The rigidity of choice: lifetime savings under information-processing constraints

Antonella Tutino

Federal Reserve Board

21. June 2008

The rigidity of choice:
Lifetime savings under information-processing constraints

Antonella Tutino *

First version: November 2007 ; This version: July 2009

Abstract

This paper studies the implications of information-processing limits on the consumption and savings behavior of households through time. It presents a dynamic model in which consumers rationally choose the size and scope of the information they want to process about their financial possibilities, constrained by a Shannon channel. The model predicts that people with higher degrees of risk aversion rationally choose higher information. This happens for precautionary reasons since, with finite processing rate, risk averse consumers prefer to be well informed about their financial possibilities before implementing consumption plan. Moreover, numerical results show that consumers with processing capacity constraints have asymmetric responses to shocks, with negative shocks producing more persistent effects than positive ones. This asymmetry results into more savings. I show that the predictions of the model can be effectively used to study the impact of tax reforms on consumers spending. The results are qualitatively consistent with the evidence on tax rebates (2001, 2008).

*COMMENTS WELCOME. E-mail: `Antonella.Tutino@frb.gov`. I am deeply indebted to Chris Sims whose countless suggestions, shaping influence and guidance were essential to improve the quality of this paper. I thank Ricardo Reis for valuable advice, enduring and enthusiastic support. I am grateful to Per Krusell for his insightful advice and stimulating discussions. I would also like to thank Mike Kiley, Nobu Kiyotaki, Angelo Mele, Philippe-Emanuel Petalas, John M. Roberts, Charles Roddie, Sam Schulhofer-Wohl, Tommaso Treu, Mark Watson and Mirko Wiederholt. Finally I thank numerous seminar participants for many helpful comments and discussions. Any remaning errors are my own. The views in this paper are solely the responsibility of the author and should not be interpreted as reflecting the views of the Federal Reserve Board or any other person associated with the Federal Reserve System.

*[...]Multa esse probabilia quae quamquam non perciperentur tamen quia uisum quemdam haberent insignem et illustrem his sapientis vitae regeretur. (*M. Tulli Ciceronis*, De Natura Deorum, Liber I, 5, 12)*

1 Introduction

Looking at individual and aggregate evidence, we observe delayed responses of consumption to shocks to income even when information about these shocks is fully available. For example, consider tax rebates in the recent U.S. experience. Both 2001 and 2008 rebates have been advertised before the checks were mailed to the beneficiaries. Rational expectations permanent income models predicts that rational agents should react by increasing spending as soon as the rebates were announced.[1] Moreover, the increase in spending should be small since rational consumers discount the temporary changes over their life-cycle. However, data tell us that consumers did not increase spending at the time of the announcement. Data also tell us that for the rebate of 2001 two-thirds of the rebates where spent when the check was mailed and in the following three-month period.[2] Further rejection of the neoclassical consumption theory comes from the empirical evidence on asymmetric response to predictable income changes. Using the Panel Study of Income Dynamics (PSID), Shea (1995) reports that consumption responds more strongly to a negative predictable change in income than a positive one.[3] A similar finding comes from the switching regression model of Garcìa, Lusardi and Ng (1997) applied to data drawn from the Consumer Expenditure Survey (CEX). Using both PSID and CEX, Dynarski and Gruber (1997) document asymmetric response of consumption -especially durables and in the CEX data- to earning changes, with earnings reductions producing stronger reaction than earnings increase. Finally, there is evidence on asymmetric response of macroeconomic aggregates to monetary changes. Using quarterly U.S. post-war data, Cover (1992) shows that negative money-supply shocks have contractionary effects on output whereas positive monetary shocks do not affect output. De Long and Summers (1988) show that Cover's findings holds true at annual frequency as well as in pre-war U.S. data.

Macroeconomists have recognized the necessity of matching these data within rational expectations framework with a number of modelling strategies. Restrictions on the information available to the agents -such as costly acquisition and diffusion of information-rely on ad-hoc assumptions to generate smooth and delayed responses of, e.g., consumption to shocks to income consistent with observed data. Moreover, these restrictions often do not provide a way to account for asymmetries in speed and amount of reactions of consumers to positive or negative changes in income.

This paper proposes a model in which optimal consumption of rational agents is

[1] Johnson, Parker and Souleles 2001, Broda and Parker, 2008.

[2] cfr. Johnson, Parker and Souleles 2001.

[3] Shea (1995) also points out that the asymmetry in consumption response to predictable changes in income cannot be reconciled with liquidity constraints.

sluggish and reacts asymmetrically to changes in personal income. Following *Rational Inattention* (Sims, 2003, 2006), my model assumes that the agents maximize their lifetime savings subject to a bound on their ability to process information at infinite rate, given by Shannon's capacity.

With this information processing limit, people choose a signal that conveys information about their wealth. The signal can provide any kind of information so long as its overall content is within Shannon's capacity. Consumers form expectations given the signals and decide how much to consume. The bound on information flow prevents people from reacting quickly and precisely to fluctuations in income. Trading off effort of processing information for speed of reactions to signals produces endogenous delayed and smoothed response of consumption to changes in income. Combining iso-elastic preferences, ex-ante uncertainty -not necessarily Gaussian- and information processing constraints produces endogenous asymmetric response of consumption to shocks to income.

My paper shows how to embed utility maximization with information processing limits formally in an intertemporal setting. I assume that people do not know the exact value of their wealth due to income fluctuations but they have a belief -prior- on their wealth. A way of thinking about this assumption is that people do not constantly check their account balance and even when they do check it, they cannot precisely map the balance into current and future consumption possibilities. They need to think hard of what that number means -i.e., process information- to sharpen their knowledge of how much current and future consumption their wealth is worth. In such a framework, it is possible to study how choices of information play out with people's preferences when they decide on savings consumption throughout their life.

The challenge of this model and, more generally, of models of rational inattention is dealing with infinite dimensional state space implied by having a prior as the state. For this reason, applications of rational inattention have been limited to either a linear quadratic framework where Gaussian uncertainty has been considered[4] or a two-period consumption-saving problem (Sims 2006) where the choice of optimal ex post uncertainty is analyzed for the case of log utility and two CRRA utility specifications. The linear quadratic Gaussian (LQG) framework can be seen as a particular instance of rational inattention in which the optimal distribution chosen by the household turns out to be Gaussian. Gaussianity has two main advantages. First, it allows an explicit analytical solution for these kinds of model. Second it is easy to compare the results to a signal extraction problem: when looking at consumers' behavior of rational inattentive consumers, it is impossible to tell apart an exogenously given Gaussian noise in the signal extraction model from endogenous noise that is optimally chosen to be Gaussian. Although the Gaussian assumption can be a good approximation when uncertainty is small, data on income fluctuations suggest that uncertainty at individual level might actually be large. Most importantly, Gaussian uncertainty prevents rational inattention LQG models from delivering endogenously asymmetric response to shocks. Thus, to fully assess the joint

[4]Such as Sims (1998, 2003), Luo (2008), Mackowiak and Wiederholt (2008a, 2008b), Mondria,(2006) and Moscarini (2004).

contribution of information constraints and people's preferences, it is important to let consumers select their information from a wider set of distributions that includes but it is not limited to the Gaussian family.

The theoretical contribution of this paper is to provide the analytical and computational tools necessary to apply information theory in a dynamic context with optimal choice of ex-post uncertainty. I propose a methodology to handle the additional complexity without the LQG setting. I propose a discretization of the framework and derive its theoretical properties. Then, I provide a computational strategy to solve the model and an efficient algorithm to handle the complexity associated with the high-dimensional state space.

Several predictions emerge from the model. Evaluating the unconditional moments of the time series of consumption for a given degree of risk aversion, the first result of the paper is that higher shadow costs of processing information are associated with more persistence and higher volatility. The intuition is that before modifying his consumption profile a person might decide to wait and have more information about wealth. If the person waits long enough, he might realize that he has saved more than he wanted to , therefore he increases his consumption by a significant amount. The combination of waiting while processing information and sharp changes once information has been processed through time generates sluggishness and volatility in consumption.

Second, by looking at the life-cycle profile of consumption I find that the behavior of consumption is smooth and persistent along the simulated path. Moreover, people with low processing capacity delay changes in consumption more than people with high processing capacity. When a variation in consumption occur, people with low processing capacity, who have waited to respond to wealth fluctuations over time, change sharply their consumption profile. These results combined are suggestive of a precautionary motive for savings driven by information processing limits.

Third, I find that consumers with processing capacity constraints have asymmetric responses to income fluctuations, with negative shocks producing sharper and more persistent effects than positive ones. This effect is stronger the higher the degree of risk aversion. Compared with a situation in which there are no information-processing limits, in a rational inattention consumption-savings model, an adverse temporary income shock makes consumers reduce their consumption for a prolonged period of time. This happens because risk-averse people who receive bad news about their finances save right away to hedge against the possibility of running out of wealth in the future. Once they have enough savings and information, they gradually increase their consumption and smooth the remaining effect of the shock over time. This result also points towards precautionary motive due to information-processing limits. Moreover, the predictions of the model can be used to address important policy questions. In order to understand how initial values of wealth affect the asymmetric response, I focus on groups of people who start off life with low, middle and high wealth. I assume that all groups have the same income and receive a temporary positive shock to their income. This experiment can be thought of as a one-time tax rebate. My model predicts that the policy will have the faster effect on individual with low income. However the policy will be most effective for people with low and middle income after several quarters, when the savings triggered by the rebates will be sizeable. These predictions appear to be consistent with the evidence on the 2001 tax rebate and the preliminary evidence on the 2008 rebate.[5]

My results are observational distinct from the previous literature on consumption and the one on consumption and information . The distinguished feature of my model with respect to previous works is its ability to generate endogenously asymmetric response of consumption to shocks. In particular, for a given degree of risk aversion and magnitude of a shock, the response of consumption to a negative shock is stronger on impact and more persistent than the one to a positive shock. This is consistent with empirical evidence in Shea (1995).

The paper is organized as follows. Section 2 lays out the theoretical basis of rational inattention and introduces informally the model. Section 3 states the problem of consumers as a discrete stochastic dynamic programming problem, while Section 4 derives the properties of the Bellman function. Section 5 provides the numerical methodology used to solve the model. Section 6 delivers its main results. Section 7 concludes. The main propositions of the paper are in Appendix A and B. Appendix C provides the pseudocode whereas the mathematical details on the information theoretic apparatus are in Appendix D.

2 Foundations of Rational Inattention

Rational inattention (Sims 1988,[6] 1998, 2003, 2005, 2006) blends information theory and economics. The first draws mainly on the work of Shannon (1948). The main contribution is to define a measure of the choice involved in the selection of the message and the uncertainty regarding the outcome. The measure used is entropy. Details on this part are in Appendix D. Based on Shannon's apparatus, the economic contribution lies in the use of Shannon's capacity as a technological constraint to capture individuals' inability of processing information about the economy at infinite rate. Given these limits, people reduce their uncertainty by selecting the focus of their attention. The resulting behavior depends on the choices of what to observe about the environment once the information-processing frictions are acknowledged.

2.1 The Economics of Rational Inattention

Consider a person who wants to buy lunch. He does not know exactly his wealth but he knows that he has some cash and a credit card in his pocket. Not recalling the expenses charged on the credit card up to that point, he can go to the bank or simply check his

[5] Johnson, Parker and Souleles (2001), Broda and Parker, (2008).

[6] The bulk of the idea of *rational inattention* can be found in C. Sims' 1988 comment in the Brooking Papers on Economic Activity .

wallet. Going to the bank to figure out the exact wealth for lunch is beyond his time and interest, so he decides to check his wallet. He browses it thinking what he wants and can afford for lunch. Mapping dollar bills into his knowledge of prices from previous consumption, he realizes he can only afford a sandwich instead of his favorite sushi roll. Then, he uses the receipt to update his prior on the price of sandwiches, what he thinks he has left in his wallet and, ultimately, his wealth. This updated knowledge will be used for his next purchase. Such a story can be directly mapped into a rational inattention framework using mutual information (Shannon, 1948) as the technology that regulates the flow of information that passes through the channel.

First, the person does not know his wealth, W, but he has a prior on it, $p(W)$. Before processing any information, his uncertainty about wealth is the entropy of his prior, $\mathcal{H}(W) \equiv -E[\log_2(p(W))]$, where $E[.]$ denotes the expectation operator.[7] Before processing any information, lunch too is a random variable, C, ranging from sandwiches to sushi To reduce entropy, he can choose whether to have a detailed report from the bank or to look at his wallet. The two options differ in amount of information and effort in processing their content. The choice of the option (signal) together with consumption result in a joint probability $p(c,w)$. Both dollar bills in the wallet and knowledge of prices of sandwiches and sushi contribute to the reduction of uncertainty in wealth of an amount equal to $\mathcal{H}(W|C) = -\int p(w,c)\log_2 p(w|c)\,dcdw$, which is the entropy of W that remains given the knowledge of C. The information flow, or maximum reduction of uncertainty about the prior on wealth, is bounded by the information that the selected signal conveys. In formulae:

$$I(C;W) = \mathcal{H}(W) - \mathcal{H}(W|C) \leq \kappa \tag{1}$$

where κ is measured in number of bits transmitted. Finally, the signal -peeking at the wallet, $p(w,c)$- and the receipt for the sandwich, $\bar{c}$, are used to update the prior on wealth via Bayes' rule and then the update is carried over for future purchases.

The example illustrates how people handle everyday decision weighting the effort of processing all the available information -personal net worth- against the precision of the information they can absorb -walking to the bank versus checking the wallet- guided by their interest -buying lunch-. This is the core of rational inattention: information is freely available but people can process it at finite rate. Information-processing limits make attention a scarce resource. As for any other scarce resource, rational people use attention optimally according to what they have at stake. By appending an information-processing constraint to an otherwise standard optimization framework, the theory explains why people react to changes in the economic environment with delays and errors.

The appeal of Shannon capacity as a constraint to attention is that it provides a measure of uncertainty which does not depend on the characteristics of the channel.

[7] Entropy is a universal measure of uncertainty that can be defined for a density against any base measure. The standard convention is to use base 2 for the logarithms, so that the resulting unit of information is binary and called a *bit,* and to attribute zero entropy to the events for which $p = 0$. Formally, given that $s\log(s)$ is a continuous function on $s \in [0,\infty)$, by l'Hopital Rule $\lim_{s\to 0} s\log(s) = 0$.

The quantity (1) is a probabilistic measure of the information shared by two random variables and it applies to any channel. Thus, Shannon capacity does not require explicit modelling of how individuals process information. Moreover, treating processing capacity as a constraint to utility maximization produces inertial reactions to the environment as a result of individual's rational choices. A rational person may not find it worth it to look beyond his wallet when deciding what to buy for lunch. The dollar bills in the wallet provide little information about current and future activities of his balance. Thus, if something happened to his current account -say, a sudden drop in his investment-, checking his wallet would give him no acknowledgement of the event Nevertheless, the signal is capable of guiding the consumer on his lunch's choice. Over time and expenses, the person would figure out the drop in investment and modify his behavior even with respect to lunch. Appendix D covers the mathematical details.

3 The Formal Set-up

3.1 The problem of the Household

In order to understand the implications of limits to information processing and contrast them with standard consumption-savings model, the starting point start is the full information problem.

Let $(\Omega, \mathcal{B})$ be the measurable space where Ω represents the sample set and $\mathcal{B}$ the event set. States and actions are defined on $(\Omega, \mathcal{B})$. Let $\mathcal{I}_t$ be the $\sigma-$algebra generated by $\{c_t, w_t\}$ up to time t, i.e., $\mathcal{I}_t = \sigma\left(c_t, w_t; c_{t-1}, w_{t-1}; ...; c_0, w_0\right)$. Then, the collection $\{\mathcal{I}_t\}_{t=0}^{\infty}$ such that $\mathcal{I}_t \subset \mathcal{I}_s \ \ \forall s \geq t$ is a filtration. Let $u\left(c\right)$ be the utility of the household defined over a consumption good, c. I assume that the utility belongs to the CRRA family, $u\left(c\right) = c^{1-\gamma}/\left(1-\gamma\right)$ with γ the coefficient of relative risk aversion. Consumer's problem is:

$$\max_{\{c_t\}_{t=0}^{\infty}} E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left(\frac{c_t^{1-\gamma}}{1-\gamma} \right) \middle| \mathcal{I}_0 \right\} \tag{2}$$

s.t.

$$w_{t+1} = R\left(w_t - c_t\right) + y_{t+1} \tag{3}$$

$$w_0 \text{ given} \tag{4}$$

where $\beta \in [0, 1)$ is the discount factor and $R = 1/\beta$ is the (constant) interest on savings, $(w_t - c_t)$. I assume that $y_t \in Y \equiv \left\{y^1, y^2, .., y^N\right\}$ follows a stationary Markov process with constant mean $E_t\left(\left(y_{t+1}\right)|\, \mathcal{I}_t\right) = \bar{y}$.

Consider now a consumer who cannot process all the information available in the economy to track precisely his wealth. This not only adds a constraint to the decision problem but fundamentally affects the set-up (2)-(4).

First, since the consumer doesn't know his wealth, (4) no longer holds. His uncertainty about wealth is given by the prior $g\left(w_0\right)$. Second, before processing any information, consumption is also a random variable. This is because the uncertainty about wealth translates into a number of possible consumption profiles with various levels of affordability. It follows that to maximize lifetime utility, consumers needs jointly to reduce uncertainty about wealth and to choose consumption. Hence, when information cannot flow at infinite rate the choice of the consumer is the distribution $p(w, c)$ as opposite to the stream of consumption $\{c_t\}_{t=0}^{\infty}$ in (2). Another way of looking at this is that the consumer chooses a noisy signal on wealth where the noise can assume any distribution selected by the consumer. Given that the agent has a probability distribution over wealth, choosing this signal is akin to choosing $p(c, w)$. The optimal choice of this distribution is the one that makes the distribution of consumption conditional on wealth as close to the wealth as the limits imposed by Shannon's capacity allow.

