t-1})}{\text{var}(\Delta Y_{a,t-1})} \\ &= \frac{\text{covar}\left(\frac{R}{R-\rho} \frac{\theta \varepsilon_t}{1-(1-\theta)R \cdot L}, \frac{\varepsilon_{t-1}}{1-\rho \cdot L}\right)}{\omega^2/(1-\rho^2)} = (1-\rho^2) \frac{R^2}{R-\rho} \frac{\theta(1-\theta)}{1-(1-\theta)R\rho} \end{aligned} \quad (2.39)$$

and then we can use this expression to calibrate κ . Given the same values for R and ρ , based on table 1 one standard deviation from the point estimate of β encompasses the range from

$\widehat{\beta} - 0.045$ to $\widehat{\beta} + 0.045$, it is easy to show that the range of θ is around $[0.87, 0.95]$ or $[0.04, 0.1]$,³¹ that is, $\kappa \in [1.46 \text{ bits}, 2.15 \text{ bits}]$ or $[0.03 \text{ bits}, 0.1 \text{ bits}]$. Similarly, since two standard deviations from the point estimates encompasses the range $[\widehat{\beta} - 2 \cdot 0.045, \widehat{\beta} + 2 \cdot 0.045]$, the range of θ is around $[0.83, 0.98]$ or $[0.01, 0.13]$, which means $\kappa \in [1.27 \text{ bits}, 2.8 \text{ bits}]$ or $[0.007 \text{ bits}, 0.1 \text{ bits}]$. Unfortunately, we do not have prior knowledge about which range of θ or κ is the better one.

Note that in any confidence interval given above, there is no overlap between two ranges of θ or κ from matching excess smoothness and excess sensitivity, respectively. For example, when $\kappa = 0.25$ bits, RI can solve the excess smoothness puzzle by predicting λ correctly, but it cannot resolve the excess sensitivity puzzle completely (too low $\widehat{\beta}$) or generates consumption process that is too sensitive to income shocks (too high $\widehat{\beta}$). Furthermore, when $\kappa = 1.5$ bits, RI can resolve the excess sensitivity puzzle, but it cannot generate enough smoothness in consumption growth. In the next section, I will discuss how combining information and rational inattention can help better explain two puzzles simultaneously.

3. Incomplete Aggregate Information and Rational Inattention

Pischke (1995) and Ludvigson and Michaelides (2001) show that incomplete information about the aggregate component of individual income (henceforth, IAI) may be an important factor in explaining the twin consumption excesses in both the standard PIH model and the buffer stock model, respectively. However, as shown in Ludvigson and Michaelides (2001), even the buffer stock model with incomplete information creates a smoothness puzzle that remains to be explained.³² Although RI also gives rise to imperfect information due to information processing constraints, it is different from the specification of incomplete information in those papers. The idea of incomplete information in their papers is that individuals can only observe the individual component and cannot distinguish the aggregate and the individual component. As a result, the income process looks like an MA(1) process for income growth if aggregate component is a random walk and individual component is i.i.d.³³ However, the RI idea says that individuals cannot digest all information about the relevant state that includes both individual and aggregate components due to finite Shannon capacity. Therefore, the two ideas capture different aspects of information imperfection, and they can be either modeled separately or combined into one model. For example, an agent who can distinguish the individual and aggregate components might still face information processing constraints, an agent with unlimited channel capacity might not have incentive to distinguish the two components, or an agent both faces information processing constraint and lacks the incentive to distinguish the two components. In this section, I will combine these two ideas in the standard PIH model and show how the combination of RI and IAI can go a long way toward resolving the smoothness of aggregate consumption.

³¹This is because solving the quadratic equation for θ yields two values of θ .

³²They report that aggregate consumption growth in their buffer stock model is, at best, only about 80% as volatile as aggregate income growth, compared to around 50% in the US data.

³³Pischke (1995) rationalizes this assumption by showing that ignoring aggregate information in individual income is not very costly for most agents and thus rational agents do not have strong incentive to acquire aggregate information.

Specifically, suppose that agents face both unobservable aggregate shocks and information processing constraints. Given income process specified in (2.9), IAI means that the income process individuals observe can be written as an MA(1) process

$$\Delta Y_t = \nu_t - \phi \nu_{t-1} \quad (3.1)$$

where the innovation, ν_t , is not a fundamental driving process but it contains information on current and past aggregate and individual income shocks and the parameter, ϕ , is determined by the relative variances of the aggregate and individual income shocks.³⁴ Following the standard procedure, it is straightforward to show that the consumption function is

$$C_t = H_1(W_t + \frac{1}{R-1}Y_t - \frac{\phi}{R-1}\nu_t) + \frac{1}{R}\bar{Y} \quad (3.2)$$

and consumption growth is

$$\Delta C_t = \frac{R-\phi}{R}\nu_t = \frac{R-\phi}{R} \frac{\Delta Y_t}{1-\phi \cdot L}. \quad (3.3)$$

where the change in individual income $\Delta Y_t = \varepsilon_t + \eta_t - \eta_{t-1}$. To incorporate RI and solve the model explicitly, as above we need to define a new state variable that is just a linear combination of all elements in the state vector, that is, $S_t = W_t + \frac{1}{R-1}Y_t - \frac{\phi}{R-1}\nu_t$ and the consumption function becomes $C_t = H_1 S_t + \frac{1}{R}\bar{Y}$. Further, we can write the budget constraint in terms of the new state as

$$S_{t+1} = R(S_t - C_t) + \frac{R-\phi}{R-1}\nu_{t+1} + \bar{Y}. \quad (3.4)$$

In the context of LQG, certainty equivalence applies and introducing RI yields the following consumption function

$$C_t = H_1 \hat{S}_t + \frac{1}{R}\bar{Y} \quad (3.5)$$

where $\hat{S}_t = E[S_t | \mathcal{I}_t]$ and the conditional distribution of S_t given information at time t , \mathcal{I}_t is $N(\hat{S}_t, \sigma_{\hat{S}_t}^2)$. The following Kalman filter equation characterizes the evolution of \hat{S}_t

$$\hat{S}_{t+1} = (1-\theta)\hat{S}_t + \theta(S_{t+1} + \xi_{t+1}). \quad (3.6)$$

where θ and ξ_{t+1} have the same definitions as that in section 2. Combining (3.4), (3.5) with (3.6) gives the expression for the change in individual consumption as

$$\Delta C_t = \theta \frac{R-\phi}{R} \left[\frac{1}{1-(1-\theta)R \cdot L} \frac{\Delta Y_t}{1-\phi \cdot L} \right] + \theta H_1 \left[\xi_t + \frac{-\theta R \xi_{t-1}}{1-(1-\theta)R \cdot L} \right], \quad (3.7)$$

³⁴As shown in Pischke (1995), $\phi = -(1 - \sqrt{1 - 4\psi^2})/2\psi$ where $\psi = -\sigma^2/(\omega^2 + 2\sigma^2)$.

and aggregating over all individuals yields the change in aggregate consumption

$$\begin{aligned}\Delta C_{a,t} &= \theta \frac{R-\phi}{R} \left[\frac{1}{1-(1-\theta)R \cdot L} \frac{1}{1-\phi \cdot L} \varepsilon_t \right] \\ &= \theta \frac{R-\phi}{R} \sum_{j=0}^{\infty} \sum_{k=0, j \geq k}^{\infty} [\phi^{j-k} ((1-\theta)R)^k] \varepsilon_{t-j}\end{aligned}$$

where ε_t is aggregate income shock. Hence, the excess smoothness ratio can be written as

$$\lambda = \frac{sd(\Delta C_{a,t})}{sd(\varepsilon_t)} = \sqrt{\frac{(1+\phi(1-\theta)R)\theta^2}{(1-\phi(1-\theta)R)[(1+\phi(1-\theta)R)^2 - (\phi+(1-\theta)R)^2]}} \frac{R-\phi}{R} \quad (3.8)$$

Note that when $\theta = 1$ (no information processing constraints), (3.8) reduces to

$$\lambda = \sqrt{\frac{1}{1-\phi^2}} \frac{R-\phi}{R}, \quad (3.9)$$

which approaches to $\sqrt{\frac{1-\phi}{1+\phi}}$ when R is close to 1. Further, when $\phi = 0$ (observable aggregate shocks), the ratio is just (2.27) that approaches to $\sqrt{\frac{\theta}{2-\theta}}$ when R is close to 1. It is easy to see from (2.27) and (3.9) that both RI and IAI can produce aggregate consumption growth that is less volatile than the standard PIH model. Figure 4 shows that how combining them can generate larger smoothness of aggregate consumption, which makes the model fit the data better. Ludvigson and Michaelides (2001) report that aggregate consumption growth is only about 80% as volatile as aggregate income growth, while the empirical counterpart in the US data is about 50%. In the figure, we can see that when $\phi = 0.22$ and $\theta = 1$, the smoothness ratio is about 80%. When considering RI, for example, $\theta = 0.5$, this ratio will be around 50% that is close to the empirical one.

