The Rigidity of Choice.
Life cycle savings with information-processing limits*

ANTONELLA TUTINO
Princeton University

December 2007

Abstract

This paper studies the implications of information-processing limits on consumption and savings behavior of households through time. It possesses a dynamic model in which consumers rationally choose the size and scope of the information they want to process about their financial possibilities. Their ability to process information is constrained by Shannon channels. The model predicts that people with higher degrees of risk aversion rationally choose higher information flows and have higher lifetime consumption. If they have limited access to information flows, risk averse agents prefer to allocate their attention in reducing the volatility of consumption in exchange for lower mean of consumption throughout their life. 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. The findings suggest information-processing limits as an additional motive for precautionary savings.

*PRELIMINARY. COMMENTS WELCOME. E-mail: atutino@princeton.edu. I am deeply indebted to Chris Sims whose counteless 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 Alan Blinder, Vasco Curdía, Ryo Jinnai, Nobu Kiyotaky, Angelo Mele, Stephen Morris, Tamas Papp, Philippe-Emanuel Petalas, Charles Roddie, Esteban Rossi-Hansberg, Sam Schulhofer-Wohl, Felipe Swartzman, Andrea Tambalotti, Tommaso Treu and Mark Watson for many helpful comments and discussions at various stages of the paper. In particular, I want to thank Carlos Carvalho for countless discussions and comments about my work. Lastly, I thank seminar partcipants to the Princeton Macro workshop and Princeton Macro Students Workshop. Remaining errors are my own.

*Information is, we must steadily remember, a measure of one's freedom of choice in selecting a message. The greater this freedom of choice, and hence the greater the information, the greater is the uncertainty that the message actually selected is some particular one. Thus greater freedom of choice, greater uncertainty, greater information go hand in hand. (*Claude Shannon, *sic.)*

1 Introduction

People face everyday an overwhelming amount of data. Imagine a consumer who wants to choose optimal plans for consumption and savings through his lifetime. He must think through the quality and quantity of his current consumption, his income possibilities, the rate of returns on his investments, etc. Considering the amount of available information to attend to, a simple task such as day-by-day shopping seems difficult and perhaps hopeless if optimality of the lifetime plan is the goal. Yet, people gather information and make decisions every day reacting to shifts in their economic environment.

Macroeconomists have realized the need to incorporate in their models slow, smooth low-frequency responses together with hump-shaped high frequency consumption behavior in order to match observed data. This has led to a number of modelling strategies that inject inertia into rational expectation framework by assuming costly acquisition of information and/or delays in the diffusion of information. While these devices have managed to track observed patterns in some dimensions, in the majority of cases they do not provide a rationale for the randomness they imply which remains exogenous to the model.

What drives people to react to some events and not others? Is it possible to relate inertial behavior in consumption and savings to people's preferences?

My paper offers a micro-founded explanation of the nature of inertia in consumption and savings behaviors.

Following *Rational Inattention Theory*, (Sims, 2003, 2006) this paper contributes to the consumption and savings literature by modelling expectations of consumers as the endogenous choice of the scope of the information they want to gather. In my framework, inertial responses to movements of economic environment are the outcome of consumers' choice and their resulting expectations.

The core idea is simple. By taking explicitly into account limits in information-processing of people in an otherwise standard dynamic optimization model, this paper studies the amount of information individuals want to process on the basis of what they can process and their preferences. Combining the standard objective of maximizing utility subject to a budget constraint with information-processinig limits leads to a departure from rational expectations. My paper shows how to model this formally in an intertemporal setting. Moreover and perhaps most importantly, it provides a framework to investigate the interaction of choices of information and degrees of risk aversion of people and their implications on consumption and savings throughout their life. The challenge of solving general rational inattention models is that they involve infinite dimensional state spaces. I address this issue by discretizing the model. I derive the properties of the discretized framework and I propose and implement a computational strategy for its solution.

Several predictions emerge from the model. First I find that more risk averse individuals choose to delay consumption until they are better informed about their wealth. This is expected: risk averse consumers react to uncertainty by processing more information regarding low values of wealth and keep their consumption low until the uncertainty is diminished. Second, the choice of information affects the expectations the agent has about his current and future wealth and this in turn leads to a more (less) conservative consumption behavior the higher (lower) is risk aversion. Moreover, if information processing is costly for the consumer, he focusses his effort more in reducing the volatility of his lifetime consumption than increasing its mean.

Third, ceteris paribus, different combinations of risk aversion and information flow lead to different consumption/savings paths. Numerical comparative-static analysis between a benchmark model and one with a strict bound on information-processing reveals that consumers save more the lower the information flow they have access to. This is suggestive of a precautionary motive for savings driven by information processing limits.

Fourth, I find that consumers with processing capacity constraints have asymmetric responses to income fluctuations, with negative fluctuations producing more persistent effects than positive ones.

Furthermore, in a limited information processing economy stickiness in consumption is persistent and path dependent. Comparing my framework with another one equivalent in all respects but information-processing limits, a favorable temporary income shock makes people modify their lifetime consumption more slowly and persistently with information-processing limits than without. This model also predicts that a temporary adverse shock makes risk averse agents reset their lifetime consumption immediately downward and the effect of this kind of shock dies out slower than the effect of a positive shock. The intuition for the path dependence result is that in my model consumers never see their wealth but they have a prior on it. The endogenous noise created by the imperfect observation due to information capacity constraints carries over for many periods, creating a path dependence of consumption..

My results are observational distinct from previous literature on consumption and information (e.g., Reis(2006)):in both my model and in previous models consumption growth moves back towards its original path after a given number of periods, but in my model it takes much longer. In my setting reversion of consumption to its pre-shock pattern occurs after many more periods than in the literature. A risk averse agent that receives a signal indicating an increase in wealth, may decide to wait and have more information in future periods about the actual consistency of his wealth, push forward consumption and in the meanwhile increase his savings to spread the windfall throughout his lifetime. Likewise, a risk averse agent immediately decides to decrease his consumption when he processes information about a reduction in his financial possibilities to avoid taking any chances on his wealth. He reverts back to his original consumption plan only after collecting a sequence of information that points him towards a re-establishment of his finances.

Relative to the literature on consumption and imperfect information (e.g., Prishke (1995)), in my models consumers select the scope of their signal about their wealth. In particular I do not constrain signals to have any specific distribution -such as Gaussian, as assumed by the literature- : the nature of the signal is the outcome of the optimal decision of the consumer. Hence, 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.

Sims (1998, 2003, 2006) pioneered the idea that individuals have limited capacity for processing information. The applications of rational inattention have been limited to either a linear quadratic framework where Gaussian uncertainty has been considered (such as Sims 1998, 2003, Luo 2007, Mackowiak and Wiederholt 2007, Mondria, 2006, Moscarini 2004) 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. One can show that the problem can be solved in two steps: first the information gathering scheme is found and then, given the optimal information, the consumption profile. The second insights of this approach is its immediate comparison to rational expectation theory based on signal extraction. The solution derived from a LQG rational inattention model is indeed observational equivalent to a signal extraction problem. This is because just by looking at the behavior of rational inattentive consumers it is impossible to tell apart an exogenously given Gaussian noise of the signal extraction model from endogenous noise that comes from information processing ehich is optimally chosen to be Gaussian.

The tractability of rational inattention LQG comes at the cost of restrictive assumptions on preferences and the nature of the signal. Constraining uncertainty of the individual to a quadratic loss / certainty equivalent setting does not take into account the possibility that the agent is very uncertain about his economic environment: ceteris paribus, more uncertainty generates second order effects of information that have first order impact on the decisions of the individuals. In this sense, rational inattention LQG models are subjected to the same limits as models that use linear approximation of optimality conditions to study stochastic dynamic models.[1] With little uncertainty about

[1] Since the work of Hall (1982), the assumption of certainty equivalnce has also been questioned in the consumption savings literature with no information friction, starting from e.g. Blanchard and Mankiw (1988).

the economic environment, linear approximations of the optimality conditions provide a fairly adequate description of the exact solution of the system. However, it seems reasonable to think that individuals may choose not to spend all their time in tracking the economy. This suggests that uncertainty at the individual level might actually be large, undermining the accuracy of both linearized and rational inattention LQG models.

To assess the importance of information choices on people' expectations, it is crucial to let consumers select their information from a wider set of distributions than includes but it is not limited to Gaussian family. In this sense, my paper contributes to the literature that models how people form their expectations and react to the economy. A necessarily non-exhaustive list of papers that address the issue of modeling consumers' expectations includes the absent-minded consumer model proposed by Ameriks, Caplin and Leahy (2003), together with Mullainathan (2002) and Wilson (2005), whose models feature imperfect recall of the agents. Mankiw and Reis (2002) develop a different model in which information disseminates slowly. They assume that every period an exogenous fraction of firms obtain perfect information concerning all current and past disturbances, while all other firms continue to 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 but the bounds on information processing given by the Shannon channel force consumers to choose the scope of their information to the limit of their capacity. This difference makes my setting and the one of inattentiveness observationally distinct in describing consumers' responses to shocks. inattentive consumers decide when to process all the information -in the wording of my model, when to have full capacity- and process nothing in the remaining periods while rational inattentive consumers always react to a probabilistic environment by gathering some information. The latter implies that rational inattentive consumers may take a very long time to revert back to their original path after a one-time shock if the information they collect is not sharp enough to justify a change in the after-shock behavior.

The paper is organized as follows. Section 2 lays down the theoretical basis of rational inattention. The first part describes the economics of rational inattention and introduces informally the concept of entropy and information applied to an economic model. The second part focuses more on the mathematics of rational inattention, with particular emphasis on the statistical properties of entropy and information. Section 3 formulates the model. It states the problem of the consumer as a discrete stochastic dynamic programming problem, while Section 4 derives the properties of the Bellman function. Section 5 shows the optimality conditions. Section 6 provides the numerical methodology used to solve the model, while Section 7 delivers its main predictions and results. Section 8 discusses some extensions and Section 9 concludes.

2 Foundations of Rational Inattention

The goal of this section is to introduce the technology I employ in my model and to discuss its implications on households' optimizing behavior. This section is divided into two parts. The first discusses how information theory can be used for modeling based on optimizing behavior. It also illustrates how and to what extent the outcomes of the resulting model depart from those postulated by a standard framework. Moreover, it lays down an informal description of my model and hints to its predictions.

The second part establishes the mathematical apparatus upon which rational inattention stands. The mathematical foundation for communication has been formally stated in the seminal work of Claude Shannon (1948).[2] The rigorous application of information and communication theories to economics and their guidelines as microeconomic foundations for modeling based on optimizing behavior is due to Christopher A. Sims (1988, 1998, 2003, 2005, 2006)[3].

The work of Shannon focussed on measuring the information content of a message selected at one point from a source located in another point. His main contribution is to define a measure of the choice involved in the selection of the message and the uncertainty regarding the outcome. This synthetic measure of how uncertain the decision maker still is, after choosing his message, goes by the name of *entropy.* Rational inattention stems from Information theory and uses Shannon capacity as a technological constraint to capture individuals' limits in processing information about the economy. People attempt to reduce their uncertainty by selecting the focus of their attention, constrained by information-processing limits. The resulting behavior of the agents depends on the choices of what to observe of the environment once the processing limits are acknowledged.

2.1 The Economics of Rational Inattention

The kernel of rational inattention stands on a probabilistic argument. Consider a consumer who wants to choose his lifetime consumption and saving plans but has limited knowledge of his wealth. To figure out precisely how wealth evolves, the consumer needs to process an amount of information that is beyond his skill, time and, equally importantly, his interest. The household enters the world with a probability distribution over his wealth that corresponds to his uncertainty about his financial possibilities. With the aim of maximizing his lifetime utility, he goes through life by selecting information necessary to sharpen his knowledge within the limits of his capacity.

The consumer chooses the size and scope of a signal about his wealth. Taking into account that processing information costs effort and utilities. Intuitively, he needs to give up leisure time to monitor his wealth whether this is looking up his account on internet,

[2] Readers not interested in the mathematical details of information theory may skip Section 2.2.

[3] As early as 1988, the bulk of the idea of *Rational Inattention* can be found in C. Sims' comment of *L. Ball*, *N.G. Mankiw* and *D. Romer* in the Brooking Papers on Economic Activity 1:1988.

figuring out the expenditures, making sure his checks do not bounce. Hence, he faces a trade off between the precision of the signal he wants to achieve and the time and efforts spent in processing the content of the signal. Moreover, no matter how precisely he wants to track his wealth, he cannot process all the information available since that goes way beyond his skills. After choosing the signal and understand its content, he shops and uses his consumption to infer the state of his wealth for future purchases. Note that from the above argument, it is clear that before processing information, consumption is a random variable for the household and, once realized, observed consumption behavior is used to reduce uncertainty in the following periods.

To make the discussion concrete, consider a person has a prior over the possible realizations of wealth, W. Define such a prior by $p(W)$.[4] The uncertainty that this probability contains before the consumer processes any information is measured by its entropy, $-E[\log(p(W))]$, where $E[.]$ denotes the expectation operator. Entropy is a universal measure of uncertainty which can be defined for a density against any base measure. The standard convention is to attribute zero entropy to the events for which $p = 0$,[5] and to use base 2 for the logarithms so that the resulting unit of information is binary and called a *bit*.

In the terms of my model, the initial prior over wealth is passed through a channel which represents the mechanism that processes information. At this point it is important to introduce a distinction between *capacity* of the channel and channel itself. I refer to capacity of a Shannon channel as the technological constraint on the maximum amount of information that can be processed. I refer to the channel as the mental device of processing information available and mapping it into real-life decisions. In my setting, Shannon channel captures the way the human brain works under the limits of its capacity.

The following example might foster intuition on this argument. Consider an individual who realizes his car needs refill. While driving to the gas station, he hears on the radio that there has been an increase in the price of gasoline. That news placed while focussing on another activities is likely to have no effect on his quest for full thank's worth of gas. Upon signing his credit card receipts he realizes he has spent more than for his previous refill. Given the information conveyed by the purchase of gasoline, figuring out the incidence of that increase on his wealth requires the individual to think or, in the wording of my model, using his channel -i.e., his brain-.

The output of this thought, or information flow, is an update on individuals' financial possibilities given the credit card receipts.

More formally, the credit card receipt can be thought of as an error-ridden datum. Before any information is processed, this datum is the random variable consumption of gas, C, whose probability distribution conditional on the random variable wealth, W, is $p(C|W)$. Knowing how much he spent in gas and thinking about the incidence of the expenditure on his wealth makes the consumer update his knowledge of his financial

[4] In this section, $p(\cdot)$ is used to denote a generic density function.

[5] Formally, given that $s\log(s)$ is a continuous function on $s \in [0,\infty)$, by l'Hopital Rule $\lim_{s\to 0} s\log(s) = 0$.

possibility. The update is an application of Bayes' rule:

$$p(W|C) = \frac{p(W)p(C|W)}{\int p(w)p(C|w)\,dw}. \tag{1}$$

The information flow resulting from sending data through the brain of the consumer is measured by the increase in knowledge *before* seeing what he was charged for a full thank of gas and *after* thinking of what the expenditure implies on his wealth.

Even if the consumer decides to pay closer attention to the incidence of gas prices on his overall expenditure, he cannot process all the information available as this would go beyond his skills and time.

The average difference in entropy before and after processing the information he acquires over time is therefore bound by a finite rate of transmission of information, say κ. In formulae:

$$E[(\log_2(p(W|C))|\,C] - E\left[\log_2\left(\int p(c|W)p(c)dc\right)\right] \leq \kappa. \tag{2}$$

The expression in (2) tells that the average reduction in uncertainty about his wealth given the information he acquires through C is bounded by a maximum number of bits, κ.

Note that the LHS of (2) is necessarily positive. This is because averaging across all the possible realizations of C sharpens the knowledge of the person about his wealth. However, some realization of C might make the individual more uncertain about W. An example would be if the person wrongly attributes the expenditure in gasoline to an excessive usage of the car rather than an increase in price. He could then decide to pay more attention to car mileage rather than prices of oil. But if the credit card bill of the following months still displays consistent withdraws at the gas station, he might detect that the increase in the expenditure is due to a price effect rather than a quantity effect and decide to pay more attention to gas prices.

Note that it is not possible to retrieve perfect information on consumer's wealth only on the basis of the observation of the behavior towards consumption since this knowledge would imply an infinite processing capacity of the consumer in finite time. This is because had the mental processing effort -i.e., the LHS in (2)- transmitted the exact value of consumer's wealth, the rate of information transmission would have been infinite. The intuitive reason is that the channel, or human brain, needs time and effort to map the information acquired into the understanding of one's financial possibilities.

Moreover, note that constraining the mapping between W and C to represent *finite* information flow as displayed by the technology in (2), naturally leads to a smooth and delayed reactions of the agents to the shifts in the consistency of their financial possibilities acquired through the observations of C.

In my example above, only after observing several gasoline purchases, the consumer may decide to pay more attention to oil prices and perhaps modify his behavior (e.g., he might decide to use the car less). This shows how a person trades off his interest to the content of the message (e.g., news on the radio, the display of prices at the gasoline station) and his current activities. Moreover, the example involves a slow and delayed responsiveness of the behavior of the individual to the content of the message even though such a delay involves potentially large costs.

Next, one may wonder how a consumer with information constraints differ behavior-wise from one with full information and one with no information. To illustrate this point, consider the following baby model of consumer's choice.

Suppose the household has three wealth possibilities $w \in W \equiv \{2, 4, 6\}$ and three consumption possibilities $c \in C \equiv \{2, 4, 6\}$. Before any observation is made, the consumer has the following prior on wealth, $\Pr(w = 2) = .5$, $\Pr(w = 4) = .25$, $\Pr(w = 6) = .25$. Moreover the consumer knows that he cannot borrow, $c \leq w$ and, if his check bounces he will have to pay a fine in terms of consumption $c = 0$. The utility he derives from consumption where utility is defined as $u(c) \equiv \log(c)$. His payoff matrix is summarized in **Figure a**.

$c\backslash w$	2	4	6
2	0.7	0.7	0.7
4	$-\infty$	1.38	1.38
6	$-\infty$	$-\infty$	1.8

Figure a: Payoff Matrix with $u(c) \equiv \log(c)$

If uncertainty in the payoff can be reduced at no cost, the consumer would set $c = 2$ whenever he knows that $w = 2$, $c = 4$ whenever $w = 5$ and, finally, $c = 6$ if $w = 8$.

In contrast, if there is no possibility of gathering information about wealth besides the one provided by the prior, the consumer will avoid infinite disutility by setting $c = 2$ whatever the wealth. The difference in bits in the two policies is measured by the mutual information between C and W. I measure the ex-ante uncertainty embedded in the prior for w by evaluating its entropy in bits, i.e., $\mathcal{H}(W) \equiv -\sum_{w \in W} p(w) \cdot \log_2(p(w)) = 0.5 \cdot \log_2\left(\frac{1}{0.5}\right) + 0.25 \cdot \log_2\left(\frac{1}{0.25}\right) + 0.25 \cdot \log_2\left(\frac{1}{0.25}\right) = 1.5$ bits. Since observation of c provides information on wealth, conditional on the knowledge of consumption uncertainty about w is reduced by the amount $\mathcal{H}(W|C) \equiv \sum_{w \in W} \sum_{c \in C} p(c, w) \log_2(p(w|c))$. The mutual information between C and W, i.e., the remaining uncertainty about the wealth after observing consumption is the difference between ex-ante uncertainty of W ($\mathcal{H}(W)$) and the knowledge of W provided through C ($\mathcal{H}(W|C)$). In formulae, the mutual information or capacity of the channel amounts to:

$$\begin{aligned} MI(C; W) &\equiv \mathcal{H}(W) - \mathcal{H}(W|C) = \\ &= \sum_{w \in W} \sum_{c \in C} p(c, w) \log\left(\frac{p(c, w)}{p(c)\, p(w)}\right) \end{aligned}$$

To see what this formula implies in the two cases proposed, consider first the situation in which information can flow at infinite rate.

First notice that in this case ex-post uncertainty will be fully resolved. Moreover, note that $(p(w|c)) = 1$, $\forall c \in C$, $\forall w \in W$ since the consumer is setting positive probability on one and only one value of consumption per value of wealth. This in turns implies $\mathcal{H}(W|C) = 0$. Thus the mutual information in this case will be $MI(C;W) = \mathcal{H}(W)$.

On the other hand, if consumer has zero information flow or, equivalently, if processing information would be prohibitively hard for him, his optimal policy of setting $c = 2$ at all times makes consumption and wealth independent of each other. This implies that $\mathcal{H}(W|C) = \sum_{w\in W}\left(\sum_{c\in C} p(c)\,p(w)\log_2\left(\frac{p(c)p(w)}{p(c)}\right)\right) = \mathcal{H}(W)$. Hence, in this case $MI(C;W) = 0$ and no reduction in the uncertainty about wealth occurred by observing consumption. This makes intuitive sense. If a consumer decides to spend the same amount in consumption regardless of his wealth level, his purchase will tell him nothing about his financial possibilities. The expected utility in the first case is $E^{FullInfo}(u(c)) = (\log(2))\cdot(.5) + (\log(4) + \log(6))\cdot(.25) = 1.14$ while in the second case $E^{NoInfo}(u(c)) = 0.7$. Now, assume that the consumer can allocate some effort in choosing size and scope of information about his wealth he wants to process, under the limits imposed by his processing capacity. Note the occurrence of two elements equally essential for the rational choice of the consumer. The first, limits in processing capacity, is a technological constraint: the information flow that his brain allows is bounded. The second is the interest of the consumer captured by his utility function. A rational consumer takes into account his limits and chooses the scope of information about his wealth accordingly guided by his preferences, $u(c)$.