Third, with respect to the program (2)-(4), there is a new constraint on the amount of information the consumer can process. The reduction in uncertainty conveyed by the signal depends on the attention allocated by the consumer to track his wealth. Paying attention to reduce uncertainty requires spending some time and utility to process information. I model the task of focussing attention by appending a Shannon's channel to the constraint sets. Limits in the capacity of the consumers are captured by the fact that the reduction in uncertainty conveyed by the signal cannot be higher than a given number, $\bar{\kappa}$. The information flow available to the consumer is a function of the signal, i.e., the joint distribution $p(\cdot_{c_t}, \cdot_{w_t})$. In formulae:[8]

$$\bar{\kappa} \leq \kappa_t \equiv I\left(p\left(\cdot_{c_t}, \cdot_{w_t}\right)\right) = \int p\left(c_t, w_t\right) \log\left(\frac{p\left(c_t, w_t\right)}{p\left(c_t\right) g\left(w_t\right)}\right) dc_t dw_t \tag{5}$$

Fourth, the update of the prior replaces the law of motion of wealth using the budget constraint in (3). To describe the way individuals transit across states, define the operator $E_{w_t}\left(E_t\left(x_{t+1}\right)\middle| c_t\right) \equiv \hat{x}_{t+1}$, which combines the expectation in period t of a variable in period $t+1$ with the knowledge of consumption in period t, c_t, and the remaining uncertainty over wealth. Applying $E_{w_t}\left(E_t(\cdot)\middle| c_t\right)$ to equation (3) leads to:[9]

$$\hat{w}_{t+1} = R\left(\hat{w}_t - c_t\right) + \widehat{\widehat{y}} \tag{6}$$

where,

$$\begin{aligned}
\widehat{\widehat{y}} &= E_{w_t}\left(E\left(y_{t+1}\right)\middle| c_t\right) \\
&\equiv E_{w_t}\left(E\left(\left(y_{t+1}\right)\middle| \mathcal{I}_t\right)\middle| c_t\right) + \left[E_{w_t}\left(E_t\left(y_{t+1}\right)\middle| c_t\right) - E_{w_t}\left(E\left(\left(y_{t+1}\right)\middle| \mathcal{I}_t\right)\middle| c_t\right)\right] \\
&\overset{LIE}{=} \bar{y} + E_{w_t}\left[\left(E\left(y_{t+1}\right)\middle| c_t\right) - \left(E\left(y_{t+1}\right)\middle| c_t\right)\right] \\
&= \bar{y}.
\end{aligned}$$

[8]In (5), I use the expression $I\left(p\left(\cdot_{c_t}, \cdot_{w_t}\right)\right)$ instead of the one in (1) to make explicit the dependency of the mutual information on the joint distribution of C and W.

[9]The exogenous stochasticity of the model is given by fluctuations of income. For wealth not to be known at each point in time, it has to be that current and past realizations of income are not known to the consumer. However, I assume that the consumer does know that the income follows a Markov process and he also knows the value of the mean, $\bar{y}$. Moreover, since consumers care about the linear combination of savings and income as shown in (3) and information about such linear combination is freely available, their state reduces to w.

To fully characterize the transition from the prior $g(w_t)$ to its posterior distribution, the consumer needs to take into account how the choice $p(w_t, c_t)$ at time t affects the distribution of consumer's belief after observing c_t. Given the initial prior state, $g(w_0)$, the successor belief state, denoted by $g'_{c_t}(w_{t+1})$ is determined by revising each state probability as displayed by the expression:

$$g'\left(w_{t+1}|_{c_t}\right) = \int \tilde{T}\left(w_{t+1}; w_t, c_t\right) p\left(w_t|c_t\right) dw_t \tag{7}$$

known as Bayesian conditioning. In (7), the function $\tilde{T}$ is the transition function representing (6). Note that the belief state itself is completely observable. Meanwhile, Bayesian conditioning satisfies the Markov assumption by keeping a sufficient statistics that summarizes all information needed for optimal control.[10] Thus, (7) replaces (3) in the limited information-processing world.

Let θ be the marginal value of using the channel (5). Note that in the problem there is a one-to-one mapping between the shadow cost of processing information, θ, and the maximum processing capacity, $\bar{\kappa}$. Fixing θ allows one to study how consumers varies the optimal amount of information flow κ_t according to whether he wants to save more in a particular point in time than another.[11]

Combining (5)-(7), the program of the household under information frictions is:

$$\max_{\{p(w_t,c_t)\}_{t=0}^{\infty}} E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \int \left(\frac{c_t^{1-\gamma}}{1-\gamma} \right) p\left(c_t, w_t\right) \mu\left(dc_t, dw_t\right) \middle| \mathcal{I}_0 \right\} \tag{8}$$

s.t.

(θ)

$$\kappa_t = I_t\left(p\left(\cdot_{c_t}, \cdot_{w_t}\right)\right) = \int p\left(c_t, w_t\right) \log \left(\frac{p\left(c_t, w_t\right)}{\left(\int p\left(\hat{w}_t, c_t\right) d\hat{w}_t\right) g\left(w_t\right)} \right) dc_t dw_t \tag{9}$$

$$p\left(c_t, w_t\right) \in \mathcal{D}\left(w, c\right) \tag{10}$$

$$g'\left(w_{t+1}|_{c_t}\right) = \int \tilde{T}\left(w_{t+1}; w_t, c_t\right) p\left(w_t|c_t\right) dw_t \tag{11}$$

$$g\left(w_0\right) \text{ given} \tag{12}$$

[10] See Astrom, K. (1965).

[11] To clarify this point, suppose that at some point the agent is considering whether or not to increase his consumption profile. Before making a change in consumption, he wants to gather information to make sure that his wealth is high enough to support the change for some time. In the current period, he might find it optimal to have a better assessment of his wealth at the expenses of exercise more effort in processing information. Suppose that he collected enogh information and changes consumption the following period. After accounting for the increase in consumption, the agent might think that it is unlinkely that wealth has changed significantly due to exogeneous shocks to income. Therefore, with fixed shadow cost of processing information, the agent might find it optimal to process very little information the periods following the changes and enjoy consumption. Sims (2006) and most recently Maćkowiak and Wiederholt (2008b) also follow this way of accounting for information processing constraint.

where $\mu(\cdot)$ in (8) is the Dirac measure that accounts for discreteness in the optimal choice $p(c,w)$ and $\mathcal{D}(w,c) \equiv \left\{p(c,w): \int p(c,w)\,dcdw = 1,\ p(c,w) \geq 0, \forall (c,w)\right\}$ in (10) restricts the choice of the agent to be drawn from the set of distributions.[12]

This problem is a well-posed mathematical problem with convex objective function and concave constraint sets. What makes it hard to solve is that both state and control variables are infinite dimensional. To make progress in solving it, a) I discretize the framework and b) I show that the resulting setting admits a recursive formulation. Then I study the properties of the Bellman recursion and solve the problem. Before turning to the solution, a brief digression on how constraint (9) operates and the difference between this model and the existing literature on rational inattention may help building up intuition for the solution methodology and the results.

3.2 The role of Shannon's capacity constraint

3.2.1 Shannon's constraint in action

To get a sense of how Shannon's capacity constraints affect the decision of households, I contrast the optimal policy function $p^*(c,w)$ for consumers that have identical characteristics but differ in their bound on processing capacity.

A caveat is due. In order to explore the interaction between information flow and coefficient of risk aversion, I solve the model in (8)-(12) by fixing the shadow cost of processing information, θ, attached to (9) and let κ vary endogenously every period. In this section I follow a different route. In order to clarify the mechanism behind Shannon's capacity as a constraint for information transmission, I focus on the optimal probability distribution for the problem (8)-(12). I fix the number of bits, κ, across utilities and adjust the shadow cost θ to map different coefficients of risk aversion to the same information flow.[13] First consider $u(c) = \log(c)$. In the full information case,[14] the distribution $g(w)$

[12] Note that in this model is possible to assume that individuals cannot borrow and append a constraint of the kind $c \leq w$.to (2)-(4). To encode this constraint without complicating the model, one may assume that κ_t in (18) is the capacity left after the consumer has processed his spending limits.

[13] To be more specific, I solve the model with CRRA consumer assuming the same parameters as the baseline model $(\beta, R, \bar{y}) \equiv (0.9881, 1.012, 1)$ and the same simplex point (prior) $g(\tilde{w})$. I then adjust the shadow cost of processing capacity, θ, to get roughly the same information capacity ($\kappa_{\log} = 2.08$ and $\kappa_{crra} = 2.13$). The latter implies that the difference in allocation of probabilities within the grid are attributable solely on the coefficient of risk aversion γ. As I will explain in more details in the solution methodology, the same shadow cost (θ) does deliver different information flow (κ) according to the degree of risk aversion of the agents, with more risk averse agents having higher κ for a given θ than less risk averse ones do. To get $\kappa_{\log} \approx \kappa_{crra}$, I set $\theta_{\log} = 0.02$ in Figure **3** while $\theta_{crra} = 0.08$ in Figure **4**.

[14] Or, in the wording of the model, when information flows at infinite rate, $\kappa \to \infty$ in (9).

is degenerate, the choice of $p(c_t, w_t)$ reduces to that of $c(w_t)$ in (8).[15] The resulting optimal policy is given by

$$c_t^*(w_t) = (1-\beta) w_t + \beta \bar{y}. \tag{13}$$

For comparison with the case with finite κ, I plot the policy function for the (discretized) full information case as the joint distribution $p(c, w)\, \delta_{c^*(w)}(c, w)$ with $\delta_{c^*(w)}$ the Dirac measure.[16] Figure **1** plots such a distribution for a $20x20$ grid where the equi-spaced vector c ranges from 0.8 to 3 and w is also equi-spaced with support in $[1, 10]$.[17]

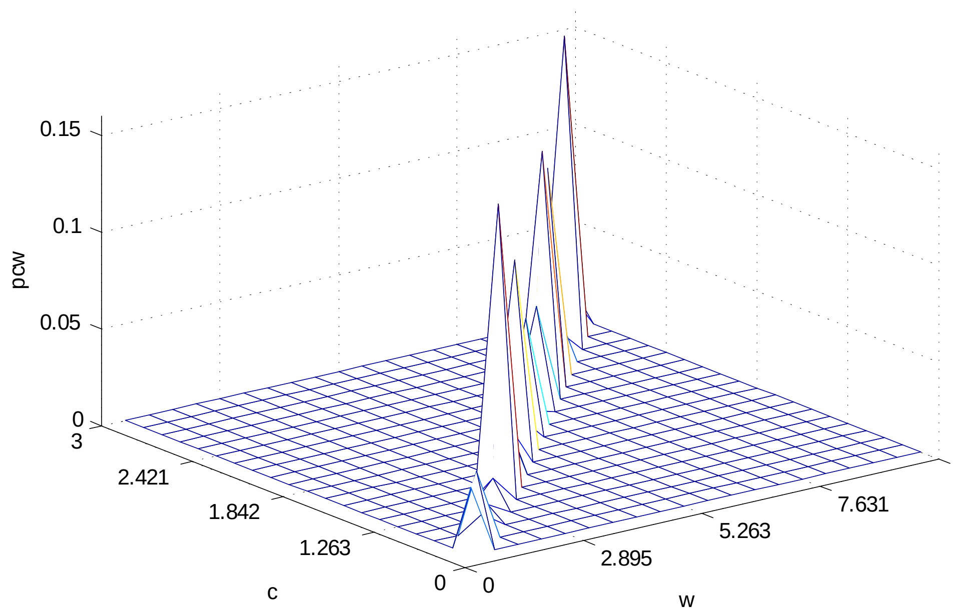

Figure 1: Joint pdf $p(c, w)$, high capacity.

Suppose now that capacity is low. In this case, rational consumers limit their processing effort by concentrating probability on the highest feasible value(s) of consumption.

[15] More formally, for $I(p(\cdot_w; \cdot_c)) \to \infty$, the probabilities $g(w)$ and $p(\cdot_w, \cdot_c)$ are degenerate. Using Fano's inequality (Thomas and Cover 1991),

$$c(\mathcal{I}(p(\cdot_w; \cdot_c))) = c(w)$$

which makes the first order conditions for this case the full information solution.

[16] Section **6** shows the optimal solution in terms of the optimal distribution of consumption conditional on wealth, i.e., $p^*(c|w)$.

[17] The distribution of wealth presented in Figure **1-4** corresponds to the same simplex point $g(w)$. In section 4, I provide details on how I construct the whole simplex. The way the simplex is build explains the reason why multimodality in the marginal distribution of wealth occurs. Here I want to emphasize that the selection of this particular simplex point is based on choosing the simplex points that put probability mass on most of the realizations considered.

To see why, recall that consumers are risk averse (log-utility). They process the necessary information to learn how much they can consume paying the least possible attention to wealth. Since Shannon's capacity places high restriction on information-processing, these people consume roughly the same amount each period independently of their level of wealth. This case captures situations in which people have a vague idea of their wealth and prefer default savings/spending options (whether it is pension plan or health care) rather than figuring out their exact net worth. Figure **2** displays the resulting optimal policy.

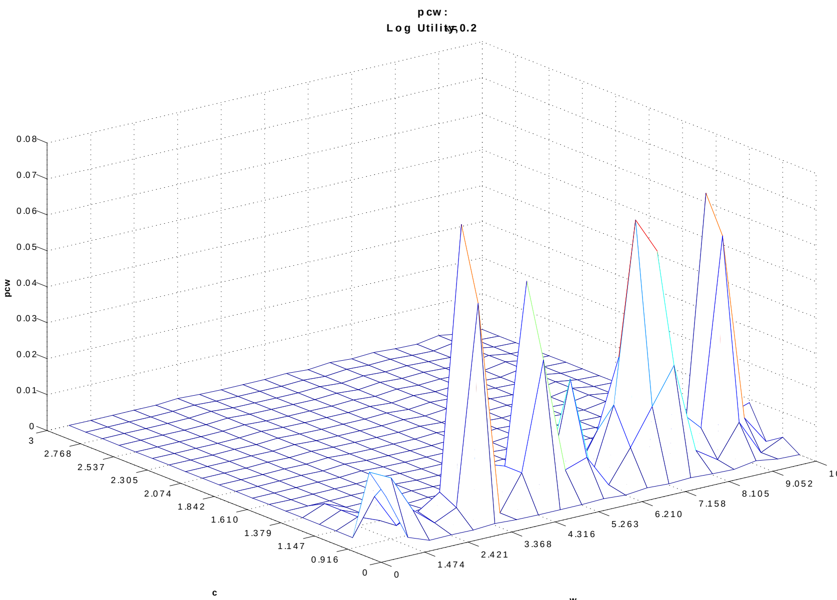

Figure 2: Joint pdf $p(c, w)$, low capacity.

Given the grid on c and w, log-utility and for the simplex point chosen, consumers find it optimal to set their consumption roughly constant and equal to the maximum value of consumption compatible with a value of wealth that allows them to keep consuming the same amount without running out of wealth. This result is robust to different specifications of the grid for c and w. [18]

Finally, Figure **3** shows the optimal joint distribution for an intermediate case, $0 < \kappa < \infty$. The first observation is that a person with finite information flow tries to make

[18] In particular, fixing all the other parameters equal and varying only the $20x20$ equi-spaced grid on c to $[0, 1.5]$, the solution puts 0.52 probability on the 0.95 value of consumption (vs. 0.46 on the highest value of consumption compatible to the lowest possible realization of wealth under the $[0.8; 3]$-consumption grid, $c = 1.031$ in Figure **1**). Moreover, the optimal solution gives 0.14 and 0.23 probability to the realizations $c = 1.026$ and $c = 0.8684$ respectively whereas the corresponding values for the $[0.8; 3]$-consumption grid, that is $c = 1.147$ and $c = 0.915$ get probabilities 0.18 and 0.25, respectively.

$p(c|w)$ as close to w as information process allows him to.

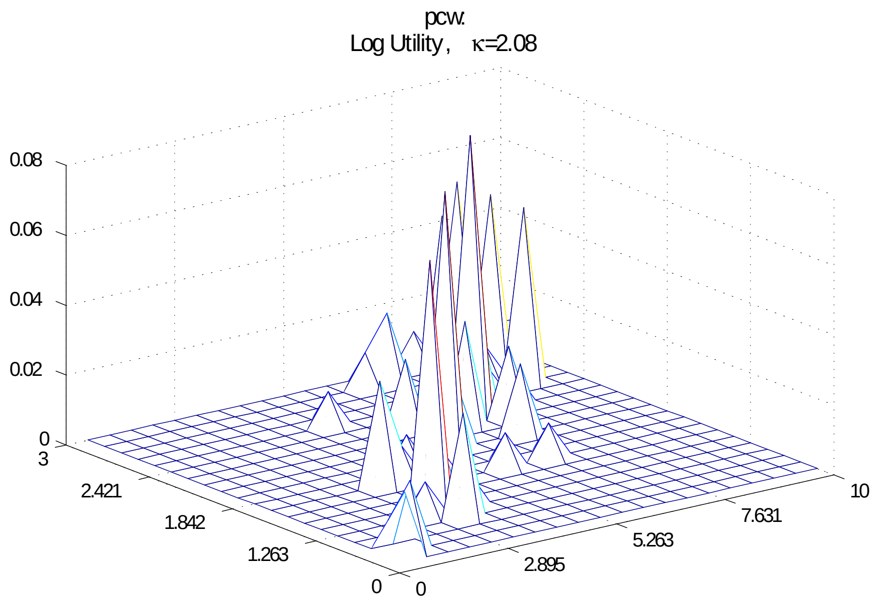

Figure 3. Joint distribution $p(c,w)$, intermediate capacity.

