

**Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.**

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

Antonella Tutino

2008-62

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

The Rigidity of Choice.

Lifecycle savings with information-processing limits

ANTONELLA TUTINO*

First version: November 2007 ; This version: October 2008

Abstract

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

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

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

Every day people face an overwhelming amount of data. Every day, though, people use these for their decisions. In selecting useful information, people face a trade off between reacting quickly and precisely to news about their financial possibilities and not spending time crunching numbers to figure out their exact net worth. To match these facts, macroeconomists have adopted a number of modelling strategies able to inject inertia within the rational expectation framework. These devices, such as the costly acquisition and diffusion of information, largely rely on ad-hoc technology to generate smooth and delayed responses of consumption to a shock to income consistent with observed data. Contrary to this approach, this paper proposes a way to relate inertial behavior in consumption and savings based on people's preferences.

To this end, the paper offers a micro-founded explanation on the nature of inertia in consumption and savings. Following *Rational Inattention* (Sims, 2003, 2006), I model the limits of people to process information at an infinite rate by using Shannon channels.

Under this information processing constraint, individuals choose a signal that conveys information about their financial possibilities. The signal can provide any kind of information as long as its overall content is within the channel's capacity. Consumers base their expectations of the economic conditions on the signal and decide how much to consume. Thus, in my framework, the delayed and smoothed responses of savings to changes in wealth are the result of a slow information flow due to processing constraints. Combining the standard utility maximization framework subject to a budget constraint with information processing limits leads to a departure from rational expectations. My paper shows how to model this formally in an intertemporal setting. In particular, I assume that people do not know the exact value of their wealth but have an idea of their net worth. A way of thinking about this hypothesis is that people do not know exactly of what the dollar value of their paycheck (nominal) corresponds to in terms of cups of coffee (real), assuming that this is what they care about. People process information to sharpen their knowledge of how much consumption their wealth can purchase. I model initial uncertainty as a probability distribution over the possible realizations of wealth. In such a framework, it is possible to study how choices of information play out with people's preferences when they decide on consumption throughout their life time.

The challenge of this model and, more generally, of models of rational inattention is dealing with the infinite dimensional state space implied by having a prior as state. For this reason, 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 Constant Relative Risk Aversion (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 models. 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. Second, it is easy to compare the results to a signal extraction problem. When looking at the behavior of rational inattentive consumers, it is impossible to separate an exogenously given Gaussian noise in the signal extraction model from an endogenous noise that is optimally chosen to be Gaussian.

The tractability of rational inattention LQG models 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 individuals' decisions. In this sense, rational inattention LQG models are subject to the same limits as methods that use linear approximation of optimality conditions to study stochastic dynamic models.¹ With little uncertainty about the economic environment, linear approximations of the optimality conditions may provide a fairly adequate description of the exact solution of the system. This fact suggests that the 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 for people' expectations, it is important to let consumers select their information from a wider set of distributions that includes but it is not limited to the Gaussian family.

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

Several predictions emerge from the model. Evaluating the unconditional moments of the time series of consumption for a given degree of risk aversion, the first result of the paper is that higher information costs are associated with more persistence and higher volatility. The seemingly paradoxical results of having sluggish and volatile consumption

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

at the same time can be reconciled if one considers that information-processing constraints prevent the consumers to respond promptly to fluctuations in wealth. To make a concrete example, suppose a person starts off with low wealth and initially chooses to consume a little. If he is risk averse, he may decide not to modify his consumption profile until he acquires more information about his wealth. As he processes information through time, he gets more and more data about his high value of wealth and changes his consumption when he is sure that he has saved enough to afford a higher consumption expenditure. The more risk averse the consumer is, the longer he waits. The longer the wait, the more wealth grows because of the accumulation of savings and current income. The combination of waiting while processing information and sharp changes once information has been processed through time generates sluggishness and volatility in consumption.