Given the risk aversion of the consumer and since the consumer has always the option to set $c = 2$ and use no capacity, with small but positive information flow available he will choose the distribution $p(c|w)$ as dependent on w as his capacity allows him to.

Let $\bar{\kappa}$ be the maximum amount of information flow that the consumer can process. Let the probability matrix of the consumer be described by:

$c\backslash w$	$P(w=2)$	$P(w=5)$	$P(w=8)$
$P(c=2)$	0.5	p_1	p_2
$P(c=4)$	0	$.25 - p_1$	p_3
$P(c=6)$	0	0	$.25 - p_2 - p_3$

Figure b: Probability Matrix

where the zero on the South-West corner of the matrix encodes the no borrowing constraint $c \leq w$.

The program of the consumer amounts to:[6]

$$\max_{\{p_1,p_2,p_3\}} E^{\kappa}(u(c))$$

[6] The explicit formulation takes up the form:

s.t.

$$\bar{\kappa} \geq MI\left(C;W\right).$$

Given $\bar{\kappa} = 0.3$,[7] the optimal policy sets $p_1^* = 0.125$, $p_2^* = 0.125$, $p_3 = 0.125$. This correspond to a mass distribution of $\Pr\left(C=c\right) : \begin{cases} 0.75 & \text{if } c=2 \\ 0.25 & \text{if } c=4 \\ 0.0 & \text{if } c=6 \end{cases}$, which leads to an expected utility of $E^{\kappa}\left(u\left(c\right)\right) = 0.87$ and a mutual information of $MI\left(C;W\right) = \bar{\kappa} = 0.3$ bits . Hence, consumers who invest effort in tracking their wealth using the channel are better off than in the no information case -higher expected utility- even though they cannot do as well as in the constrained case.

Note that the result of trading off information on the highest value for more precise knowledge of lower value of wealth is driven by the choice of utility. For instance, if I had chosen a consumer with the same bound of processing capacity but higher degree of risk aversion, say one for which $u\left(c\right) = \frac{c^{1-\gamma}}{1-\gamma}$ with $\gamma = 5$, he would have chosen a probability $\Pr\left(C=2\right)$ even higher than his log-utility counterpart. On the other hand, a consumer with degree of risk aversion $\gamma = 0.3$ would have shift probability mass from $\Pr\left(C=2\right)$ to $\Pr\left(C=6\right)$. The intuition behind this result is that since the attention of the consumer within the limits of Shannon capacity is allocated according to his utility, the degree of risk aversion plays an important role in choosing the direction of this attention and, in turns, the scope of the signal. A log-utility consumer wants to be well informed about middle values of his wealth, whether an high risk aversion consumer selects a signal which provides sharper information on the lower values of wealth, so that he can avoid $-\infty$ disutility. The opposite direction is taken by a risk averse agent characterized by $\gamma = 0.3$.

$$\begin{aligned} \max_{\{p_1,p_2,p_3\}} E^{\kappa}\left(u\left(c\right)\right) &= \left(\log\left(2\right)\right)\cdot\left(.5+p_1+p_2\right) + \\ &+ \left(\log\left(4\right)\right)\cdot\left(.25-p_1+p_3\right) + \\ &+ \left(\log\left(6\right)\right)\cdot\left(.25-p_2-p_3\right) \end{aligned}$$

s.t.

$$\begin{aligned} \bar{\kappa} &\geq MI\left(C;W\right) = \\ &= .5\log_2\left(\frac{.5}{.5\left(.5+p_1+p_2\right)}\right) + .p_1\log_2\left(\frac{p_1}{.25\left(.5+p_1+p_2\right)}\right) + \\ &\quad +p_2\log_2\left(\frac{p_2}{.25\left(.5+p_1+p_2\right)}\right) + \left(.25-p_1\right)\log_2\left(\frac{\left(.25-p_1\right)}{.25\left(.25-p_1+p_3\right)}\right) + \\ &\quad + \left(.25-p_2-p_3\right)\log_2\left(\frac{\left(.25-p_2-p_3\right)}{.25\left(.25-p_2-p_3\right)}\right). \end{aligned}$$

[7] Note that such a bound of information flow is unrealistically low. However I decided to trade off realism for simplicity in this example.

2.2 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 provide an axiomatic characterization of entropy and mutual information and show the main theoretical features of these two pivotal quantities that set the stage for a rigorous basis of information theory.

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 form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics[8] where 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:

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 that amounts for the choice of the unit measure.

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

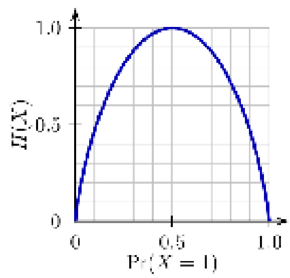


Figure c: Entropy of two choices with probability p and q=1−p as function of p.

There are certain distinguished features that make entropy a suitable measure of uncertainty.

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[9].

Remark 5. Given two random variables X and Y as in 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

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

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

The *conditional entropy* of Y, is then defined as $\mathcal{H}_X(Y)$ and amounts to 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 Axiom 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 of such a channel 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 a telegraph where all symbols are of the same duration, and any sequence of the 32 symbols is allowed . Each symbol represents five bits of information. If the system transmits n symbols per second it is natural to say that the channel has a capacity of $5n$ bits per second. This does not mean that the teletype channel will always be transmitting information at this rate — this is the maximum possible rate and whether or not the actual rate reaches this maximum depends on 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 , hence it depends 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, $E(\mathcal{H}_y(x)) = E(\mathcal{H}(X(|Y))) = 0$. The definition is then equivalent to that already given for a noiseless channel since the maximum entropy for the channel is its capacity..

3 The Formal Set-up

3.1 The problem of the Representative Household

To understand the implications of limits to information processing, let me first focus on the program of on household who can process infinite amount of information about his wealth.

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., $\mathcal{I}_t = \sigma(c_t, w_t; c_{t-1}, w_{t-1}; ...; c_0, w_0)$. 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 $u(c)$ be the utility of the household defined over a consumption good c. I assume that the utility belongs to the CRRA family. In particular $u(c) = \frac{c_t^{1-\gamma}}{1-\gamma}$ with γ the coefficient of risk aversion. If the consumer process information about his financial possibility, he can observe at each time t his wealth, w_t. The program in this case amounts to:

$$\max_{\{c_t\}_{t=0}^{\infty}} E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left[\left(\frac{c_t^{1-\gamma}}{1-\gamma} \right) \right] \middle| \mathcal{I}_0 \right\} \tag{3}$$

s.t.

$$w_{t+1} = R(w_t - c_t) + y_{t+1} \tag{4}$$

$$c_t \leq w_t \tag{5}$$

$$w_0 \text{ given} \tag{6}$$

where $\beta \in [0, 1)$ is the discount factor and $R = \beta^{-1}$ is the interest on savings $(w_t - c_t)$. The constraint (5) prevents the household from borrowing. I assume that $y_t \in Y \equiv \{y^1, y^2, .., y^N\}$ follows a stationary Markov process with mean $E_t((y_{t+1})|\mathcal{I}_t) = \bar{y}$.

Assume now that the consumer cannot process all the information available in the economy to track precisely his wealth. At time zero, his uncertainty about the wealth is summarized by a prior $g(w_0)$ which replaces (6) above.

The consumer can reduce uncertainty about the prior by choosing any joint distribution of consumption and wealth that he can process. That is, the consumer will rationally choose any distribution that makes $p(c|w)$ as dependent on w as his information processing constraint will allow him to. When information cannot flow at infinite rate the choice of the consumer is $p(c|w)$ as opposite to the stream of consumption $\{c_t\}_{t=0}^{\infty}$ in (3). Another way of looking at this is that the consumer chooses a noisy signal on wealth where the noise distribution-wise 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 jointly $p(c, w)$. The optimal choice of this distribution is the one that makes the distribution of consumption conditional on wealth close to wealth under the limits imposed by Shannon capacity. Hence, the choice of consumption in my setting corresponds to $\{c(\mathcal{I}_t)\}_{t=0}^{\infty}$.

Next, I turn to the information constraint. 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 the consumer to spend some time and utility to process information. I model the arduous task of thinking by appending a Shannon channel to the constraints set, and by assuming that the agent associates a cost to his effort in terms of utils. 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 described by:

$$\bar{\kappa} \geq I(C_t; W_t) = \int p(c_t, w_t) \log \left(\frac{p(c_t, w_t)}{p(c_t)\, g(w_t)} \right) dc_t dw_t \tag{7}$$

To describe the way individuals transit across states, define the operator $E_{w_t}\left(E_t(x_{t+1})\middle| c_t\right) \equiv \hat{x}_{t+1}$, 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 $\mathcal{E}$ to equation (4) leads to:

$$\hat{w}_{t+1} = R(\hat{w}_t - c_t) + \widehat{\widehat{y}} \tag{8}$$

where, note that

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

To fully characterize the transition from the prior $g(w_t)$ to its posterior distribution, I need to take into account how the choice in time t, $p(w_t, c_t)$ 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:

$$g'\left(w_{t+1}|_{c_t}\right) = \int \tilde{T}(w_{t+1}; w_t, c_t)\, p(w_t|c_t)\, dw_t \tag{9}$$

known as Bayesian conditioning. In (9), the function $\tilde{T}$ is the transition function representing (8).

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. Thus, (9) replaces (4) in the limited processing world.

Combining all these ingredients, the program of the household under information frictions amounts to

$$\max_{\{p(w_t, c_t)\}_{t=0}^{\infty}} E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left(\frac{c_t^{1-\gamma}}{1-\gamma} \right) \middle| \mathcal{I}_0 \right\} \tag{10}$$

s.t.

$$\bar{\kappa} \geq I_t\left(C_t; W_t\right) = \int p\left(c_t, w_t\right) \log\left(\frac{p\left(c_t, w_t\right)}{p\left(c_t\right) g\left(w_t\right)}\right) dc_t dw_t \tag{11}$$

$$p\left(c_t, w_t\right) \in \mathcal{D}\left(w, c\right) \tag{12}$$

$$g'\left(w_{t+1}|_{c_t}\right) = \int \tilde{T}\left(w_{t+1}; w_t, c_t\right) p\left(w_t|c_t\right) dw_t \tag{13}$$

$$g\left(w_0\right) \text{ given} \tag{14}$$

where $\mathcal{D}\left(w, c\right) \equiv \left\{\left(c, w\right) : \int p\left(c, w\right) dcdw = 1,\ p\left(c, w\right) \geq 0, \forall \left(c, w\right)\right\}$ in (12) restricts the choice of the agent to be drawn from the set of distributions. Note that this problem is a well-posed mathematical problem with convex objective function and concave constraint sets. What makes it an hard problem to solve is that both state and control variables are infinite dimensional. To make progress in solving it, I implement two simplifications: 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.

3.2 Comparing the model with the literature

Before I turn to the solution, I want to compare my model with the literature of rational inattention. In this digression, I write down the first order conditions specifying utility functions for several degrees of risk aversion and levels of information flows. The benchmark model is the standard consumption saving problem of a consumer who anticipated perfectly his income, y_t. Abstracting from borrowing constraints for now but assuming only a No-Ponzi scheme, the problem of the agent amounts to

$$\max_{\{c_t, w_t\}} E_0\left[\sum_{t=0}^{\infty} \beta^t u\left(c_t\right)\right] \tag{15}$$

subject to:

$$w_{t+1} = R\left(w_t - c_t\right) + y_{t+1} \tag{16}$$

where $u\left(c_t\right) = c_t - \frac{1}{2}c_t^2$, β is the discount factor and and the flow budget constraint has a Gaussian i.i.d income process with mean $\bar{y}$ and variance ω^2. Then the optimal consumption delivers the well known result that consumption is a martingale process and equals permanent income:

$$c_t^* = \left(1 - \beta\right) w_t + \beta \bar{y}$$

For the Gaussian income case with quadratic utility, Sims (2003) shows that the rational inattention equivalent of the above problem amounts to

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

subject to (16),

$$w_{t+1}|\mathcal{I}_{t+1} \tilde{\ } D_{t+1}, w_t|\mathcal{I}_t \tilde{\ } D_t$$

and given $w_0|\mathcal{I}_0\tilde{~}Gauss\left(\hat{w}_0, \sigma_0^2\right)$ and

$$\kappa = \tfrac{1}{2}\left(\log(R^2\sigma_t^2 + \omega^2) - \log(\sigma_{t+1}^2)\right) : .$$

Given the LQG specification, Sims (2003) shows that the optimal distribution D_t is Gaussian with mean $\hat{w}_t = E_t\left[w_t\right]$ and variance $\sigma_t^2 = var_t\left(w_t\right)$.

Note that I assume a constant borrowing constraint, i.e., $c_t < w_t$ $\forall t$ and $c_t > 0$. Therefore, the conventional solution to the benchmark model is no longer correct, nor Gaussianity of the optimal posterior distribution of consumption and wealth for the rational inattention version of the problem is preserved. The failure of both the martingale solution for the in the standard model and Gaussianity in the optimal policy of its rational inattention version is due to the break of the LQ framework implied by the inequality in the borrowing constraint. In particular in a rational inattention setting, numerical simulations reveal that preventing excessive borrowing forces to zero some regions of the optimal joint distribution of consumption and wealth. Moreover, the support of the distribution is truncated by the limit on c_t.

4 Solution Methodology

4.1 Discretizing the Framework

Let me start by assuming that wealth and consumption are defined on compact sects. In particular, admissible consumption profiles belongs 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 $g : \Omega_w \to \mathbb{R}$ fulfilling $g\left(w\right) \geq 0$ for all $w \in \Omega_w$ and $\sum_{w\in\Omega_w} g\left(w\right) = 1$ 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 approximate the distribution $g\left(w\right)$ and let $\Gamma \equiv \left\{ g \in \mathbb{R}^{|S|} : g\left(i\right) \geq 0 \text{ for all } i \ \sum_{i=1}^{|S|} g\left(i\right) = 1 \right\}$ denote the set of all probability distribution on Π. The initial condition for the problem is $g\left(w_0\right)$.

The consumer enters each period choosing the joint distribution of consumption and financial possibilities. Arguments exactly symmetrical to the ones of the previous section lead to specify the control variable for the discretized set up as the probability mass function $\Pr\left(w, c\right)$ where $c \in \Omega_c$ and $w \in \Omega_w$. This choice of the control variable is also constrained to be drawn from the set of distributions. Given $g\left(w_0\right)$ and $\Pr\left(c_t, w_t\right)$ and the observation of c_t consumed in period t, the belief state is updated using Bayesian conditioning:

$$g'_{c_j}(\cdot) = \sum_{w_t \in \Omega_w} T(\cdot; w_t, c_t) \Pr(w_t | c_t) \tag{17}$$

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 density function on a discrete set with counting measure. The term "$(\cdot)$" stands for all the possible values that w_{t+1} can assume in its support Ω_w given a pair (w_t, c_t).

Next I turn to the processing constraint. Given the setting, limits in information capacity information need to be defined in terms of the discrete mutual information between state and actions . The maximum reduction in uncertainty lies in:

$$\begin{aligned} \bar{\kappa} &\geq \mathcal{I}_t(C_t; W_t) \equiv \mathcal{H}(W_t) - \mathcal{H}(W_t | C_t) \\ &= \sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_c} \Pr(c_t, w_t) \left(\log \frac{\Pr(c_t, w_t)}{p(c_t) g(w_t)} \right) \end{aligned} \tag{18}$$

The interpretation of (18) 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[10] in time t. In (18), $\mathcal{H}(W_t)$ is the entropy of the random variable W_t. The entropy of W is a succinct representation of the amount of uncertainty embedded in the variable. Formally, the entropy is a functional of the distribution of W which does not depend on the actual values taken by the random variables but only on the probabilities. If $W \sim g(W)$, then its entropy is defined as $\mathcal{H}(W_t) = -\sum_{w_t \in \Omega_w} g(w_t) \log(g(w_t))$, where the logarithm is taken in base 2 so that the unit of measure of the entropy is *bits.* Upon knowledge of C_t, $\mathcal{H}(W_t) - \mathcal{H}(W_t | C_t)$ accounts for the uncertainty remaining in W_t after observing C_t. Thus, the mutual information $\mathcal{I}_t(W_t, C_t)$ can be interpreted as the (average) amount of uncertainty in W_t resolved per period by the observation of C_t. The capacity of the channel that allows information from C_t to flow into the knowledge of W_t constrains the maximum reduction in uncertainty captured by $\mathcal{I}_t(W_t, C_t)$. From this argument, it is clear that such a reduction, cannot occur at a rate grater than $\bar{\kappa}$. Finally, the objective function (10) in the discrete world amounts to

$$\max_{\{p(w_t, c_t)\}_{t=0}^{\infty}} E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left[\sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_w} \left(\frac{c_t^{1-\gamma}}{1-\gamma} \right) \Pr(c_t, w_t) \right] \middle| \mathcal{I}_0 \right\}. \tag{19}$$

4.2 Recursive Formulation

The purpose of this section is to show that the discrete dynamic programming problem has a solution and to recast it into a Bellman recursion. To show that a solution exists,

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

first note that the set of constraints for the problem is a compact-valued concave correspondence. Second, I need to show that the state space is compact. Compactness comes from the assumption of CRRA 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 (17) has the Feller property. Then the conditions for applying the Theorem of the Maximum in Stockey et al. (1989) are fulfilled which guarantees the existence of a solution. In the next section I will provide sufficient conditions to guarantee uniqness.

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

$$V\left(g\left(w_t\right)\right) = \max_{\Pr(c_t, w_t)} \left[\begin{array}{c} \sum_{w_t \in \Omega_w} \left(\sum_{c_t \in \Omega_c} u\left(c_t\right) \Pr\left(c_t, w_t\right) \right) + \\ +\beta \sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_c} V\left(g'_{c_t^j}\left(w_{t+1}\right)\right) \Pr\left(c_t, w_t\right) \end{array} \right] \tag{20}$$

subject to:

$$\bar{\kappa} \geq \mathcal{I}_t\left(C_t; W_t\right) = \sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_c} \Pr\left(c_t, w_t\right) \left(\log \frac{\Pr\left(c_t, w_t\right)}{p\left(c_t\right) g\left(w_t\right)}\right) \tag{21}$$

$$g'_{c_j}\left(\cdot\right) = \sum_{w_t \in \Omega_w} T\left(\cdot; w_t, c_t\right) \Pr\left(w_t | c_t\right) \tag{22}$$

$$\sum_{c_t \in \Omega_c} \Pr\left(c_t, w_t\right) = g\left(w_t\right) \tag{23}$$

$$1 \geq \Pr\left(c_t, w_t\right) \geq 0 \; \forall \left(c_t, w_t\right) \in B, \forall t \tag{24}$$

where $B \equiv \{(c_t, w_t) : w_t \geq c_t, \ \forall c_t \in \Omega_c, \forall w_t \in \Omega_w, \ \forall t\}$.

The Bellman equation in (20) 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 (20) is the utility function $u\left(.\right)$ which is assumed to be of the CRRA family with coefficient of risk aversion $\gamma > 0$. The second term, $\sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_c} V\left(g'_{c_t^j}\left(w_{t+1}\right)\right) \Pr\left(c_t, w_t\right)$, represents the expected continuation value of being in state $g\left(.\right)$ discounted by the factor $\beta \in (0,1)$. The expectation is taken with respect to the endogenously chosen distribution $\Pr\left(c_t, w_t\right)$. I have discussed at lenght the relations in (21)-(24) earlier. Moreover, I appended the equation in (23) which constraints the choice of the distribution to be consistent with the initial prior $g\left(w_t\right)$. Before turning to the optimality conditions that characterize the solution to the problem (20)-(24), I will first analyze the main properties of the Bellman recursion (20) and derive conditions under which is a contraction mapping and show that the mapping is isotone.

4.3 Properties of the Bellman Recursion

Definitions. 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 (21)-(24).To make the notation more compact, let $p \equiv \Pr(c_j|w_i)$, $\forall c_j \in \Omega_c$, $\forall w_i \in \Omega_w$ and let Γ be the set that contains (21)-(24).

D1. A control probability distribution $p \equiv \Pr(c_i, w_j)$ is ***feasible*** for the problem (20)-(24) if $p \in \Gamma$. Let $|W|$ be the cardinality of Ω_w and let

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

denote the set of all probability distributions on Ω_w. An optimal policy has a value function that satisfies the Bellman optimality equation in (20):

$$V^*(g) = \max_{p\in\Gamma} \left[\sum_{w\in\Omega_w} \left(\sum_{c\in\Omega_c} u(c)\, p(c|w) \right) g(w) + \beta \sum_{w\in\Omega_w} \sum_{c\in\Omega_c} \left(V^*\left(g'_c(\cdot)\right)\right) p(c|w)\, g(w) \right] \tag{25}$$

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 $0 \leq \beta < 1$.

Next, I prove that the value function recursion is an isotonic contraction. From these results, it follows that this recursion converges to a single fixed point corresponding to the optimal value function V^*.

These theoretical results establish that in principle there is no barrier in defining value iteration algorithms for the Bellman recursion under rational inattention. All the proofs are in appendix A.

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 contraction mapping.

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.*

5 Optimality Conditions

In this section I incorporate explicitly the constraint on information processing and derive the Euler Equations that characterize its solution.