The second observation is that the optimal policy function for the information-constrained consumer places low weight, even zero, on low values of consumption for high values of wealth. and high value of consumption for low values of wealth. The reason why this happens depends on the utility function. A consumer with log-utility wants to maintain a fairly smooth consumption profile throughout his lifetime, as can be seen from (13). To avoid values of consumption that are either too low or too high, he needs to be well informed about such events in order to reduce the probability of their occurrence. The resulting optimal policy places higher probability mass on the central values of consumption and wealth.

To see how the allocation of probability changes with the utility function, consider a consumer that differs from the previous one only in the utility specification which now assumes a CRRA form, $u(c) = c^{1-\gamma}/(1-\gamma)$ with $\gamma = 2$. As in the previous case, the optimal policy function still places close-to-zero probability on low values of consumption for high values of wealth but now the CRRA consumer trades off probabilities about modest values of consumption and wealth for increasing the likelihood of high values of consumption for high values of wealth.

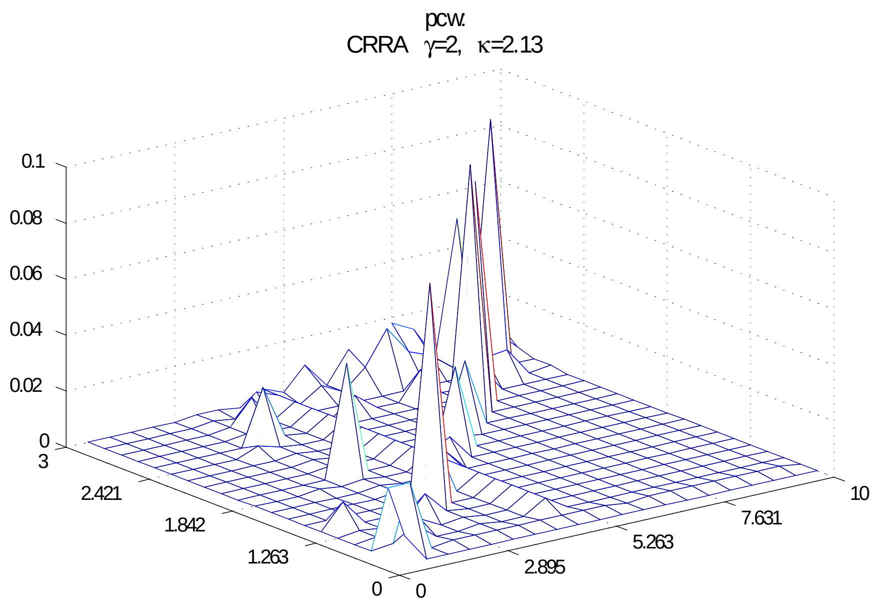

Figure 4. Joint distribution $p\,(c, w)$, CRRA utility.

In other words, with CRRA preferences, people want to be better informed on low and middle values of wealth to enjoy high consumption every period. Figure **4** illustrates this case. Also the results are robust for different specifications of the grid and all the simplex points. [19]

3.2.2 Shannon's channel through the economic literature

The goal of this section is to compare my model with the literature in rational inattention. The first comparison is with the consumption-saving model in the linear quadratic Gaussian (LQG) case [20] Sims(2003) fully characterizes the analytical solution of a consumption saving model where utility is quadratic, $u\,(c) = c - 0.5^{*}c^{2}$, constraints are linear and ex-ante uncertainty is Gaussian. In this LQG setting, the optimal distribution of ex-post uncertainty is also Gaussian. The Gaussian solution makes a model with rational inattention in the LQG case observationally equivalent to a signal extraction problem à la Lucas.

[19] In particular for the example in Figure **4**, when changing the realizations of consumption to be in an equispaced $20x20$ $[1, 12]$ grid, the optimal solution palces high probability on high value of wealth for high value of consumption assigning 0.73 to $p\,(c = 10|w = 10)$ instead of 0.61 for $p\,(c = 3|w = 10)$ under $[0.8; 3]$-grid, 0.11 to $p\,(c = 9.47|w = 10)$ instead of 0.12 for $p\,(c = 2.88|w = 10)$ under $[0.8; 3]$-grid, 0.01 to $p\,(c = 8.94|w = 10)$ instead of 0.07 for $p\,(c = 2.76|w = 10)$ under $[0.8; 3]$-grid. Moreover, the optimal solution sets $p\,(c = 11.42|w = 10) = p\,(c = 12|w = 10) = 0$.

[20] See Sims (1998, 2003), Luo (2008), Lewis (2007).

Note also that the analytical solution in Sims (2003) might not hold if one assumes a restriction in the support of either c or w (e.g., the conventional $c > 0$) or a no-borrowing constraint (e.g., $c_t \leq w_t \ \forall t$). This is because both constraints break the LQ framework, necessary to obtain Gaussianity in the optimal ex-post uncertainty.

The second issue with the LQG approach is that the linear quadratic approximation gives valid predictions when uncertainty is small. This is similar to the argument for linearizing the first order condition of a problem and getting locally a good approximation. However, if one wants to explain observed consumption and savings time series through limited processing constraints, the inertial behavior that we see in the data suggests that uncertainty is fairly big. Thus, the tractability of the LQG framework comes at the expenses of effectiveness in matching the data.

The third issue -the most important for the purpose of this paper- is that rational inattention LQG models do not allow to explain different speed and amounts of reactions of people to different news about their wealth. For instance, consumption drops fast following a sudden layoff while it increases slowly in the event of a tax break. The certainty equivalence framework that arises with Gaussian ex ante uncertainty and quadratic utility does not allow to differentiate endogenously among these events. This paper pushes forward another approach. The model in the paper assumes that information is freely available and it does not constraint ex-ante uncertainty to be Gaussian. Moreover, the paper explores the link between risk aversion and information-processing limits by allowing utility specifications of the CRRA family.

Previous to this paper, Sims (2006) solves a two period model with non-Gaussian ex-ante uncertainty and CRRA preferences. Sims (2006) assumes that agents live two periods, the first of which they are inattentive while the second period their uncertainty is resolved. This paper focusses on a fully dynamic rational inattention model. Two main contribution stems from departing from the work of Sims (2006). The first contribution is conceptual. A fully dynamic model with rational inattention allows to investigate time series properties of consumption and savings. The resulting behavior is characterized by endogenous noise and delays of consumption in response to shocks to income, with negative income shocks producing faster reactions the higher the risk aversion. The intuition for this result is that a risk averse individual reacts fast to negative news about wealth by dropping his consumption. Information-processing limits and prudence might want the person to wait before increasing his consumption in the event of a positive shock. Complementary to these findings, richer dynamic makes the model suitable to address policy questions such as reaction to a fiscal policy stimulus as the last section shows. This paper is also distinct from the one of Lewis (2008). The most prominent differences are that in Lewis (2008) households do not see consumption over time and they optimize over a finite horizon. Not observing consumption in turn implies that once the stream of probabilities is chosen at the beginning of time, the update of the beliefs is deterministic in the choice of the signal. While this framework does deliver upward-sloping age profiles as average consumption over a fixed time length, it does not allow to study unconditional moments of consumption nor conditional response of consumption to shocks as in this framework.

The second contribution is methodological. A fully dynamic rational inattention model involves facing an infinite dimensional problem as displayed in (8)-(12). To work with this framework, I develop analytical and computation tools suitable to address the dynamics of a non-LQG model.

Moreover, my results are observational distinct from the previous literature on sticky information (Mankiw and Reis, 2002) and consumption and information (Reis, 2006)). Mankiw and Reis(2002) assume that every period an exogenous fraction of agents (firms) obtain perfect information concerning all current and past disturbances, while all other firms set prices based on old information. Reis (2006) shows that a model with a fixed cost of obtaining perfect information can provide a microfoundation for this kind of slow diffusion of information. My model differs from the literature on inattentiveness in that I assume that information is freely available in each period. Only the bounds on information-processing given by Shannon's capacity limit consumers in the informativeness of the signals about wealth that they want to acquire. In this setting, the interaction of information flow and risk aversion delivers endogenous asymmetry in the response of consumption to shocks both in terms of speed and amount. This prediction constitutes a distinguished feature of my model with respect to the literature of inattentiveness and, more generally, to the consumption-saving literature.

4 Solution Methodology

4.1 Discretizing the Framework

I consider wealth and consumption to be defined on compact sets. In particular, admissible consumption profiles belong to $\Omega_c \equiv \{c_{\min}, ..., c_{\max}\}$. Likewise, wealth has support $\Omega_w \equiv \{w_{\min}, ..., w_{\max}\}$. I identify by j the elements of set Ω_c and by i the elements in Ω_w. I approximate the state of the problem, i.e., the distribution of wealth by using the simplex:

Definition The set Π of all mappings $g : \Omega_w \rightarrow \mathbb{R}$ fulfilling $g(w) \geq 0$ for all $w \in \Omega_w$ and $\sum_{w \in \Omega_w} g(w) = 1$ is called a **simplex**. Elements w of Ω_w are called *vertices* of the simplex Π, functions g are called *points* of Π.

Let $|S|$ be the dimension of the *belief simplex* which approximates the distribution $g(w)$ and let $\Gamma \equiv \left\{ g \in \mathbb{R}^{|S|} : g(i) \geq 0 \text{ for all } i \ \sum_{i=1}^{|S|} g(i) = 1 \right\}$ be the set of all probability distribution on Π. The initial condition for the problem is $g(w_0)$.

The consumer enters each period choosing the joint distribution of consumption and wealth. From the previous section, the control variable for the discretized set up is the probability mass function $\Pr(w, c)$ where $c \in \Omega_c$ and $w \in \Omega_w$. The restriction on

$\Pr(w,c)$ is that it has to belong to the set of distributions. Given $g(w_0)$ and $\Pr(c_t, w_t)$ and the observation of c_t consumed in period t, the belief state is updated using Bayesian conditioning:

$$g'\left(w_{t+1}|_{c_t}\right) = \sum_{w_t \in \Omega_w} T\left(w_{t+1}; w_t, c_t\right) \Pr\left(w_t | c_t\right) \tag{14}$$

where $T(.)$ is a discrete counterpart of the transition function $\tilde{T}(.)$. Note that $\tilde{T}(.)$ is a density function on the real line while $T(.)$ is a discrete probability function on a compact set with counting measure. The processing constraint, in terms of the discrete mutual information between state and actions, is.:

$$\mathcal{I}_t\left(p\left(\cdot_{c_t}, \cdot_{w_t}\right)\right) = \sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_c} \Pr\left(c_t, w_t\right) \left(\log \frac{\Pr\left(c_t, w_t\right)}{p\left(c_t\right) g\left(w_t\right)}\right) \tag{15}$$

The interpretation of (15) is akin to its continuous counterpart. The capacity of the agents to process information is constrained by a number, $\bar{\kappa}$, which denotes the upper bound on the rate of information flow between the random variables C and W [21] at time t. Finally, the objective function (8) in the discrete world is:

$$\max_{\{p(w_t, c_t)\}_{t=0}^{\infty}} E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left[\sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_w} \left(\frac{c_t^{1-\gamma}}{1-\gamma}\right) \Pr\left(c_t, w_t\right) \right] \middle| \mathcal{I}_0 \right\}. \tag{16}$$

4.2 Recursive Formulation

The purpose of this section is to show that the discrete dynamic programming problem under rational inattention has a solution and to recast it into a Bellman recursion. In order to show that a solution exists, first note that the set of constraints for the problem is a compact-valued concave correspondence. Second, note that the state space is compact. Compactness comes from the curvature of the utility function and the fact that the belief space has a bounded support in $[0,1]$. Compact domain of the state and the fact that Bayesian conditioning for the update preserves the Markovianity of the belief state ensures that the transition $Q : (\Omega_w \times Y \times \mathcal{B} \to [0,1])$ and (14) has the Feller property. Then the conditions for applying the Theorem of the Maximum are fulfilled which, in turn, guarantee the existence of a solution. In the next section, I provide sufficient conditions for uniqueness.

Casting the problem of the consumer in a recursive Bellman equation formulation, the full discrete-time Markov program is:

$$V\left(g\left(w_t\right)\right) = \max_{\Pr(c_t, w_t)} \left[\begin{array}{c} \sum_{w_t \in \Omega_w} \left(\sum_{c_t \in \Omega_c} u\left(c_t\right) \Pr\left(c_t, w_t\right) \right) + \\ + \beta \sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_c} V\left(g'_{c_t^j}\left(w_{t+1}\right)\right) \Pr\left(c_t, w_t\right) \end{array} \right] \tag{17}$$

[21] Recall from the argument in *Secton 2.1* that both W *and* C are random variables before the household has processed any information.

subject to:

$(\theta :)$

$$\kappa_t = \mathcal{I}_t\left(p\left(\cdot_{c_t}, \cdot_{w_t}\right)\right) = \sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_c} \Pr\left(c_t, w_t\right)\left(\log \frac{\Pr\left(c_t, w_t\right)}{p\left(c_t\right) g\left(w_t\right)}\right) \tag{18}$$

$$g'\left(w_{t+1}|_{c_t}\right) = \sum_{w_t \in \Omega_w} T\left(w_{t+1}; w_t, c_t\right) \Pr\left(w_t | c_t\right) \tag{19}$$

$$\sum_{c_t \in \Omega_c} \Pr\left(c_t, w_t\right) = g\left(w_t\right) \tag{20}$$

$$1 \geq \Pr\left(c_t, w_t\right) \geq 0 \; \forall \left(c_t, w_t\right) \in B, \forall t \tag{21}$$

where $B \equiv \{(c_t, w_t) : w_t \geq c_t,\ \forall c_t \in \Omega_c, \forall w_t \in \Omega_w,\ \forall t\}$ and θ is the shadow cost associated to (18).

The Bellman equation in (17) takes up as argument the marginal distribution of wealth $g\left(w_t\right)$ and uses as control variable the joint distribution of wealth and consumption, $\Pr\left(c_t, w_t\right)$. The latter links the behavior of the agent with respect to consumption (c), on one hand, and income (w) on the other, hence specifying the actions over time. The first term on the right hand side of (17) is the utility function $u\left(.\right)$. The second term, $\sum\limits_{w_t \in \Omega_w} \sum\limits_{c_t \in \Omega_c} V\left(g'_{c_t^j}\left(w_{t+1}\right)\right) \Pr\left(c_t, w_t\right)$, represents the expected continuation value of being in state $g\left(.\right)$ discounted by the factor $\beta = 1/R$. The expectation is taken with respect to the endogenously chosen distribution $\Pr\left(c_t, w_t\right)$. I have discussed the relations in (18)-(21) earlier. Moreover, I appended the equation in (20) which constrains the choice of the distribution to be consistent with the initial prior $g\left(w_t\right)$.

Next, I analyze the main properties of the Bellman recursion (17) and derive conditions under which it is a contraction mapping and show that the mapping is isotone. The optimality conditions that characterize the solution to the problem (17)-(21), along with a special case that admits close form solution are in Appendix D in the addendum.

4.3 Properties of the Bellman Recursion

To prove that the value function is a contraction and isotonic mapping, I shall introduce the relevant definitions. Let me restrict attention to choices of probability distributions that satisfy the constraints (18)-(21).To make the notation more compact, let $p \equiv \Pr\left(c_j | w_i\right)$, $\forall c_j \in \Omega_c$, $\forall w_i \in \Omega_w$ and let Γ be the set that contains (18)-(21). I introduce the following definitions:

D1. A control probability distribution $p \equiv \Pr\left(c_i, w_j\right)$ is ***feasible*** for the problem (17)-(21) if $p \in \Gamma$. Let $|W|$ be the cardinality of Ω_w and let

$$\mathcal{G} \equiv \left\{ g \in \mathbb{R}^{|W|} : g\left(w_i\right) \geq 0,\ \forall i,\ \sum_{i=1}^{|W|} g\left(w_i\right) = 1 \right\}$$

denote the set of all probability distributions on Ω_w. An optimal policy has a value function that satisfies the Bellman optimality equation in (17):

$$V^*(g) = \max_{p\in\Gamma}\left[\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c} u(c)\, p(c|w)\right) g(w) + \beta \sum_{w\in\Omega_w}\sum_{c\in\Omega_c}\left(V^*\left(g'_c(\cdot)\right)\right) p(c|w)\, g(w)\right] \tag{22}$$

The Bellman optimality equation can be expressed in value function mapping form. Let $\mathcal{V}$ be the set of all bounded real-valued functions V on $\mathcal{G}$ and let $h : \mathcal{G} \times \Omega_w \times (\Omega_w \times \Omega_c) \times \mathcal{V} \to \mathbb{R}$ be defined as follows:

$$h(g, p, V) = \sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c} u(c)\, p(c|w)\right) g(w) + \beta \sum_{w\in\Omega_w}\sum_{c\in\Omega_c}\left(V\left(g'_c(\cdot)\right)\right) p(c|w)\, g(w).$$

Define the value function mapping $H : \mathcal{V} \to \mathcal{V}$ as $(HV)(g) = \max_{p\in\Gamma} h(g, p, V)$.