[Insert Figure 4 about here]

4. Welfare Implications of Information Processing Constraints

In this section, I examine the welfare effects of income fluctuations under RI. Specifically, I will calculate how much utility consumer will lose if the actual consumption path deviates from the first-best instantly adjusted consumption path due to RI. To examine the welfare effects of RI, we first need to establish a welfare criterion. As usual, the value functions are used to measure individuals' intertemporal welfare. In appendix A, I derive the value functions in both the full-information case and the RI case.

Since information processing constraints cannot help in individual's optimization, the average welfare difference between the two economies should be greater than 0. One way to evaluate the effect of RI on welfare is to compute the difference of the two value functions

around $\bar{S}_t = E_t(S_{t+1}) = E_t(\hat{S}_{t+1})$.³⁵

$$\Delta J = J(\bar{S}_t) - \hat{J}(\bar{S}_t) = \frac{1}{2} \left(\frac{R^2 - 1}{R} \right) \bar{\sigma}_S^2.$$

Note that welfare loss due to RI, ΔJ , converge to 0 when channel capacity κ increases to ∞ . As shown in the preceding section, the actual consumption path will deviate from the first-best immediately adjusted consumption path due to information processing constraints. Thus, we are interested in how much utility the consumer will lose in a RI economy. Specifically, we will measure the loss by computing the direct reduction of welfare in steady state or a money metric which is equivalent with the welfare loss.

First, following Cochrane (1989), Pischke (1995), and others, I use a money metric to measure welfare cost of deviating from the first-best full information rational expectation solution. As we just showed, we can easily derive the level of expected lifetime utility, $J(S_t)$ and $\hat{J}(\hat{S}_t)$, by following different decision rules, and then we get the expected welfare losses as

$$E(\Delta J) = E(J(S_t) - \hat{J}(\hat{S}_t))$$

To convert it to dollars per quarter, following Cochrane (1989) and Pischke (1995), we divide ΔV by the marginal utility of a dollar at time t and then convert it to quarterly rates by multiplying $\frac{R-1}{R}$,

$$\begin{aligned} \$ \text{ Loss/quarter} &= \frac{R-1}{R} \frac{E(\Delta J)}{u'(\bar{Y})} = \frac{R-1}{R} \frac{E(\Delta J)}{\bar{C} - \bar{Y}} \\ &= \frac{R-1}{R} \gamma \frac{\frac{1}{2} \left(\frac{R^2-1}{R} \right) \bar{\sigma}_S^2 + A_2(\text{var}(S_t) - \text{var}(\hat{S}_t))}{\bar{Y}} \end{aligned} \quad (4.1)$$

where \bar{C} is the bliss point of consumption (it is normalized to 1 in our setup), \bar{Y} is the mean income, γ is the local coefficient of relative risk aversion and equal to $\frac{\bar{Y}}{\bar{C} - \bar{Y}}$ for the utility function $u(\bar{Y})$. Using the quarterly data 1954-1990 provided in Pischke (pp. 830, 1995), set the parameters as follows: $\bar{Y} = \$6,929$, the standard deviation of individual income $\omega_\zeta = \$2,470$, $R = 1.01$, and $\gamma = 4$. Using the expression (4.1), table 2 reports utility costs for several values of the local CRRA γ and channel capacity κ .

[Insert Table 2 here]

It is clear from the table that the welfare losses due to RI are trivial. For example, for $\gamma = 4$ and $\kappa = 1$ bit, the loss only amounts to 12 cents per quarter. This result is similar to the findings by Pischke (1995) who calculated utility losses in the no-aggregate-information model.³⁶ Furthermore, even for a high value of the CRRA, $\gamma = 10$, and a low channel capacity, $\kappa = 0.3$ bits, the welfare loss is still minor: only around \$1.74 per quarter.

³⁵Since both S_{t+1} and \hat{S}_{t+1} follow random walk that do not have unconditional mean, we will use the conditional mean of S_{t+1} and \hat{S}_{t+1} at t around which welfare is evaluated.

³⁶He found that in most cases the utility losses due to no information about aggregate shock to income are less than \$1 per quarter.

We may also adopt another simple way to measure the welfare loss due to RI. Suppose that the welfare loss due to RI is equivalent with a reduction in steady state wealth, $(1 - \varphi)\bar{S}_t$, where φ is the maximum fraction of steady state wealth that the consumer would like to pay to avoid his welfare loss due to RI. Put it mathematically, in steady state, $\bar{S}_t = E_t(S_{t+1}) = E_t(\hat{S}_{t+1})$, we have

$$J((1 - \varphi)E_t(S_{t+1})) = \hat{J}(E_t(\hat{S}_{t+1}))$$

Solving this quadratic equation for φ yields

$$\varphi = \frac{2 + A_1/(A_2\bar{S}_t) + \sqrt{[2 + A_1/(A_2\bar{S}_t)]^2 - 4(A_0 - B_0)/(A_2^2\bar{S}_t)}}{2}$$

where we ignore the negative eigenvalue since it implies that RI generates positive welfare effect. We consider the following numerical example to evaluate welfare effect of RI. We assume that $R = 1.01, \omega_\zeta^2 = 0.001, \kappa = 1$, and $Y_{i,t}$ follows a uniform distribution over $[0, 1]$ with mean 0.5. Note that in this case the steady state permanent income \bar{S}_t should be less than $(1 - H_0)/H_1$.³⁷ Figures 5a and 5b show that the welfare loss in this case is tiny. For example, when $\bar{S}_t = 5, \kappa = 1$, and $\omega_\zeta^2 = 0.001$, welfare loss φ is only around $4 \cdot 10^{-6}$. The figures also show that the welfare loss due to RI increases with income volatility and decreases with channel capacity. The intuition is that since consumers are risk averse, they dislike uncertainty and risk. Hence, since both high income volatility (which has direct effect on welfare) and low capacity (which has indirect effect on welfare) induce high steady state conditional variance of the state, consumers are willing to pay more to avoid this induced uncertainty.

[Insert Figure 5a and 5b about here]

Based on the above calculations, we can conclude that for plausible assumptions, the welfare losses due to RI are negligible.³⁸ This provides some evidence that it is quite reasonable for consumers to devote low channel capacity in observing and processing information, as the welfare improvement from adopting the first-best decision rule is trivial. In other words, although consumers can devote much more capacity in processing economic information and improve their optimal consumption decisions, it is rational for them not to do so since the welfare improvement is so tiny. Consequently, this rational inertial behavior can generate interesting dynamics of consumption and savings and thus make the model fit the data better in some important aspects.

5. Comparison with other PIH Models

In this section, I compare the RI hypothesis with two alternative popular hypothesis (habit formation in consumption and rule-of-thumb consumption) within the PIH framework.