Second, by looking at the life-cycle profile of consumption I find that the behavior of consumption is smooth and persistent with several peaks along the simulated path. These peaks in consumption occur later in life for people that have access to low information flow. This effect is stronger as risk aversion increases. The logic behind this result is that risk averse consumers react to uncertainty by processing more information on their low values of wealth and keep their consumption low as a precaution until the uncertainty is diminished. They accumulate more savings throughout early adult life than their infinite-information-processing counterparts. They keep saving until the accumulation of wealth and information indicates that they can enjoy a high consumption profile.

This leads also to the finding that individual consumption can have more than one hump along its path as wealth accumulates through time. The key point is that individuals can vary their information flow during their life time. To see why, suppose that a person receives signals that his wealth is low. In this case, he wants to pay attention to his expenses and closely monitor the activity of his account. Once he makes sure that he has saved enough, he may decide to spend less effort monitoring his balance and enjoy consumption. Decumulation of savings continues until he receives information that he has emptied his checking account. This news call for his attention again, so he starts saving and monitor his balance more frequently than before. These results combined are suggestive of a precautionary motive for savings driven by information processing limits.

Third, I find that consumers with processing capacity constraints have asymmetric responses to income fluctuations, with negative shocks producing sharper and more persistent effects than positive ones. This effect is stronger as the degree of risk aversion increases. Compared with a situation in which there are no information-processing limits, in a rational-inattention consumption-savings model, an adverse temporary income shock makes consumers reduce their consumption for a longer period of time. This happens because risk-averse people who receive bad news about their finances save right away to hedge against the possibility of running out of wealth in the future. Once they have enough savings and information, they gradually increase their consumption and smooth the remaining effect of the shock over time. This result also points toward precautionary motive due to information-processing limits.

Finally, I find that the predictions of the model can be used to address important policy questions. In the context of fiscal reforms of consumer spending, I show that, as wealth decreases, rationally inattentive consumers respond faster to a tax rebate that increases their income by 10%. For a given level of wealth, the lower the processing capacity, the longer it takes for consumption to react to shocks to disposable income. These findings make intuitive sense. A tax rebate matters more for people with lower income and, as a result, tighter budgetary constraints than for wealthy people.² As a result, poorer people acknowledge and react faster to the positive income shocks. By contrast, wealthy people do not perceive the increase in disposable income as a significant change in their financial position. Thus, consumption for wealthy people does not change significantly, instead it adjusts slowly over time. Consider an individual that has wealth and infinite processing capacities. The reaction of consumption to a temporary positive income shock would be to adjust immediately to a new higher value of consumption so to smooth out the effect of the shock throughout time. With limited processing capacity, the individual smooths consumption slowly over time because the effect of the increase in disposable income on wealth spread out slowly through time. These predictions are in line with the empirical evidence on tax rebate (e.g., Johnson, Parker, and Soules (2006)).

My results are observational distinct from the previous literature on consumption and information (e.g., Reis (2006)). The distinguishing feature of my model with respect to previous works is its ability to generate endogenously asymmetric response of consumption to shocks.³ Finally, my paper contributes to the literature that models how people form endogenously expectations and react to the economy on the basis of their rationally chosen information.⁴

The paper is organized as follows. Section 2 lays out the theoretical basis of rational inattention and informally introduces the model. Section 3 states the problem of the consumers as a discrete stochastic dynamic programming problem, while Section 4 derives the properties of the Bellman function. Section 5 provides the numerical methodology used to solve the model. Section 6 delivers its main results. By comparing the predictions of the model on the preliminary evidence on tax rebates, I find that the model can be a valid instrument to address the impact of tax reforms on consumer spending. Section 7 concludes.

²Another way of looking at it is that people with lower income are generally more liquidity constrained. This makes their marginal propensity to consume to a positive shock closer to one than the wealthier people.

³In particular, for a given degree of risk aversion and magnitude of a shock, the response of consumption to a negative shock is stronger on impact and more persistent than the one to a positive shock.