D2. A value function V ***dominates*** another value function U if $V(g) \geq U(g)$ for all $g \in \mathcal{G}$.

D3. A mapping H is ***isotone*** if $V, U \in \mathcal{V}$ and $V \geq U$ imply $HV \geq HU$.

D4. A ***supremum norm*** of two value functions $V, U \in \mathcal{V}$ over $\mathcal{G}$ is defined as

$$||V - U|| = \max_{g\in\mathcal{G}} |V(g) - U(g)|$$

D5. A mapping H is a ***contraction under the supremum norm*** if for all $V, U \in \mathcal{V}$,

$$||HV - HU|| \leq \beta\, ||V - U||$$

holds for some $0 \leq \beta < 1$.

Endowed with these notion, it is possible to derive some properties of the solution to the Bellman equation.

First, note that uniqueness of the solution to which the value function converges to requires concavity of the constraints and convexity of the objective function. It is immediate to see that all the constraints but (18) are actually linear in $p(c, w)$ and $g(w)$. For (18), the concavity of $p(c, w)$ is guaranteed by Theorem (16.1.6) of Thomas and Cover (1991). Concavity of $g(w)$ is the result of the following:

Lemma 1. For a given $p(c|\, w)$, the expression (18) is concave in $g(w)$.

Proof. See Appendix B. ■

Next, I need to prove convexity of the value function and the fact that the value iteration is a contraction mapping. All the proofs are in Appendix A

Proposition 1. *For the discrete Rational Inattention Consumption Saving value recursion H and two given functions V and U, it holds that*

$$||HV - HU|| \leq \beta \, ||V - U|| \, ,$$

with $0 \leq \beta < 1$ and $||.||$ the supreme norm. That is, the value recursion H is a contraction mapping.

Proposition 1 can be explained as follows. The space of value functions defines a *vector space* and the contraction property ensures that the space is *complete*. Therefore, the space of value functions together with the supreme norm form a *Banach space* and the *Banach fixed-point theorem* ensures (a) the existence of a single fixed point and (b) that the value recursion always converges to this fixed point (see Theorem 6 of Alvarez and Stockey, 1998 and Theorem 6.2.3 of Puterman, 1994).

Corollary *For the discrete Rational Inattention Consumption Saving value recursion H and two given functions V and U, it holds that*

$$V \leq U \Longrightarrow HV \leq HU$$

that is the value recursion H is an isotonic mapping.

The isotonic property of the value recursion ensures that the value iteration converges *monotonically.*

These theoretical results establish that in principle there is no barrier in defining value iteration algorithms for the Bellman recursion for the discrete rational inattention consumption-savings model.

5 Numerical Technique and its Predictions

I solve the model by transforming the underlying partially observable Markov decision process into an equivalent, fully observable, Markov decision process with a state space that consists of all probability distributions over the core [22] state of the model (wealth).

For a model with n core states, $w_1, .., w_n$, the transformed state space is the $(n-1)$-dimensional simplex, or *belief simplex.* Expressed in plain terms, a belief simplex is a point, a line segment, a triangle or a tethraedon in a single, two, three or four-dimensional space, respectively. Formally, a belief simplex is defined as the *convex hull*[23] of belief

[22] The state of the model is a probability distribution of wealth, i.e., $g(w)$. In lack of a better alternative, I call *core* state the random variable w whose distribution is the state of the model. This nomenclature is borrowed from information theory and AI literature. See Puterman (1994) .

[23] A *convex hull* of a set of points is defined as the closure of the set under convex combination.

states from an *affinely independent*[24] set B. The points of B are the vertices of the belief simplex. The convex hull formed by any subset of B is a face of the belief simplex. To address the issue of dimensionality in the state space of my model, I use a grid-based approximation approach. The idea of a grid based approach is to use a finite grid to discretize the uncountably infinite continuous state space. The implementation has the following steps: (1) place a finite grid over the simplex point, (2) compute the values for points in the grid. I use a kernel regression to interpolate solution points that fall outside the grid.

5.1 Belief Simplex and Dynamic Programming

Had infinite information-processing rate been available, previous history of the process would have been irrelevant to the problem. However, since the consumer cannot completely observe wealth, he may require all the past information about the system to behave optimally. The most general approach is to keep track of the entire history of previous consumption purchases up to time t, denoted $H_t = \{g_0, c_1, .., c_{t-1}\}$. For any given initial state probability distribution g_0, the number of possible histories is $(|\mathcal{C}|)^t$ with $\mathcal{C}$ denoting the set of consumption behavior up to time t. This number goes to infinity as the decision horizon approaches infinity, which makes this method of representing history useless for infinite-horizon problems.

To overcome this issue, Astrom (1965) proposed an *information state* approach. It is based on the idea that all the information needed to act optimally can be summarized by a vector of probabilities over the system, the *belief* state. Let $g(w)$ denote the probability that the wealth is in state $w \in \Omega_w$ where Ω_w is assumed to be a finite set. Probability distributions such as $g(w)$ defined on finite sets are in fact *simplices*. Let n be the possible values that w can assume. The discretization of the core state is an equi-spaced grid with $n = 20$ values of w ranging from 1 to 10. The points in the simplex Δ are n distinct values for the marginal pdf $g(w)$ in the interval $I \equiv [0, 1]$. The simplex is constructed using uniform random samples from the unit simplex. The rational for this methodology is that it is computationally faster than non-uniform grid and it is able to handle higher dimensional space.[25] In the current setting, each point in the simplex is an n-array whose column contains m random values in the $[0, 1]$ range and whose sum per row is 1. To span the simplex I use $m = (n-1)!$.[26] The distribution of values within the simplex is uniform in the sense that it has the conditional probability of a uniform distribution over the whole m-cube, given that the sum per row is 1. The algorithm calls three types of random processes that determine the placement of random points

[24] A set of belief states $\{g_i\}$, $1 \leq i \leq z$ is called *affinely independent* when the vectors $\{g_i - g_z\}$ are linearly independent for $1 \leq i \leq z$.

[25] At least compared to the ndgrid library functions in Matlab. This is because the algorithm creates the simplex directly while when using ndgrid it is necessary to define a uniform grid over the whole $n-1$ space and then sectioning the resulting grid so that each simplex point sum to one.

[26] With $n = 20$, the proposed sampling produces the same results for sample size of $m = (n-k)!$, for $k = 1, .., 5$. I have not tried cases with $k > 5$. When $k > 1$, even if the algorithm produces the same results it takes longer to converge (about 3 minutes more per iteration).

in the $n-1-$dimensional simplex. The first process considers values uniformly within each simplex. The second random process selects samples of different types of simplex in proportion to their volume. Finally, the third implements a random permutation in order to have an even distribution of simplex choices among types.

For each simplex point, I initialize the corresponding joint distribution of consumption c and wealth w. I assume $n = 20$ equi-spaced values for c ranging in $\Omega_c \equiv [0.8, 3]$. The values in Ω_c are chosen so that w is about 3 times c, roughly consistent with individual data on consumption and wealth.

Let core states and behavior state be sorted in descending order. Then, given the symmetry in the dimensionality of Ω_c and Ω_w, the joint distribution of consumption and wealth for a given multidimensional grid point is a square matrix with rows corresponding to levels of consumption . Summing the matrix per row returns the marginal distribution of consumption, $p(c)$. Likewise, the columns of the matrix correspond to levels of wealth. Evaluating the sum per columns of the matrix returns the marginal distribution of wealth, $g(w)$. Given the initial belief simplex, its successor belief state can be determined by Bayesian conditioning at each multidimensional point of the simplex. The resulting expression is:

$$g\left(w'|_c\right) = \sum_i T\left(w'; w_i, c\right) \Pr\left(w_i|c\right). \tag{23}$$

Without loss of generality, I restrict the columns of the matrix $\Pr(c, w)$ to sum to the marginal pdf of wealth in the main diagonal. Moreover, since some of the values of the marginal $g(w)$ per simplex-point are exactly zero given the definition of the envelope for the simplex, I constrain the choices of the joint distributions corresponding to those values to be zero. This handling of the zeroes implies that the parameter vectors being optimized over have different lengths for different rows of the simplex . Hence the degrees of freedom in the choice of the control variables for simplex points vary from a minimum of 0 to a maximum of $\frac{n*(n-1)}{2}$. Once the *belief simplex* is set up, I initialize the joint probability distribution of consumption and wealth per belief point and solve the program of the household by backward induction iterating on the value function $V(g(w))$. To map the finer state space into Matlab possibilities, I interpolate the value function with the new values of (23) using a kernel regression of $V(\cdot)$ into $g'\left(w'|_a\right)$. I use an Epanechnikov kernel with smoothing parameter $h = 2.7$. [27] A kernel regression approximates the exact non linear value function in (17) with a piece-wise linear function. The following propositions illustrate this point.

Proposition 2. *If the utility is CRRA with parameter of risk aversion* $\gamma \in (0, +\infty)$ *and if* $\Pr(c_j, w_i)$ satisfies (18)-(21)., *then the optimal* $n-$ *step value function* $V_n(g)$

[27] Epanechnikov kernel is optimum choice for smoothing because it minimizes asymptotic mean integrated squared error (cfr. Marron, J. S. and Nolan, D. (1988)). I use the algorithm proposed in Beresteanu, A. and C. F. Manski (2000) and experiement with smoothing paramter $h \in [0.3, .., 4.2]$. For the charactersitcs of the problem and the optimization routine used (csminwel) and for different specification of utility functions and Lagrange multiplier θ, the parameter $h = 2.7$ performs best in terms of computational time and convergence of the value function.

defined over G can be expressed as:

$$V_n(g) = \max_{\{\alpha_n^i\}_i} \sum_i \alpha_n(w_i)\, g(w_i)$$

where the $\alpha -$ vectors, $\alpha : \Omega_w \rightarrow R$, are $|W|-$dimensional hyperplanes.

Intuitively, each α_n-vector corresponds to a plan and the action associated with a given α_n-vector is the optimal action for planning horizon n for all priors that have such a function as the maximizing one. With the above definition, the value function is:

$$V_n(g) = \max_{\{\alpha_n^i\}_i} \left\langle \alpha_n^i, g \right\rangle,$$

and thus the proposition holds.

Using the above proposition and the fact that the set of all consumption profiles $\mathbf{P} \equiv \{c < w : p(c) > 0\}$ is discrete, it is possible to show directly the convex properties for the value function. For fixed α_n^i-vectors, $\langle \alpha_n^i, g \rangle$ operator is linear in the belief space. Therefore the convex property is given by the fact that V_n is defined as the maximum of a set of convex (linear) functions and, thus, obtains a convex function as a result. The optimal value function V^* is the limit for $n \rightarrow \infty$ and, since all the V_n are convex function, so is V^*.

Proposition 3. *Assuming CRRA utility function and under the conditions of Proposition 1, let V_0 be an initial value function that is piecewise linear and convex. Then the i^{th} value function obtained after a finite number of update steps for a rational inattention consumption-saving problem is also finite, piecewise linear and convex (PCWL).*

To implement the optimization of the value function at each point of the simplex, I use Sims' CSMINWEL as gradient-based search method and iterate on the value function up to convergence. The value iteration converges in about 202 iterations. Table 1 reports the benchmark parameter values and the discretization of the grid.

I simulate the model for $T = 80$ periods by drawing from the optimal policy function, $p^*(c, w)$, and generate time series path of consumption, wealth and expected wealth. For each $t = 1, .., T$, I use the joint distribution $p_t^*(c, w)$ to evaluate the time path of information flow ($\kappa_t^* \equiv \sum_i \sum_j p_t^*(c_j, w_i) \log\left(\frac{p_t^*(c_j, w_i)}{p_t^*(c_j) g_t^*(w_i)}\right)$). Finally, I derive impulse response functions for the economy by assuming temporary shocks to the mean of income, $\bar{y}$. A pseudocode that implements the procedure is in Appendix C.

Benchmark Values

	Discretization
Wealth Space W	$[1, .., 10]$
Consumption Space C	$[0.8, .., 3]$
Mean of Income, $\bar{y}$	1.1
Joint Distribution per simplex point, $p(c, w)$	20×20
Marginal C	20×1
Marginal W	20×1
Coeff. risk aversion, γ	1
Interest rate, R	1.012
Discount Factor, β [36]	0.9881

Table 1

6 Results

In this section I investigate the dynamic interplay of information flow and degree of risk aversion. In particular, I study different specifications of the model changing degrees of risk aversion, $\gamma \in \{0.5, 1, 2, 5, 7\}$, and different shadow costs of processing information, $\theta \in (0, 4]$, attached to (18).[37] Time paths for each individual are average across simplex points. For the time series of the aggregate economy, I perform 10,000 Monte Carlo runs and simulate the model for each path for $T = 80$ periods. Then, I take average across runs and simplex-points. Sample statistics are calculated based on these averages. I choose this way of calculating averages to compare my model, tailored for individual behavior, to aggregate data. I divide the results into three parts: 1. interaction of information flow and risk aversion; 2. implications of information constraint on lifetime

[37] I choose the set of $\gamma = \{0.5, 1, 2, 5, 7\}$ as these values for the coefficient of relative risk aversion appear to be the most common in the literature. As far as the choices of θ are concerned, first note that $\theta = 0$ corresponds to the full information case and for $\theta > 3$ the results are the ones states in section 3.1: consumers choose the maximum consumption compatible with the minimum value of wealth and avoid processing information. For the values of θ within the interval $(0, 4]$ and in order to get a sense of the magnitude involved, the difference in expected lifetime consumption with respect to the full information case for a log utility person is $E(c|\,\theta = 0) - E(c|\,\theta = 2) = 1.124 - 1.08 = 0.044$ units of consumption or 4% $(= [E(c|\,\theta = 0) - E(c|\,\theta = 2)]/E(c|\,\theta = 0))$ when $\theta = 0.2$. Under $\theta = 2$ the difference becomes 0.214 units of consumption or 19%.

consumption; 3. consumption's reactions to temporary income shocks.

Statistics

θ=0.2	CRRA $\gamma = 7$	CRRA $\gamma = 5$	Log Utility	CRRA $\gamma = .5$
$E(C)$	1.14	1.09	1.08	1.02
$std(C)$	0.08	0.09	0.11	0.14
κ	2.03	1.99	1.87	1.72

θ=2	CRRA $\gamma = 7$	CRRA $\gamma = 5$	Log Utility	CRRA $\gamma = .5$
$E(C)$	1.01	0.98	0.91	0.83
$std(C)$	0.15	0.18	0.21	0.33
κ	1.41	1.20	0.86	0.78

Table 2

Result 1. Information flow and risk aversion *In the discrete rational inattention consumption-savings model, higher degrees of risk aversion result in higher amount of information processed for a given information-processing cost. Moreover, for a given degree of risk aversion, the relation between information-processing cost and volatility of consumption is non-monotonic.*

With respect to the non monotonic relation between shadow cost of processing information (θ) and volatility of consumption, fixing the degree of risk aversion, γ, consider Figure 5b. When the cost of processing information tends to zero or exceeds the upper bound of the consumption grid ($\theta > \bar{c}$, $\bar{c} = 3$, for the current grid), the volatility of consumption tends to zero. When θ ranges in the interval $(0, \bar{c}]$, the higher the cost of processing information, the higher the volatility of consumption is. This result is robust to different specifications of consumption grid -i.e., different upper bound, $\bar{c}$-. The finding is documented in Table 2 and Figures 5(a,b)-6. Figure 5a plots the difference between the mean of the time series of consumption between $\theta = 0$ and $\theta > 0$. After deriving the time path of consumption, I compute the mean of the average across paths and subtract the corresponding mean for the full information solution ($\theta = 0$).[38] Figure 5a shows how this difference changes as θ varies and when utility is logarithmic. Figure 5b plots the corresponding difference in standard deviation of consumption as a function of θ. For high shadow cost of processing information, $\theta > 3$, consumption does not vary over time.[39] When θ is strictly bigger than the upper bound on consumption, an agent with concave utility rationally prefers to give up processing information and consume very little rather than spending all his efforts processing information without enjoying consumption.[40] For

[38] For the parameter of the model, the solution with infinite rate of information-processing ($\theta = 0$), $c_t^f = \beta w_t + (1-\beta)\bar{y}$, has mean $E\left(c_t^f\right) = 1.124$ and standard deviation $std\left(c_t^f\right) = 0.0713$.

[39] Recall that $\bar{c} = 3$ is the maximum amount of consumption in the grid proposed.

[40] As an example, consider a person who check constantly his account balance in order to increase his current consumption profile. If he spends all his time and effort crunching numbers, not only he will not

$0 < \theta < 3$, volatility of consumption increases with θ. This result makes sense. To see why, consider the full information version of the model. With $R\beta = 1$, the agent wants to smooth consumption. People's will to smooth consumption in full information is limited by a finite flow of information available. When deciding the precision of their signals, risk averse people trade off lower volatility in consumption for better knowledge of low value of wealth.

Loss in consumption due to increasing processing effort

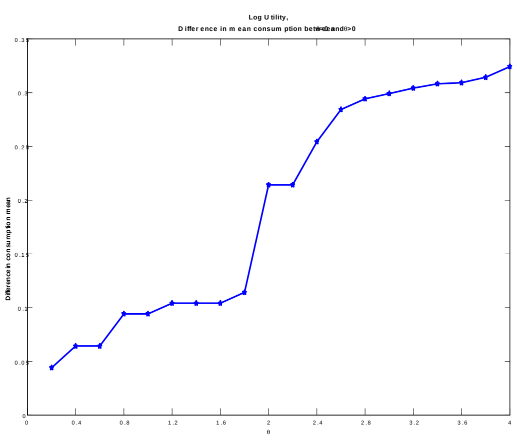

Figure 5a.