³⁷That is because in steady state $\bar{C}_t = H_1\bar{S}_t + H_0 \leq 1$ (1 is the bliss point).

³⁸This conclusion is also consistent with that in Cochrane (1989), Pischke (1995), and others.

5.1. Comparison with the PIH Model with Habit Formation

Habit formation in consumption is widely used in economics and finance to study consumption and wealth accumulation behavior, asset pricing, and optimal monetary policy. A short listing of these applications includes: Constantinides (1990), Deaton (1992), Fuhrer (2000), and Boldrin, Christiano, and Fisher (2001). Intuitively, the RI model should have some similarity with the habit-formation model since the habit formation models also imply that slow adjustment in consumption is optimal because consumers smooth both consumption level and consumption growth, while the RI model predicts that slow and delayed adjustment in consumption is optimal because capacity constraints make consumers take more time to observe and process information. Hence, although these two arguments have distinctive mechanisms for consumption adjustment, they might have similar impacts on the dynamics of consumption, savings, and wealth accumulation. One of the main differences between the two models is that habit formation could affect the dynamics of consumption and saving in both certain cases and uncertain cases, whereas RI affects the dynamics through its interaction with the fundamental uncertainty and it has no impact in the absence of uncertainty.

Specifically, consider a simple model with internal and additive habit formation,³⁹ in which we assume that higher consumption in last period creates a habit that lowers utility in this period. Under this assumption, a typical consumer with habit formation have the following preference:

$$J(S_0) = \max_{C_t} E_0 \sum_{t=0}^{\infty} \beta^t u(C_t - \gamma C_{t-1})$$

where $u(C_t - \gamma C_{t-1})$ is also a quadratic function in terms of $\tilde{C}_t = C_t - \gamma C_{t-1}$, that is, $u(\tilde{C}_t) = \tilde{C}_t - \frac{1}{2}\tilde{C}_t^2$. In this model, in making their consumption decisions, consumers take into account the effects of their current consumption on their future reference stock. Following Alessie and Lusardi (1997), we have the following recursive solution for consumption

$$C_t = \frac{\gamma}{R} C_{t-1} + (1 - \frac{\gamma}{R})(H_1 S_t + H_0) \quad (5.1)$$

where S_t is defined by (2.14) and is a linear combination of two state variables. individual consumption is also a weighted average of past consumption and permanent income; the stronger the habit, the more weight will be assigned to past consumption. Apparently, equation (2.19) and (5.1) deliver very similar dynamics of individual consumption in the PIH model except that in the RI case there is an endogenous i.i.d. noise term (ξ_t) with mean zero in the recursive equation. When θ (or κ) that governs the ability of information processing and γ that governs the degree of habit formation satisfy the following equality,

$$1 - \theta = \frac{\gamma}{R}, \quad (5.2)$$

the dynamic behavior of consumption in these two different models are a little similar in that

³⁹ “Internal” means that individuals care about how their consumption in the current period compares to their own consumption in the past.

current consumption depends on both lagged consumption and permanent income.

To further examine the implications of the two hypotheses for both individual and aggregate consumption dynamics, it would be helpful to express the changes in consumption in MA(∞) processes. In the above habit formation model, we can rewrite the change in consumption as

$$\Delta C_t = (1 - \frac{\gamma}{R})H_1 \frac{\zeta_t}{1 - \gamma \cdot L} \quad (5.3)$$

where $H_1 = \frac{R-1}{R}$, $\zeta_t = \frac{R}{R-1}\varepsilon_t + \eta_t$, and we use the fact that $S_t - S_{t-1} = \frac{-\frac{\gamma}{R}\zeta_{t-1} + \zeta_t}{1 - \gamma \cdot L}$. Hence, the standard deviation of individual consumption is

$$sd(\Delta C_t) = \sqrt{\frac{(1 - \frac{\gamma}{R})^2}{1 - \gamma^2}(\omega^2 + H_1^2\sigma^2)}$$

Similarly, aggregate consumption dynamics in this case can be obtained by aggregating over all individuals and the standard deviation of aggregate consumption can be written as⁴⁰

$$sd(\Delta C_{a,t}) = \frac{1 - \frac{\gamma}{R}}{\sqrt{1 - \gamma^2}}\omega$$

and the ratio of the standard deviation of aggregate consumption with respect to aggregate income, i.e., the excess smoothness ratio, is

$$\lambda_{HF} = \frac{sd(\Delta C_{a,t})}{sd(\varepsilon_t)} = \frac{1 - \frac{\gamma}{R}}{\sqrt{1 - \gamma^2}}. \quad (5.4)$$

which is equal to (2.27), the excess smoothness ratio with RI, when $1 - \theta = \frac{\gamma}{R}$.

By contrast, In the presence of RI, (2.21) implies that the standard deviation of individual consumption becomes

$$sd(\Delta C_t) = \sqrt{\frac{\theta^2}{1 - ((1 - \theta)R)^2}(\omega^2 + H_1^2\sigma^2) + (\theta H_1)^2(1 + \frac{(\theta R)^2}{1 - ((1 - \theta)R)^2})\text{var}(\xi_t)}$$

Similarly, the standard deviation of aggregate consumption can be written as

$$sd(\Delta C_{a,t}) = \frac{\theta}{\sqrt{1 - ((1 - \theta)R)^2}}\omega$$

and the excess smoothness ratio is defined in (2.27).

Hence, even when $1 - \theta = \frac{\gamma}{R}$, RI and HF imply different properties of individual and aggregate consumption. Specifically, individual consumption growth is more volatile in the presence of RI. Further, in the presence of HF, the ratios of the standard deviation of consumption to the standard deviation of income are the same for both individual level and aggregate level, whereas in the RI case the ratios are different due to the presence of the noise terms in individ-

⁴⁰Here idiosyncratic income shocks are cancelled out after aggregating over all individuals.

ual consumption. It is also worth emphasizing that when the uncertainty due to fundamental income shocks reduces, the properties of the dynamics of individual consumption in the two models become more and more similar given $1 - \theta = \frac{\gamma}{R}$.

As usual, we can rewrite the PIH model in terms of savings rather than consumption. Following Campbell (1987) and Deaton (1992), define the expression for savings D_t as⁴¹

$$D_t = \frac{R-1}{R}W_t + \frac{1}{R}Y_t - C_t = \frac{R-1}{R}S_t + \frac{1}{R}Y_{i,t} - C_t. \quad (5.5)$$

Hence, in the presence of RI, we have the following recursive saving function

$$D_t = (1 - \theta)RD_{t-1} + (1 - \theta)\Delta Y_t + \theta\frac{1}{R}(Y_{i,t} - \bar{Y}) - \frac{R-1}{R}\theta\xi_t \quad (5.6)$$

where $Y_{i,t} - \bar{Y} = \eta_t$ and $\Delta Y_t = \eta_t - \eta_{t-1} + \varepsilon_t$. Note that in the presence of RI saving depends on (i) past saving, (ii) a linear combination of current income change ΔY_t and the discounted present value of future income changes⁴², and (iii) noise terms due to RI. The weights in the linear combination depend on the degree of the consumer's inattention θ . The stronger the inattention, the less importance of future income changes and the more important of past savings.

By contrast, with habit formation the saving function can be characterized by the following recursive form:

$$D_t = \gamma D_{t-1} + \frac{\gamma}{R}\Delta Y_t + (1 - \frac{\gamma}{R})\frac{1}{R}(Y_{i,t} - \bar{Y}) \quad (5.7)$$

Hence, in the habit formation model, saving depends on past saving and on a linear combination of current transitory income change $\Delta Y_{i,t}$ and the discounted present value of future income changes. Again, when $1 - \theta = \frac{\gamma}{R}$, the two saving functions, (5.6) and (5.7), are similar too. This conclusion is not surprising since the same condition holds in terms of the consumption functions.