⁴A 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 agents with imperfect recall. Mankiw and Reis (2002) develop a different model in which information disseminates slowly due to infrequent update of information.

2 Foundations of Rational Inattention

Rational inattention (Sims 1988,⁵ 1998, 2003, 2005, 2006) blends two main fields: Information Theory and Economics. The first draws mainly on the work of Shannon (1948). The main contribution is to define a measure of the choice involved in the selection of the message and the uncertainty regarding the outcome. The measure used is entropy. Details on this part are in Appendix F. Based on Shannon’s apparatus, the economic contribution is that of using Shannon capacity as a technological constraint to capture individuals’ inability of processing information about the economy at infinite rate. Given these limits, people reduce their uncertainty by selecting the focus of their attention. The resulting behavior depends on the choices of what to observe of the environment once the information-processing frictions are acknowledged.

2.1 The Economics of Rational Inattention

Consider a person who wants to buy lunch. He doesn’t know his exact wealth but he knows that he has some cash and a credit card. Not recalling the expenses charged on the credit card up to that point, he can go to the bank or simply check his wallet. Going to the bank to figure out his wealth for lunch is beyond his time and interest, so he decides to check his wallet. He browses through it thinking about what he wants and what he can afford to buy for lunch. Mapping dollar bills into his knowledge of prices from previous consumption, he realizes he can only afford a sandwich instead of his favorite sushi roll. Then, he uses the receipt to update his prior on the price of sandwiches, what he thinks he has left in his wallet and, ultimately, his wealth. This updated knowledge will be used for his next purchase. Such a story can be directly mapped into a rational inattention framework.

First, the person does not know his wealth, W , but he has a prior on it, $p(W)$. Before processing any information, his uncertainty about wealth is the entropy of his prior, $\mathcal{H}(W) \equiv -E[\log_2(p(W))]$, where $E[\cdot]$ denotes the expectation operator.⁶ Before processing any information, lunch too is a random variable, C , ranging from sandwiches to sushi. To reduce entropy, he can choose whether to have a detailed report from the bank or to look at his wallet. The two options differ in amount of information and effort in processing their content. The choice of the option (signal) together with consumption result in a joint probability $p(c, w)$. Both dollars in the wallet and knowledge of prices of sandwiches and sushi contribute to the reduction of uncertainty in wealth of an amount equal to $\mathcal{H}(W|C) = -\int p(w, c) \log_2 p(w|c) dc dw$, which is the entropy of W that remains given the knowledge of C . The information flow, or maximum reduction of uncertainty

⁵The bulk of the idea of *rational inattention* can be found in C. Sims’ 1988 comment in the *Brookings Papers on Economic Activity* .

⁶Entropy is a universal measure of uncertainty that can be defined for a density against any base measure. The standard convention is to use base 2 for the logarithms, so that the resulting unit of information is binary and called a *bit*, and to attribute zero entropy to the events for which $p = 0$. Formally, given that $s \log(s)$ is a continuous function on $s \in [0, \infty)$, by l’Hopital Rule $\lim_{s \rightarrow 0} s \log(s) = 0$.

about the prior on wealth, is bounded by the information that the selected signal conveys. In formulae:

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

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

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

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

3 The Formal Set-up

3.1 The problem of the household

To understand the implications of the limits to information processing, I start with the full information problem.

Let (Ω, \mathcal{B}) be the measurable space where Ω represents the sample set and \mathcal{B} the event set. States and actions are defined on (Ω, \mathcal{B}) . Let \mathcal{I}_t be the σ -algebra generated by $\{c_t, w_t\}$ up to time t , i.e., $\mathcal{I}_t = \sigma(c_t, w_t; c_{t-1}, w_{t-1}; \dots; c_0, w_0)$. 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, $u(c) = c^{1-\gamma}/(1-\gamma)$ with γ the coefficient of risk aversion. Consumer's problem is:

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

s.t.