Difference in std. of consumption due to processing effort.

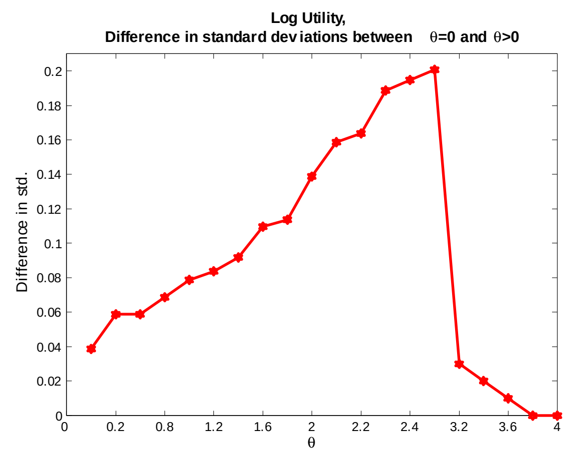

Figure 5b.

To appreciate how preferences toward risk play out with processing limits ($\theta > 0$), consider Figure 6b. It plots the distribution of consumption that results from the optimal choice $p^*(c, w)$ conditional on three particular values of w -$w = 1$, $w = 5.3$ and $w = 9.5$- for two individuals : $\gamma \to 1$ -represented by bars in the figure- and $\gamma = 5$ -represented by dotted line in the figure- when information is very costly to process ($\theta = 3$). In this case, a rational agent places high probabilities on consuming a given amount every period compatible with his wealth. Such a choice implies little attention and little dependence of consumption to wealth, for it requires figuring out the limits of wealth and consume the same amount from then on. The first panel of Figure 6b shows the optimal distribution of consumption when $w = 1$.Note how a person with log utility places higher probability on low values of consumption with respect to a person with CRRA utility and $\gamma = 5$. This result owns to two factors. First, agents with higher degree of risk aversion tend to process more information than people with lower degree of risk aversion do. Thus, more risk averse types have access to more precise signals about wealth than the other group. Second, people with CRRA, $\gamma = 5$ tend to avoid lower values of consumption for a given value of wealth whereas a log utility person prefers keeping his consumption steady along

be able to increase his consumption but also by sitting on his computer and working all day he will not consume at all.

his life-time, even at the expenses of some consumption units. Given $w = 1$, preferences' specification and choice of information flow jointly lead the consumer with $\gamma \rightarrow 1$ to assign probability 0.63 to the lowest value of consumption ($c = 0.8$) and 0.28 to $c = 0.92$ whereas the agent with $\gamma = 5$ sets probability 0.18 to $c = 0.8$ and 0.71 to $c = 0.92$. The second and third panels of Figure 6b show the optimal distribution of consumption conditional on $w = 5.3$ and $w = 9.5$ respectively. Both panels show a pattern akin to the one discussed for the optimal distribution of consumption conditional on $w = 1$. The expected value of consumption is higher in all cases for the person with $\gamma = 5$ given that he assigns more probability mass to higher values of consumption than the log-utility person does.

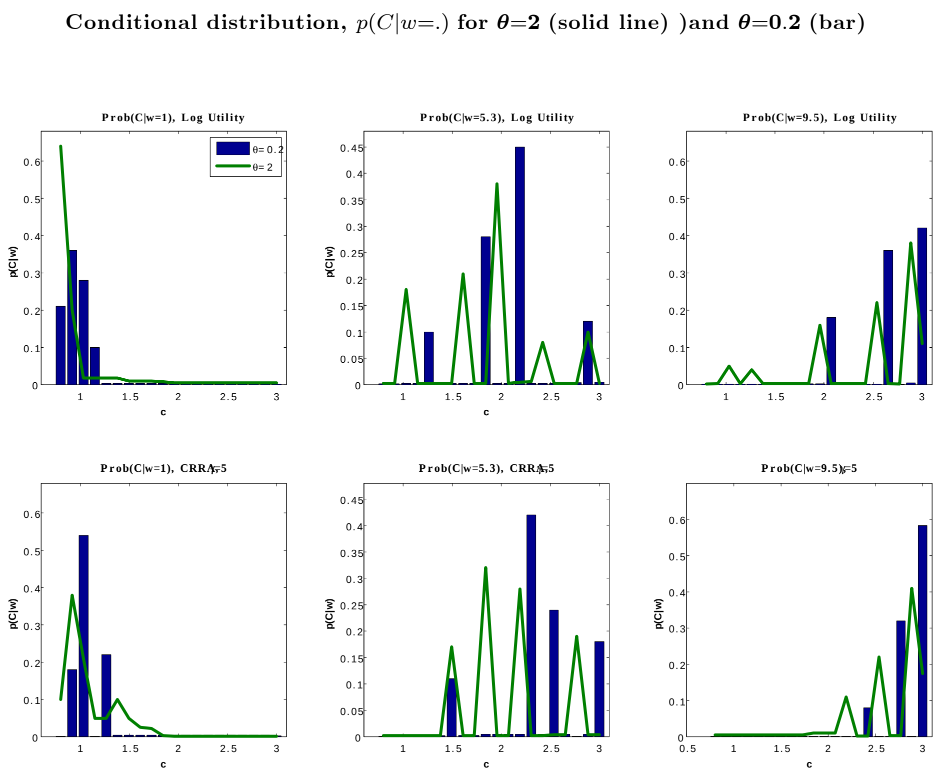

Figure 6a

Conditional distribution of Consumption, $p(C|w)$**,** θ=3. Log utility (bar) and CRRA γ=5 (dotted line)

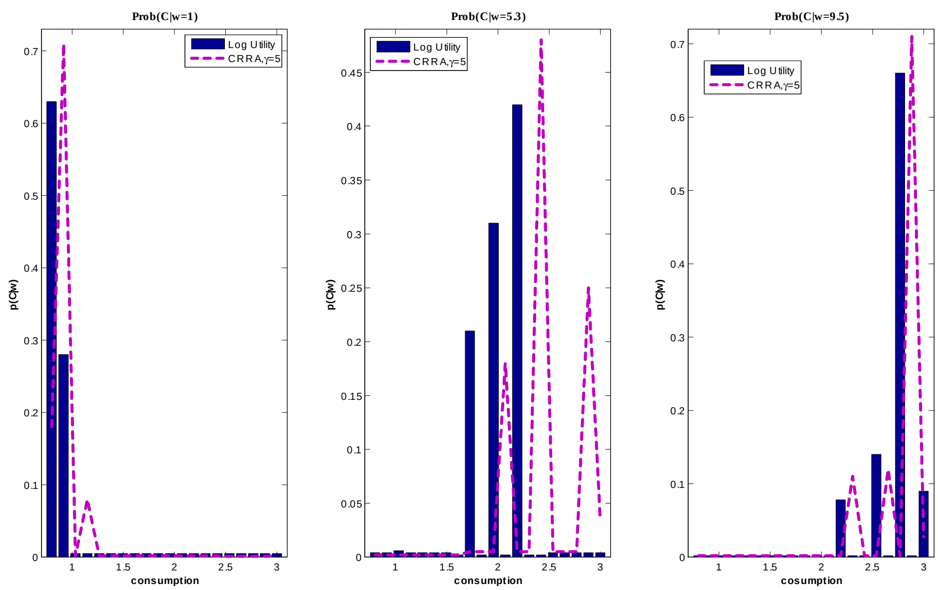

Figure 6b

Now consider Table 2 and Figures 6a The top panel of figure 6a displays the conditional distribution of consumption for an agent with log utility whereas the bottom panel displays the same distribution for an agent with CRRA utility, $\gamma = 5$. The first row of the figure shows the optimal distribution of consumption conditional on $w = 1$, the second row shows the optimal distribution conditional on $w = 5.3$ and in the third row shows the conditional distribution of consumption for a value of wealth $w = 9.5$. In both panels, bars identify the optimal choices when $\theta = 0.2$ while solid lines represent $\theta = 2$. Consider the case with $\theta = 2$. For this case, the higher the degree of risk aversion, the higher the information flow (κ) is. This is exactly what Table 2 shows. In the table, the higher the coefficient of risk aversion, γ, the higher the information processed by the agent, κ, the higher the mean of consumption. As Figures 6a displays, the same story can be told in terms of conditional probability distributions. In the top panel, people with low shadow costs of processing information ($\theta = 0.2$, bars in the figure) and log utility enjoy higher value of consumption for a given $w \in (\{1\}, \{5.2\}, \{9.5\})$ than their $\theta = 2$ counterparts. This result occurs because lower information processing costs allow for more precise signals on wealth. For instance, when $w = 5.3$ a person with log utility and $\theta = 0.2$ knows that he can afford consuming on average $E(C|w = 5.3) = 2.1$ whereas the same person with $\theta = 2$ expects to consume about 14% less than the better informed type -or, $E(C|w = 5.3) = 1.8$-. Likewise, in the bottom panel of Figure 6a, a person with CRRA utility, $\gamma = 5$ and a shadow cost of processing information of $\theta = 0.2$ has an expected value of consumption $E(C|w = 5.3) = 2.4$ when $w = 5.3$ whereas had he had $\theta = 2$, his expected value would have dropped to $E(C|w = 5.3) = 2.0$.

To sum up, for a given level of θ, a person with log utility will be better informed on extreme values of wealth in order to avoid such values. This knowledge makes it possible to assign high probability to middle value of consumption for a given value of wealth. By contrast, a consumer with CRRA, $\gamma = 5$, wants to avoid low values of consumption for high values of wealth. Processing information about these events decreases the likelihood of their occurrence and makes it possible to place high probability on high value of consumption.

The results in Figures 6a and 6b and Table 2 build up intuition on the time series properties of consumption, wealth and information flow drawn from the optimal distributions. Figure 6a tells us that people with higher information flow tend to concentrate probability mass on fewer values of consumption whereas the optimal distribution of people with higher shadow costs, θ, have wider dispersion than that of less information constrained types. Moreover, people with higher degree of risk aversion tend to avoid lower values of consumption for a given value of wealth. This mechanism results in higher expected value of consumption for those people (see Figure 6a-6b and Table 2). Combining these findings, one expects people with lower information flow and lower risk aversion to vary their consumption profile less frequently than the other types. Moreover, when these types decide to change consumption after waiting to process information about wealth, they do so by a sizeable amount since in the meanwhile they have accumulated (or decreased) savings. Such a behavior produces higher volatility of consumption for people with lower information flow.

Time series paths of consumption, wealth and information flow drawn from the optimal policy $p^*(c, w)$ confirm this intuition and offer further insights on the properties of the model.

Aggregate Consumption

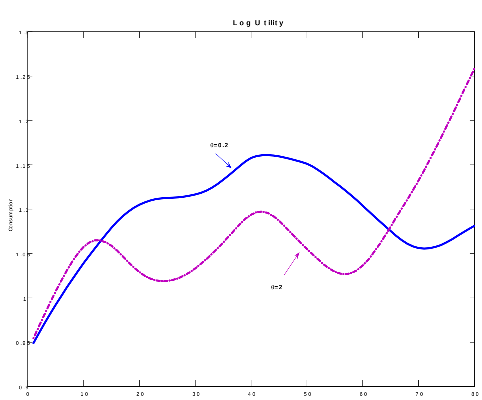

Figure 7a.

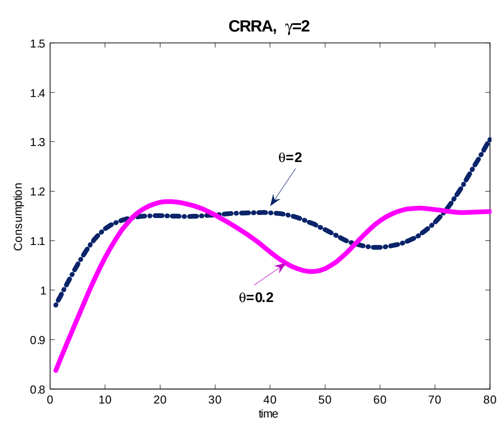

Figure 7b.

Aggregate Consumption and Information Flow

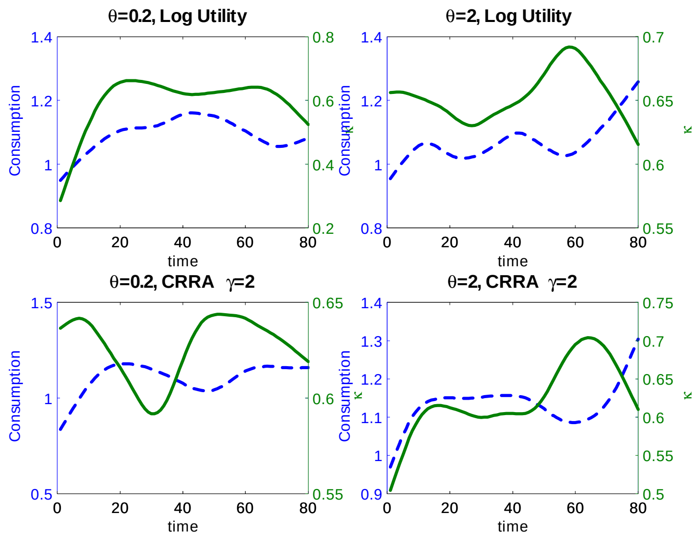

Figure 7c.

Result 2. Time path of consumption and savings. *Changes in consumption over time are infrequent and significant. Moreover:*

1. *People with low information flow delay substantial increases in consumptions more than people with high processing capacity. The effect is stronger the higher the degree of risk aversion.*

2. *People with high information flow have savings behavior that follows closely their wealth because they have sharper signals on wealth. Furthermore, the lower the degree of risk aversion, the higher the fluctuations of savings per period.*

3. *People with low information flow tend to consume a constant amount every period. They increase their consumption only if the information they process points them towards a significant increase in wealth. The higher the degree of risk aversion, the less volatile the time path of consumption for those types is.*

Figures 7(a,b,c)-8(a,b) illustrate these points for aggregate and individual time series behaviors, respectively. The simulations are derived by drawing time path of consumption and wealth from $p^*(c, w)$, after the value iteration has converged.[41] To have some

[41] Figures 7a-7c plot average across Monte Carlo runs and simplex points -i.e., initial beliefs about wealth-. Figure 7a shows aggregate consumption for a person with log utility with $\theta = 2$ -dotted line- and $\theta = 0.2$ - full line- and Figure 7b displays aggregate consumption under high -$\theta = 2$, dotted line- and low -$\theta = 0.2$, solid line- information processing cost for a person with CRRA utility, $\gamma = 2$. Figure 7c shows aggregate consumption - left axes, dotted line- and information flow -right axes, solid line-.

interesting transitional dynamics, I begin the simulation with an initial condition for wealth far from the steady state[42].

Individual consumption

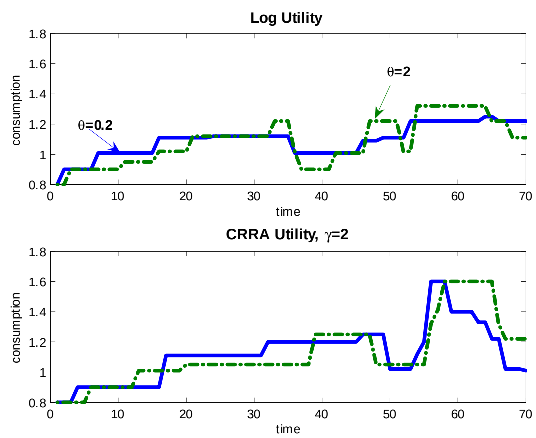

Figure 8a.

Individual savings and wealth

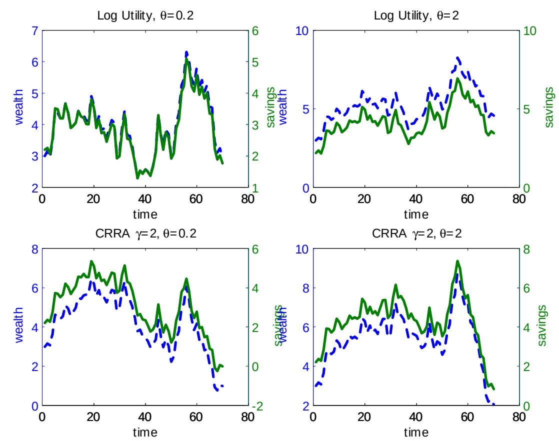

Figure 8b.

To appreciate the results, consider what would happen under infinite processing rate. In such a case, consumption smoothing ($R\beta = 1$) implies an immediate ($T = 1$) adjustment of consumption to its long-run optimal value and no transient behavior. Thus, from $T = 2$ onwards, the simulations lead to a constant time path. Now consider Figures 7(a-c)-8(a,b). The hump in consumption comes from Result 1 and a simple intuition: information-constrained people are cautious (degree of risk aversion $\gamma \geq 1$), consume a little and collect information about wealth before they change consumption. For a fixed θ, the more risk averse they are (see Figure 7a with log utility and Figure 7b with CRRA, $\gamma = 2$), the longer they wait before increasing consumption. This inertial behavior in consumption leads to an increase in savings and, as a result, in wealth (see Figure 8a-8b).