The intuition behind the equality (5.2) is that since the parameter γ in the habit formation model also governs the speed with which reference stock adjusts, the larger γ is, the higher dependence of current consumption on past consumption, while in the RI model, the more the inattentiveness, (i.e., less θ), the less attention the consumer will pay to “news” and thus wealth level in the current period and then depend larger on past effective wealth level which can be expressed by past consumption level. Thus, the effects of RI and habit formation on consumption dynamics and thus savings behaviors are very similar except that the recursive consumption and saving equations in the RI model are affected by endogenous noise term ξ_t .

Finally, both models also generate similar wealth accumulation dynamics. In the RI model,

⁴¹In Campbell (1987), he defines that saving $D_t = \frac{r}{1+r}A_t + Y_t - C_t$, where A_t is a single asset carried over from $t-1$. Here we just define $W_t = A_t + Y_t$ as state variable, hence we have the following expression for savings.

⁴²Note that in the i.i.d. case, this expression for saving is just $\frac{1}{R}(Y_{i,t} - \bar{Y})$.

combining (2.16) and (2.17) above gives the dynamics of permanent income S_t ,

$$S_{t+1} - S_t = \zeta_{t+1} + (R-1) \frac{(1-\theta)\zeta_t - \theta\xi_t}{1 - (1-\theta)R \cdot L} \quad (5.8)$$

where $\zeta_{t+1} = \frac{R}{R-1}\varepsilon_{t+1} + \eta_{t+1}$ is a linear combination of permanent and transitory shocks to income and ξ_t is the noise term. Since $S_t = W_t + \frac{1}{R-1}Y_{a,t}$ and $Y_{a,t}$ is a random walk, the above equation of permanent income can be easily transformed to an equation for wealth dynamics, that is,

$$W_{t+1} - W_t = (\varepsilon_{t+1} + \eta_{t+1}) + (R-1) \frac{(1-\theta)\zeta_t - \theta\xi_t}{1 - (1-\theta)R \cdot L}. \quad (5.9)$$

Hence, in the RI model, wealth also follows a random walk.

By contrast, in the above habit-formation model, combining the flow budget constraint (2.16) and (5.1) yields⁴³

$$W_{t+1} - W_t = (\varepsilon_{t+1} + \eta_{t+1}) + (R-1) \frac{\gamma}{R} \frac{\zeta_t}{1 - \gamma \cdot L}. \quad (5.10)$$

Hence, once $1 - \theta = \frac{\gamma}{R}$, individual wealth dynamics in both models are similar except that in the RI model the individual dynamics is driven by noise terms. However, the impacts of these terms may be largely reduced when aggregating over all individual. As a result, we expect that these two models can generate more similar dynamics in aggregate level.

Finally, although we do not compare the welfare costs of fluctuations in both models here, intuitively, they would have different implications for welfare since in the habit formation model, the preference is changed and then has first-order impact on welfare losses, while in the RI model the preference is unchanged, consumers change their decision rule due to information processing constraints, which has second-order impact on welfare.

5.2. Comparison with the PIH Model with Rule-of-thumb Consumers

Campbell and Mankiw (1989) provides compelling evidence for the existence of rule-of-thumb consumers, i.e., consumers whose current consumption is set to current income. This greatly deviates from the permanent income hypothesis and can be used to interpret the excess sensitivity findings on aggregate data. Specifically, they assume that the economy consists of two groups of consumers. Half the consumers are forward-looking and consume their permanent income and half the consumers follow the rule-of-thumb of consuming their current income. Hence, the change in aggregate consumption derived from the Campbell-Mankiw model can be written as

$$\Delta C_{a,t} = \varkappa \Delta Y_{a,t} + (1 - \varkappa) \zeta_t \quad (5.11)$$

where \varkappa is the fraction of the rule-of-thumb consumers,⁴⁴ $\Delta Y_{a,t}$ is the change in current aggregate income, and ζ_t is the unforecastable innovation in permanent aggregate income com-

⁴³Note here we assume that $\gamma R < 1$.

⁴⁴Campbell and Mankiw (1989) estimated that \varkappa is around 40% or 50% in the US economy.

ponent. As before, assume that aggregate income component follows a difference-stationary process and is positively serially correlated in first-differences. Hence, we know that $\Delta Y_{a,t} = \frac{\varepsilon_t}{1-\rho L}$ and $\zeta_t = \frac{R}{R-\rho}\varepsilon_t$. Substituting them into (5.11) yields an MA(∞) process for the changes in aggregate consumption

$$\Delta C_{a,t} = [\varkappa + (1 - \varkappa)\frac{R}{R - \rho}]\varepsilon_t + \varkappa\frac{\rho\varepsilon_{t-1}}{1 - \rho \cdot L} \quad (5.12)$$

Hence, as in the PIH model with RI, the change in consumption can also be expressed as an MA(∞) process with decreasing coefficients. In other words, aggregate consumption is sensitive to past shocks to income. Furthermore, since $1 < \varkappa + (1 - \varkappa)\frac{R}{R-\rho} < \frac{R}{R-\rho}$, aggregate consumption becomes more smooth in the rule-of-thumb model, but still cannot resolve the excess smoothness puzzle. The intuition is that as λ increases, $\Delta Y_{a,t}$ would have greater weight in the consumption change. As a result, consumption becomes smoother because the change in current income $\Delta Y_{a,t}$ is less variable than the innovation to permanent income ζ_t . Note that in the RI model, we have

$$\Delta C_{a,t} = \theta\frac{R}{R - \rho}\frac{\varepsilon_t}{1 - (1 - \theta)R \cdot L}.$$

Comparing the coefficients in these two MA(∞) processes, it is easy to see that both models may generate the same initial responses of consumption with respect to income innovations when

$$\varkappa = (1 - \theta)\frac{R}{\rho},$$

but the initial response in the rule-of-thumb model is always greater than 1. Hence, although the rule-of-thumb model and the RI model have potential to resolve the excess sensitivity puzzle, only RI has a potential to resolve the excess smoothness puzzle. Of course, it is obvious that combining these two hypothesis in a model (for example, suppose that half the consumers are RI and half the consumers follow the rule of thumb) may go a long way to resolve the excess smoothness puzzle.

6. Conclusions

In this paper, I examine the effects of information processing constraints on the intertemporal allocation of consumption and welfare in the otherwise canonical PIH model. It is shown that incorporating RI into the standard model can generate different individual and aggregate consumption dynamics and thus make the models better explain the data in several aspects. Specifically, in the context of the LQG PIH framework, introducing RI can generate: (1) the gradual responses of consumption to various shocks to wealth, (2) the excess sensitivity of consumption to lagged shocks, and (3) the excess smoothness of aggregate consumption. Furthermore, I analyze welfare consequences due to RI and found that the welfare loss due to deviating from the first-best instantly adjusted consumption path is very tiny. I then discuss endogenizing channel capacity. Finally, I compare our RI model with the habit-formation model and the rule-of-thumb model.

So far, I have not considered costly attention and just set channel capacity κ exogenously.

Since attention is a scarce economic resource and there are many competing demands, it is natural to ask how to determine optimal channel capacity endogenously. In future research we should consider how to model endogenous inattention and examine its implications for consumption and saving behavior. In addition, given the findings in this paper, it seems also promising to incorporating RI into other frameworks, e.g., the monetary models or the full-fledged dynamic general equilibrium models. Specifically, it is well known that in the baseline New Keynesian monetary models, the dynamics of output and inflation that are characterized by the “IS” equation and the “New Keynesian” Phillips curve do not depend on any lagged variables but entirely depend on expectations of future variables, monetary policy and exogenous force processes, that is, they are completely forward-looking. However, as have been documented in many empirical studies, these forward looking models cannot match some important features of the data. For example, they are unable to replicate the high serial correlations found in both output and inflation. Hence, since we have demonstrated in this paper that RI can generate the persistence of wealth shocks endogenously in the basic PIH model, it might be a potential explanation for the persistence problem in the monetary models. Second, since RI largely affects the optimal intertemporal allocation of consumption, it is also worthwhile examining if incorporating RI into the Heterogeneous-agent dynamic stochastic general equilibrium models (e.g. Krusell and Smith 1998) 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.? or can RI increase aggregate saving in general equilibrium? I leave these interesting topics for future research.