The main feature of this section is to relate the link between the output of the channel -consumption- with the capacity chosen by the agent. In deriving the optimality conditions, I incorporate the consistency assumption (23) in the main diagonal of the joint distribution to be chosen, $\Pr_t(c_j, w_i)$. Note that such a restriction is *WLOG.*

5.1 First Order Conditions

To evaluate the derivative of the Bellman equation with respect to a generic distribution $\Pr(c_{k_1}, w_{k_2})$, define the differential operator $\Delta_k v(l) \equiv v(l_{k_1}) - v(l_{k2})$ and θ as the shadow cost of processing information. Then, the optimal control for the program (20)-(24) amounts to:

$\partial p^*(c_{k_1}, w_{k_2}):$

$$\Delta_k u(c) + \beta \Delta_k V(g'_c(.)) = p^*(c_{k_1}, w_{k_2}) \left(-\Delta_k u'(c)\,\theta p^*(w_{k_2}|c_{k_1}) - \beta \Delta_k V'_{p^*}(g'_c(.))\right) \quad (26)$$

This expression states that the optimal distribution depends on the weighted difference of two consumption profiles, c_{k_1} and c_{k_2} where the weights are given by current and future discounted utilities. Note that the differential of the marginal utility of current consumption is also weighted by the conditional optimal distribution of consumption and wealth.

The interpretation of (26) is that the consumer allocates probabilities by trading-off current and future. utilities levels between two consumption profiles, feasible given his prior on wealth, with the corresponding intertemporal difference in marginal utilities. To illustrate the argument, suppose a consumer believes that his wealth is w_{k_2} with high probability. Suppose for simplicity that w_{k_2} allows him to spend c_{k_1} or c_{k_2}. The decision of shifting probability from $p(c_{k_2}, w_{k_2})$ to $p(c_{k_1}, w_{k_2})$ depends on four variables. First, the current difference in utility levels, $\Delta_k u(c)$ which tells the immediate satisfaction of consuming c_{k_1} rather than c_{k_2}. However, consuming more today has a cost in future consumption and wealth levels tomorrow, $\beta \Delta_k V(g'_c(.))$. This is not the end of the story. Optimal allocation of probabilities requires trading off not only intertemporal levels of utility but also marginal intertemporal utilities where now the current marginal utility of consumption is weighted by the effort required to process information today.

To explore this relation further, I evaluate the derivative of the continuation value for a given optimal $p^*(c_{k_1}, w_{k_2})$, that is $\Delta_k V'_{p^*}(g'_c(.))$. To this end, define the ratio between differential in utilities (current and discounted future) and differential in marginal current utility as $\Psi^\kappa \equiv \frac{\Delta_k u(c(\kappa)) + \beta \Delta_k V(g'_c(.))}{-\theta\left[u'\left(c_{k_1}(\kappa)\right) - u'\left(c_{k_2}(\kappa)\right)\right]}$. Also, let Φ^κ be the ratio Ψ^κ when current level of utilities are equalized and future differential utilities are constant, i.e., $\Delta_k u(c) = 0$ and

$\Delta_k V\left(g'_c\left(.\right)\right)=1$ or, $\Phi^\kappa \equiv \frac{\beta}{-\theta\left[u'\left(c_{k_1}(\kappa)\right)-u'\left(c_{k_2}(\kappa)\right)\right]}$. Then, an application of Chain rule and point-wise differentiation leads to

$$p^*\left(c_{k_1}, w_{k_2}\right)=\Lambda\left(k_1, k_2\right) p^*\left(c_{k_1}\right) \tag{27}$$

where

$$\Lambda\left(k_1, k_2\right) \equiv \Lambda_1\left(\Psi^\kappa, p^*\left(c_{k_1}, w_{k_2}\right)\right) \cdot \Lambda_2\left(\Phi^\kappa, g'_{c_{k_1}}\left(\cdot\right), p^*\left(c_{k_1}, w_{k_2}\right)\right) \cdot \Lambda_3\left(\Phi^\kappa, g'_{c_{k_2}}\left(\cdot\right), p^*\left(c_{k_1}, w_{k_2}\right)\right)$$

The details of the derivation of (43) are in Appendix C. Here I focus on the explanation for the terms $\Lambda\left(k_1, k_2\right)$ which characterize the optimal solution of the conditional distribution $p^*\left(w_{k_2}|c_{k_1}\right)$.

The first term $\Lambda_1\left(\Psi^\kappa, p^*\left(c_{k_1}, w_{k_2}\right)\right) \equiv e^{\left(\frac{\Psi^\kappa}{p^*\left(c_{k_1}, w_{k_2}\right)}\right)}$ states that the optimal choice of the distribution balances differentials between current and future levels of utilities between high (k_2) and low (k_1) values of consumption. In case of log utility, the term $\exp\left(\Psi^\kappa\right)$ is a likelihood ratio between utilities in the two states of the word (k_1 and k_2) and the interpretation is that the higher is the value of the state of the world k_2 with respect to k_1 as measured by the utility of consumption, the lower is the optimal $p^*\left(c_{k_1}, w_{k_2}\right)$. This matches the intuition since the consumer would like to place more probability on the occurrence of k_2 the wider the difference between c_{k_1} and c_{k_2}. A perhaps more interesting intertemporal relation is captured by the terms Λ_2 and Λ_3, both of which display the occurrence of the update distribution $g'_{c_{k_i}}\left(\cdot\right)$, $i=1,2$. To disentangle the contribution of each argument of Λ_2 and Λ_3, I combine the derivative of the control with the envelope condition. Let Λ'_1 be the term Λ_1 led one period and define the differential between transition from one particular state to another and transition from one particular state to all the possible states as $\tilde{\Delta} T_j \equiv T\left(\cdot; w_{k_2}, c_{k_j}\right)-\left(\sum_i T\left(\cdot; w_i, c_{k_1}\right) p^*\left(w_i|c_{k_j}\right)\right)$ for $j=1,2$. Evaluating the derivative with respect to the state *almost surely* reveals that $\Lambda_2 \equiv \exp\left(-\Phi^\kappa \frac{\Lambda'_1}{p\left(c_{k_1}\right)} \tilde{\Delta} T_1\right)$ while $\Lambda_3 \equiv \exp\left(-\Phi^\kappa \frac{\Lambda'_1}{p\left(c_{k_2}\right)} \tilde{\Delta} T_2\right)$. The terms Λ_2 and Λ_3 reveal that in setting the optimal distribution $p^*\left(c_{k_1}, w_{k_2}\right)$ consumers take into account not only differential between levels and marginal utilities but also how the choice of the distribution shrinks or widens the spectrum of states that are reachable after observing the realized consumption profile.

An interesting special case that admits a close form solution is when the agent is risk neutral. Consider the framework in Section (3.2) and let utility take up the form $u\left(c\right)=c_t$, then in the region of admissible solution $c_t<w_t$, the optimal probability distribution makes c independent on w. To see this, it is easy to check that in the two period case with no discounting, the utility function reduces to $u\left(c\right)=w$, which implies $c|w \propto U\left(w_{\min}, w_{\max}\right)$. That is, since all the uncertainty is driven by w, the consumer does not bother processing information beyond the knowledge of where the limit of $c=w$ lies. In other word, the constraint on information flow does not bind. With continuation

value, exploiting risk neutrality, the optimal policy function amounts to:

$$p^{*}\left(w_{k_2}|c_{k_1}\right)=\frac{e^{\left(\frac{\left[\left(c_{k_1}-c_{k_2}\right)+\beta\Delta_k\bar{V}\left(g_c'(.)\right)\right]}{\theta}\right)}}{\sum_j\tilde{\Delta}T_j} \tag{28}$$

The solution uncovers some important properties of the interplay between risk neutrality and information flow. First of all, households with linear utility do not spend extra consumption units in sharpening their knowledge of wealth. This is due to the fact that since the consumer is risk neutral and, at the margin, costs and benefits of information flows are equalized amongst periods, there is no necessity to gather more information than the boundaries of current consumption possibilities. In each period, the presence of information processing constraint forces the consumer to allocate some utils to learn just enough to prevent violating the non-borrowing constraint. Once those limits are figured out, consumption profiles in the region $c < w$ are independent on the value of wealth.

Another special case that admits close form soluition when consumers are risk averse and have information-processing limits is the $3-points$ distribution illustrated in Appendix D.

6 Numerical Technique and its Predictions

I solve the discrete dynamic rational inattention consumption-saving model is to transform 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 states of the model (i.e., wealth) and solve it using dynamic programming.

For a model with n cores 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*[11] of belief states from an *affinely independent*[12] 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 continuous state space which is uncountably infinite. The implementation amounts to: I place a finite grid over the simplex point, I compute the values for points in the grid, and I use interpolation to evaluate all other points in the simplex.

In the following subsections I will fix some definitions, describe the techniques in details and discuss the results.

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

[12] A set of belief states $\{g_i\}$, $1 \leq i \leq |S|$ is called *affinely independent* when the vectors $\left\{g_i - g_{|S|}\right\}$ are linearly independent for $1 \leq i \leq |S|$.

6.1 Belief Simplex and Dynamic Programming

As mentioned previously, if I were to model wealth as the state of a Markov system directly accessible to the consumer, previous history of the process would be irrelevant to its optimal control. However, since the consumer does not know or 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 his 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 the history useless for infinite-horizon problems. To overcome this issue, Astrom (1965) proposed an *information state* approach. The latter is based upon the idea that all the information needed to act optimally can be summarized by a vector of probabilities over the system, called *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 can be looked up as *simplex.* The following definitions provide the formal basis for the construction of the grid for the simplex of the state.

Recall that $|S|$ is the dimension of the *belief simplex* which approximates the distribution $g(w)$ and $G \equiv \left\{ g \in \mathbb{R}^{|S|} : g(i) \geq 0 \text{ for all } i \ \sum_{i=1}^{|S|} g(i) = 1 \right\}$ is the set of all probability distribution on the simplex.

The discretization of the core states and the belief states amount to an equi-spaced grid with $n = 6, 7, 8$ values for w ranging from 1 to n i.e., $w \in \Omega_w \equiv \{w_1, .., w_n\}$ and $|S| \equiv 8, 9$ and 10 distinct values for the marginal pdf $g(w)$ in the interval $I_\Gamma \equiv [0, 1]$. Hence, the simplex result into a 1296x6 matrix for $(n; |S|) = (\{6\}; \{10\})$, 3003x7 matrix for $(n; |S|) = (\{7\}; \{9\})$ and, finally 11364x8 matrix for $(n; |S|) = (\{8\}; \{10\})$. Given the initial belief simplex, its successor belief states can be determined by using Bayesian conditioning at each multidimensional point of the simplex and amounts to the expression:

$$g'_c(\cdot) = \sum_i \sum_j T(\cdot; w_i, c_j) \Pr(w_i | c_j) = \Pr(\cdot | c). \tag{29}$$

Next, let me turn to the action space. Imposing the constraint that consumption cannot exceed wealth in each period, that is $c_t < w_t$, $\forall t$, I perform the discretization of the behavior space in a fashion similar to the core states, that is an equi-spaced grid where $c = \frac{1}{3}w$. As a result, the behavior space is the compact set $\Omega_c \equiv \{c_1, .., c_n\} = \left\{\frac{1}{3}w_1, .., \frac{1}{3}w_n\right\}$. Let core states and behavior states be sorted in descending order. Then, given the symmetry in the dimensionality of core space and behavior space and the constraint $c < w$, the the joint distribution of consumption and wealth for a given multidimensional point on the grid of the simplex is a square matrix with rows correspond to levels of consumption and 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 amounts to the marginal pdf of wealth, $g(w)$. Let $\mathcal{V}$ be the set of all bounded real-valued function V on G. Then the Bellman optimality equation of the household amounts to:

$$V(g(W)) = \max_{\Pr(c_j,w_i)} \left\{ \begin{array}{c} \sum_i \sum_j u(c_j) \Pr(c_j,w_i) + \\ +\beta \left\{ \sum_i \sum_j V\left(g'_{c_j}(\cdot)\right) \Pr(c_j,w_i) \right\} \end{array} \right\}$$

s.t

$$\kappa = \mathcal{I}(C;W) = \sum_i \sum_j \Pr(c_j,w_i) \log\left(\frac{\Pr(c_j,w_i)}{p(c_j)\,g(w_i)}\right)$$

Without loss of generality, I place the restriction that the columns of the matrix $\Pr(c_j,w_i)$ need 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 zeros makes the parameter vector 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 $\frac{n*(n-1)}{2}$.[13] 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 $V(g(W))$. I evaluate the value function that takes as argument the updated distribution of the wealth in (29), i.e., $V\left(g'_{c_j}(\cdot)\right)$ using linear interpolation.

A linear interpolant approximates the exact non linear value function in (20) with a piece-wise linear function. The following propositions illustrate this point.

Proposition 2. *If the utility is CRRA and if* $\Pr(c_j,w_i)$ satisfying (21)-(24)*, then the optimal* $n-step$ *value function* $V_n(g)$ *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 $\alpha - vectors$*,* $\alpha : \Omega_w \to R$*, are* $|W| - dimensional$ *hyperplanes.*

Intuitively, each $\alpha_n -$vector corresponds to a plan and the action associated with a given $\alpha_n -$vector is the optimal action for planning horizon n for all priors that have such

[13] To illustrate this point, two example in which the 0-degree of freedom and the $\frac{n*(n-1)}{2}$-degree of freedom occur are as follows. Suppose for simplicity that $n = 3$. Then, if a simplex point has realization $g \equiv \{1,0,0\}$ the joint pdf of consumption and wealth turns out to be $p(c,w) = \begin{bmatrix} 1 & 0 & 0 \\ & 0 & 0 \\ & & 0 \end{bmatrix}$ leaving zero degrees of freedom. If, instead, e.g., $g \equiv \left\{\frac{1}{3},\frac{1}{3},\frac{1}{3}\right\}$, the consumers has to choose $\frac{3*(2)}{2} = 3$ points on the joint distribution, $\{p_1,p_2,p_3\}$ placed as:

$$p(c,w) = \begin{bmatrix} \frac{1}{3} & p_1 & p_2 \\ & \frac{1}{3} & p_3 \\ & & \frac{1}{3} \end{bmatrix}.$$

a function as the maximizing one. With the above definition, the value function amounts to:

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

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 $n \to \infty$ 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 numerical the optimization of the value function at each point of the simplex, I use a via gradient-based search methods using Chris Sims's CSMINWEL and iterate on the value function up to convergence.

Finally, I draw from the optimal policy function -i.e., ergodic posterior joint distribution of consumption and wealth, $p^*(c, w)$ - and generate time series path of consumption, wealth and expected wealth evaluated by combing the core states with the posterior distribution of wealth which results from the optimization. Moreover I use the joint posterior $p^*(c, w)$ to draw the time path of Information Flow ($\kappa_t^* \equiv \sum_i \sum_j p_t^*(c_j, w_i) \log \left(\frac{p_t^*(c_j, w_i)}{p_t^*(c_j) g_t^*(w_i)} \right)$). A pseudocode that implements the procedure is in Appendix E.

7 Results

In this section, by varying the shadow cost of information flow and utility specifications, I investigate the dynamic interplay of information flows and degrees of risk aversion. In particular, I study three different models where each model is characterized by a given processing effort, θ, and different degrees of risk aversion, γ, where $\gamma = (\{7\}, \{5\}, \{3\}, \{0\}, \{0.5\}, \{0.3\})$.

The patterns that emerge are the following.

Result 1. Restricted Support *The optimal policy function for the information-constrained consumer places low weight, even zero, on low values of consumption for high values of wealth. This effect is more pronounced the higher the information flow.*

Figure 1a illustrates this point by comparing optimal policy function for two individuals with Log Utility and different information costs ($\theta = 0.2$ and $\theta = 7$). Both policy functions are drawn from a prior on wealth whose entropy is $\mathcal{H}(W) = 1.75$.

The consumer with high information flow ($\theta = 0.2$) allocates probability to avoid consuming too little for any given wealth. The more constrained agent ($\theta = 7$) has less informative signals and uses as little information as it takes to figure out the limits of his budget constraint.

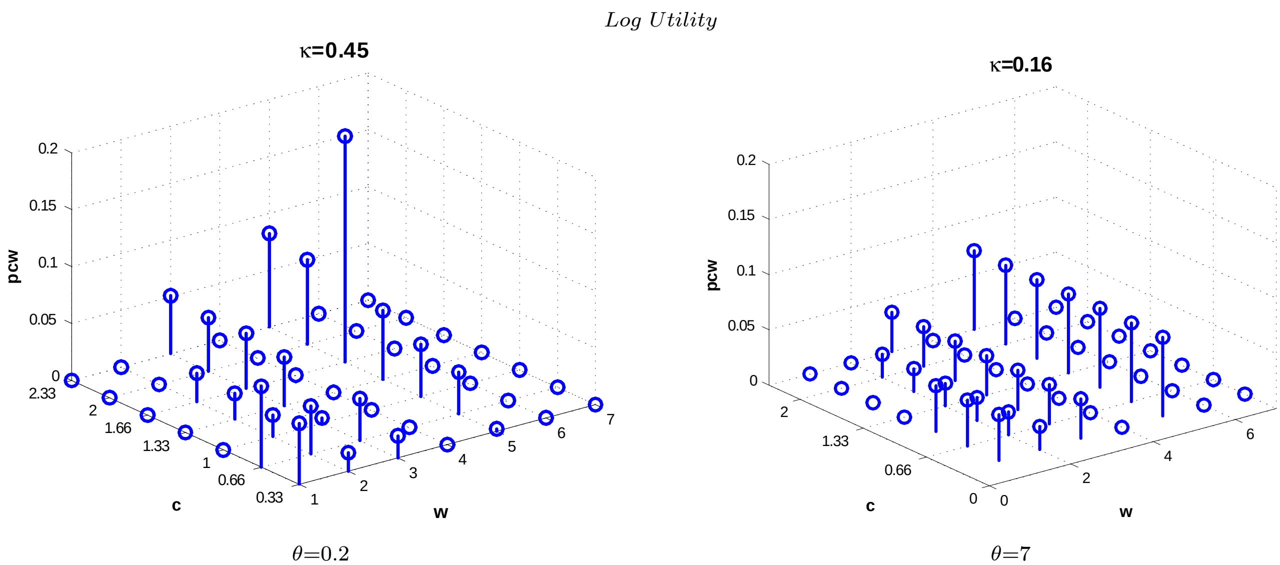
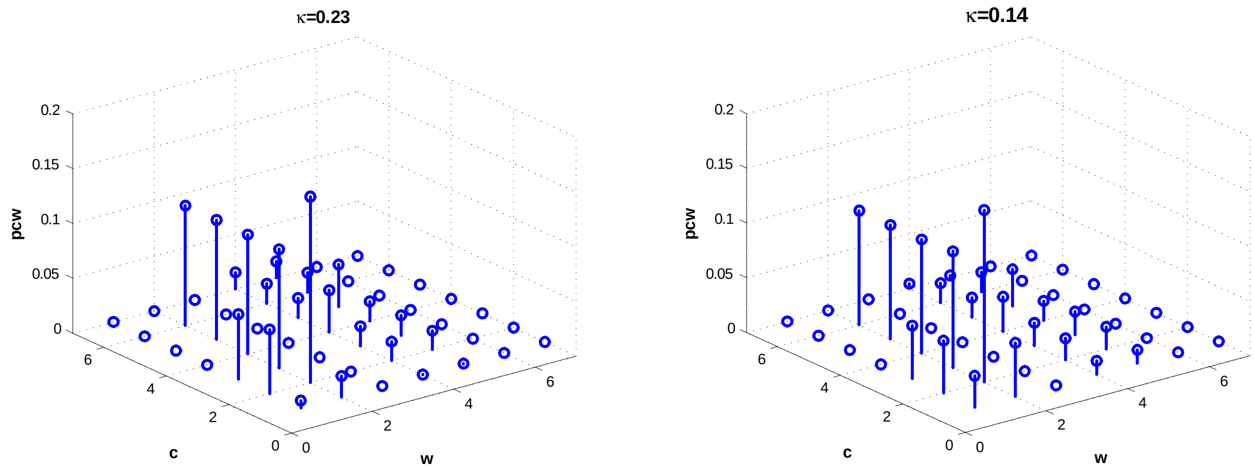


Figure 1a

Figure 1b shows how consumers with a CRRA utility specification with $\gamma = 5$ as coefficient of risk aversion allocates their capacity when the cost is $\theta = 0.2$ and $\theta = 2$ respectively.

CRRA, γ=5



θ=0.2 θ=2

Figure 1b

To compare the implications of different information processing costs on several utility functions, consider figures 2a-3d.

Figure 1c and 1d displays the conditonal distribution $p(c|w)$ when utility is logarithmic and $\mathcal{H}(W) = 2.50$. Consider Figure 2a. When $\theta = 0.2$, the distribution of consumption for a given value of wealth is centered around the deterministic optimal value of consumption $c \simeq \frac{(1-\beta)}{R}w + \beta\bar{y}$.