In the graph on the top left corner I show consumption and information flow for log utility and $\theta = 0.2$. while in the same row, the graph on the top right corner shows $\theta = 2$ and log utility. On the bottom row, the graph on the bottom left corner displays the results for CRRA, $\gamma = 2$ and $\theta = 2$ while the graph at the bottom right corner shows the corresponding results for CRRA with $\gamma = 2$ and $\theta = 0.2$. Individual time series (Figures 8a-8b) are average of initial beliefs. Figure 8a shows individual consumption for a log utility person (graph on the top) with high cost of processing information ($\theta = 2$, dotted line) and low cost of processing information ($\theta = 0.2$, solid line). The bottom graph of Figure 8a provides the same information for a person with CRRA utility, $\gamma = 2$. Figure 8b shows individual savings - solid line, right axes- and wealth - dotted line, left axes-. The top left graph displays savings and wealth for a person with log utility and $\theta = 0.2$ while the bottom left graph shows savings and wealth for a person with CRRA, $\gamma = 2$, utility and $\theta = 0.2$. The two graphs on the right of Figure 8b display savings and wealth when $\theta = 2$. The top graph on the right shows time paths of savings and wealth for a log utility person whereas the bottom graph shows the same time paths for a CRRA person.

[42]For the grid in the model, the steady state value of wealth is $\cong 5.65$ and I initialize the simulation with $w_0 = 3$.

The hump in consumption is the mirrored image of the rise (until people know they rich) and fall (once people know they are rich) in wealth. Note that, depending on the history of income shocks, consumption can have more than one hump in its path. To see why, consider a high realization of income occurring after an hump in consumption. Over time, signals about wealth convey such information, consumers start savings and history as well as humps repeat themselves. These effects are enhanced by the shadow cost of processing information, θ, with higher costs forcing long periods of inertia in consumption followed by sizeable changes. Note also the relationship between consumption and information flow (Figure 7c): risk averse agents would rather push forward consumption in times in which they are processing information about wealth. Finally, note from 7(a-b)-8(a,b) how the peak in consumption occurs later for an individual with higher degree of risk aversion and lower information flow. The rationale for this result is that more cautious people wait to be better informed about wealth before modifying their consumption behavior. In particular, since a consumer with CRRA utility ($\gamma = 2$) chooses to be better informed about low values of wealth than a log utility one (see Figures 7a and 7b), he processes news about high value of wealth slower than his log counterpart. The resulting additional savings for precautionary motives are triggered by both the curvature of the utility function and the limits on information-processing constraint.

The last result comes from studying how consumers with limited processing capacity react to temporary changes in income (y). Before stating the result, it is worth comparing to the predictions of standard consumption-saving literature. Under infinite rate of information, the response of consumption to either negative or positive temporary income shocks are immediate: consumption adjusts in period $T = 0$ of an amount exactly equal to the discounted present value of the shock, $|\Delta y|$. This is the case regardless whether the shock is adverse or favorable, so long as the absolute value of these shocks match each other. The same holds true under certainty-equivalence with a linear constraints and quadratic utility (LQ) framework. With risk averse agents and information-processing limits, it happens that:

Result 3. Persistent and asymmetric response to shocks. *Consumption's response to temporary fluctuations of wealth is asymmetric: Negative shocks trigger a sharper reaction and higher persistence of consumption than positive ones.*

Result 3 is the main finding of the paper and it is novel to the consumption-saving theoretical literature. Most importantly, Result 3 agrees with empirical evidence on consumption documented in, e.g., Shea (1995).

The logic behind this finding is the following. A risk averse person is more likely to be alerted by negative events than positive ones. A risk averse person with limited capacity rationally choose to pay more attention to signals that point to a decrease in wealth than to news about high values of wealth. As soon as he receives signals that his wealth is lower than what he thought, he reacts by decreasing consumption. The change in behavior and its persistence are more conspicuous the more risk averse and information-processing constrained the person is as this person awaits to gather more information before changing behavior and, in the meanwhile, he builds up a savings buffer. Thus, the temporary change in income propagates slowly over time. A positive temporary income shock triggers the opposite behavior in a risk averse uninformed person. The intuition is that this type of consumers is concerned about negative wealth fluctuations and allocate most of information capacity to prevent this events. A signal that indicates positive wealth may be ignored, generating extra savings in the meanwhile. Once this is acknowledged, a prudent consumer distributes the additional consumption driven by the income shock plus savings throughout his lifetime.

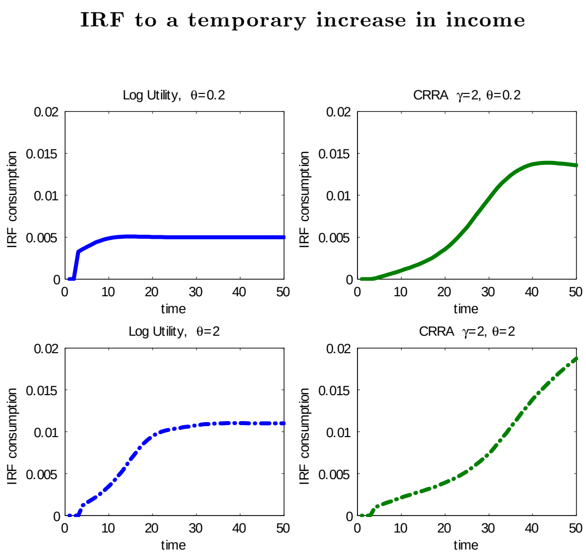

Figure 9a.

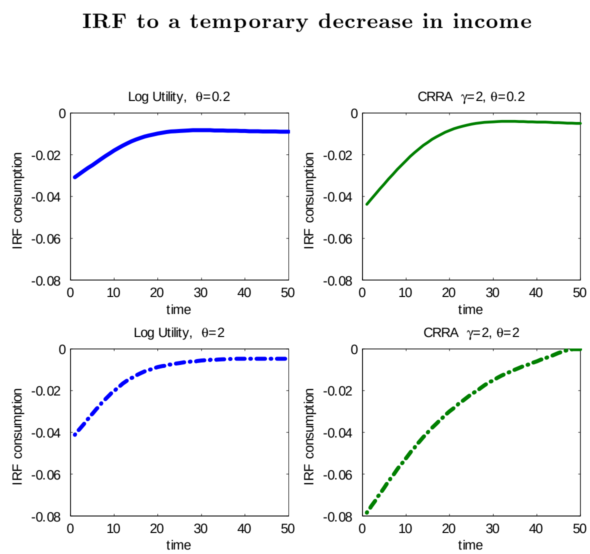

Figure 9b.

The impulse response functions are plotted in Figures 9a-9b.[43] They display a positive (Figure 9a) and a negative (Figure 9b) shock to income. Note that for both log and CRRA $\gamma = 2$ and for different value of the shadow cost ($\theta = 0.2 \vee \theta = 2$) the reaction to a negative shocks ($\Delta y = |1|$) starts from the very first period. However, the extent of the reaction varies across utilities and information costs. When $\theta = 0.2$, a log utility -type consumer reacts on impact by increasing savings of an amount lower than that of the change in income. He then adjusts savings and consumption so to distribute the averse shock throughout time. The same log-type but with $\theta = 1$ decreases consumption more on impact than his $\theta = 0.2$ counterpart. He increases consumption slowly over time until it reaches its new long-run value. Likewise, a consumer with risk aversion $\gamma = 2$ varies

[43] To generate impulse response for this non-linear model, I simulate the model drawing $10,000$ times from the same optimal policy distribution under two scenarios. In the first I draw from a distribution with constant mean of the shock to income. In the second, I assume that the mean of the shocks increase/decrease in the very first period (one-time shocks) and then revert back to its original distribution. Impulse responses of consumption are the difference between the two income paths averaged over simplex-points and $10,000$ Monte Carlo draws of income.

his saving when the shock hits. The amount of consumption's change depends on the person's information flow. In particular, note that for $\theta = 2$ the decrease in consumption on impact and in the following periods is so significant that consumers can use the accumulated savings to restore their original consumption plan. The endogenous asymmetric response to shocks makes rational inattention models observationally distinct from any other standard macroeconomic consumption-saving models or rational inattention LQG models. Furthermore, this asymmetry makes the theory of rational inattention appealing from an empirical standpoint and suitable to study the impact of different shocks to income and effectiveness of policy changes on consumers spending. Tax rebates provide one example.

6.1 Policy Implications and Sensitivity Analysis

6.1.1 The effect of a one-time increase in income

A feature of the model worth exploring is how consumption's reaction to shocks depend on the initial value of wealth. Drawing time series from the probability distribution that solves the model, it is natural that the farther away wealth is from its steady state, the more consumption reacts to changes to wealth. The asymmetric response of consumption to shocks presented in section 6.1 holds true whether we start from a value of wealth above or below the steady state. In both cases, the reactions are faster in case of a negative shock than a positive one. However, extent and timing are different. Exploring these two dimensions -extent and timing- in the contest of this framework uncovers some additional features of the model that can be used to address important policy questions. In particular, it can be used to analyze the effectiveness of tax policy reforms on individual consumption and savings decisions. Figure 10 displays the impulse response function of consumption to a stimulus payment which increases income by 2% with respect to its (constant) long run level. [44]

[44] The discretized solutions are generated using equi-spaced grid of consumption and wealth, with 50 points each . Consumption takes up value in $[0.5, 3]$ while wealth ranges from 1 to 10. I use the same parameters ($R = 1.012$ and $\beta = 1/R$) of the baseline model and a simplex of size $(50!) * (49)$ and two specifications of utility functions.

Impulse Response function as a function of wealth, Δy=0.02

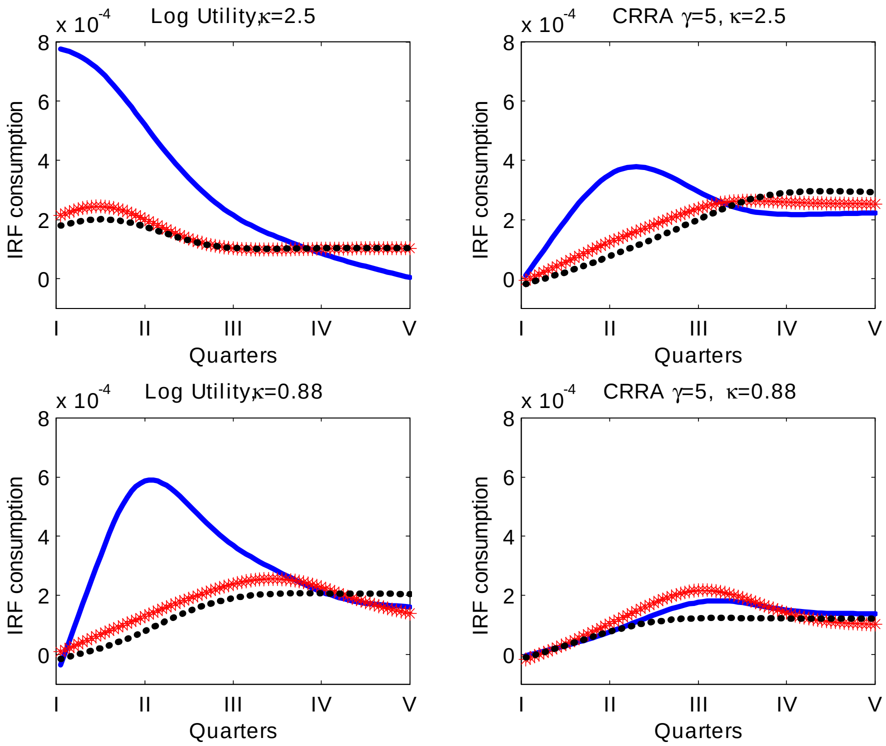

Figure 10. Solid: $w_0 = 1.94$; Star Dashed:$w_0 = 3.3$; Dotted: $w_0 = 5.2$

In both cases I choose θ so that the capacities corresponds to $\cong 2.5$ bits and 0.88 bits[45]. Once the value iteration has converged, I generate the impulse response function by simulating time series path of consumption and wealth with $10,000$ Monte Carlo runs for each initial condition on wealth. I consider three initial values of wealth as a proxy of population with low, middle and high wealth. Then, I average the time series per quarters and simplex points. Figure 10 gives interesting insights on the effect of the stimulus on consumer spending. For the degrees of risk aversion and information capacity considered, the reaction of the stimulus is higher the lower the initial wealth. This is not surprising, as the stimulus payments have bigger impact on the disposable income of credit constrained consumers than richer people. For a given amount of information capacity and wealth, the higher the risk aversion the lower the spending in the first quarters. This result also makes sense. If a consumer is risk averse and have no credit frictions, it allocates more attention in processing information about low values of wealth. This leads to processing slower and, in turn, reacting slower to positive news to income (Result 3). Finally for a given wealth and degree of risk aversion, the lower the information processing capacity,

[45]The constraint $\kappa = 2.5$ corresponds to $\theta_{\log} = 0.01$ and $\theta_{crra} = 0.05$ for the log case and the crra, $\gamma = 2$ case respectively, while $\kappa = 0.88$ is given by $\theta_{\log} = 0.1$ and $\theta_{crra} = 0.9$.

the slower the response of consumption spending to the rebate. The findings in Figure 10 can be summarizes as:

Result 4. Economic stimulus and rational inattention. *The impact of a one time tax rebate on rationally inattentive consumers:*

1. *is faster the lower the initial net worth.*

2. *is more delayed the higher the degree of risk aversion.*

3. *is more persistent and more delayed the lower the information-processing capacity.*

The insights one can gather from the model have strong policy implications on the effectiveness of tax reform on people's behavior. Not surprisingly, the model predicts that such a policy has the fastest impact for people that are credit constrained. However, the strongest impact of the policy will be on people who are concerned about their wealth. As Figure 10 suggests, the effects of the policy for households with middle- (star-dashed line) and high- (dotted line) income households will be spread out through several quarters. Initially these people do not vary consumption when they acknowledge the rebate. After savings accumulate, they start spending and increase consumption profile permanently. Notwithstanding the stylized nature of the model, it is possible to map U.S. household data on realized consumption into the current framework. Showing how this mapping is done and how the results of this paper can be formally tested are the objects of a companion paper. Here, I provide a qualitative comparison of evidence and model's results meant to suggest how rational inattention can be useful for policy analysis. Table 3 shows the fraction of the rebate, Δy, spent in consumption for the 2001 tax rebate -1st column-, the 2008 tax rebate -2nd column- and a one time 2% increase in income for the model with capacity $\kappa = 2.5$ and log utility -3rd column- and CRRA utility, $\gamma = 5$ -last column. Table 3 summarizes the effect of the additional income on impact for the full sample and low income sample as estimated by Johnson, Parker and Souleles (2001) and Broda and Parker(2008).[46]

	2001 Tax Rebate	2008 Tax Rebate	Log Utility,	CRRA
			$\kappa = 2.5$	$\gamma = 5$, $\kappa = 1.5$
Perc. tax rebate spent:				
Full Sample:				
1^{st} Quar.	27%*	19%**	37%	12%
2^{nd} and 3^{rd} Quar.	66%*	33%**	79%	21%
Low Income:				
1^{st} Quar.	76%*	32%**	62%	24%

Table 3: Fraction of $\mathbf{\Delta y}$ spent

*Source: Johnson, Parker and Souleles (2006); **Source: Broda and Parker(2008)

[46] The choice of time span and low-income households is due to the availability of evidence for this income group.

Table 3 shows that the data and the results of the model are roughly consistent both on impact -1st Quarter- and for the medium run -cumulative 2nd and 3rd Quarters-. One point worth mentioning is how to relate the results for 2008 tax rebate to the ones of the model with CRRA with $\gamma = 5$. The model predicts that under rational inattention people with relatively high risk aversion do not translate positive changes in income into positive changes in consumption prior to be well informed about their wealth. As a result, this group of people fails to adjust promptly their consumption in response to the stimulus. Moreover, consumption remains sluggish for a long period of time even for people of this group with low income.

The mild effect of the stimulus for the 2008 tax rebate might reflect mostly the uncertainty in the economic condition associated to this particular period. Even though the model does not account for other sources of uncertainty besides wealth(/income) fluctuations, it is interesting to notice how the interplay of risk aversion and information processing constraint seems to capture the reluctance of people to revise their consumption plans upwards despite the predictable increase in their income.

6.1.2 The importance of risk aversion

In this section, I analyze the previous consumption-savings model under Shannon processing limits with a quadratic utility in order to assess the role of the utility function in choosing the optimal conditional distribution of consumption given wealth. To this end, I use the same solution methodology and the same specification for the grid as in Sections 4 and 5, varying only the utility function, $u(c_t) = (c_t - .5c_t^2)$. [47] The bars in Figure 11 display the optimal distribution of c conditional on $w = 1$ -first column-, $w = 5.3$ -second column- and $w = 9.5$ -third column for the quadratic utility. For comparison, the solid lines in Figure 11 correspond to the case $u(c_t) = \log(c_t)$. The top panel of Figure 11 displays the solution for $\theta = 0.2$ whereas the bottom panel shows the corresponding

[47] Note that even with quadratic utility, using discrete distributions prevents one to use the result in Sims (2003) where the optimal solution has Gaussian distribution. If one assumes continuous distributions and full support for c and w, it is possible to show that even when the prior $g(w)$ is not Gaussian, the optimal $p(c, w)$ will be close to a Gaussian distribution (see Sims (2003)).

solution when $\theta = 2$.