7. Appendix

7.1. Appendix A: Deriving the Value Function in the RI Model

Before deriving the value function in the RI economy, it is helpful to first derive the value function in the corresponding full-information case. In the standard full-information PIH model, we have the consumption rule: $C_t = H_0 + H_1 S_t$. As usual, conjecture that the value function has the following form

$$J(S_t) = A_0 + A_1 S_t + A_2 S_t^2 \quad (7.1)$$

where A_0, A_1 , and A_2 are undetermined coefficients. Following the standard procedure, substituting the optimal consumption function into the Bellman equation, $J(S_t) = \max_{C_t} E_t[u(C_t) + \beta J(S_{t+1})]$, we can then pin down the undetermined coefficients:

$$A_1 = \frac{1}{1-\beta} H_1 (1 - H_0); A_2 = \frac{1}{1-\beta} \left(-\frac{1}{2} H_1^2\right); A_0 = \frac{1}{1-\beta} \left(H_0 - \frac{1}{2} H_0^2 + \beta A_2 \omega_\zeta^2\right)$$

By contrast, in the RI economy, we have the following Bellman equation

$$\widehat{J}(\widehat{S}_t) = \max_{\{C_t, \mathcal{D}_t\}} E_t[u(C_t) + \beta \widehat{J}(\widehat{S}_{t+1})] \quad (7.2)$$

where the expectation is formed under the assumption that the current and future consumption are chosen under information processing constraints. Similarly, conjecture that

$$\widehat{J}(\widehat{S}_t) = B_0 + B_1\widehat{S}_t + B_2\widehat{S}_t^2. \quad (7.3)$$

Substituting the optimal consumption rule (2.17), the Kalman equation (2.18), and the guessed value function (7.3) into the above Bellman equation, we have the following equality

$$B_0 + B_1\widehat{S}_t + B_2\widehat{S}_t^2 = (H_0 + H_1\widehat{S}_t) - \frac{1}{2}(H_0 + H_1\widehat{S}_t)^2 + \beta(B_0 + B_1E_t(\widehat{S}_{t+1}) + B_2E_t(\widehat{S}_{t+1}^2)) \quad (7.4)$$

Using the following facts

$$\begin{aligned} E_t(\widehat{S}_{t+1}) &= \widehat{S}_t, \\ \text{var}_t(\widehat{S}_{t+1}) &= \theta^2(R^2 \text{var}_t(S_t) + \omega_\zeta^2 + \text{var}_t(\xi_t)), \\ E_t(\widehat{S}_{t+1}^2) &= \text{var}_t(\widehat{S}_{t+1}) + \widehat{S}_t^2 \\ &= (R^2 - 1) \text{var}_t(S_t) + \omega_\zeta^2 + \widehat{S}_t^2, \end{aligned}$$

we can then easily pin down the coefficients B_0 , B_1 , and B_2 as follows:

$$B_1 = \frac{1}{1-\beta}H_1(1-H_0); B_2 = \frac{1}{1-\beta}(-\frac{1}{2}H_1^2); B_0 = \frac{1}{1-\beta}(H_0 - \frac{1}{2}H_0^2 + \beta B_2((R^2-1)\text{var}_t(S_t) + \omega_\zeta^2))$$

where $\omega_\zeta^2 = (\frac{R}{R-1})^2\omega^2 + \sigma^2$ and $\text{var}_t(S_t) = \bar{\sigma}_S^2 = \frac{\omega_\zeta^2}{\exp(2\kappa) - R^2}$. Note that $A_1 = B_1$ and $A_2 = B_2$.

7.2. Appendix B: Deriving the Gaussianity of Optimal Signal

Given the LQG framework, we define the loss function at t due to imperfect information as $\Delta J = J(S_t) - \widehat{J}(\widehat{S}_t)$. Hence, the expected loss function is

$$\begin{aligned} E_t(\Delta J) &= E_t(J(S_t) - \widehat{J}(\widehat{S}_t)) \\ &= E_t((A_0 - B_0) + A_1(S_t - \widehat{S}_t) + A_2(S_t^2 - \widehat{S}_t^2)) \\ &= (A_0 - B_0) + A_2E_t(S_t - \widehat{S}_t)^2 \\ &= -RA_2E_t(S_t - \widehat{S}_t)^2 \end{aligned}$$

where we use the facts that $J(S_t) = A_0 + A_1S_t + A_2S_t^2$, $\widehat{J}(\widehat{S}_t) = B_0 + B_1\widehat{S}_t + B_2\widehat{S}_t^2$, $A_1 = B_1 > 0$, $A_2 = B_2 < 0$, $\widehat{S}_t = E_t(S_t)$, and $\text{covar}_t(S_t - \widehat{S}_t, \widehat{S}_t) = 0$.

Since $A_2 < 0$, choosing the signal to minimize the above expected loss function above is equivalent to choosing the signal to minimize $E_t(S_t - \widehat{S}_t)^2$. Hence, minimizing this loss function

under the IPC can be characterized by the following optimization problem⁴⁵

$$\min_{p(S_t, \hat{S}_t)} E_t(S_t - \hat{S}_t)^2$$

subject to

$$-\mathcal{H}(S_t, \hat{S}_t) + \mathcal{H}(\hat{S}_t) + \mathcal{H}(S_t) < \kappa \quad (7.5)$$

where \mathcal{H} is entropy and κ is the bound (channel capacity) on the mutual information between S_t and \hat{S}_t .

We could reformulate the above optimization problem as follows

$$\min_{p(S_t, \hat{S}_t)} \int \int (S_t - \hat{S}_t)^2 p(S_t, \hat{S}_t) dS_t d\hat{S}_t \quad (7.6)$$

subject to

$$\int \int \log(p(S_t, \hat{S}_t)) p(S_t, \hat{S}_t) dS_t d\hat{S}_t - \int \log p_0(S_t) p_0(S_t) dS_t \quad (7.7)$$

$$- \int [\log(\int p(S_t, \hat{S}_t) dS_t) \cdot \int p(S_t, \hat{S}_t) dS_t] d\hat{S}_t < \kappa;$$

$$\int p(S_t, \hat{S}_t) d\hat{S}_t = p_0(S_t) \quad (7.8)$$

where $p_0(S_t)$ is the prior pdf of the true state and $p(S_t, \hat{S}_t)$ is the joint pdf of the true state and the observed signal.

The Lagrangian multipliers for the above two constraints are λ and $\mu(S_t)$, respectively. The Lagrangian function is then

$$L = \int \int (S_t - \hat{S}_t)^2 p(S_t, \hat{S}_t) dS_t d\hat{S}_t \quad (7.9)$$

$$+ \lambda \left\{ \int \log p_0(S_t) p_0(S_t) dS_t + \int [\log(\int p(S_t, \hat{S}_t) dS_t) \cdot \int p(S_t, \hat{S}_t) dS_t] d\hat{S}_t \right.$$

$$- \int \int \log(p(S_t, \hat{S}_t)) p(S_t, \hat{S}_t) dS_t d\hat{S}_t \left. \right\}$$

$$- \mu(S_t) \left[\int p(S_t, \hat{S}_t) d\hat{S}_t - p_0(S_t) \right]$$

Since there are no derivatives of $p(S_t, \hat{S}_t)$ involved here, we can solve this problem point by point. Differentiating the Lagrangian function with respect to the pdf $p(S_t, \hat{S}_t)$ gives us the first order condition

$$(S_t - \hat{S}_t)^2 - \lambda [\log p(S_t, \hat{S}_t) - \log(\int p(S_t, \hat{S}_t) dS_t)] - \mu(S_t) = 0 \quad (7.10)$$

⁴⁵The following derivations are largely based on Sims's notes on "Optimality of Gaussian Observation Error" that is available from <http://sims.princeton.edu/yftp/AdvMacro03/GaussianOptNotes.pdf>. Any errors here are solely mine.