Log Utility, θ=0.2

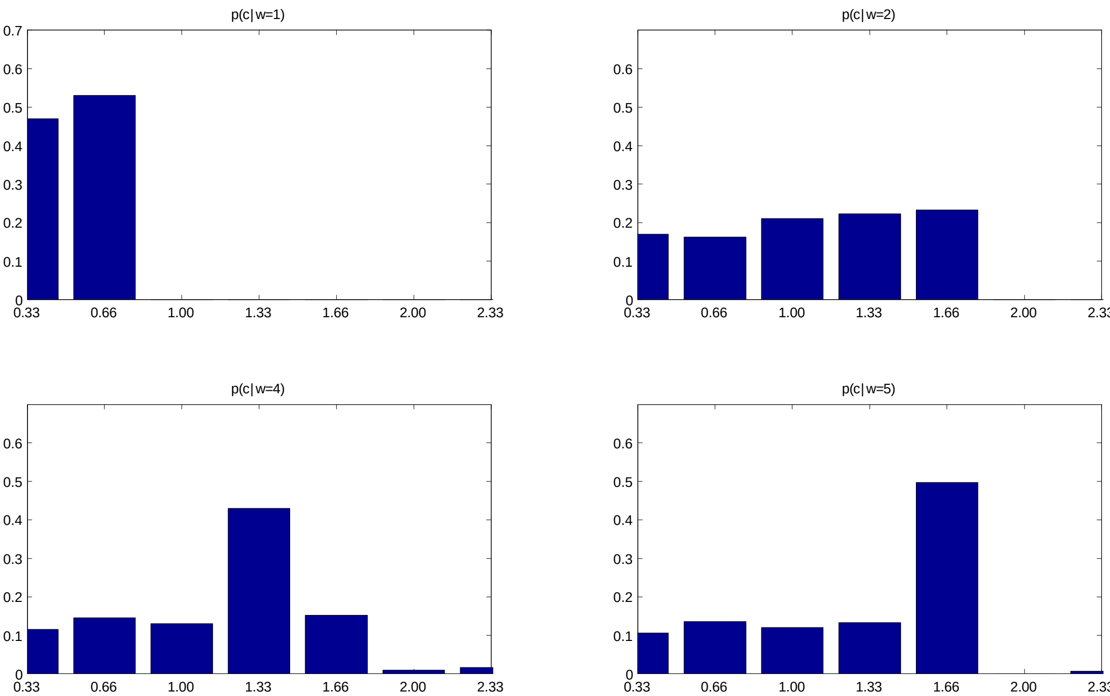


Figure 1c

Figure 1c shows what happens when information is costly. The probability distribution of consumption conditional on wealth is uniform over a range of c for given values of wealth. In particular, since information flows at a very low rate, any value of wealth reveals a much more disperse distribution of consumption for a $\theta = 7$-type than a $\theta = 0.2$-type.

Log Utility, θ=7

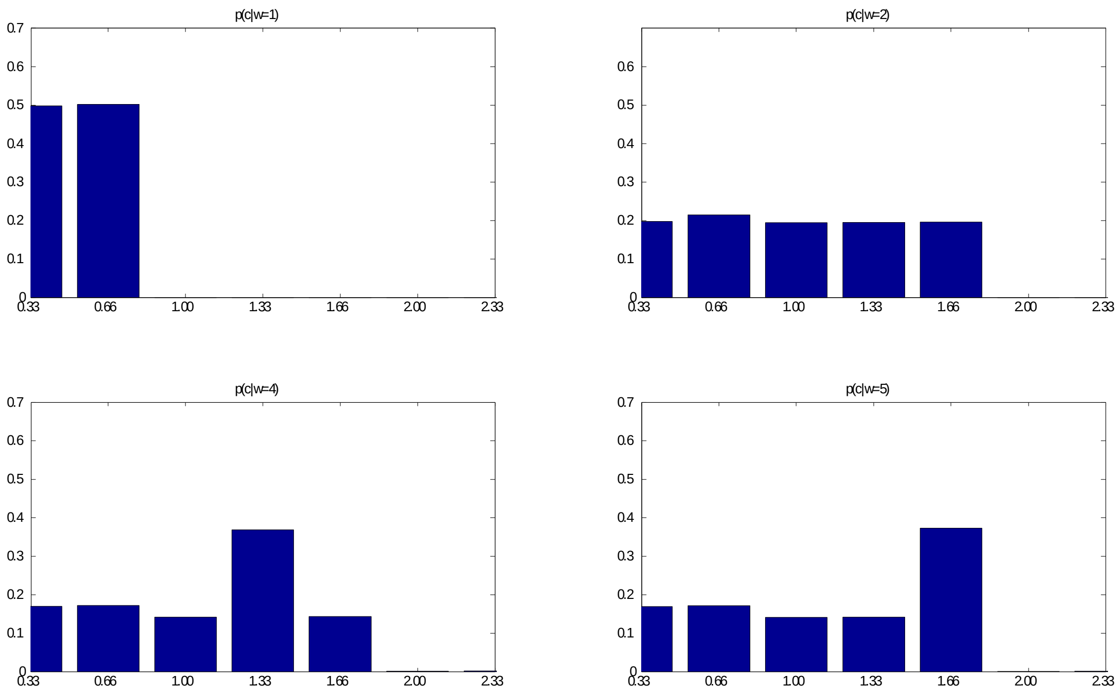


Figure 1d

Figures 1e-1f display the conditional distribution of wealth given consumption $p(w|c)$ for $\theta = 0.2$ and $\theta = 7$ respectively. Low signals (c) for the less information-constrained agent $(\theta = .2)$ are much more informative about lower values of wealth. High signals for this type are weakly more informative about high values of wealth. The highest signal is equally informative for the two $\theta's$. This is because once consumption is at its highest, the optimal policy for both the types is to shift all probability onto the lowest possible value of wealth that supports such consumption and the highest possible value of wealth which garantees sustaining such a consumption in the future.

Log Utility, θ=0.2

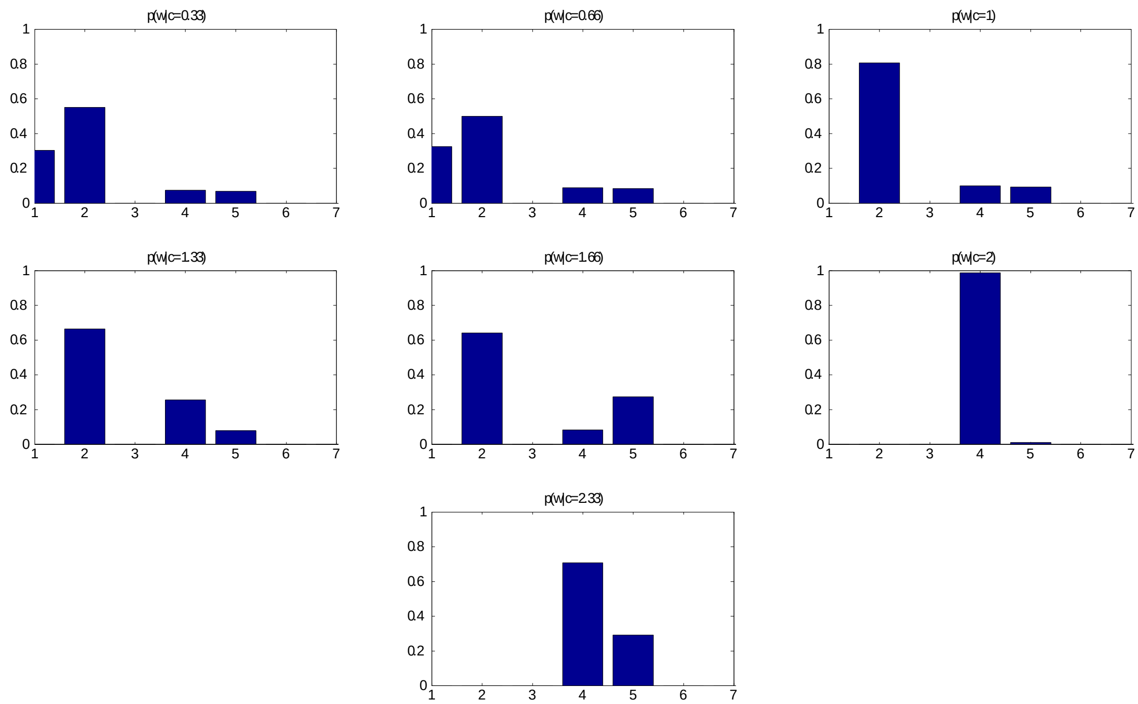


Figure 1e

Log Utility, θ=7

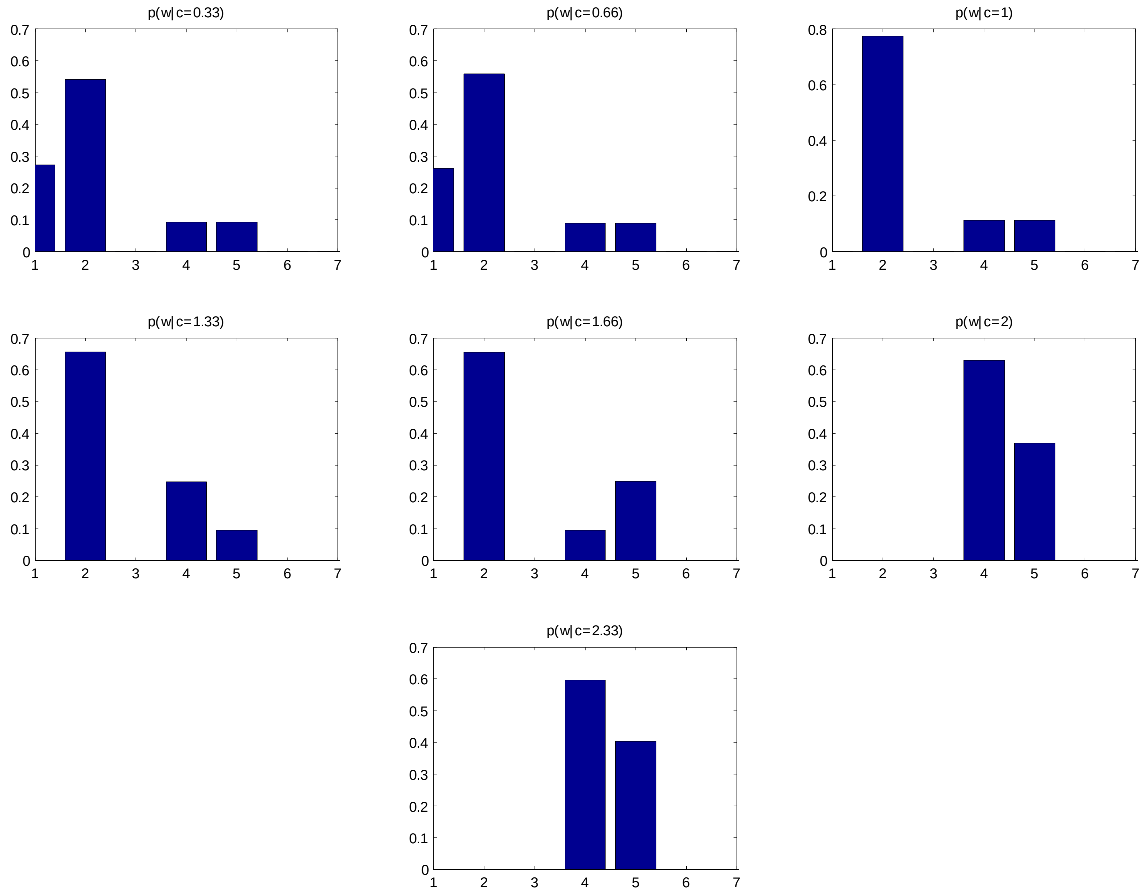
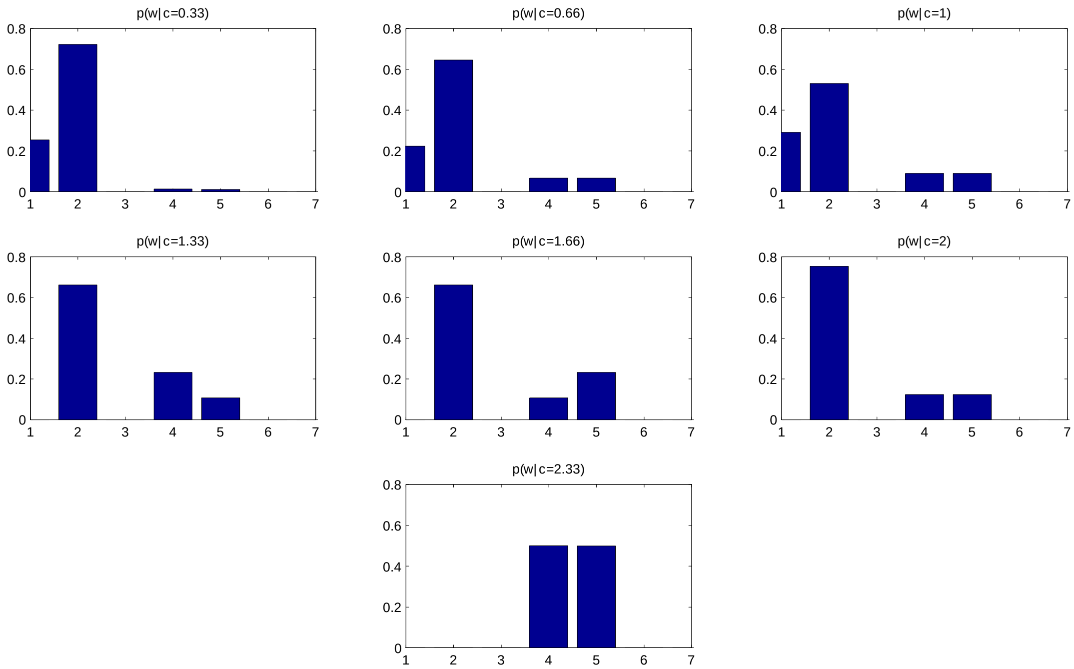


Figure 1f

Figure 1g and 1h display the optimal policy function with CRRA utility specification with coefficient of risk aversion $\gamma = 5$ and $\theta = 0.2$ and $\theta = 2$ repsectively. The initial prior is the same as before. Figure 2c and 2d show the conditional $p(w|c)$.

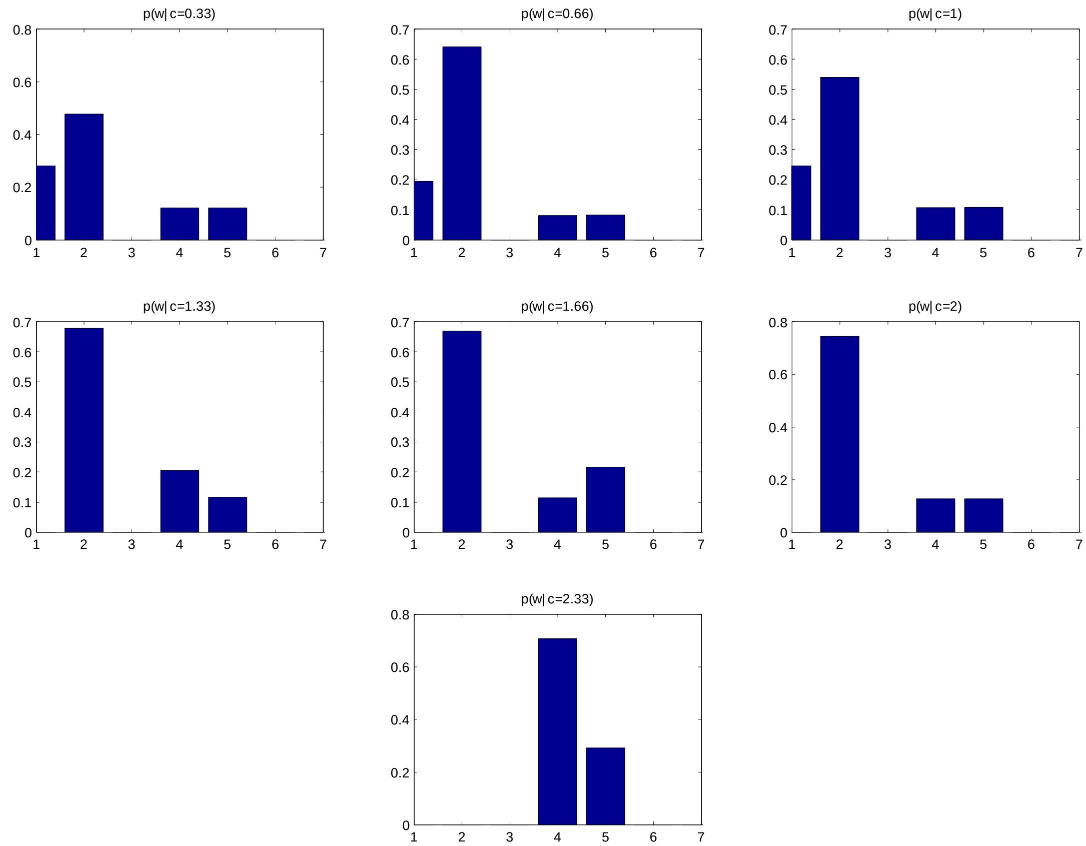
CRRA Utility, γ=5



θ=0.2

Figure 1g

CRRA Utility, γ=5

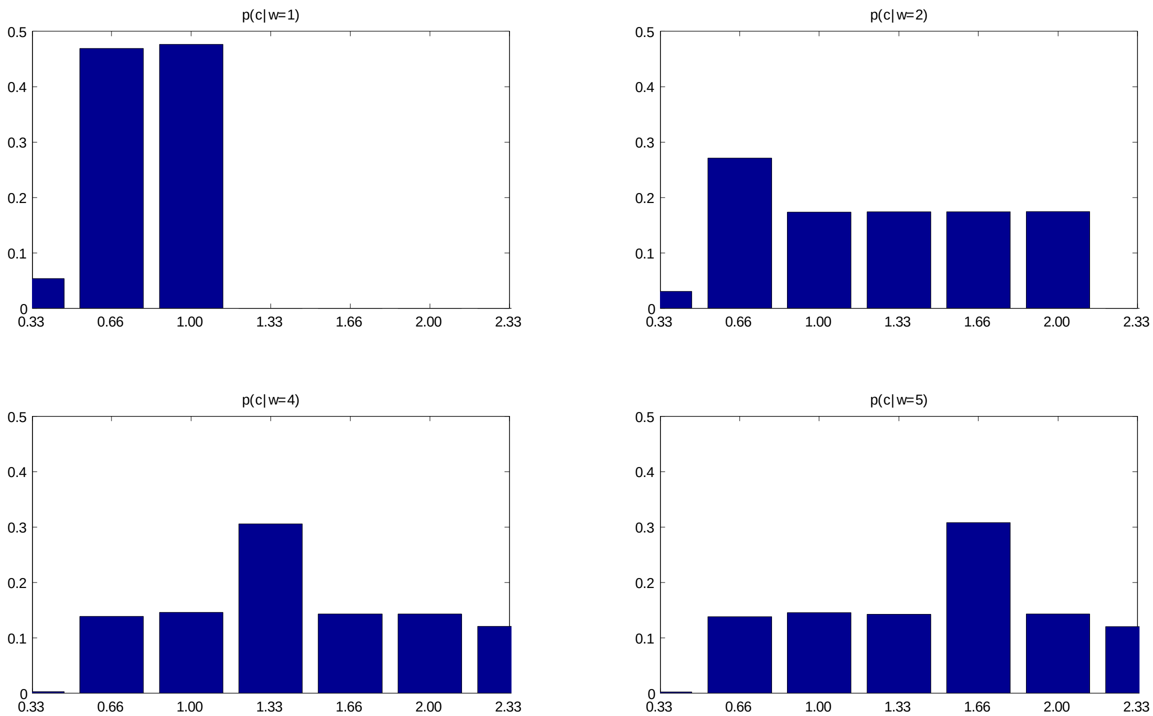


θ=2

Figure 1h

Figure 1i and 1j plots the corresponding $p(c|w)$ for the risk averse consumer.

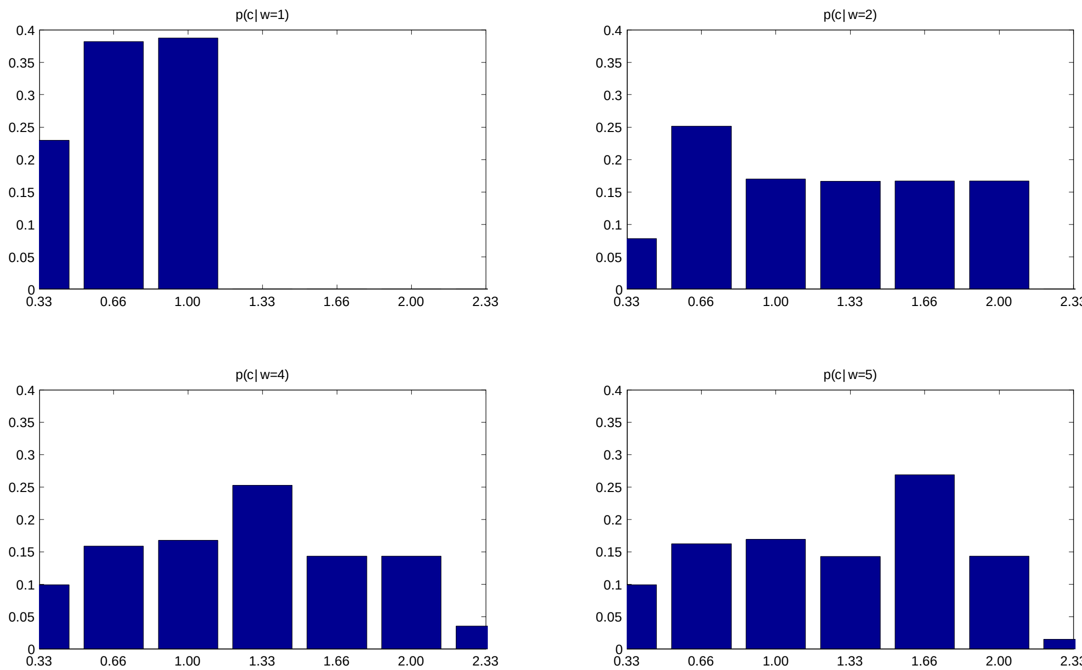
CRRA Utility, γ=5



θ=0.2

Figure 1i

CRRA Utility, γ=5



θ=2
Figure 1j

Result 2. Information Flow and Risk Aversion *In the discrete Rational Inattention Consumption-Saving model, the lower the optimal information flow, the lower is consumption. Moreover, when the cost of using the channel increases, more risk averse consumers pay more attention to decreasing the volatility of consumption rather than to increasing its mean.*

The results are documented in Tables 1a-1c. The intuition goes as follows. If processing information is relatively easy, i.e., information flow is large, then a risk averse consumer would spend his capacity in having both higher mean and lower variance of consumption throughout his life. However, when information processing becomes costly, the consumer would trade off information about the mean of the consumption for a reduction in its volatility.