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

$$w_0 \text{ given} \quad (4)$$

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

Consider now a consumer who cannot process all the information available in the economy to track his wealth precisely. This not only adds a constraint to the decision problem but fundamentally affects each constraint (3)-(4).

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

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

$$\kappa_t \geq I(p(\cdot_{c_t}, \cdot_{w_t})) = \int p(c_t, w_t) \log \left(\frac{p(c_t, w_t)}{p(c_t) g(w_t)} \right) dc_t dw_t \quad (5)$$

Fourth, the update of the prior replaces the law of motion of wealth by using the budget constraint in (3). To describe the way individuals transit across states, define

the operator $E_{w_t} (E_t (x_{t+1}) | c_t) \equiv \hat{x}_{t+1}$, which combines the expectation in period t of a variable in period $t + 1$ with the knowledge of consumption in period t , c_t , and the remaining uncertainty over wealth. Applying $E_{w_t} (E_t (\cdot) | c_t)$ to equation (3) leads to:

$$\hat{w}_{t+1} = R(\hat{w}_t - c_t) + \hat{\bar{y}} \quad (6)$$

where,

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

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:

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

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

Let θ be the shadow cost of using the channel (5), and combine all these four ingredients. Then, the program of the household under information frictions is:

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

s.t.

(θ)

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

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

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

⁷See Astrom, K. (1965).

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

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

This problem is a well-posed mathematical problem with convex objective function and concave constraint sets. What makes it hard to solve is that both the state and the 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.

Before turning to the solution, I present a brief digression about how constraint (9) operates and how the difference between this model and the existing literature on rational inattention may help to build up the intuition for the solution methodology and the results.

3.2 The role of Shannon’s capacity constraint

3.2.1 Shannon’s constraint in action

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

A caveat is in order. In order to explore the interaction between information flow and coefficient of risk aversion, I solve the model in (8)-(12) information flow by fixing the shadow cost of processing information, θ , attached to (9) and let κ vary endogenously every period. In this section, I follow a different route. In order to clarify the mechanisms behind Shannon capacity as a constraint for information transmission, I fix the number of bits, κ , across utilities and adjust the shadow cost θ to map different coefficients of risk aversion to the same information flow.⁸ First consider $u(c) = \log(c)$. In the full information case,⁹ the distribution $g(w)$ is degenerate, the choice of $p(c_t, w_t)$ reduces to that of $c(w_t)$ in (8).¹⁰ The resulting optimal policy is given by

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

⁸To be more specific, I solve the model with CRRA consumer assuming the same parameters as the baseline model $(\beta, R, \bar{y}) \equiv (0.9881, 1.012, 1, 1)$, the same simplex point (prior) $g(\tilde{w})$ and adjusting the shadow cost of processing capacity, θ , to get roughly the same information capacity ($\kappa_{\log} = 2.08$ and $\kappa_{crra} = 2.13$). The latter implies that the difference in allocation of probabilities within the grid are attributable solely on the coefficient of risk aversion γ . As I will explain in more details in the solution methodologies, the same shadow cost (θ) does deliver different information flow (κ) according to the degree of risk aversion of the agents with more risk averse agents having higher κ for a given θ than less risk averse ones. To get $\kappa_{\log} \approx \kappa_{crra}$, I set $\theta_{\log} = 0.02$ in Figure 3 while $\theta_{crra} = 0.08$ in Figure 4.

⁹Or, in the wording of my model, when information flows at infinite rate, $\kappa \rightarrow \infty$ in (9).