We define the pdf of the signal observed through the limited channel as

$$q(\widehat{S}_t) = \int p(S_t, \widehat{S}_t) dS_t,$$

then the above FOC implies that

$$\begin{aligned} p(S_t|\widehat{S}_t) &= \frac{p(S_t, \widehat{S}_t)}{q(\widehat{S}_t)} \\ &= \exp\left\{\frac{1}{\lambda}[\mu(S_t) - (S_t - \widehat{S}_t)^2]\right\} \\ &= A(S_t) \exp[B(S_t - \widehat{S}_t)^2] \end{aligned}$$

where $A(S_t) = \exp[\frac{1}{\lambda}\mu(S_t)]$ and $B = -\frac{1}{\lambda}$. By setting the simple function A and B , it is clear that the right-hand side is a Gaussian pdf.

Hence, if the signal \widehat{S}_t follows a Gaussian distribution, then the joint distribution of S_t and \widehat{S}_t is also Gaussian. This is exactly what Shannon (1948) showed that Gaussian variables minimize the variance for a given entropy.

References

- [1] Adam, K., 2004. Optimal monetary policy in the presence of imperfect common knowledge, Manuscript, European Central Bank.
- [2] Alessie, R., Lusardi, A., 1997. Consumption, saving and habit formation, *Economics Letters* 55, 103-108.
- [3] Ash, R. 1990. *Information Theory*. Dover Publications, Inc.
- [4] Ball, L., Mankiw, G., Reis, R., 2003. Monetary policy for inattentive economies, *Forthcoming Journal of Monetary Economics*.
- [5] Boldrin, M., Christiano, L., and Fisher, J., 2001. Habit persistence, asset returns, and the business cycle, *American Economic Review* 91, 149-166.
- [6] Caballero, R., 1990. Consumption puzzles and precautionary savings, *Journal of Monetary Economics* 25, 136-136.
- [7] Campbell, J., 1987. Does saving anticipate declining labor income? An Alternative Test of the Permanent Income Hypothesis, *Econometrica* 55, 1249-1273.
- [8] Campbell, J., Deaton A., 1989. Why is consumption so smooth?, *Review of Economic Studies* 56, 357-374.
- [9] 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.

- [10] Carroll, C. 1997. Buffer stock saving and the life cycle/permanent income hypothesis, *Quarterly Journal of Economics* 112 (1), 1-55.
- [11] Christiano, L., 1987. Why is consumption less volatile than income?, *Fed. Reserve Bank Minneapolis Quarterly Review* 11, 2-20.
- [12] Cochrane, J., 1989. The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives, *American Economic Review* 79, 319-337.
- [13] Constantinides, G., 1990. Habit formation: a resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519-543.
- [14] Cover, T.M., Thomas, J.A., 1991. *Elements of Information Theory*. New York: John Wiley and Sons.
- [15] Deaton, A., 1992. *Understanding Consumption*. Oxford University Press, Oxford.
- [16] Flavin, M., 1981. The adjustments of consumption to changing expectations about future income, *Journal of Political Economy* 89, 974-1009.
- [17] Fuhrer, J., Habit formation in consumption and its Implications for monetary policy, *American Economic Review* 90, 367-390.
- [18] Gali, J., Budget constraints and time series evidence on consumption, *American Economic Review* 81, 1238-1253.
- [19] Hall, R., 1978. Stochastic implications of the life cycle-permanent income hypothesis: Theory and Evidence, *Journal of Political Economy* 91, 249-265.
- [20] Krusell, P., Smith, A.Jr., 1998. Income and wealth heterogeneity in the macroeconomy, *Journal of Political Economy* 106, 867-896.
- [21] Ludvigson, S., Michaelides, A., 2001. Does buffer stock saving explain the smoothness and excess sensitivity of consumption? *American Economic Review* 91 (3), 631-647.
- [22] Luo, Y., Young, E.R., 2006. Rational inattention and aggregate fluctuations, Manuscript, University of Hong Kong and University of Virginia.
- [23] Lucas, R., 1987. *Models of Business Cycles*, Basil Blackwell Press.
- [24] Maćkowiak, B., Wiederholt, M., 2005. Optimal sticky prices under rational inattention, Manuscript, Humboldt University Berlin.
- [25] Mankiw, G., Reis, R., 2002. Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve, *Quarterly Journal of Economics* 117 (4), 1295-1328.
- [26] Moscarini, G., 2004. Limited information capacity as a source of inertia, *Journal of Economic Control and Dynamics* 28(10), 2003-2035.

- [27] Parker, J., 1999. The reaction of household consumption to predictable changes in social security taxes, *American Economic Review*, Vol 89, No.4, 959-973.
- [28] Peng, L., Xiong, W., 2001. Time to digest and volatility dynamics, Manuscript, Baruch College and Princeton University.
- [29] Pischke, J.S., 1995. Individual income, incomplete information, and aggregate consumption, *Econometrica* 63 (4), 805-840.
- [30] Reis, R., 2003. Inattentive consumer, Manuscript, Harvard University.
- [31] Shannon, C.E., 1948. A mathematical theory of communication, *The Bell System Technical Journal* 27, 623-656.
- [32] Sims, C.A., 1998. Stickiness, *Carnegie-Rochester Conference Series on Public Policy*, Vol 49(1), 317-356.
- [33] Sims, C.A., 2003. Implications of rational inattention, *Journal of Monetary Economics* 50, 3, 665-690.
- [34] Storesletten, K., Telmer, C.I., Yaron A., 2001. Cyclical dynamics in idiosyncratic labor market risk," *Journal of Political Economy* 112 (3), 695-717.
- [35] Whittle, P., 1982. *Optimization Over Time Volume 1*. Chichester: John Wiley & Sons.
- [36] Whittle, P., 1996. *Optimal Control: Basics and Beyond*. John Wiley & Sons.
- [37] Woodford, M., 2001. Imperfect common knowledge and the effects of monetary policy, Princeton University, in *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds.

Table 1
 Aggregate Stylized Facts on $\Delta Y_{a,t}$ and $\Delta C_{a,t}$ (quarterly)
 (standard errors in parentheses)

sample period	AR(1) coef. of $\Delta Y_{a,t}$	s.d. of income innovations	
1954 – 1990	0.288 (0.08)	\$50 (mean income is \$7000)	
sample period	coef. of $\Delta C_{a,t}$ on $\Delta Y_{a,t-1}(\hat{\beta})$	AR(1) coef. of $\Delta C_{a,t}$	ES ratio
1954 – 1990	0.110 (0.045)	0.200 (0.082)	0.578 (0.055)

where “ES” ratio means the excess smoothness ratio.