Conditional distribution, $p(C|w\text{=}.)$ **for θ=0.2 (top panel) and θ=2 (bottom panel)**

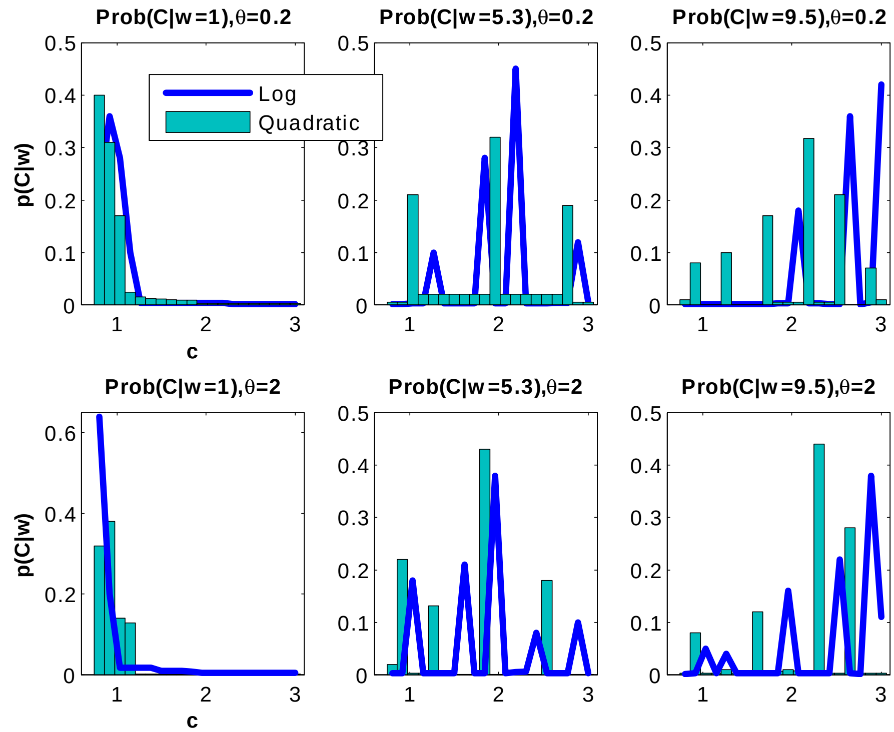

Figure 11: Optimal solution under Quadratic (bars) and Log (solid line) $u(c)$.

For both values of shadow costs considered, the optimal distribution under quadratic utility assigns more probability to central values of consumption for a given value of wealth. Moreover, the optimal distribution appears to be more spread out than the corresponding choice of a person with log-utility. For $\theta = 0.2$, conditional on $w = 1$, the optimal solution assigns higher probability to the lowest value of consumption for quadratic utility $(\Pr(c = 0.8|w = 1) = 0.4)$ than it does for log utility $(\Pr(c = 0.8|w = 1) = 0.2)$. When $\theta = 2$, risk aversion for the person with log utility induces more precaution in consumption than quadratic utility by setting 0.84 probability on the lowest two values of consumption instead of 0.7 as the optimal solution under quadratic utility does. Conditional on $w = 5.2$, the optimal distribution under quadratic utility assigns probabilities 0.38 to $c = 1.96$ and 0.43 to $c = 1.84$ when $\theta = 0.2$ and $\theta = 2$ respectively. With log utility, the optimal distribution places probability 0.45 on $c = 2.2$ and about 0.60 probabilities on values of consumption $c \geq 1.96$. Finally, when $w = 9.5$, for each value of θ, the optimal conditional distribution of consumption is skewed towards higher values of consumption for the person with log utility than it is for the person with quadratic utility with about 0.8 (0.5) probability mass on values of $c \geq 2.20$ versus 0.10 (0.3) for $\theta = 0.2$ $(\theta = 2)$.

Table 4 summarizes the average across simplex points of expected value and variance of consumption and information flow for the cases of CRRA $\gamma = 5, \gamma \to 1$, $\gamma = 0.5$ and quadratic utilities.

Statistics

θ=0.02

	CRRA $\gamma = 5$	Log Utility	CRRA $\gamma = .5$	Quadratic
$E(C)$	2.14	2.08	1.91	1.11
$std(C)$	0.046	0.07	0.14	0.37
κ	2.12	1.97	1.82	0.94

θ=0.2

	CRRA $\gamma = 5$	Log Utility	CRRA $\gamma = .5$	Quadratic
$E(C)$	1.09	1.08	1.02	0.69
$std(C)$	0.09	0.11	0.14	0.42
κ	1.99	1.87	1.72	0.78

θ=2

	CRRA $\gamma = 5$	Log Utility	CRRA $\gamma = .5$	Quadratic
$E(C)$	0.98	0.91	0.83	0.45
$std(C)$	0.18	0.21	0.33	0.52
κ	1.20	0.86	0.78	0.21

Table 4

Table 4 shows how a person with quadratic utility tends to acquire less information - lower κ- than people with higher degrees of risk aversion, on average for $\theta = 0.02$ -top panel- $\theta = 0.2$ and $\theta = 2$ -middle and lower panels, respectively- This choice implies lower expected value and higher variance for the consumption's profile of a person with quadratic utility than for a person with CRRA.

7 Conclusions

This paper applies rational inattention to a dynamic model of consumption and savings. Consumers rationally choose the nature of the signal they want to acquire subject to the limits of their information processing capacity. The dynamic interaction of risk aversion and endogenous choice of information flow enhances precautionary savings.

I showed that for a given degree of risk aversion, the lower the information flow, the flatter the consumption path. The model predicts that for a given information flow, the higher the degree of risk aversion, the more persistent consumption is. Also, for a given degree of risk aversion, the lower the information flow, the more volatile consumption is.

Furthermore, the model predicts that consumption path has humps. Under information-processing constraints, an hump occurs when people consume a little and save a lot while collecting information about wealth. When consumers realize that they are rich, they increase consumption and decrease savings. This increase stops when they acknowledge that their wealth is low again: they start savings and process more information. Thus, consumption decrease. Consistent with the previous two results, I find that the peak in consumption is delayed the more risk averse and information-constrained the individual is. Different from other life-cycle models, in my setting there could be more than one hump in the consumption path. Depending on the history of the income shocks, a very low or very high realization of income affects consumers' signal through its effect on wealth. Consumer react to the news by varying savings and information over time, thereby generating another hump.

Finally, the model predicts that consumers with processing capacity constraints have asymmetric responses to shocks, with negative shocks producing stronger and more persistent effects than positive ones. This asymmetry, observed in actual data, is novel to the theoretical literature of consumption and savings. Studying the reactions of rational inattentive people to temporary income shocks can also be used to assess the effectiveness of policy reforms on consumption spending. The model predicts that, for a given level of wealth, the speed and magnitude of the adjustment of consumption to the income shock depends on their processing capacity. Moreover, consumers with low wealth react faster to temporary tax relief than wealthier people. The results agree with both intuition and evidence on consumer spending for the 2001 and 2008 tax rebates.

The results suggest that enriching the standard macroeconomic toolbox with rational inattention theory is a step worth making.

References

[1] Allen, F., S. Morris and H. S. Shin, (2006), *Beauty Contests, Bubbles and Iterated Expectations in Asset Markets* Review of Financial Studies, 19, pp. 719-752.

[2] Alvarez, F. and N. Stockey, (1998), *Dynamic Programming with Homogeneous Functions*, Journal of Economic Theory, 82, pp.167-189.

[3] Angeletos G.-M. and A. Pavan, (2004), *Transparency of Information and Coordination in Economies with Investment Complementarities*, American Economic Review, papers and proceedings, 94 (2), pp.91-98.

[4] Astrom, K. (1965), *Optimal Control of Markov decision process with incomplete state estimation.* Journal of Mathematical Analysis and Applications 10, pp.174-205

[5] Blanchard, O. J. and Mankiw, N G., (1988) *Consumption: Beyond Certainty Equivalence* American Economic Review, American Economic Association, vol. 78(2), pp. 173-77.

[6] Broda, C. and J. Parker (2008), *A preliminary analysis of how household spending changed in response to the receipt of a 2008 economic stimulus payment.* Mimeo.

[7] Caballero, R. J. (1990), *Expenditure on Durable Goods: a case for slow adjustment,* The Quarterly Journal of Economics, Vol. 105, No. 3. pp. 727-743.

[8] Caballero, R. J. (1995), *Near-rationality, Heterogeneity and Aggregate Consumption*, Journal of Money, Credit and Banking, 27 (1), pp29-48.

[9] Campbell, J.Y. (1987), *Does Savings Anticipate Declining Labor Income? An alternative Test of the Permanent Income Hypothesis*, Econometrica, 55, pp.1249-1273.

[10] Campbell, J.Y. and A. Deaton (1989), *Why is Consumption so Smooth?* Review of Economic Studies, 56, pp.357-374.

[11] Campbell, J.Y. and N. G. Mankiw (1989), *Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence*, NBER Macroeconomic Annual 4, pp.185-216.

[12] Campbell, J.Y. and N. G. Mankiw (1989), *Permanent Income, Current Income and Consumption*, Journal of Business and Economic Statistics, 8 (3), pp.265-279.

[13] Carroll, C. D. (2003) *Macroeconomic Expectations of Households and Professional Forecasters*, Quarterly Journal of Economics, 118 (1), pp. 269-298.

[14] Cochrane, J.H. (1989), *The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives*, American Economic Review, 90, pp.319-337.

[15] Cover, J. P., (1992), *Asymmetric Effects of Positive and Negative Money-Supply shocks,* The Quarterly Journal of Economics, Vol. 107, No. 4, pp.1261-1282.

[16] Cover, T.M. and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991

[17] Deaton, A. (1992), *Understanding Consumption*, Oxford, University Press.

[18] DeLong, J., and L. Summers (1988). *How does macroeconomic policy affect output?*, Brookings Papers on Economic Activity, 2: pp. 433-480.

[19] Dynarski, S. and J. Gruber (1997), *Can Families Smooth Variable Earnings?*, Brookings Papers on Economic Activity, Vol.1997 No. 1, pp.229-284.

[20] Flavin, M. A. (1981), *The Adjustment of Consumption to Changing Expectations about Future Income*, Journal of Political Economy, 89, pp. 974-1009.

[21] Friedman, M. (1957), *A Theory of the Consumption Function*, Princeton, Princeton University Press.

[22] Garcia, R., Lusardi, A. and S. Ng, (1997), *Excess sensitivity and asymmetries in consumption: an empirical investigation*, Journal of Money, Credit and Banking, Blackwell Publishing, vol. 29 (2), pp. 154-76.

[23] Gourinchas, P-O, and J.A. Parker (2002), *Consumption over the Life-Cycle*, Econometrica, vol. 70 (1), pp.47-89.

[24] Hall, R. (1978), *Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence*, Journal of Political Economy, 86, pp.971-987..

[25] Johnson, D., Parker, J. and N. Soules (2006) *Household expenditure and the income tax rebates of 2001.* American Economic Review, Vol. 96, No. 5.

[26] Kim, J., Kim, S., Schaumburg, E. and C. Sims (2005). *Calculating and Using Second Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models.* Mimeo.

[27] Lewis, K. (2007), *The life-cycle effects of information-processing constraints.* Working Paper, University of Iowa.

[28] Lewis, K. (2008), *The Two-Period Rational Inattention Model: Accelerations and Analyses*, Computational Economics, Forthcoming. Currently available as Federal Reserve Financial and Economics Discussion Series Paper, No. 2008-22, Board of Governors .

[29] Lucas, R.E. (1972), *Expectations and the Neutrality of Money.* Journal of Economic Theory, 4, 103-124.

[30] Luo, Y. (2008): *Consumption Dynamics under Information Processing Constraints.* Review of Economic Dynamics, 11, pp, 366—385

[31] MacKay, David J. C (2003). *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003

[32] Maćkowiak, B., and M. Wiederholt (2008a), *Optimal Sticky Prices under Rational Inattention*, American Economic Review, forthcoming.

[33] Maćkowiak, B., and M. Wiederholt (2008b), *Business cycles dynamics under rational inattention*, Mimeo.

[34] Mankiw, N.G. and R. Reis (2002), *Sticky information versus sticky prices: A proposal to replace the New Keynesian Phillips curve*, Quarterly Journal of Economics 117, 1295–1328.

[35] Mondria, J. (2006), *Financial Contagion and Attention Allocation*, Working paper, University of Toronto.

[36] Moscarini, G. (2004), *Limited Information Capacity as a Source of Inertia*, Journal of Economic Dynamics and Control, pp. 2003–2035.

[37] Mullainathan, S. (2002), *A Memory-Based Model of Bounded Rationality*, Quarterly Journal of Economics, 117 (3), pp. 735-774.31.

[38] Parker, J. (1999) *The Reaction of Household Consumption to Predictable Changes in Social Security Taxes*, American Economic Review, 89 (4), pp. 959-973.

[39] Pischke, J.-S. (1995) *Individual Income, Incomplete Information, and Aggregate Consumption*, Econometrica, 63 (4), pp. 805-840.

[40] Puterman, M.L. (1994), *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, Inc.

[41] Reis, R. A. (2006) *Inattentive Consumers*, Journal of Monetary Economics, 53 (8), 1761-1800.

[42] Romer, C. and D. H. Romer (2007). *The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks*. NBER Working Paper No. 13264, July 2007.

[43] Shea, J. (1995), *Union Contracts and the Lifecycle/Permanent-Income Hypothesis.* American Economic Review, 85(1):186-200.

[44] Sims, C. A. (1998), *Stickiness*, Carnegie-Rochester Conference Series On Public Policy, 49(1), 317–356.

[45] Sims, C. A. (2003), *Implications of Rational Inattention*, Journal of Monetary Economics, 50(3), 665–690.

[46] Sims, C. A.(2005), *Rational Inattention: a Research Agenda*, American Economic Review Papers and Proceedings, 96, 158-163..

[47] Woodford, M. (2002), *Imperfect Common Knowledge and the Effects of Monetary Policy*, in P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, eds., Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps, Princeton: Princeton University Press.

[48] Woodford, M. (2003), *Interest and Prices*, Princeton. Princeton University Press

[49] Zeldes, S. (1989) *Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence*, Quarterly Journal of Economics, May 1989, 275-98..

8 Appendix A

8.1 Proof of Proposition 1.

The Bellman Recursion in the discrete Rational Inattention Consumption-Saving Model is a Contraction Mapping.

Proof. The H mapping displays:

$$HV(g) = \max_p H^p V(g),$$

with

$$H^p V(g) = \left[\sum_{w \in \Omega_w} \left(\sum_{c \in \Omega_c} u(c)\, p(c|w) \right) g(w) + \beta \sum_{w \in \Omega_w} \sum_{c \in \Omega_c} \left(V\left(g'_c(\cdot)\right)\right) p(c|w)\, g(w) \right].$$

Suppose that $||HV - HU||$ is the maximum at point g. Let p_1 denote the optimal control for HV under g and p_2 the optimal one for HU

$$HV(g) = H^{p_1} V(g),$$
$$HU(g) = H^{p_2} U(g).$$

Then it holds

$$||HV(g) - HU(g)|| = H^{p_1} V(g) - H^{p_2} U(g).$$

Suppose *WLOG* that $HV(g) \leq HU(g)$. Since p_1 maximizes HV at g , I get

$$H^{p_2} V(g) \leq H^{p_1} V(g).$$

Hence,

$$\begin{aligned}
||HV - HU|| &= \\
||HV(g) - HU(g)|| &= \\
H^{p_1} V(g) - H^{p_2} U(g) &\leq \\
H^{p_2} V(g) - H^{p_2} U(g) &= \\
\beta \sum_{w \in \Omega_w} \sum_{c \in \Omega_c} \left[\left(V^{p_2}\left(g'_c(\cdot)\right)\right) - \left(U^{p_2}\left(g'_c(\cdot)\right)\right)\right] p_2 g(w) &\leq \\
\beta \sum_{w \in \Omega_w} \sum_{c \in \Omega_c} \left(||V - U|| \right) p_2 g(w) &\leq \\
\beta\, ||V - U||.
\end{aligned}$$

Recalling that $0 \leq \beta < 1$ completes the proof. ∎

8.2 Proof of Corollary.

The Bellman Recursion in the discrete Rational Inattention Consumption-Saving Model is an Isotonic Mapping.

Proof. Let p_1 denote the optimal control for HV under g and p_2 the optimal one for HU

$$HV(g) = H^{p_1}V(g),$$
$$HU(g) = H^{p_2}U(g).$$

By definition,

$$H^{p_1}U(g) \leq H^{p_2}U(g).$$

From a given g, it is possible to compute $g'_c(\cdot)|_{p_1}$ for an arbitrary c and then the following will hold

$$V \leq U \Longrightarrow$$

$\forall g(w), c,$

$$V\left(g'_c(\cdot)|_{p_1}\right) \leq U\left(g'_c(\cdot)|_{p_1}\right) \Longrightarrow$$
$$\sum_{c\in\Omega_c} V\left(g'_c(\cdot)|_{p_1}\right)\cdot p_1 g \leq \sum_{c\in\Omega_c} U\left(g'_c(\cdot)|_{p_1}\right)\cdot p_1 g \Longrightarrow$$

$$\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c} u(c)\,p_1\right) g(w) + \beta\sum_{c\in\Omega_c} V\left(g'_c(\cdot)|_{p_1}\right)\cdot p_1 g$$
$$\leq \sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c} u(c)\,p_1\right) \Longrightarrow$$

$$H^{p_1}V(g) \leq H^{p_1}U(g) \Longrightarrow$$
$$H^{p_1}V(g) \leq H^{p_2}U(g) \Longrightarrow$$
$$HV(g) \leq HU(g) \Longrightarrow$$
$$HV \leq HU.$$

Note that g was chosen arbitrarily and, from it, $g'_c(\cdot)|_{p_1}$ completes the argument that the value function is isotone. ■

8.3 Proof of Proposition 2.

The Optimal Value Function in the discrete Rational Inattention Consumption-Saving Model is Piecewise Linear and Convex (PCWL).