The second result comes from evaluating the relation of the value function in steady state with the expected value of wealth.

Result 3. Optimal Value and Expected Wealth *In the discrete Rational Inattention Consumption-Saving model, the fixed-point solution of the value function,* $V^*\left(g^*\left(w\right)\right)$, *is positively related with the expected wealth,* $E^*\left(w\right) \equiv \sum_i w_i g^*\left(w_i\right)$ *evaluated under the ergodic distribution of the posterior for wealth. Moreover:*

1. *For a given degree of risk aversion, γ, the dispersion between $V^*(g^*(w))$ and $E(g^*(w))$ is higher the lower is the information flow.*

2. *For a given information flow, the lower the degree of risk aversion, the steeper the correlation between $V^*(g^*(w))$ and $E^*(w)$.*

The scatter plots in Figures 3a-3b illustrate the proposition. While it is expected that the fixed point solution of the value function covaries positively with the expected value of wealth, the numerical regularities expressed in (1.) and, especially, in (2.) are less obvious.

The result in (1.) says that for a given degree of risk aversion, an agent that can process information inexpensively (high κ_t) will attribute an higher value of having a signal $g(w)$ than an individual who finds costly to process information since that signal conveys more information. On the other extreme, consumers who have high cost of processing information have less informative signals about wealth. This in turn implies that the optimal value function defined on the distribution $g^*(w)$ when information flow is low is more concentrated around lower (expected) wealth capturing the remaining ex-post uncertainty after processing information.

The message of (2) is more subtle. As the numerical simulations show, for a given cost of processing information, a more risk adverse consumer (say, one for which $\gamma = 2$) will place low probability on having high wealth than an individual whose degree of risk aversion is $\gamma = .5$. The reason is that since when the degree of risk aversion is high, utility goes to minus infinity whenever consumption is close to zero. Thus, a consumer with $\gamma = 2$ would like to acquire more information or, for fixed κ, be more prudent in his expenditures whenever he fears that his wealth is low.

Next, let me turn to the time series properties of consumption and wealth. Analyzing the scatter plots in 5a-5h leads to the following:

Result 4. Time Path of Consumption. *In the discrete Rational Inattention Consumption-Saving Model, consumption over time displays:*

1. *an hump-shaped responses to fluctuations in expected wealth. Moreover the lower the information flow, the more responses of consumption to fluctuation of wealth are delayed and noisy.*

2. *an asymmetric response to shocks. Consumers reacts faster and sharper to signals about low wealth while signals about high wealth are absorbed slower over time.*

Figures 11a-11b illustrate the result via impulse response function. The pictures display the difference in consumption between two household with the same characteristics but wealth level. One household has a very low wealth while the other has higher wealth. They have the same policy function and the same information flow. Figure 11a depicts the difference in consumption when the two households has log-utility and information costs of $\theta = 0.2$ and $\theta = 1$. Such difference is evaluated as an average of 20000 Monte Carlo draws. The picture desplays that the difference in consumption has an hump-shaped and delayed acknowledgement that the wealth is low. Figure 11b shows the result for the same exercise when the utility belongs to the CRRA family with coefficient of risk aversion $\gamma = 5$.

Risk averse agents would rather push forward consumption in times in which they are processing information about wealth. Therefore, processing information in a given period requires a loss in consumption for that period .

Agents who have low processing capacities would rather select a flat consumption profile than to spending utils of consumption for tracking wealth. This is due to the fact that the information they gather provides them with an imperfect signal of wealth. Thus, low-κ type consumers smooth consumption over their life time by saving a fixed amount each period unless their signals on high wealth is sharp enough to justify a change in their behavior.

The more risk averse consumers are, the less responsive to information is their consumption path. This pattern of consumption behavior matches what we observe in macro data on consumption and documented in the literature as *excess smoothness*. Furthermore, the discrete rational inattention consumption-saving model provides a rationale for *excess sensitivity* in response to news on wealth.[14] High-κ type consumers react to signals about changes in wealth by modifying their consumption. The lower their risk aversion, the more keen these types of agents are to adjust their consumption in accordance to the signals on wealth they process. The histograms of consumption across different types of risk averse individuals under different limits of information processing is displayed in Figures 4a-4b.

8 Computational Complexity and Extensions

The main computational challenge in solving dynamic rational inattention problems is that they require evaluating every one of an uncountably infinite number of belief states and control distributions. The proposed algorithm based on the approximation of the value function results into an exponential growth of number of dimensions as I increase the precision of the approximation of states and control distribution and the number of belief points representing the states in the grid. While it is not tractable for all but trivial problems to find exact solutions for these models, it is worth investigating more sophisticated approximation methods that may work in these cases.

[14] *Excess sensitivity* (Flavin, 1981) of consumption refers to the empirical evidence that aggregate consumption reacts with delays to anticipated changes in income while *excess smoothness* (Deaton, 1987) documents that aggregate consumption reacts less than one-to-one to unanticipated income changes. There is no univoque explanation in economic theory for these two features of U.S. data on aggregate consumption.

One suitable extension of the methodology I propose is a grid-based approximation with fixed- and variable- resolution regular grids. Design of a grid-approximation strategy requires addressing two main questions. The first pertains the trade-off between time and quality that the interpolation method provides. The second question amounts to asking what belief states should be included in the grid. Every grid must contain the corner points of the belief simplex to ensure that a convex combination for the approximation of the value function can be found. There are different strategies for adding other points to the grid based on the way in which grids are generated. The methodologies split into two subclasses: fixed-grid and variable-grid methods. Fixed grids are generally constructed only once in a regular pattern; whereas variable grids are revised during calculations to achieve better approximation schemes. Each type of grid has its strengths and weaknesses.

The advantage of regular grid is that it allows an elegant and efficient method of interpolation based on Freudenthal triangulation. Lovejoy (1991). The disavantages of a fixed-resolution grid is that increasing the resolution of the value function approximation in one region of the belief simplex requires increasing its resolution everywhere. This causes an exponential explosion in the size of the grid and makes this approach intractable for all but small-sized problems. Lovejoy (1991) also suggested a variable-resolution generalization of his method where the mesh is adjusted as the algorithm progresses to add where it is most needed and subtract where it is least needed. However Lovejoy did not developed this extension. Other have explored the use of non-regular grids that allow grid points to be placed everywhere in the belief simplex. Although a non-regular grid avoids the exponential explosion of a fixed-resolution regular grid, its serious drawback is that it relies on methods of interpolation more complex and less efficient than triangulation.

Although the complexity of interpolation of regular grid does not depend on the size of the grid, the complexity of value iteration in both regular and not regular grid does. Values for grid belief states are computed using value iteration. For a grid-based upper-bound function, the complexity of the value iteration is $O\left(|G|\,|C|\,|G'|\,|S|^2 \lg\lg M\right)$ where $|G|$ is the number of grid points, $|C|\,|G'|$ is the number of successor belief states of each grid point, $|S|^2$ is the worst-case complexity of computing successor belief states by Bayesian conditioning and $\lg\lg M$ reflects the added complexity of interpolation with M the number of subintervals that each edge of the state simplex is divided up into (*resolution* of the grid). Since both success belief states and coordinates used for the interpolation can be cached for re-use the complexity of each subsequent iteration of value iteration in a variable-resolution grid is $O\left(|G|\,|C|\,|G'|\,|S|\right)$.

If a variable-resolution regular grid is not refined during value iteration, then each iteration uses the same grid belief states and coordinates for interpolation. In non-regular grids, the coordinates for interpolation are re-computed each iteration. Since the complexity of interpolation in non regular grids depend on $|G|$, the complexity of each iteration of value iteration is quadratic in $|G|$. This underscores the advantage of using a regular grid.

In addition to limiting the complexity of interpolation, it is helpful to limit the factor $|G|\,|C|\,|G'|$, which is the number of times interpolation (and other calculations involved in computing a backup) are performed in each iteration of value iteration. The trivial way to limit this factor is to limit the size of the grid. This adjusts a trade-off between quality of approximation and grid size, and a variable-resolution grid lets us space the point of the grid to achieve the best possible approximation for a given grid size. This procedure needs to be extended to the controls too to limit $|C|\,|G'|$. It is possible to reduce the size of the latter by ignoring some observations and the evaluate the successor belief states on a coarser set of probability controls. However, this impairs the quality of the approximation even though it can be regarded as an opportunity to adjust a time-quality trade-off. This trade-off can be adjusted by only considering the most probable observation or by only considering observation when the Mutual Information is above some threshold[15]. However, if only the successor belief states of control probability are considered in grid-based interpolation, the resulting approximation may no longer be an upper bound since information provided in some region of the joint pdf $p^*(c, w)$ is ignored.

Finally, it is worth mentioning a rich set of algorithms drawn from Artificial Intelligence that algorithms perform point-based value iteration. The idea is to identify a coarse set of belief points, say B. At each value backup stage, the algorithms improve the value of all points in B by only updating the value and its gradient of a (randomly selected) subset of points. In each backup stage, given a value function V_n, the next step value function V_{n+1} improves the value of all the point in B. The algorithm keeps iterating until some convergence criterion is met. Although point-based value iteration could be a promising avenue to explore in future work on rational inattention, as of now it has been experiencing mixed results in terms of time-quality in the experiments run so far. Furthermore, since the framework in the AI literature deals with infinite state space but finite and extremely limited control space, the mapping between that and rational inattention model is not immediate nor likely to be useful in the immediate future.

9 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 degree of risk aversion and endogenous choice of information flow enhances precautionary savings.

Numerical analyses show that for a given degree of risk aversion, the lower the information flow, the flatter consumption path. The model further predicts that for a given information flow, the more risk averse the consumer is, the more he would like to acquire information on his wealth to reduce volatility on consumption rather than increasing his lifetime consumption average.

[15] Note the analogy of this methodology with that of the *Reverse water-filling* and *Nyquist-Shannon* Theorem for independent Gaussian. random variables. See Thomas and Cover (1991).

The model predicts that consumers with processing capacity constraints have asymmetric responses to shocks, with negative shocks producing more persistent effects than positive ones. This asymmetry is novel to the literature of consumption and savings and it implies that, in my model, consumption displays excess smoothness with respect to positive income shocks and excess sensitivity when the shock is averse.

The first priority for future research includes extending the dimensionality of the discretization by improving on the grid proposed. An approach that seems promising for optimizing the discretization which might help in facing the curse of dimensionality is to use a variable resolution grid instead of a fixed resolution one. Another approach would be to use point-based value iteration which minimizes the simplex points over which the value function is computed. The second priority would be to enrich the model by allowing more general income processes and augmenting choice space of the household with labor. Moreover, it will be interesting given this extended setting to study the (strategic) interactions of consumers characterized by different degrees of risk aversion and information flow. This will make my model the demand side of a DSGE model where risk averse consumers have limited information capacity and producers have higher processing capacity then their costumers and compete strategically. Given this setting, it would be interesting to explore the role of a fully informed Central Bank in driving the expectations of the private sectors via information conveyed by its policy.

There is still a long way to go to solve this latter problem but my paper suggests that this road, although challenging, is worth (rational) attention.

10 Appendix A

10.1 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,

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

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

10.2 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

$$\begin{aligned}
HV(g) &= H^{p_1}V(g), \\
HU(g) &= H^{p_2}U(g).
\end{aligned}$$

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

$$V \leq U \Longrightarrow$$

$\forall g(w), c,$

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

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

$$\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c}u(c)\,p_1\right)g(w)+\beta\sum_{c\in\Omega_c}V\left(\left.g'_c(\cdot)\right|_{p_1}\right)\cdot p_1 g$$
$$\leq\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c}u(c)\,p_1\right)\Longrightarrow$$
$$H^{p_1}V(g)\leq H^{p_1}U(g)\Longrightarrow$$
$$H^{p_1}V(g)\leq H^{p_2}U(g)\Longrightarrow$$
$$HV(g)\leq HU(g)\Longrightarrow$$
$$HV\leq HU.$$

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

10.3 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:

$$V_0(g)=\max_{p\in\Gamma}\left[\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c}u(c)\,p\right)g(w)\right] \tag{30}$$

and therefore if I define the vectors

$$\left\{\alpha_0^i(w)\right\}_i\equiv\left(\sum_{c\in\Omega_c}u(c)\,p\right)_{p\in\Gamma} \tag{31}$$

I have the desired

$$V_0(g)=\max_{\left\{\alpha_0^i(w)\right\}_i}\left\langle\alpha_0^i,g\right\rangle \tag{32}$$

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 (25):

$$V_n(g)=\max_{p\in\Gamma}\left[\begin{array}{c}\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c}u(c)\,p(c|w)\right)g(w)+\\+\beta\sum_{w\in\Omega_w}\sum_{c\in\Omega_c}\left(V_{n-1}\left(g'_c(\cdot)_c\right)\right)p(c|w)\,g(w)\end{array}\right] \tag{33}$$

by the induction hypothesis

$$V_{n-1}\left(g\left(\cdot\right)|_{c}\right)=\max_{\left\{\alpha_{n-1}^{i}\right\}_{i}}\left\langle \alpha_{n-1}^{i},g_{c}^{\prime}\left(\cdot\right)\right\rangle \tag{34}$$

Plugging into the above equation (22) and by definition of $\langle .,.\rangle$,

$$V_{n-1}\left(g_{c}^{\prime}\left(\cdot\right)\right)=\max_{\left\{\alpha_{n-1}^{i}\right\}_{i}}\sum_{w^{\prime}\in\Omega_{w}}\alpha_{n-1}^{i}\left(w^{\prime}\right)\left(\sum_{w\in\Omega_{w}}\sum_{c\in\Omega_{c}}T\left(\cdot;w,c\right)\frac{\Pr\left(w,c\right)}{\Pr\left(c\right)}\right) \tag{35}$$

With the above:

$$\begin{aligned} V_{n}\left(g\right) &= \max_{p\in\Gamma}\left[\begin{array}{c}\displaystyle\sum_{w\in\Omega_{w}}\left(\sum_{c\in\Omega_{c}}u\left(c\right)p\right)g\left(w\right)+\\ \displaystyle+\beta\max_{\left\{\alpha_{n-1}^{i}\right\}_{i}}\sum_{w^{\prime}\in\Omega_{w}}\alpha_{n-1}^{i}\left(w^{\prime}\right)\left(\sum_{w\in\Omega_{w}}\left(\sum_{c\in\Omega_{c}}\tfrac{T\left(\cdot;w,c\right)}{\Pr\left(c\right)}\cdot p\right)g\left(w\right)\right)\end{array}\right] \\ &= \max_{p\in\Gamma}\left[\left\langle u\left(c\right)\cdot p,\ g\left(w\right)\right\rangle+\beta\sum_{c\in\Omega_{c}}\frac{1}{\Pr\left(c\right)}\max_{\left\{\alpha_{n-1}^{i}\right\}_{i}}\left\langle\sum_{w^{\prime}\in\Omega_{w}}\alpha_{n-1}^{i}\left(w^{\prime}\right)T\left(\cdot;w,c\right)\cdot p,\ g\right\rangle\right] \end{aligned} \tag{36}$$

At this point, it is possible to define

$$\alpha_{p,c}^{j}\left(w\right)=\sum_{w^{\prime}\in\Omega_{w}}\alpha_{n-1}^{i}\left(w^{\prime}\right)T\left(\cdot:w,c\right)\cdot p. \tag{37}$$

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

$$V_{n}\left(g\right)=\max_{p\in\Gamma}\left[\left\langle u\left(c\right)\cdot p,\ g\right\rangle+\beta\sum_{c\in\Omega_{c}}\frac{1}{\Pr\left(c\right)}\max_{\left\{\alpha_{p,c}^{j}\right\}_{j}}\left\langle\alpha_{p,c}^{j},g\right\rangle\right], \tag{38}$$

and define:

$$\alpha_{p,c,g}=\arg\max_{\left\{\alpha_{p,c}^{j}\right\}_{j}}\left\langle\alpha_{p,c}^{j},g\right\rangle. \tag{39}$$

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

$$\begin{aligned} V_{n}\left(g\right) &= \max_{p\in\Gamma}\left[\left\langle u\left(c\right)\cdot p,\ g\right\rangle+\beta\sum_{c\in\Omega_{c}}\frac{1}{\Pr\left(c\right)}\left\langle\alpha_{p,c,g},g\right\rangle\right] \\ &= \max_{p\in\Gamma}\left\langle u\left(c\right)\cdot+\beta\sum_{c\in\Omega_{c}}\frac{1}{\Pr\left(c\right)}\alpha_{p,c,g},\ g\right\rangle. \end{aligned} \tag{40}$$

Now

$$\left\{\alpha_{n}^{i}\right\}_{i}=\bigcup_{\forall g}\left\{u\left(c\right)\cdot p+\beta\sum_{c\in\Omega_{c}}\frac{1}{\Pr\left(c\right)}\alpha_{p,c,g}\right\}_{p\in\Gamma} \tag{41}$$

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

10.4 Proof of Proposition 3.

Proof. The first task is to prove that $\{\alpha_n^i\}_i$ 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 (31), the set of vectors $\{\alpha_0^i\}_i$,

$$\{\alpha_0^i\}_i \equiv \left(\sum_{w\in\Omega_w}\left(\sum_{c\in\Omega_c}\frac{c^{1-\gamma}}{1-\gamma}p(c|w)\right)g(w)\right)_{p\in\Gamma}$$

is discrete. For the general case, observe that for discrete controls and assuming $M = \left|\left\{\alpha_{n-1}^j\right\}\right|$, the sets $\left\{\alpha_{p,c}^j\right\}$ are discrete, for a given action p and consumption c, I can only generate $\alpha_{p,c}^j-$vectors. Now, fixing p it is possible to select one of the M $\alpha_{p,c}^j-$vectors for each one of the observed consumption c and, thus, $\{\alpha_n^j\}_i$ 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 $\{\alpha_n^j\}_i$ set is of finite cardinality. It follows that V_n is defined as a finite set of linear functions. ■

11 Appendix B

11.1 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. ■

12 Appendix C

12.1 Optimality Conditions

Derivative with Respect to Controls In the main text, I state that the optimal control amounts to :

$\partial p^{*}\left(c_{k_1}, w_{k_2}\right)$:

$$\Delta_k u\left(c\left(\kappa\right)\right) + \beta \Delta_k V\left(g_c'\left(.\right)\right) = p^{*}\left(c_{k_1}, w_{k_2}\right)\left(-\theta \Delta_k u'\left(c\left(\kappa\right)\right) + \beta \Delta_k V_{p^{*}}\left(g_c'\left(.\right)\right)\right) \tag{42}$$

which can be rewritten, opening up the operator Δ_k as:

$$\varphi^{\kappa}_{\left(c_{k_1}, c_{k_2}\right)} = \Pr\left(c_{k_1}, w_{k_2}\right)\left(\psi^{\kappa}_{\left(c_{k_1}, c_{k_2}, \theta\right)} \ln \frac{\Pr\left(c_{k_1}, w_{k_2}\right)}{\Pr\left(c_{k_1}\right)} + \beta \left[\frac{\partial V'\left(g'_{c_{k_1}}\left(\cdot\right)\right)}{\partial \Pr\left(c_{k_1}, w_{k_2}\right)} - \frac{\partial V'\left(g'_{c_{k_2}}\left(\cdot\right)\right)}{\partial \Pr\left(c_{k_1}, w_{k_2}\right)}\right]\right)$$

where

- $\varphi^{\kappa}_{\left(c_{k_1}, c_{k_2}\right)} \equiv -\left[u\left(c_{k_1}\left(\kappa\right)\right) - u\left(c_{k_2}\left(\kappa\right)\right) + \beta\left(V\left(g'_{c_{k_1}}\left(\cdot\right)\right) - V\left(g'_{c_{k_2}}\left(\cdot\right)\right)\right)\right]$, and
- $\psi^{\kappa}_{\left(c_{k_1}, c_{k_2}, \theta\right)} \equiv -\theta\left[u'\left(c_{k_1}\left(\kappa\right)\right) - u'\left(c_{k_2}\left(\kappa\right)\right)\right].$