Table 2

Utility Losses due to RI (\$/quarter)			
	$\gamma = 1$	$\gamma = 4$	$\gamma = 10$
$\kappa = 0.3$ bits	0.174	0.69	1.74
$\kappa = 0.6$ bits	0.067	0.27	0.67
$\kappa = 1$ bits	0.028	0.12	0.29
$\kappa = 2$ bits	0.006	0.02	0.06
$\kappa = 3$ bits	0.001	0.01	0.01

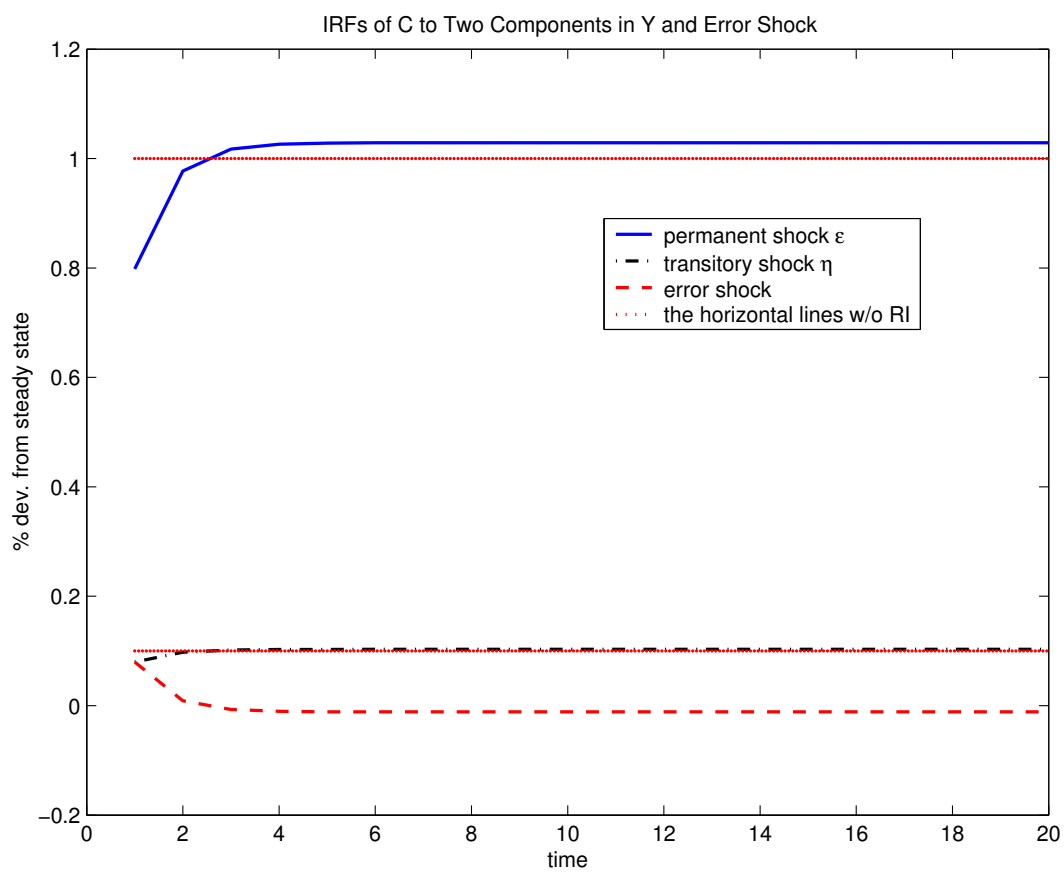