¹⁰More formally, for $I(p(\cdot_w; \cdot_c)) \rightarrow \infty$, the probabilities $g(w)$ and $p(\cdot_w, \cdot_c)$ are degenerate. Using

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

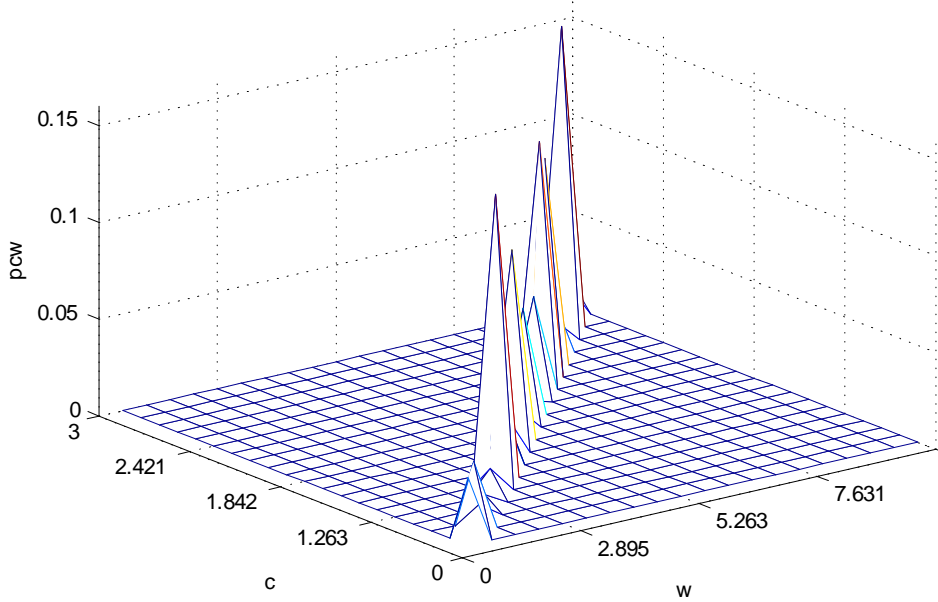


Figure 1: Joint pdf $p(c, w)$, high capacity.

Suppose now that capacity is low. In this case, rational consumers limit their processing effort by concentrating probability on the highest feasible value(s) of consumption. To see why, recall that consumers are risk averse (log-utility). They process the necessary information to learn where the boundary $c \leq w$ is and avoid infeasible consumption bundles.¹¹ Since the Shannon capacity places high restriction on information-processing, this individual consumes roughly the same amount each period, independently of his level

Fano's inequality (Thomas and Cover 1991),

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

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

¹¹The model assumes a standard No-Ponzi condition for the model (8)-(12).

of wealth.

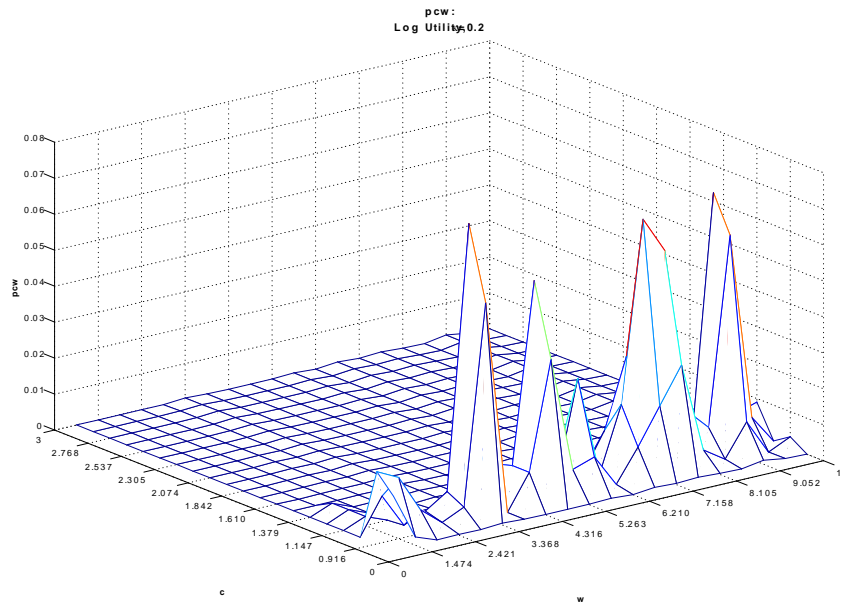


Figure 2: Joint pdf $p(c, w)$, low capacity.