Note that by Chain rule $\frac{\partial V'\left(g'_{c_{k_j}}\left(\cdot\right)\right)}{\partial \Pr\left(c_{k_j}, w_{k_2}\right)} = \frac{\partial V'\left(g'_{c_{k_j}}\left(\cdot\right)\right)}{\partial\left(g'_{c_{k_j}}\left(\cdot\right)\right)} \frac{\partial\left(g'_{c_{k_j}}\left(\cdot\right)\right)}{\partial \Pr\left(c_{k_j}, w_{k_2}\right)}$, for $j = 1, 2$. Plug (26) in the second term of the above expression and evaluating pointwise the derivatives delivers

In $c_j = c_{k_1}$,

$$\Longrightarrow \frac{\partial g\left(\cdot|_{c_{k_1}}\right)}{\partial \Pr\left(c_{k_1}, w_{k_2}\right)} = \frac{\partial\left[\frac{1}{p\left(c_{k_1}\right)}\left(\sum_i T\left(\cdot; w_i, c_{k_1}\right) \Pr\left(w_i, c_{k_1}\right)\right)\right]}{\partial \Pr\left(c_{k_1}, w_{k_2}\right)} =$$

$$\frac{1}{p\left(c_{k_1}\right)}\left(T\left(\cdot; w_{k_2}, c_{k_1}\right) - \frac{\left(\sum_i T\left(\cdot; w_i, c_{k_1}\right) \Pr\left(w_i, c_{k_1}\right)\right)}{p\left(c_{k_1}\right)}\right)$$

Define $\Psi^{\kappa} \equiv \frac{\varphi^{\kappa}_{\left(c_{k_1}, c_{k_2}\right)}}{\psi^{\kappa}_{\left(c_{k_1}, c_{k_2}\right)}}$ and $\Phi^{\kappa} \equiv \frac{\beta}{\psi^{\kappa}_{\left(c_{k_1}, c_{k_2}, \theta\right)}}$, and to get rid of cumbersome notation, let $\left(k_1, k_2\right) \equiv \left(\Psi^{\kappa}, \Phi^{\kappa}, g'_{c_{k_1}}\left(\cdot\right), g'_{c_{k_2}}\left(\cdot\right), \Pr\left(c_{k_1}, w_{k_2}\right),\right)$. Then the first order conditions result into

$$\Pr\left(c_{k_1}, w_{k_2}\right) = \Lambda\left(k_1, k_2\right) \Pr\left(c_{k_1}\right) \tag{43}$$

where

$$\Lambda\left(k_1, k_2\right) \equiv \Lambda_1\left(\Psi^{\kappa}, \Pr\left(c_{k_1}, w_{k_2}\right)\right) \cdot \Lambda_2\left(\Phi^{\kappa}, g'_{c_{k_1}}\left(\cdot\right), \Pr\left(c_{k_1}, w_{k_2}\right)\right) \cdot \Lambda_3\left(\Phi^{\kappa}, g'_{c_{k_2}}\left(\cdot\right), \Pr\left(c_{k_1}, w_{k_2}\right)\right)$$

while

- $\Lambda_1 \left(\Psi^{\kappa}, \Pr\left(c_{k_1}, w_{k_2}\right)\right) \equiv e^{\left(\Psi^{\kappa} \frac{1}{\Pr\left(c_{k_1}, w_{k_2}\right)}\right)}$;
- $\Lambda_2 \left(\Phi^{\kappa}, g'_{c_{k_1}}(\cdot), \Pr\left(c_{k_1}, w_{k_2}\right)\right) \equiv e^{\left(-\Phi^{\kappa} \frac{\partial V\left(g'_{c_{k_1}}(\cdot)\right)}{\partial\left(g'_{c_{k_1}}(\cdot)\right)} \frac{1}{p\left(c_{k_1}\right)} \left(T\left(\cdot; w_{k_2}, c_{k_1}\right) - \frac{\left(\sum_i T\left(\cdot; w_i, c_{k_1}\right) \Pr\left(w_i, c_{k_1}\right)\right)}{p\left(c_{k_1}\right)}\right)\right)}$;
- $\Lambda_3 \left(k_1, k_2\right) \equiv e^{\left(\Phi^{\kappa} \frac{\partial V\left(g'_{c_{k_2}}(\cdot)\right)}{\partial\left(g'_{c_{k_2}}(\cdot)\right)} \frac{1}{p\left(c_{k_2}\right)} \left(T\left(\cdot; w_{k_2}, c_{k_2}\right) - \frac{\left(\sum_i T\left(\cdot; w_i, c_{k_1}\right) \Pr\left(w_i, c_{k_2}\right)\right)}{p\left(c_{k_2}\right)}\right)\right)}$.

Derivative with Respect to States To derive the envelope condition with respect to a generic state $g(w_k)$ for $k = 1, 2, 3$, let me start by placing the restrictions on the marginal distribution of wealth in the main diagonal of the joint distribution $\Pr(c, w)$. The derivative then amounts to:

$$\frac{\partial \Pr_{\ t}(c_j, w_k)}{\partial g(w_k)} = \frac{\partial \Pr_{\ t}(c_j)}{\partial g(w_k)} = \begin{cases} 1 & \{(j = k) \cap (j \neq \max l \in \Omega_c)\} \\ -1 & \{j = \max l \in \Omega_c\} \\ 0 & \text{o/whise} \end{cases}.$$

Let $l_{\max}$ denote the maximum indicator l belonging to Ω_c. Then the derivative of the state $g(w_k)$ displays:

$$\frac{\partial V(g(w_k))}{\partial g(w_k)} \overset{a.s}{=}$$

$$\left(u\left(c_k(\kappa)\right) + \beta\left(V\left(g'_{c_k}(\cdot)\right)\right) - \left(u\left(c_{l_{\max}}(\kappa)\right) + \beta V\left(g'_{c_{l_{\max}}}(\cdot)\right)\right)\right) +$$

$$-\theta\left(\log\left(\frac{\Pr(c_k, w_k)}{p(c_k) g(w_k)}\right) u'\left(c_k(\kappa)\right) \Pr\left(c_k, w_k\right) - \log\left(\frac{\Pr\left(c_{l_{\max}}, w_k\right)}{p\left(c_{l_{\max}}\right) g(w_k)}\right) u'\left(c_{l_{\max}}(\kappa)\right) \Pr\left(c_{l_{\max}}, w_k\right)\right) +$$

$$+\beta \sum_j \left[\frac{\partial V\left(g'_{c_{k_j}}(\cdot)\right)}{\partial\left(g'_{c_{k_j}}(\cdot)\right)} \left(\frac{\partial\left(g'_{c_{k_j}}(\cdot)\right)}{\partial g(w_k)}\right) \Pr\left(c_j, w_k\right)\right].$$

Combining first order conditions and envelopes condition after some algebra amount to the result in (43).

13 Appendix D

13.1 Analytical Results for a three-point distribution

In this section I will specialize the optimality conditions derived above for a three point distribution. The goal is to fully characterize the solution for this particular case and explore its insights.[16]

Let me assume the wealth to be a random variable that takes up values in $w \in \Omega_w \equiv \{w_1, w_2, w_3\}$ with distribution $g(w_i) = \Pr(w = w_i)$ described by:

W	w_l	w_m	w_h
$g(w_i)$	g_1	g_2	$1 - g_1 - g_2$

The equation describing the evolution of the wealth is displayed by the budget constraint

$$w_{t+1} = R(w_t - c_t) + Y_t$$

where I denote by Y_t the exogenous stochastic income process earned by the household and by $R > 0$ the (constant) interest rate on savings, $(w_t - c_t)$. Like wealth, before processing information consumption, c_t, is a random variable. It takes up a discrete number of values in the event space $\Omega_c \equiv \{c_1, c_2, c_3\}$. The joint distribution of wealth and consumption, $\Pr_t(c_j, w_i)$, amounts to:

$\Pr_t(c_j, w_i)$

$C \backslash W$	w_1	w_2	w_3
c_1	x_1	x_2	x_3
c_2	0	x_4	x_5
c_3	0	0	x_6

where the zeros in the SW end of the matrix encodes the feasibility constraint $w_i(t) \geq c_j(t)$ $\forall i \in \Omega_w$, $j \in \Omega_c$ and $\forall t \geq 0$. The additional restrictions to the above matrix are the ones commanded by the marginal on wealth. That is:

$$x_1 = g_1$$
$$x_2 + x_4 = g_2$$
$$x_3 + x_5 + x_6 = 1 - g_1 - g_2$$

Without loss of generality, I place the marginal distribution of wealth in the main diagonal of $\Pr_t(c_j, w_i)$ and I impose the restrictions above together with the condition that the resulting matrix describes a proper distribution. The joint distribution of wealth and consumption amounts to:

$\Pr(c_j, w_i)$:

$C \backslash W$	w_1	w_2	w_3
c_1	g_1	p_1	p_2
c_2	0	$g_2 - p_1$	p_3
c_3	0	0	$1 - (g_1 + g_2) - (p_2 + p_3)$

(44)

[16] Three-point distribution is indeed a special case of the more general N points distribution since two of the states in the event space Ω_w are *absorbing* states. This, in turn, sets to zero several dimensions of the problem and allows for a close form solution of the optimal policies. Although the solution for this particular case does not have a straightforward generalization, it provides useful insights on the optimal choice for the joint probability distribution of wealth and consumption and its relation with the prior distribution of wealth -$g(w)$- and the utility of the consumer.

The resulting marginal distribution of consumption that endogenously depends on the choices of p_i's, $i = 1, 2, 3$, displays:

$$\Pr(C = c_j) = \begin{cases} c_1 & \text{w.p} \quad g_1 + p_1 + p_2 \\ c_2 & \text{w.p} \quad g_2 - p_1 + p_3 \\ c_3 & \text{w.p} \quad 1 - (g_1 + g_2) - (p_2 + p_3) \end{cases}.$$

Once the consumer chooses p_i's and observes the realized consumption c_t, he updates the marginal distribution of wealth. The latter, $g'\left(\cdot|_{c_j}\right)$, is obtained combining the joint distribution of wealth and consumption and the transition probability function. In formulae, the updated marginal on wealth amounts to:

$$g'\left(\cdot|_{c_j}\right) = \sum_i T(\cdot; w_i, c_j) \Pr(w_i|c_j). \tag{45}$$

The specification of $T(\cdot; w_i, c_j)$ adopted in the analytical derivation of the discrete probability distribution as well as in the numerical simulation can be explained as follows. The transition probability function is meant to approximate the expected value of next period wealth:

$$EW' = R(w_t - c_t) + \bar{Y}. \tag{46}$$

The approximation is necessary since (46) cannot hold exactly at the boundaries of the support of the wealth, Ω_w. In the above equation, R is the interest rates assumed to be a given number while $\bar{Y}$ is the mean of the stochastic income process, Y_t. Suppose we have a three point distribution. Assume *WLOG* that the values $w_i \in \Omega_w$ are equally spaced. For a given (w_i, c_j) pair, the distribution of next period wealth is concentrated on three w_i' values closest to $R(w_i - c_j) + \bar{Y}$, which will be denoted by $\omega_1, \omega_2, \omega_3$ with respective probabilities π_1, π_2, π_3. The mean of the distribution is $-\pi_1(\omega_2 - \omega_1) + \pi_3(\omega_3 - \omega_2) + \omega_2$. Let δ be the distance between the values of w_i. Then the mean becomes $\mu_\omega \equiv -\delta(\pi_3 - \pi_1) + \omega_2$. The variance of the distribution is then $\sigma_\omega^2 \equiv \delta^2(\pi_3 - \pi_1) - (\mu_\omega - \omega_2)^2$. Since π_2 is an exact function of π_1 and π_3, the equations for mean and variance of the process constitutes two equations in two unknowns. With the additional restriction that all the π_i's are positive and sum to one, it is not possible to guarantee the existence of a solution for $R(w_i - c_j) + \bar{Y}$ close to the boundaries of the support of the distribution of wealth. To make sure that there is *always* a solution for $\mu_\omega \in (\min(w) + .5\delta, \max(w) - .5\delta)$, and the solution is continuous at points where $\mu_\omega = \frac{(w_i + w_{i+1})}{2}$, one has to choose $\sigma_\omega^2 = .25\delta^2$.

Euler Equations. Making use of the marginal distribution of wealth described above and making use of (45) together with the specifications of $T(\cdot; w_i, c_j)$ and $\Pr(w_i, c_j)$, I can explicitly evaluate $g'\left(\cdot|_{c_j}\right)$ *point-wise*. To illustrate this point, using the numerical values of $T(\cdot; w_i, c_j)$ above, the derivatives *point-wise* are as follows.

In $c_j = c_1$,

$$g'\left(\cdot|_{c_1}\right) = \frac{1}{(g_1 + p_1 + p_2)}\left(T(\cdot; w_1, c_1) g_1 + T(\cdot; w_2, c_1) p_1 + T(\cdot; w_3, c_1) p_2\right)$$

In $c_j = c_2$

$$g'\left(\cdot|_{c_2}\right) = \frac{1}{(g_2 - p_1 + p_3)} \left(T\left(\cdot; w_2, c_2\right)(g_2 - p_1) + T\left(\cdot; w_3, c_2\right) p_3\right)$$

In $c_j = c_3$

$$g'\left(\cdot|_{c_3}\right) = T\left(\cdot; w_3, c_3\right)$$

Then, the first order conditions and envelope conditions amount to

∂p_1 :

$$\begin{aligned}
&\left[u\left(c_1\right) - u\left(c_2\right) + \beta\left(V'\left(g'_{c_2}\left(\cdot\right)\right) - V'\left(g'_{c_2}\left(\cdot\right)\right)\right)\right] \\
&= p_1 \left(\begin{array}{c} \theta\left(\left[u'\left(c_1 - \theta\kappa\right) - u'\left(c_2 - \theta\kappa\right)\right]\right) \ln\left(\frac{p_1}{(g_1+p_1+p_2)}\right) + \\ + \frac{\partial V'\left(g'_{c_1}(\cdot)\right)}{\partial g'_{c_1}(\cdot)} \frac{\partial g'_{c_1}(\cdot)}{\partial p_1} - \frac{\partial V'\left(g'_{c_2}(\cdot)\right)}{\partial g'_{c_2}(\cdot)} \frac{\partial g'_{c_2}(\cdot)}{\partial p_1} \end{array}\right)
\end{aligned}$$

Note that $\frac{\partial g'_{c_j}(\cdot)}{\partial p_j} = 0$ for $j \in \{1, 2, 3\}$.[17] This result is not driven by the specification chosen for the transition function $T\left(\cdot; w_i, c_j\right)$ but it is a feature of the three point distribution. Indeed, since two of the three values of wealth are at the boundaries of Ω_w, the absorbing states w_1 and w_3 place tight restrictions on the continuation value $V'\left(g'_{c_j}\left(\cdot\right)\right)$ through the transition function and, as a result, the update for the marginal $g'_{c_j}\left(\cdot\right)$ according to (45). That is, the marginal probability on wealth $g'_{c_j}\left(\cdot\right)$ in this case tends to its ergodic value $\bar{g}_{c_j}\left(\cdot\right)$. It follows that $V'\left(\bar{g}_{c_j}\left(\cdot\right)\right) \overset{a.s.}{\longrightarrow} \bar{V}^*\left(\bar{g}_{c_j}\left(\cdot\right)\right)$ which is a constant since the functional argument is. This is what makes the 3-point distribution tractable.

For the general case, the first order condition with respect to the first control amounts to:

∂p_1 :

$$\begin{aligned}
&\left[u\left(c_1\left(\kappa\right)\right) - u\left(c_2\left(\kappa\right)\right) + \beta\left(\bar{V}\left(\bar{g}_{c_1}\left(\cdot\right)\right) - \bar{V}\left(\bar{g}_{c_2}\left(\cdot\right)\right)\right)\right] \\
&= p_1 \left(\theta\left(\left[u'\left(c_1\left(\kappa\right)\right) - u'\left(c_2\left(\kappa\right)\right)\right]\right) \ln\left(\frac{p_1}{(g_1 + p_1 + p_2)}\right)\right)
\end{aligned} \tag{47}$$

[17] To see this, plug (45) in $\frac{\partial g'_{c_j}(\cdot)}{\partial p_j}$ for $j \in \{1, 2\}$ and evaluating pointwise the derivatives delivers

$\partial g'\left(\cdot|_{c_1}\right)$:

$$\frac{1}{(g_1 + p_1 + p_2)^2} \left[\begin{array}{c} 0.81p_2 - 0.15g_1 \\ -\left(0.56p_2 - 0.15g_1\right) \\ -0.25p_2 \end{array}\right] = 0$$

$\partial g'\left(\cdot|_{c_2}\right)$:

$$\frac{p_3}{(g_2 - p_1 + p_3)^2} \left[\begin{array}{c} -0.15 \\ 0.15 \\ 0 \end{array}\right] = 0$$

Similarly, for the second control

∂p_2 :

$$\begin{aligned}&\left[u\left(c_1\left(\kappa\right)\right)-u\left(c_3\left(\kappa\right)\right)+\beta\left(\bar{V}\left(\bar{g}_{c_1}\left(\cdot\right)\right)-\bar{V}\left(\bar{g}_{c_3}\left(\cdot\right)\right)\right)\right]\\&=p_2\left(\theta\left(\left[u'\left(c_1\left(\kappa\right)\right)-u'\left(c_3\left(\kappa\right)\right)\right]\right)\ln\left(\frac{p_2}{\left(g_1+p_1+p_2\right)}\right)\right)\end{aligned}\tag{48}$$

And finally:

∂p_3 :

$$\begin{aligned}&\left[u\left(c_2\left(\kappa\right)\right)-u\left(c_3\left(\kappa\right)\right)+\beta\left(\bar{V}\left(\bar{g}_{c_2}\left(\cdot\right)\right)-\bar{V}\left(\bar{g}_{c_3}\left(\cdot\right)\right)\right)\right]\\&=p_3\left(\theta\left(u'\left(c_2\left(\kappa\right)\right)-u'\left(c_3\left(\kappa\right)\right)\right)\ln\left(\frac{p_3}{\left(g_2-p_1+p_3\right)}\right)\right)\end{aligned}\tag{49}$$

Using the result that the value function converges to V^* when the utility function belongs to the family of absolute risk aversion (CARA), I assume the utility takes up the specification:

$$u\left(c_j\left(\kappa\right)\right)=\begin{cases}-\frac{e^{-\gamma\left(c_j\left(\kappa\right)\right)}}{\gamma} & \text{for } \gamma>0\\ \log\left(c_j\left(\kappa\right)\right) & \text{for } \lim_{\gamma\to 0}\left(-\frac{e^{-\gamma\left(c_j\left(\kappa\right)\right)}}{\gamma}\right)\end{cases}$$

where γ is the coefficient of absolute risk aversion and $j\in\Omega_c\equiv\{c_1,c_2,c_3\}$. Moreover, by proposition 1, the value function is $PCWL$, that is:

$$\bar{V}\left(\bar{g}_{c_j}\left(\cdot\right)\right)=\arg\max_{\left\{\alpha_j'\right\}_j}\left\langle\alpha_j',\bar{g}_{c_j}'\left(\cdot\right)\right\rangle$$

where $\left\{\alpha_j'\right\}_j$ are a set of vectors each of them generated for a particular observation of previous values of consumption c_j and $\langle .,.\rangle$ denotes the inner product $\left\langle\alpha_j',\bar{g}_{c_j}'\left(\cdot\right)\right\rangle\equiv\sum_{w'\in\Omega_w}\alpha_j'\left(w'\right)T\left(\cdot:w,c_j\right)\cdot p\left(c_j|w\right)$. To get a close form solution, I need to represent the probability distribution of the prior. One of the possibilities is to use a particle based representation. The latter is performed by using N random samples, or particles, at points w_i and with weights ϖ_i. The prior is then

$$g_t\left(w\right)=\sum_{i=1}^{N}\varpi_i\tilde{\delta}\left(w-w_i\right)$$

where $\tilde{\delta}\left(w-w_i\right)=\text{Dirac}\left(w-w_i\right)$ is the *Dirac* delta function with center in zero. A particle-based representation can approximate arbitrary probability distributions (with an infinite number of particles in the extreme case), it can accommodate nonlinear transition models without the need of linearizing the model, and it allows several quantities of interest to be computed efficiently. In particular, the expected value in the belief update equation becomes:

$$\bar{g}'\left(\cdot|_{c_j}\right) = \Pr\left(c_j|\cdot\right)\sum_{i=1}^{N}\varpi_i T\left(\cdot; w_i, c_j\right)$$

The central issue in the particle filter approach is how to obtain a set of particles to approximate $\bar{g}'\left(\cdot|_{c_j}\right)$ from the set of particles approximating $g\left(w\right)$. The usual Sampling Importance Re-sampling (SIR) approach (Dellaert et al., 1999; Isard and Blake, 1998) samples particles using the motion model $T\left(\cdot; w_i, c_j\right)$, then it assigns a new weights in order to make all particles weights equal. The trouble with the SIR approach is that it requires many particles to converge when the likelihood $\Pr\left(c_j|\cdot\right)$ is too peaked or when there is a small overlap between prior and posterior likelihood. he main problem with SIR is that it requires many particles to converge when the likelihood is too peaked or when there is only a small overlap between the prior and the likelihood. In the auxiliary particle filter, the sampling problem is address by inserting the likelihood inside the mixture

$$\bar{g}'\left(\cdot|_{c_j}\right) \propto \sum_{i=1}^{N}\varpi_i \Pr\left(c_j|\cdot\right) T\left(\cdot; w_i, c_j\right).$$

The state $(\cdot)$ used to define the likelihood $\Pr\left(c_j|\cdot\right)$ is not observed when the particles are resampled and this calls for the following approximation

$$\bar{g}'\left(\cdot|_{c_j}\right) \propto \sum_{i=1}^{N}\varpi_i \Pr\left(c_j|\mu_\omega^i\right) T\left(\cdot; w_i, c_j\right)$$

with μ_ω^i any likely value associated with the i^{th} component of the transition density$T\left(\cdot; w_i, c_j\right)$, e.g., its mean. In this case, we have that $\mu_\omega^i = w_i + \Delta\left(c_j\right)$. Then, $\bar{g}'\left(\cdot|_{c_j}\right)$ can be regarded as a mixture of N transition components $T\left(\cdot; w_i, c_j\right)$ with weights $\varpi_i \Pr\left(c_j|\mu_\omega^i\right)$. Therefore, sampling a new particle w_j' to approximate $\bar{g}'\left(\cdot|_{c_j}\right)$ can be carried out by selecting one of the N components, say i_m, with probability $\varpi_i \cdot \Pr\left(c_j|\mu_\omega^i\right)$ and then sampling w_i' from the corresponding component $T\left(\cdot; w_{i_m}, c_j\right)$. Sampling is performed in the intersection of the prior and the likelihood and, consequently, particles with larger prior and larger likelihood (even if this likelihood is small in absolute value) are more likely to be used. After the set of states for the new particles is obtained using the above procedure, it is necessary to define the weights. This is done using

$$\varpi_m' \propto \frac{\Pr\left(c_j|w_m'\right)}{\Pr\left(c_j|\mu_\omega^{i_m}\right)}.$$

Using the sample-based belief representation the averaging operator $\langle .,. \rangle$ can be computed in close form as:

$$
\begin{aligned}
\langle \alpha, \bar{g}' \rangle &= \sum_{w \in \Omega_w} \left[\sum_k \varpi_k \tau \left(w | w_k, \Sigma_k \right) \right] \left[\sum_l \varpi'_l \tilde{\delta} \left(w - w_l \right) \right] \\
&= \sum_k \varpi_k \sum_{w \in \Omega_w} \left(\tau \left(w | w_k, \Sigma_k \right) \left[\sum_l \varpi_l \tilde{\delta} \left(w - w_l \right) \right] \right) \\
&= \sum_k \varpi_k \sum_l \varpi_l \tau \left(w_l | w_k, \Sigma_k \right) \\
&= \sum_{k,l} \varpi_k \varpi_l \tau \left(w_l | w_k, \Sigma_k \right).
\end{aligned}
$$

where $\tau(.)$ is the distribution of the r.v. W' that use the specification of the transition function above, i.e., mean $\mu_\omega \equiv -\delta(\pi_3 - \pi_1) + \omega_2$ and variance $\sigma^2_\omega \equiv \delta^2 (\pi_3 - \pi_1) - (\mu_\omega - \omega_2)^2$ with δ the (constant) distance between the values of w_i.