Figure 1

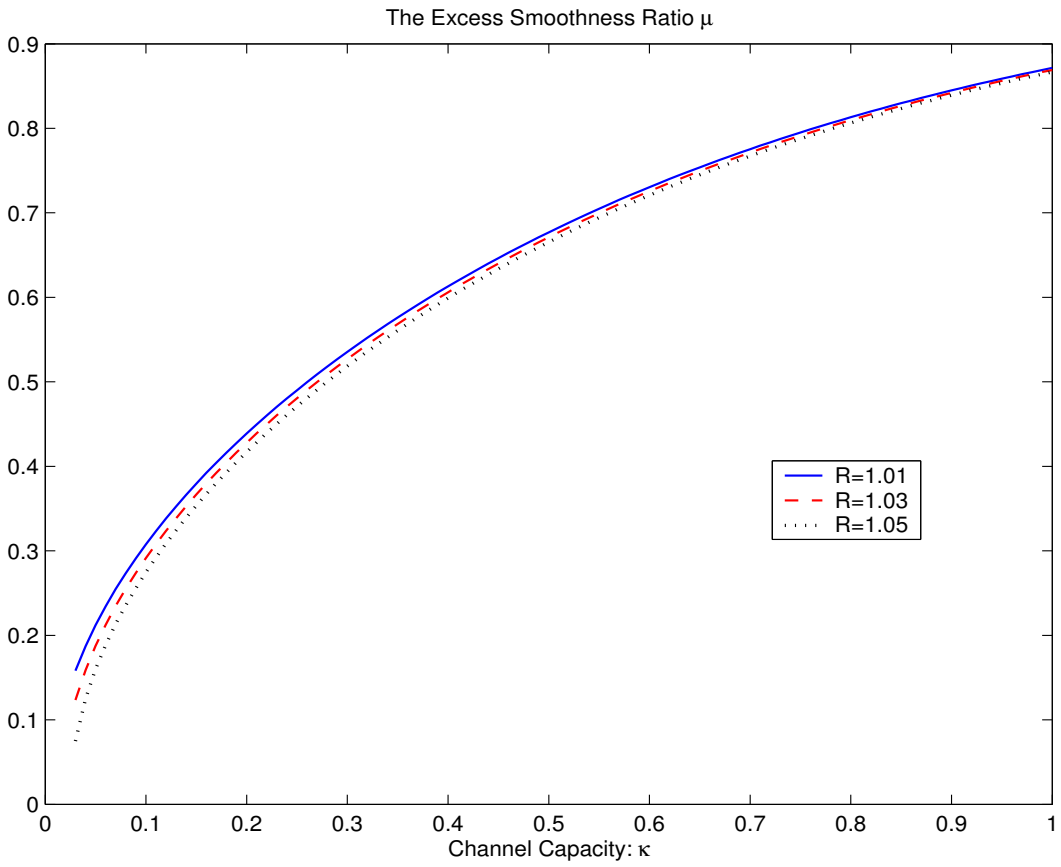


Figure 2

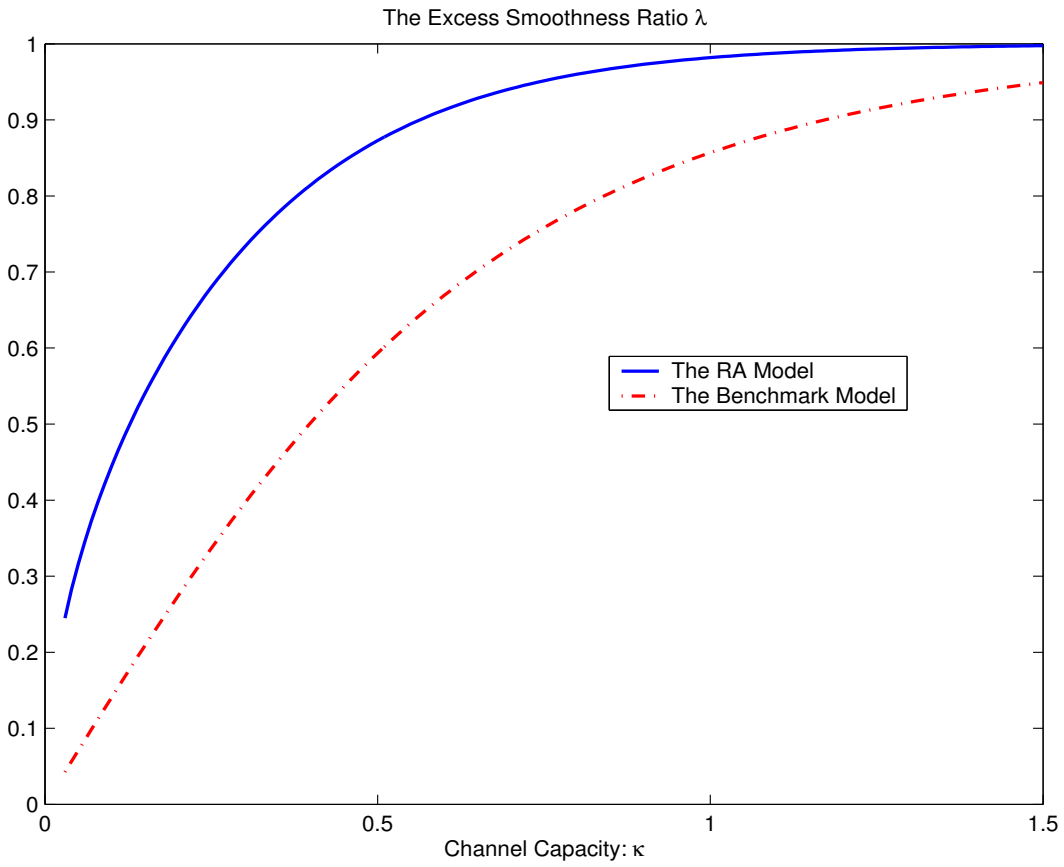


Figure 3

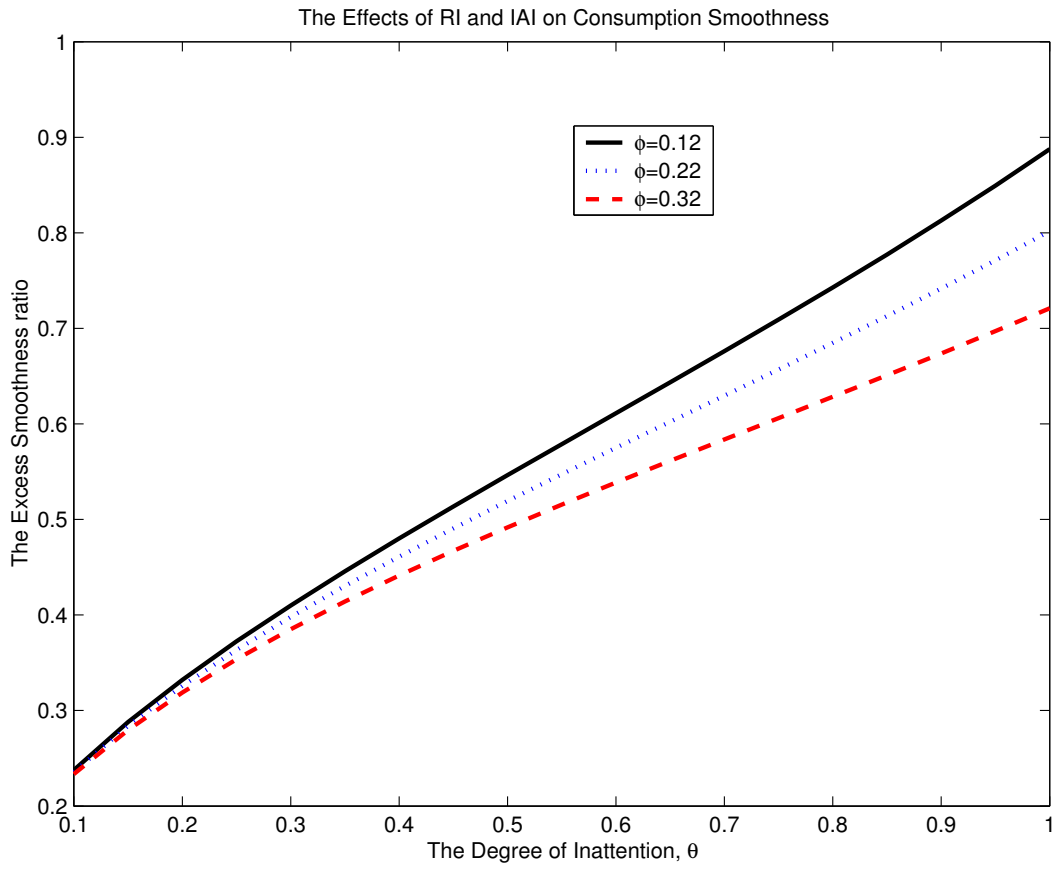


Figure 4

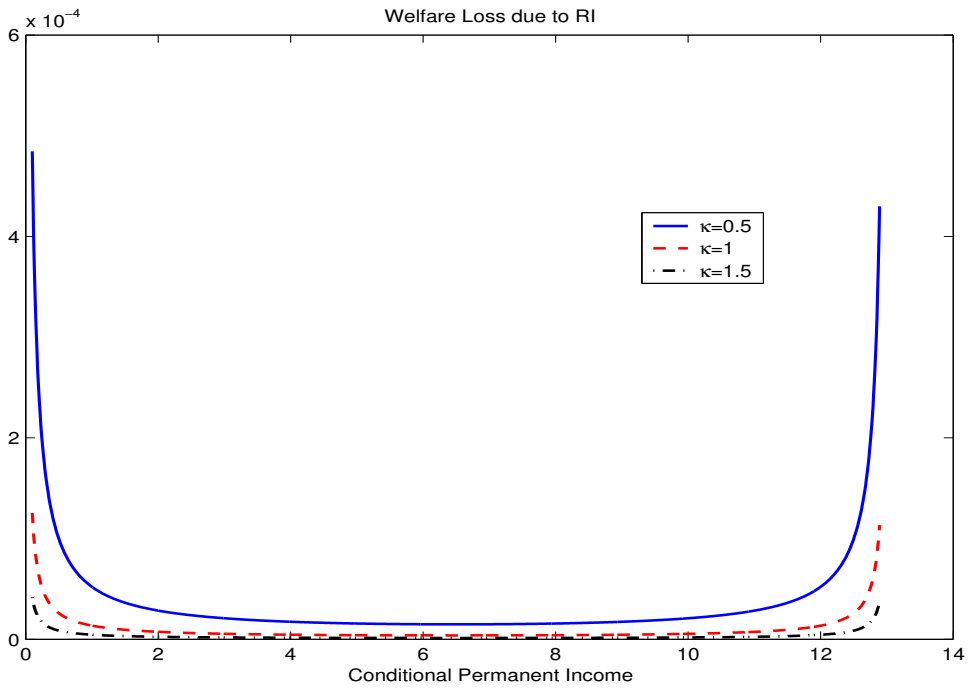


Figure 5a

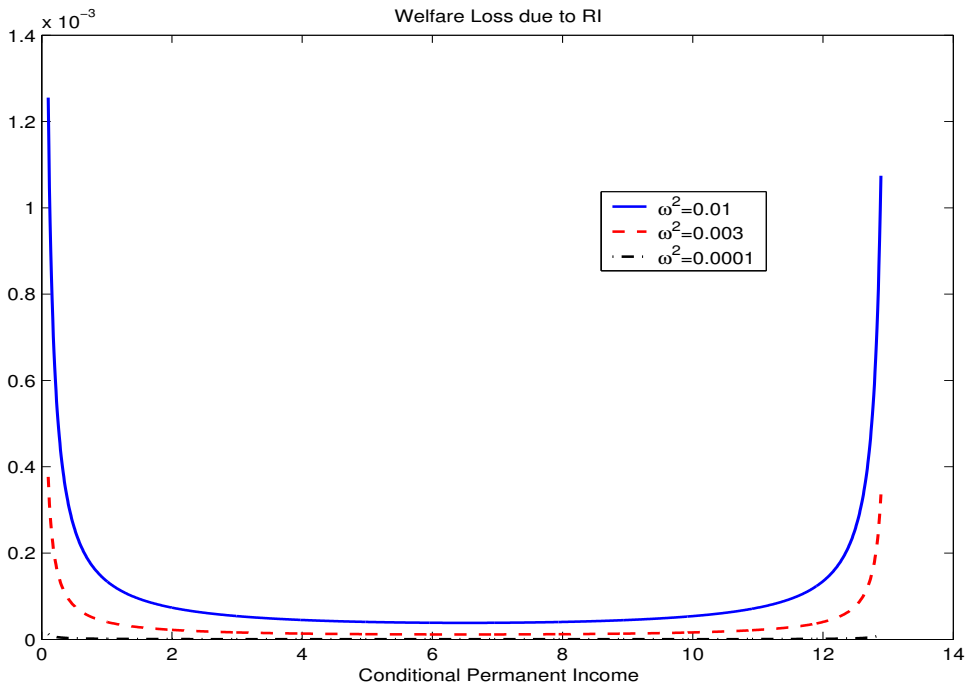


Figure 5b