This case describes situations in which people have a vague idea of their wealth and prefer default savings/spending options (whether it is a pension plan or health insurance) rather than figuring out the exact consistency of their net worth. Figure 2 displays the resulting optimal policy. Finally, Figure 3 displays the optimal joint distribution for an intermediate case, $0 < \kappa < \infty$. The first observation is that a person with a finite information flow tries to make $p(c|w)$ as close to w as the information constraint allows

him to.

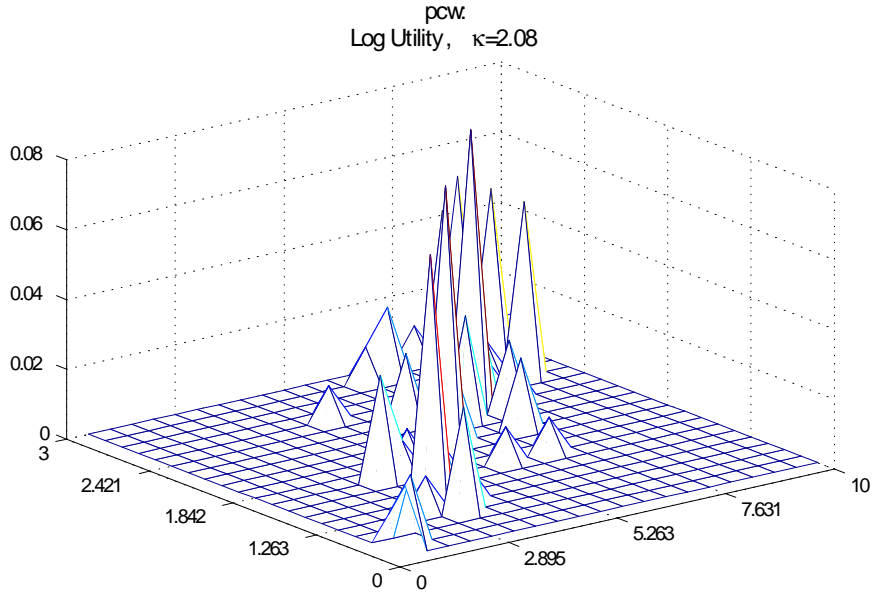


Figure 3. Joint distribution $p(c, w)$, intermediate capacity.

The second observation is that the optimal policy function for the information-constrained consumer places low weight, even no weight, on low values of consumption for high values of wealth. The reason why this happens depends on the utility function. A consumer with log-utility wants to maintain a consumption profile that is fairly smooth throughout the lifetime, as can be seen from (13). To avoid values of consumption that are either too low or too high, he needs to be well informed about such events to reduce the probability of their occurrence. The resulting optimal policy places a higher probability mass on the central values of consumption and wealth.