Proof. The proof is done via induction. I assume that all the operations are well-defined in their corresponding spaces. For planning horizon $n = 0$, I have only to take into account the immediate expected rewards and thus I have that:

$$V_0(g) = \max_{p\in\Gamma}\left[\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c} u(c)\,p\right) g(w)\right] \tag{24}$$

and therefore if I define the vectors

$$\left\{\alpha_0^i(w)\right\}_i \equiv \left(\sum_{c\in\Omega_c} u(c)\,p\right)_{p\in\Gamma} \tag{25}$$

I have the desired

$$V_0(g) = \max_{\left\{\alpha_0^i(w)\right\}_i} \left\langle \alpha_0^i, g\right\rangle \tag{26}$$

where $\langle .,.\rangle$ denotes the inner product $\langle \alpha_0^i, g\rangle \equiv \sum_{w\in\Omega_w} \alpha_0^i(w), g(w)$.For the general case, using equation (22):

$$V_n(g) = \max_{p\in\Gamma}\left[\begin{array}{c} \sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c} u(c)\,p(c|w)\right) g(w) + \\ +\beta \sum_{w\in\Omega_w}\sum_{c\in\Omega_c}\left(V_{n-1}\left(g'_c(\cdot)_c\right)\right) p(c|w)\, g(w) \end{array}\right] \tag{27}$$

by the induction hypothesis

$$V_{n-1}\left(g(\cdot)|_c\right) = \max_{\left\{\alpha_{n-1}^i\right\}_i} \left\langle \alpha_{n-1}^i, g'_c(\cdot)\right\rangle \tag{28}$$

Plugging into the above equation (19) and by definition of $\langle .,.\rangle$,

$$V_{n-1}\left(g'_c(\cdot)\right) = \max_{\left\{\alpha_{n-1}^i\right\}_i} \sum_{w'\in\Omega_w} \alpha_{n-1}^i(w')\left(\sum_{w\in\Omega_w}\sum_{c\in\Omega_c} T(\cdot; w, c)\,\frac{\Pr(w,c)}{\Pr(c)}\right) \tag{29}$$

With the above:

$$\begin{aligned} V_n(g) &= \max_{p\in\Gamma}\left[\begin{array}{c} \sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c} u(c)\,p\right) g(w) + \\ +\beta \max_{\left\{\alpha_{n-1}^i\right\}_i} \sum_{w'\in\Omega_w} \alpha_{n-1}^i(w')\left(\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c} \frac{T(\cdot;w,c)}{\Pr(c)}\cdot p\right) g(w)\right) \end{array}\right] \\ &= \max_{p\in\Gamma}\left[\left\langle u(c)\cdot p,\ g(w)\right\rangle + \beta\sum_{c\in\Omega_c}\frac{1}{\Pr(c)} \max_{\left\{\alpha_{n-1}^i\right\}_i}\left\langle \sum_{w'\in\Omega_w} \alpha_{n-1}^i(w')\, T(\cdot; w, c)\cdot p,\ g\right\rangle\right] \end{aligned} \tag{30}$$

At this point, it is possible to define

$$\alpha^j_{p,c}(w) = \sum_{w' \in \Omega_w} \alpha^i_{n-1}(w') T(\cdot : w, c) \cdot p. \tag{31}$$

Note that these hyperplanes are independent on the prior g for which I am computing V_n. Thus, the value function amounts to

$$V_n(g) = \max_{p \in \Gamma} \left[\langle u(c) \cdot p,\ g \rangle + \beta \sum_{c \in \Omega_c} \frac{1}{\Pr(c)} \max_{\left\{\alpha^j_{p,c}\right\}_j} \left\langle \alpha^j_{p,c}, g \right\rangle \right], \tag{32}$$

and define:

$$\alpha_{p,c,g} = \arg \max_{\left\{\alpha^j_{p,c}\right\}_j} \left\langle \alpha^j_{p,c}, g \right\rangle. \tag{33}$$

Note that $\alpha_{p,c,g}$ is a subset of $\alpha^j_{p,c}$ and using this subset results into

$$\begin{aligned} V_n(g) &= \max_{p \in \Gamma} \left[\langle u(c) \cdot p,\ g \rangle + \beta \sum_{c \in \Omega_c} \frac{1}{\Pr(c)} \langle \alpha_{p,c,g}, g \rangle \right] \\ &= \max_{p \in \Gamma} \left\langle u(c) \cdot + \beta \sum_{c \in \Omega_c} \frac{1}{\Pr(c)} \alpha_{p,c,g},\ g \right\rangle. \end{aligned} \tag{34}$$

Now

$$\left\{\alpha^i_n\right\}_i = \bigcup_{\forall g} \left\{ u(c) \cdot p + \beta \sum_{c \in \Omega_c} \frac{1}{\Pr(c)} \alpha_{p,c,g} \right\}_{p \in \Gamma} \tag{35}$$

is a finite set of linear function parametrized in the action set. ■

8.4 Proof of Proposition 3.

Proof. The first task is to prove that $\{\alpha^i_n\}_i$ sets are discrete for all n. The proof proceeds via induction. Assuming CRRA utility and since the optimal policy belongs to Γ, it is straightforward to see that through (25), the set of vectors $\{\alpha^i_0\}_i$,

$$\left\{\alpha^i_0\right\}_i \equiv \left(\sum_{w \in \Omega_w} \left(\sum_{c \in \Omega_c} \frac{c^{1-\gamma}}{1-\gamma} p(c|w) \right) g(w) \right)_{p \in \Gamma}$$

is discrete. For the general case, observe that for discrete controls and assuming $M = \left|\left\{\alpha^j_{n-1}\right\}\right|$, the sets $\left\{\alpha^j_{p,c}\right\}$ are discrete, for a given action p and consumption c, I can only generate $\alpha^j_{p,c}-$vectors. Now, fixing p it is possible to select one of the M $\alpha^j_{p,c}-$vectors for each one of the observed consumption c and, thus, $\{\alpha^j_n\}_i$ is a discrete set. The previous proposition, shows the value function to be convex. The *piecewise-linear* component of the properties comes from the fact that $\{\alpha^j_n\}_i$ set is of finite cardinality. It follows that V_n is defined as a finite set of linear functions. ■

9 Appendix B

9.1 Concavity of Mutual information in the Belief State.

For a given $p(c|\,w)$, Mutual Information is concave in $g(w)$

Proof. Let Z be the binary random variable with $P(Z=0)=\lambda$ and let $W=W_1$ if $Z=0$ and $W=W_2$ if $Z=1$. Consider

$$\begin{aligned} I(W,Z;C) &= I(W;C)+I(Z;C|\,W) \\ &= I(W;C|\,Z)+I(Z;C) \end{aligned}$$

Condition on W, C and Z are independent, $I(C;Z|\,W)=0$. Thus,

$$\begin{aligned} I(W;C) &\geq I(W;C|\,Z) \\ &= \lambda(I(W;C|\,Z=0))+(1-\lambda)(I(W;C|\,Z=1)) \\ &= \lambda(I(W_1;C))+(1-\lambda)(I(W_2;C)) \end{aligned}$$

Q.E.D. ■

10 Appendix C.

Pseudocode

Let θ be the shadow cost associated to $\kappa_t = I_t(C_t, W_t)$. Define a Model as a pair (γ,θ). For a given specification :

- Step 1: Build the simplex. equi-spaced grid to approximate each $g(w_t)$-simplex point.
- Step 2: For each simplex point, define $p(c_t, w_t)$. and Initialize with $V\left(g'_{c_j}(\cdot)\right)=0$.
- Step 3: For each simplex point, find $p^*(c,w)$ s.t.

$$V_0(g(w_t))|_{p^*(c_t,w_t)} = \max_{p(c_t,w_t)}\left\{\sum_{w_t\in\Omega_w}\sum_{c_t\in\Omega_c}\left(\frac{c_t^{1-\gamma}}{1-\gamma}\right)p^*(c_t,w_t)-\theta\left[I_t(C_t,W_t)\right]\right\}.$$

- Step 4: For each simplex point, compute $g'_{c_j}(\cdot)=\sum_{w_t\in\Omega_w}T(\cdot;w_t,c_t)\,p^*(w_t|c_t)$. Use a kernel regression to interpolate $V_0(g(w_t))$ into $g'_{c_j}(\cdot)$.
- Step 5: Optimize using csminwel and iterate on the value function up to convergence.

Obs. Convergence and Computation Time vary with the specification (γ,θ).

$\rightarrow$ 180-320 iterations each taking 8min-20min

- Step 6. For each model (γ, θ), draw from the ergodic $p^*(c, w)$ a sample (c_t, w_t) and simulate the time series of consumption, wealth, expected wealth and information flow by averaging over 1000 draws.
- Step 7. Generate histograms of consumption and impulse response function of consumption to temporary positive and negative shocks to income.

11 Appendix D.

11.1 The Mathematics of Rational Inattention

This part addresses the mathematical foundations of rational inattention. The main reference is the seminal work of Shannon (1948). Drawing from the information theory literature, I overview Shannon's axiomatic characterization of entropy and mutual information and show the main theoretical features of these two quantities.

Formally, the starting point is a set of possible events whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$. Suppose for a moment that these probabilities are known but that is all we know concerning which event will occur. The quantity $\mathcal{H} = -\sum_i p_i \log p_i$ is called the entropy of the set of probabilities $p_1, \ldots, p_n$. If x is a chance variable, then $H(x)$ indicates its entropy; thus x is not an argument of a function but a label for a number, to differentiate it from $H(y)$ say, the entropy of the chance variable y.

Quantities of the form $H = -\sum_i p_i \log p_i$ play a central role in Information Theory as measures of information, choice and uncertainty. The quantity H goes by the name of entropy [48] and p_i is the probability of a system being in cell i of its phase space.

The measure of how much choice is involved in the selection of the events is $H(p_1, p_2, .., p_n)$ and it has the following properties:

Axiom 1 H is continuous in the p_i.

Axiom 2 If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n. With equally likely events there is more choice, or uncertainty, when there are more possible events.

Axiom 3 If a choice is broken down into two successive choices, the original H should be the weighted sum of the individual values of H.

Theorem 2 of Shannon (1948) establishes the following results:

[48] See, for example, R. C. Tolman, *Principles of Statistical Mechanics*, Oxford, Clarendon, 1938.

Theorem 1 *The only H satisfying the three above assumptions is of the form:*

$$\mathcal{H} = -K \sum_{i=1}^{n} p_i \log p_i$$

where K is a positive constant to account for the change in unit of measurement.

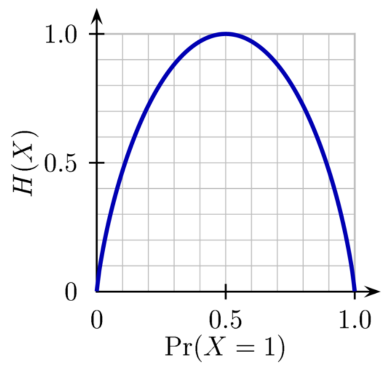

Figure a: Entropy of two choices with probability p and q=*1*$-p$ as function of p.

Remark 1. . $\mathcal{H} = 0$ if and only if all the p_i but one are zero, this one having the value unity. Thus only when we are certain of the outcome does $\mathcal{H}$ vanish. Otherwise $\mathcal{H}$ is positive.

Remark 2. For a given n, $\mathcal{H}$ is a maximum and equal to $\log n$ when all the p_i are equal (i.e., $\frac{1}{n}$). This is also intuitively the most uncertain situation.

Remark 3. Suppose there are two random variables, X and Y,

$$\mathcal{H}(Y) = -\sum_{x,y} p(x,y) \log \sum_{x} p(x,y)$$

Moreover,

$$\mathcal{H}(X,Y) \leq \mathcal{H}(X) + \mathcal{H}(Y)$$

with equality only if the events are independent (i.e., $p(x,y) = p(x)p(y)$). This means that the uncertainty of a joint event is less than or equal to the sum of the individual uncertainties.

Remark 4. Any change toward equalization of the probabilities $p_1, p_2, \ldots, p_n$ increases $\mathcal{H}$. Thus if $p_1 < p_2$ an increase in p_1, or a decrease in p_2 that makes the two probabilities more alike results into an increase in $\mathcal{H}$. The intuition is trivial since equalizing the probabilities of two events makes them indistinguishable and therefore increases uncertainty on their occurrence. More generally, if we perform any "averaging" operation on the p_i of the form $p'_i = \sum_j a_{ij} p_j$ where $\sum_i a_{ij} = \sum_j a_{ij} = 1$, and all $a_{ij} \geq 0$, then in general $\mathcal{H}$ increases[49].

[49] The only case in which $\mathcal{H}$ remains unchanged is when the transformation results in just one permutation of p_j.

Remark 5. Given two random variables X and Y as in Remark 3, not necessarily independent, for any particular value x that X can assume there is a conditional probability $p_x(y)$ that Y has the value y. This is given by

$$p_x(y) = \frac{p(x,y)}{\sum_y p(x,y)}.$$

The *conditional entropy* of Y, is then defined as $\mathcal{H}_X(Y)$ and it is the average of the entropy of Y for each possible realization the random variable X, weighted according to the probability of getting a particular realization x. In formulae,

$$\mathcal{H}_X(Y) = -\sum_{x,y} p(x,y) \log p_x(y).$$

This quantity measures the average amount of uncertainty in Y after knowing X. Substituting the value of $p_x(y)$, delivers

$$\begin{aligned}\mathcal{H}_X(Y) &= -\sum_{x,y} p(x,y) \log p(x,y) + \sum_{x,y} p(x,y) \log \sum_y p(x,y) \\ &= \mathcal{H}(X,Y) - \mathcal{H}(X)\end{aligned}$$

or

$$\mathcal{H}(X,Y) = \mathcal{H}(X) + \mathcal{H}_X(Y).$$

This formula has a simple interpretation. The uncertainty (or entropy) of the joint event X, Y is the uncertainty of X plus the uncertainty of Y after learning the realization of X.

Remark 6. Combining the results in Axiom 3 and remark 5, it is possible to recover $\mathcal{H}(X) + \mathcal{H}(Y) \geq \mathcal{H}(X,Y) = \mathcal{H}(X) + \mathcal{H}_X(Y)$.

This reads $\mathcal{H}(Y) \geq \mathcal{H}_X(Y)$ and implies that the uncertainty of Y is never increased by knowledge of X. If the two random variables are independent, then the entropy will remain unchanged.

To substantiate the interpretation of entropy as the rate of generating information, it is necessary to link $\mathcal{H}$ with the notion of a channel. A *channel* is simply the medium used to transmit information from the source to the destination, and its *capacity* is defined as the rate at which the channel transmits information. A discrete channel is a system through which a sequence of choices from a finite set of elementary symbols $S_1, \ldots, S_n$ can be transmitted from one point to another. Each of the symbols S_i is assumed to have a certain duration in time t_i seconds . It is not required that all possible sequences of the S_i be capable of transmission on the system; certain sequences only may be allowed. These will be possible signals for the channel. Given a channel, one may be interested in measuring its capacity to transmit information. In general, with different lengths of symbols and constraints on the allowed sequences, the capacity of the channel is defined as:

Definition 2 *The capacity C of a discrete channel is given by*

$$C = \lim_{T \to \infty} \frac{\log N(T)}{T}$$

where $N(T)$ is the number of allowed signals of duration T.

To explain the argument in a very simple case, consider transmitting files via computers. The speed at which one can exchange documents depends on the internet connection and it is expressed in bits per seconds. The maximum amount of bits per second that can be transmitted is negotiated with the provider. However, this does not means that the computer will always be trasmitting data at this rate — this is the maximum possible rate and whether or not the actual rate reaches this maximum depends on the usage and the source of information which feeds the channel. The link between channel capacity and entropy is illustrated by the following *Theorem 9 of Shannon*:

Theorem 3 *Let a source have entropy $\mathcal{H}$ (bits per second) and a channel have a capacity C (bits per second). Then it is possible to encode the output of the source in such a way as to transmit at the average rate $\frac{C}{\mathcal{H}} - \varepsilon$ symbols per second over the channel where ε is arbitrarily small. It is not possible to transmit at an average rate greater than $\frac{C}{\mathcal{H}}$.*

The intuition behind this result is that by selecting an appropriate coding scheme, the entropy of the symbols on a channel achieves its maximum at the channel capacity. Alternatively, channel capacity can be related to *mutual information.*

Definition 4 *The* **Mutual Information between two random variables** X **and** Y *is defined as the average reduction in uncertainty of random variable X achieved upon the knowledge of the random variable Y.*

In formulae:

$$\mathcal{I}(X;Y) \equiv \mathcal{H}(X) - E(\mathcal{H}(X|Y)),$$

which says that the mutual information is the average reduction in uncertainty of X due to the knowledge of Y or, symmetrically, it is the reduction of uncertainty of X due to the knowledge of Y. Mutual information is invariant to transformation of X and Y , depending only on their copula.

Intuitively, $\mathcal{I}(X;Y)$ measures the amount of information that two random variables have in common. The capacity of the channel is then alternatively defined by

$$C = \max_{p(Y)} (\mathcal{I}(X;Y))$$

where the maximum is with respect to all possible information sources used as input to the channel (i.e., the probability distribution of Y, $p(Y)$). If the channel is noiseless,

$E(\mathcal{H}_y(x)) = E(\mathcal{H}(X(|Y))) = 0$. To get an intuition, think about a newspaper editor who wants to maximize his sales. To do that, he has to choose the allocation of space for his articles in such a way that it is attractive for the consumers. In this example, Y is the random variable *space*, X the random variable *sales*, channel's capacity is the maximum number of pages in the newspaper and the channel itself is the best articles' allocation of space that signals that the journal is worth buying.