Representing priors in this fashion allows an explicit evaluation of the differences in the value functions in the first order conditions, since $V'\left(\bar{g}'_{c_j}(\cdot)\right) = \arg\max_{\{\alpha'_j\}_j} \left\langle \alpha'_j, \bar{g}'_{c_j}(\cdot) \right\rangle = \sum_{k,l} \tilde{\varpi}'_k \tilde{\varpi}'_l \tau \left(w_l | w_k, \Sigma_k \right)$, where $\tilde{\varpi}'_k \equiv \left(\frac{\Pr\left(c_j | w'_k\right)}{\Pr\left(c_j | \mu^k_\omega\right)} \right)$, $\tilde{\varpi}'_l \equiv \left(\frac{\Pr\left(c_j | w'_l\right)}{\Pr\left(c_j | \mu^l_\omega\right)} \right)$. Since the result of the arg max is just one of the member of the set $\left\{\alpha'_j\right\}_j$ and all the elements involved in the definition of α'_j function in $\Gamma_{(p)}$ are a finite set of linear function parametrized in the action set, so is the final result.

Let a prime" ′ " denote the variables led one period ahead, algebraic manipulation delivers the following optimal control functions:

$$
p^*_1(\vec{g}, \theta) = \frac{g_1 \left(\psi_1 - \theta \beta \nu_1 \right)}{\theta g_1 \left(\text{LambertW}(\chi_1) x_{12} - \text{LambertW}(\chi_{11}) x_{11} \right) + 2 g_1 \left(\psi_1 - \theta \beta v_1 \right)}; \tag{50}
$$

$$
p^*_2(\vec{g}, \theta) = \frac{g_1 \left(\psi_2 - \theta \beta \nu_2 \right)}{\theta g_1 \left(\text{LambertW}(\chi_2) x_{21} - \text{LambertW}(\chi_2) x_{22} \right) + 2 \psi_2 g_1 \left(\psi_2 - \theta \beta \nu_3 \right)}; \tag{51}
$$

$$
p^*_3(\vec{g}) = \frac{\psi_3 - \theta \beta v_3}{\theta x_3 \, \text{LambertW}(\chi_3)} \tag{52}
$$

where

- $\psi_1 \equiv \left(\frac{e^{-\gamma(c_2 - \theta\kappa)} e^{-\gamma(c_1 - \theta\kappa)}}{\gamma} \right)$; $\psi_2 \equiv \left(\frac{e^{-\gamma(c_3 - \theta\kappa)} e^{-\gamma(c_1 - \theta\kappa)}}{\gamma} \right)$; $\psi_3 \equiv \left(\frac{e^{-\gamma(c_3 - \theta\kappa)} e^{-\gamma(c_2 - \theta\kappa)}}{\gamma} \right)$;
- $\nu_1 \equiv g_2 \left(\psi'_3 \right) + \left(g_2 - g_1 \right) \left(\psi'_2 - \psi'_3 \right)$;

- $\nu_2 \equiv g_2\left(\psi_1'\right) + \left(g_2 - g_1\right)\left(\psi_1' - \psi_2'\right);$
- $\nu_3 \equiv \left(1 - g_2 - g_1\right)\left(\psi_2'\right) + \left(g_2 - g_1\right)\left(\psi_3' - \psi_1'\right)$
- $\chi_1 \equiv \frac{(\psi_1 - \theta\beta v_1)\psi_1}{\theta g_1\left(e^{\gamma(c_2 - c_1)}\right)};$, $x_{11} \equiv e^{-\gamma(c_1 - \theta\kappa)}$, $x_{12} \equiv e^{-\gamma(c_2 - \theta\kappa)};$
- $\chi_{21} \equiv \frac{(\psi_2 - \theta\beta v_2)\psi_2}{\theta g_1\left(e^{\gamma(c_3 - c_1)}\right)}$, $x_{21} \equiv e^{-\gamma(c_1 - \theta\kappa)}$, $x_{22} \equiv e^{-\gamma(c_3 - \theta\kappa)}$ and
- $\chi_3 \equiv \frac{\psi_3 - \theta\beta v_3}{\theta g_2\left(e^{\gamma(c_3 - c_2)}\right)}$, $x_3 \equiv e^{\gamma(c_3 - c_2)}$.

and LambertW (.) is the LambertW function that satisfies $\mathrm{LambertW}(x)\, e^{\mathrm{LambertW}(x)} = x$[18]. The argument of the LambertW is always positive for the first order conditions derived, implying that for each of the optimal policies the function returns a real solution amongst other complex roots, which is unique and positive. Since $\frac{\partial\, \mathrm{LambertW}(x)}{\partial x} = \frac{\mathrm{LambertW}(x)}{x(1 + \mathrm{LambertW}(x))}$ it is possible to calculate the derivatives of the above expression with respect to $\{\theta, g_1, g_2\}$. However, the sign of the derivatives with respect to those variables is indeterminate. The rational behind this result is quite simple. Consider the joint probability distribution $\Pr(c_i, w_j)$. The overall effect of an increase in this probability results from the interplay of several factors. In general, if θ is low -or, equivalently, the capacity of the channel, $\bar{\kappa}$, in (21) is high-, a risk averse consumer will try to reduce the off diagonal term of the joint as much as possible. That is, he would set $p_1 = \Pr(c_1, w_2)$, $p_2 = \Pr(c_1, w_3)$ and $p_3 = \Pr(c_3, w_2)$ as low as its capacity allows him to sharpen his knowledge of the state. On the opposite extreme, for very high value of the cost associated to information processing, θ, p_1 and p_2 will be higher, the higher the prior $g_1 = g(w_1)$ with respect to $g_2 = g(w_2)$ and $g_3 = g(w_3)$. This is due to the fact that when the capacity of the channel is low -or, equivalently, the effort of processing information is high-, the first order conditions indicate that it is optimal for the consumer to shift probabilities towards the higher belief state. The intuition is that when it is costly to process information, the household cannot reduce the uncertainty about his wealth. If the individual is risk adverse as implied by the CRRA utility function, in each period, he would rather specialize in the consumption associated to the higher prior

[18] Formally, the LambertW function is the inverse of the function $f : \mathbb{C} \to \mathbb{C}$ given by $f(x) \equiv xe^x$. Hence LambertW (x) is the complex function that satisfies

$$\mathrm{LambertW}(x)\, e^{\mathrm{LambertW}(x)} = x$$

for all $x \in \mathbb{C}$.. In practice the definition of LambertW requires a branch cut, which is usually taken along the negative real axis. LambertW (x) function is sometimes also called product log function.

This function allows to solve the functional equation

$$g(x)^{g(x)} = x$$

given that

$$g(x) = e^{\mathrm{LambertW}(\ln(x))}.$$

See Corless, Gonnet, Hare, Je rey and Knuth (1996).

than attempt to consume a different quantity and running out of wealth in the following periods. This intuition leads to an optimal policy of the consumer that commands high probability to one particular consumption profile and set the remaining probabilities as low as possible. To illustrate this, consider a consumer who has a high value of θ and a prior on w_1 higher than the other priors. If he cannot sharpen his knowledge of the wealth due to prohibitively information processing effort, he will optimize its dynamic problem by placing very high probability on $\Pr(c_1) = g_1 + p_1 + p_2$, i.e., increase p_1 and p_2 and decrease p_3. Likewise, if g_2 is higher than the other priors and θ is high -κ is low-, optimality commands to decrease both p_1 and p_2 and increase p_3.

14 Appendix E.

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.

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

- Step 4: For each simplex point, compute $g'_{c_j}(\cdot) = \sum_{w_t\in\Omega_w} T(\cdot; w_t, c_t)\, p^*(w_t|c_t)$. Interpolate $V_0(g(w_t))$ with $g'_{c_j}(\cdot)$.
- 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).

$\rightarrow$ 9-45 iterations each taking 5min-40min

- 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.

References

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

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

[3] Amato, J.S. Morris and H. S. Shin (2002), *Communication and Monetary Policy*, Oxford Review of Economic Policy 18, pp. 495-503.

[4] Ameriks, J., Andrew C. and J. Leahy (2003b) *The Absent-Minded Consumer*, unpublished, New York University.

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

[6] Aoki, K. (2003), *On the Optimal Monetary Policy Response to Noisy Indicators*, Journal of Monetary Economics 50, pp. 501-523.

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

[16] Deaton, A. (1987), *Life-Cycle Models of Consumption: is the evidence consistent with the Theory?* in Advances in Econometrics, Fifth World Congress, vol.2, ed. Truman Bewley, Cambridge, Cambridge University Press.

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

[18] Dynan, Karen (2000), *Habit Formation in Consumer Preferences: Evidence from Panel Data*, American Economic Review, 90, pp. 391-406.

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

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

[21] Goodfriend, Marvin (1992), *Information-Aggregation Bias*, American Economic Review, 82, pp. 508-519.

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

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

[24] Hellwig, C. (2004), *Heterogeneous Information and the Benefits of Transparency*, Discussion paper, UCLA.

[25] Hellwig, C. and L. Veldkamp, (2006) *Knowing What Others Know: Coordination Motives in Information Acquisition*, NYU, mimeo.

[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] Lorenzoni, G. (2006), *Demand Shocks and Monetary Policy*, MIT, mimeo.

[28] Lucas, R. E., J. (1973), *Some International Evidence on Output-Inflation Tradeoffs*, American Economic Review, 63(3), 326–334.

[29] Luo, Y. (2007), *Consumption Dynamics, Asset Pricing, and Welfare Effects under Information Processing Constraints*, forthcoming in Review of Dynamics.

[30] Lusardi, A.(1999), *Information, Expectations, and Savings*, in Behavioral Dimensions of Retirement Economics, ed. Henry Aaron, Brookings Institution Press/Russell Sage Foundation, New York, pp. 81-115.

[31] Lusardi, A. (2003), *Planning and Savings for Retirement*, Dartmouth College, unpublished.

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

[33] Macrowiack, B., and M. Wiederholt (2007), *Optimal Sticky Prices under Rational Inattention*, Discussion paper, Northwestern/ ECB 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] Mankiw, N. G., Reis, R. and J. Wolfers (2004) *Disagreement in Inflation Expectations*, NBER Macroeconomics Annual 2003, vol. 18, pp. 209-147.

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

[37] Morris, S., and H. S. Shin (2002), *The Social Value of Public Information*, American Economic Review, 92, 1521–1534.

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

[39] Mullainathan, S. (2002), *A Memory-Based Model of Bounded Rationality*, Quarterly Journal of Economics, 117 (3), pp. 735-774.

[40] Orphanides, A. (2003), *Monetary Policy Evaluation with Noisy Information*, Journal of Monetary Economics 50 (3), pp. 605-631.

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

[42] Peng, L. (2005), *Learning with Information Capacity Constraints*, Journal of Financial and Quantitative Analysis, 40(2), 307–329.

[43] Peng, L., and W. Xiong (2005), *Investor Attention, Overconfidence and Category Learning*, Discussion paper, Princeton University.

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

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

[46] Schmitt-Grohè, S. and M. Uribe(2004). *Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function*. Journal of Economic Dynamics and Control 28, pp.755-75

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

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

[49] Sims, C. A.(2005), *Rational Inattention: a Research Agenda*, Princeton University mimeo.

[50] Van Nieuwerburgh, S., and L. Veldkamp (2004a), *Information Acquisition and Portfolio Under-Diversication*, Discussion paper, Stern School of Business, NYU.

[51] Stokey, N. L. and R. E. Lucas Jr., with Edward C. Prescott (1989) *Recursive Methods in Economic Dynamics*, Cambridge, Harvard University Press.

[52] Van Nieuwerburgh, S., and L. Veldkamp (2004b), *Information Immobility and the Home Bias Puzzle*, Discussion paper, Stern School of Business, NYU.

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

[54] Wilson, Andrea (2003) "Bounded Memory and Biases in Information Processing," unpublished, Princeton University.

[55] 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.

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

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

15 Tables and Figure

Baseline Model	
Variables\Values	
β	0.963
R	1.03
$E(C)$	1.3333
$std(C)$	0.7201
$E(W)$	4
$std(W)$	2.1602
w_{grid}	$\Omega_w \equiv \{1, 2.., 7\}$
c_{grid}	$\Omega_c \equiv \frac{1}{3}\Omega_w$
$g(w)_{grid}$	$\Delta_{g(w)} \equiv \{3003x7\}$

Table 1

Optimal Value and Expected Wealth

15.1 Full Information Case: $\theta = 0$

	CRRA $\gamma = 7$	CRRA $\gamma = 5$	Log Utility	CRRA $\gamma = .5$	CRRA $\gamma = .3$
$E(C)$	1.0000	1.0476	1.1429	1.2381	1.2857
$std(C)$	0.1925	0.2300	0.2623	0.3171	0.3563
$Ex.Skewness$	0.0008	-0.1342	-0.85991	-0.6660	-0.5954
$Ex.Kurtosis$	3.5000	2.3900	2.3639	2.7687	2.3594

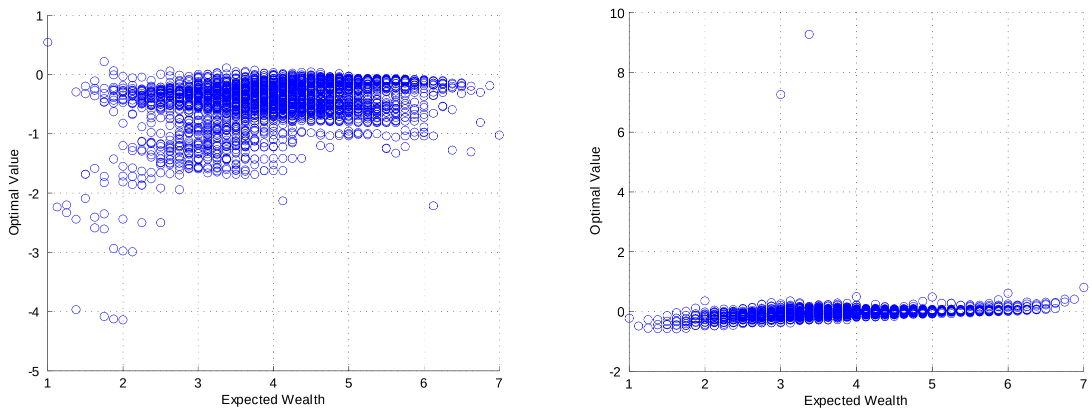
Table 1a

15.2 Model 1: $\theta = 0.2$

	CRRA $\gamma = 5$	CRRA $\gamma = 3$	Log Utility	CRRA $\gamma = .5$
$E(C)$	1.0130	1.0143	1.0156	1.0161
$std(C)$	0.0505	0.0503	0.0509	0.0514
$Ex.Skewness$	-0.0282	-0.0743	-0.0855	-0.0743
$Ex.Kurtosis$	2.9227	3.0506	3.0126	2.9689
κ	0.1630	0.1626	0.1626	0.1625

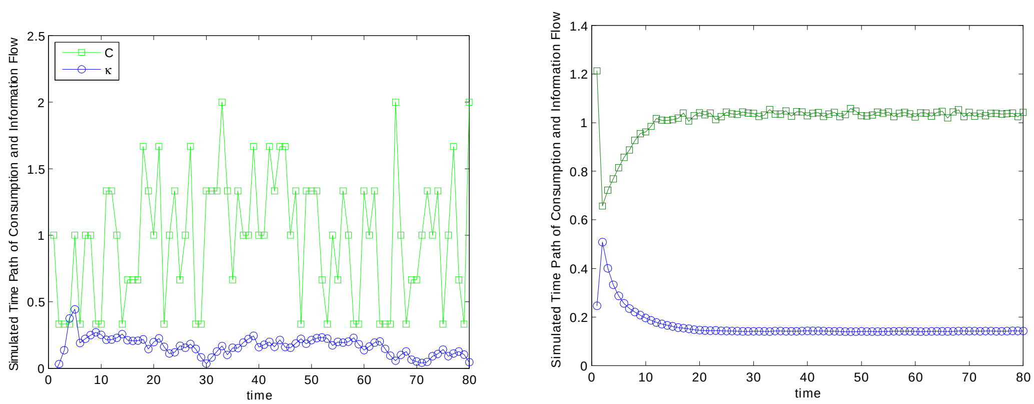
Table 1b

Optimal Value and Expected Wealth



Figures 2a-2b: Dispersion of $V^*\left(g\left(w_i\right)\right)$ and $E^*\left(g\left(w_i\right)\right)$

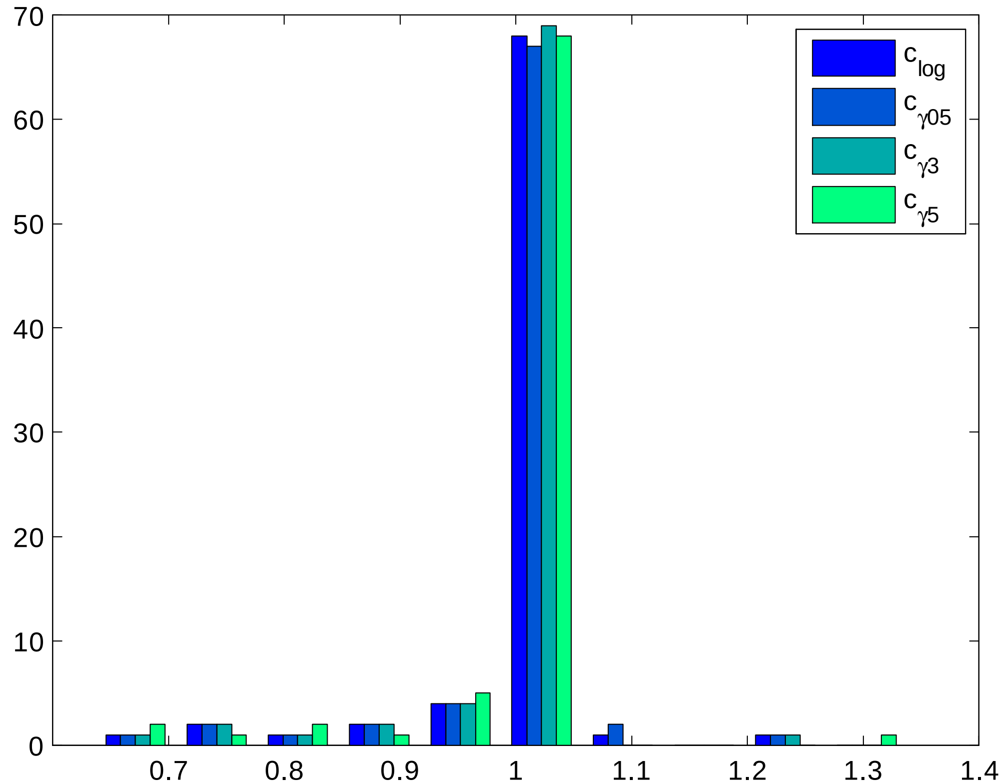
Time Series Results



CRRA γ=5 *Log Utility*

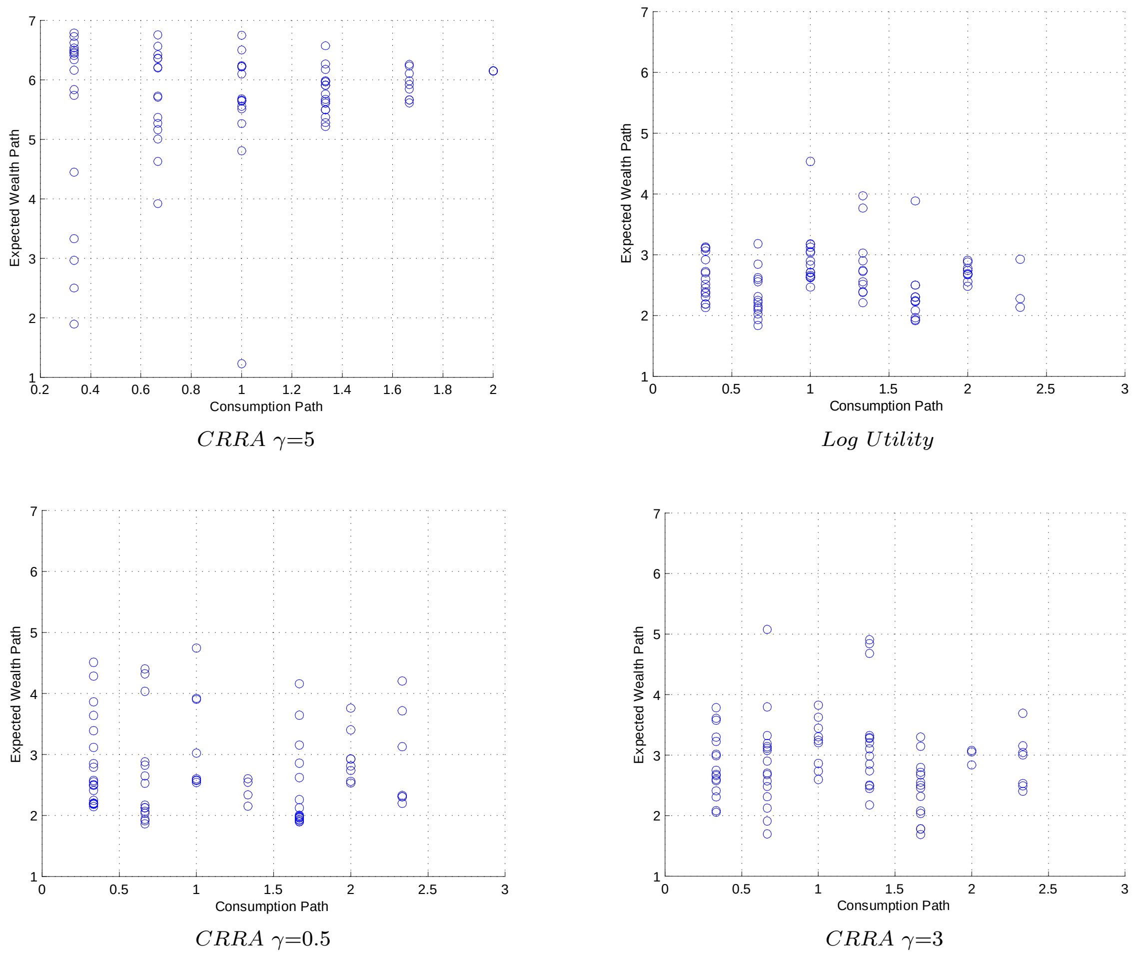
Figures 3a-3b: Consumption (C) and Information Flow (κ) Path.

15.2.1 Histogram of Consumption



Figures 4a: Consumption Path for several Utility Specifications.

15.2.2 Dispersion of Consumption and Expected Wealth Path



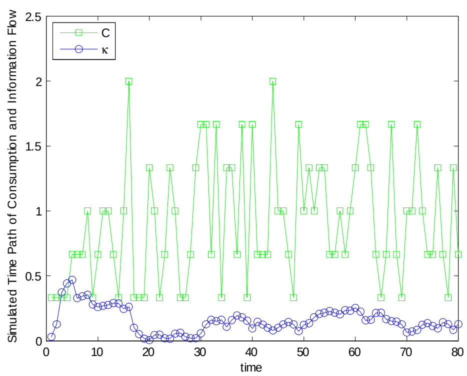
Figures 5a-5d: Scatter of Consumption Path and Expected Wealth Path.

15.3 Model 2 $\theta = 1$

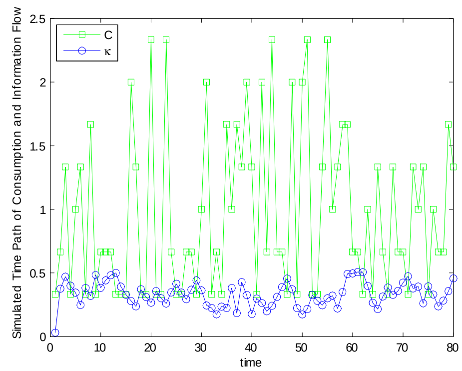
	CRRA $\gamma = 7$	CRRA $\gamma = 5$	CRRA $\gamma = 3$	Log Utility	CRRA $\gamma = .3$
$E(C)$	1.0213	1.0156	1.0148	1.0143	1.0138
$std(C)$	0.0515	0.0503	0.0513	0.0513	0.0513
$Ex.Skewness$	-0.1086	-0.0855	0.0183	0.0254	0.0109
$Ex.Kurtosis$	3.1786	3.0126	2.8750	2.9549	2.9337
κ	0.1729	0.1625	0.1622	0.1619	0.1612

Table 1c

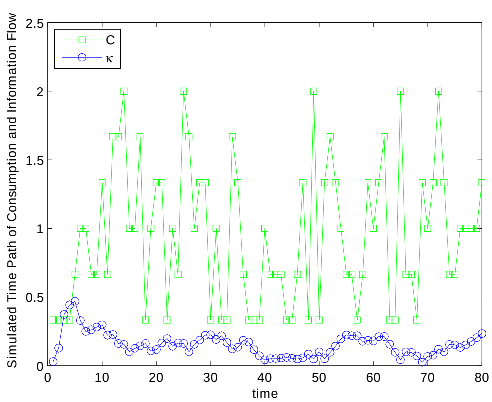
Time Series Results



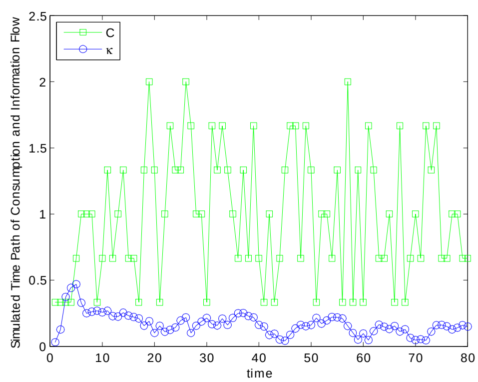
$CRRA\ \gamma$=5



$Log\ Utility$



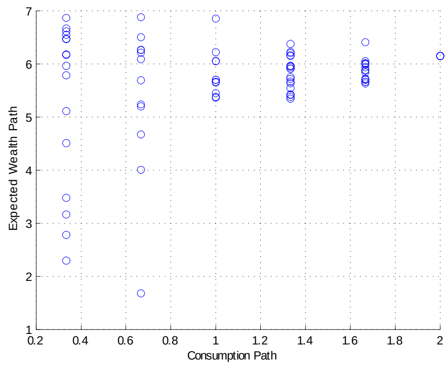
$CRRA\ \gamma$=3



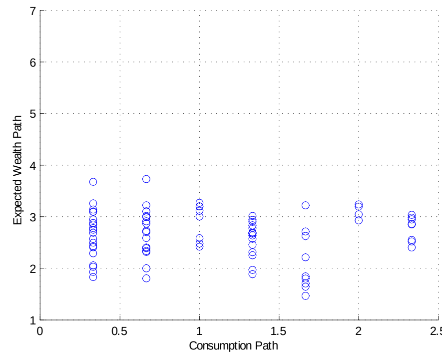
$CRRA\ \gamma$=0.3

Figures 3c-3d: Consumption (C) and Information Flow (κ) Path.

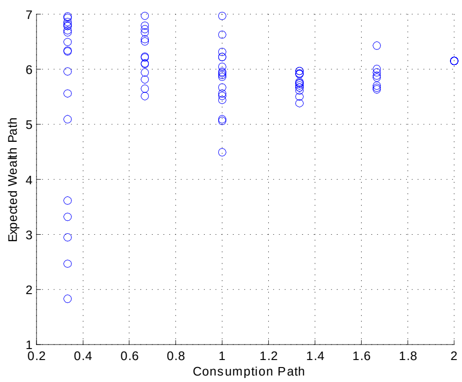
15.3.1 Dispersion of Consumption and Expected Wealth Paths



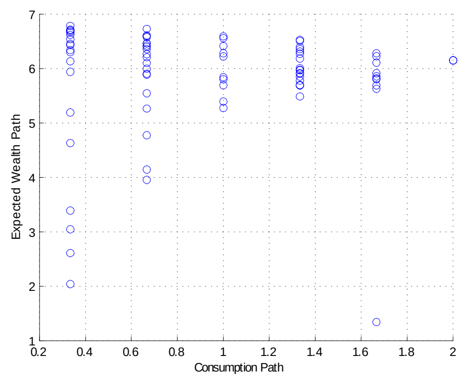
CRRA γ=0.3



Log Utility



CRRA γ=3



CRRA γ=7

Figures 5e-5h: Scatter of Consumption Path and Expected Wealth Path.

Histograms Log Utility: s_1

Histogram for s_1: several temporary shocks.

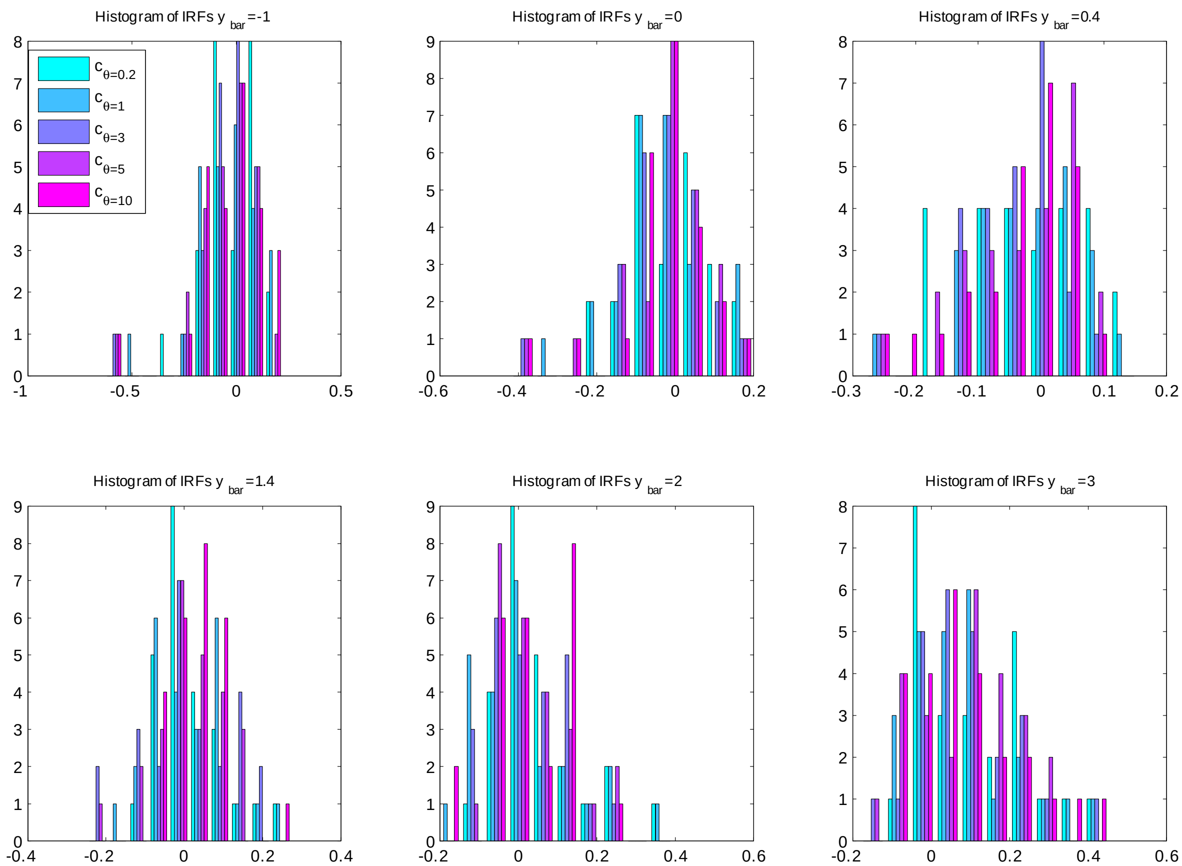


Figure 10a: Histogram of IRFs of consumption for different information flows

Histograms Log Utility: s_4

Histogram for s_4: several temporary shocks.

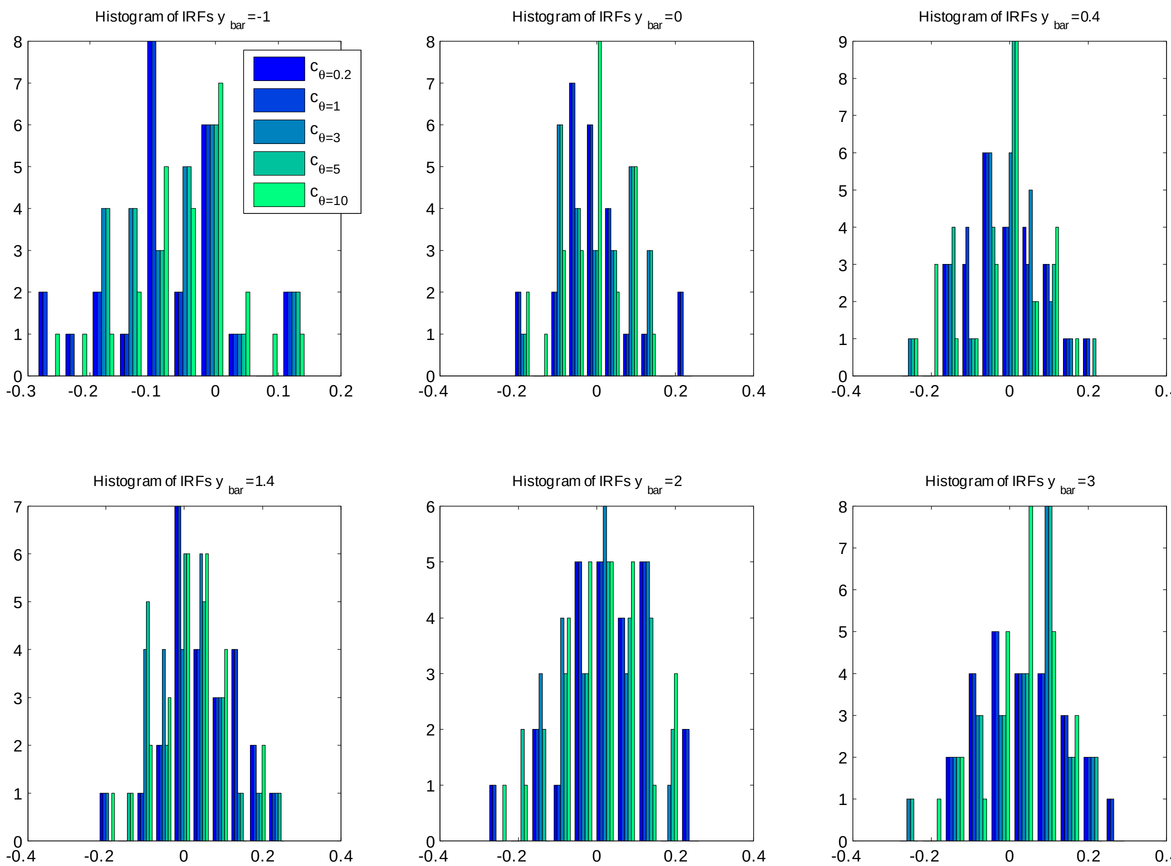


Figure 10b: Histogram of IRFs of consumption for different information flows

Impulse Response Functions

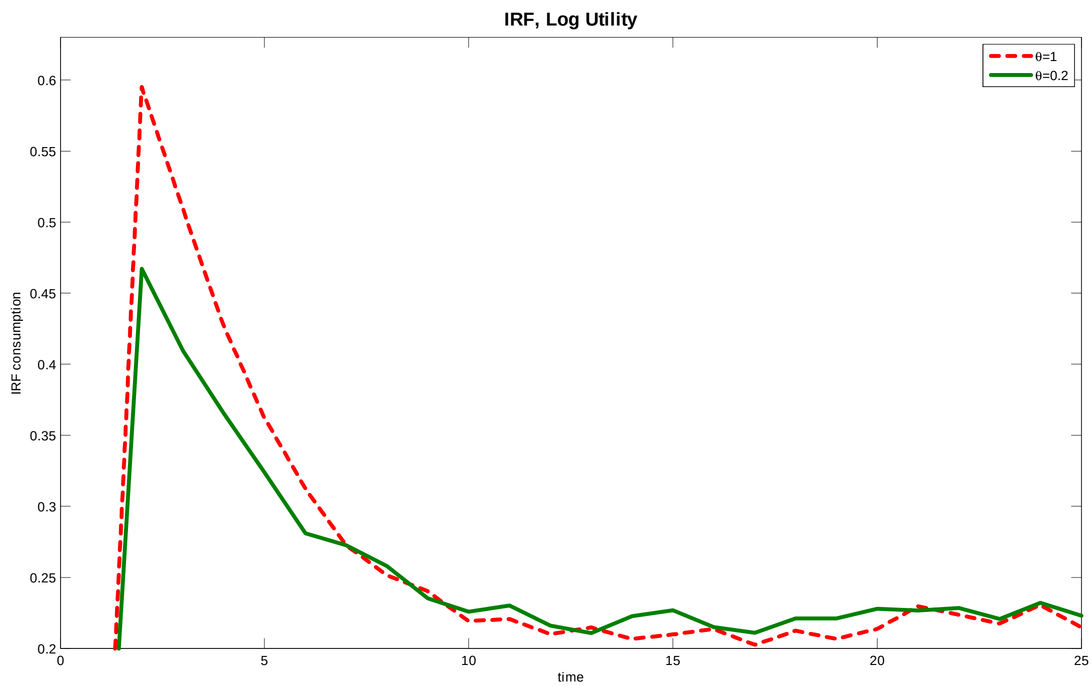


Figure 11a

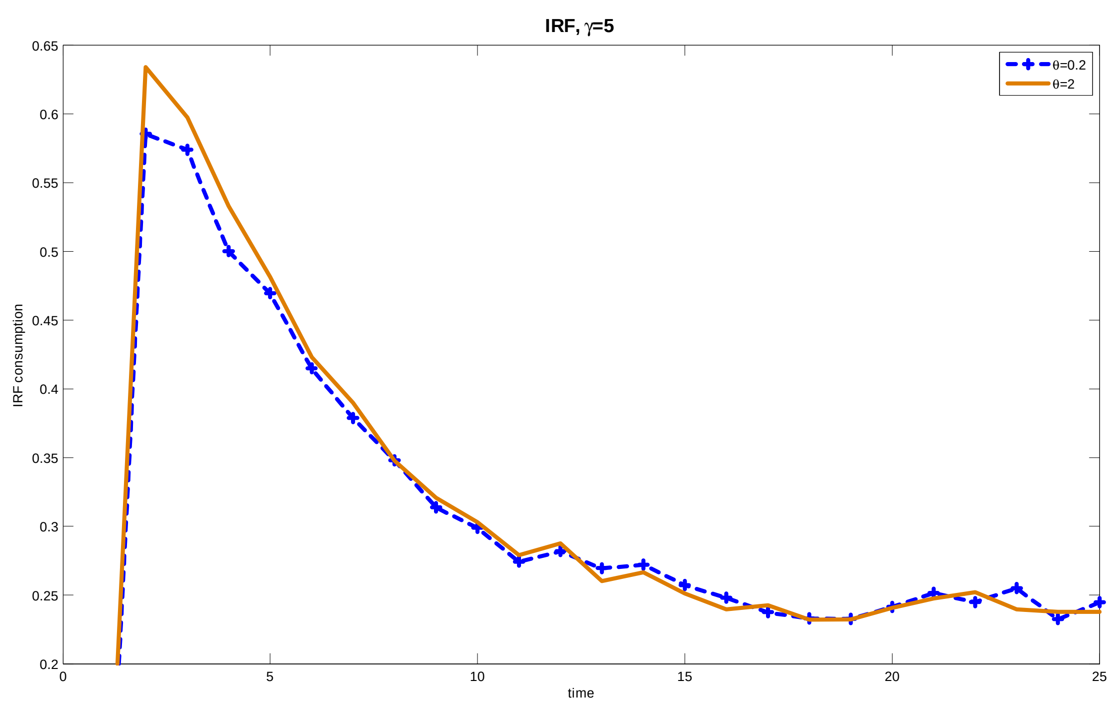


Figure 4a