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

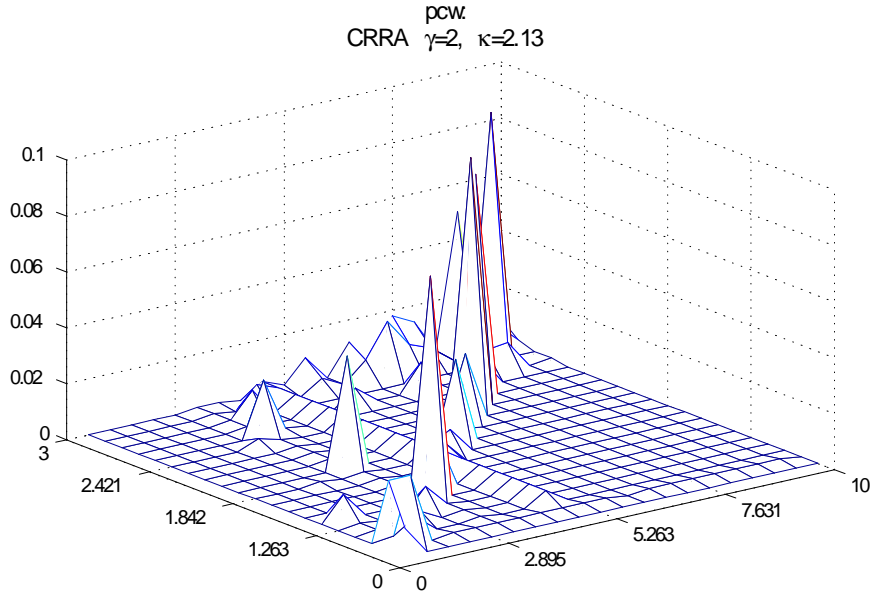


Figure 4. Joint distribution $p(c, w)$, CRRA utility.

In other words, with CRRA preferences, individuals want to be better informed on low and middle values of wealth to enjoy high consumption in every period. Figure 4 illustrates this case.

3.2.2 Shannon’s channel through the economic literature

The goal of this section is to compare my model with the literature in rational inattention. The first comparison is with the consumption saving model in the linear quadratic Gaussian (LQG) case¹² Sims(2003) fully characterizes the analytical solution of a consumption saving model where utility is quadratic, $u(c) = c - 0.5*c^2$, constraints are linear and ex-ante optimal shape of uncertainty is Gaussian. In this LQG setting, the optimal distribution of ex-post uncertainty is also Gaussian. The Gaussian solution make a model with rational inattention in the LQG case observationally equivalent to a signal extraction problem a la Lucas.

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

¹²cfr. Sims (1998, 2003), Luo (2008), Lewis (2007).

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

The third issue, which is the most important for the purpose of this paper, is that rational inattention LQG models do not allow to explain different speed and amounts of reactions of people to different news about their wealth. For instance, consumption drops faster following a sudden layoff than in the event of a tax break. Moreover, the magnitude of the change in consumption depends on people's attitude towards risk¹³ and their income level¹⁴. The certainty equivalence framework that arises with Gaussian ex ante uncertainty and quadratic utility does not allow for endogenous differentiation amongst these events. In such a setting, the speed and amount of households' reactions to different news are created by sources of inertia exogenous to the model. This has been one of the criticisms to signal extraction models a la Lucas and applies also to rational inattention LQG.¹⁵ For instance, different reactions are generated by assuming that people have immediate access to some signals and not others, as in Lucas (1973) or they receive independent information about different news, as in Maćkowiak and Wiederholt (2008). In this paper, I choose another approach. I assume that information is freely available and I do not constrain ex-ante uncertainty to be Gaussian. Moreover, I explore the link between risk aversion and information-processing limits by allowing utility specifications of the CRRA family.

Before this paper, Sims (2006) solves a two period model with non-Gaussian ex-ante uncertainty and CRRA preferences. Sims (2006) assumes that agents live two periods, the first of which they are inattentive while the second period their uncertainty is resolved. This paper focuses on a fully dynamic rational inattention model. I depart from the work of Sims (2006) in two main dimensions. The first is conceptual. A fully dynamic model with rational inattention allows the researcher to investigate time series properties of consumption and savings. The resulting behavior reveals endogenous noise and delays of consumption in response to shock to income, with negative income shocks producing faster reactions effects as the risk aversion increases. The intuition for this result is the reaction of risk adverse individuals to signals that indicate a reduction in wealth is to immediately decrease their consumption for precautionary motives while collecting information over time about the consistency of their net worth. Complementary to these findings, richer dynamic makes the model suitable to address policy questions such as reaction to fiscal policy stimulus as I will show in the last section. This paper is also distinct from the one of Lewis (2008) . The most prominent differences are that, in Lewis (2008), households do not see consumption over time and they optimize over a finite

¹³cfr., e.g. Gourinchas and Parker, 2001.

¹⁴cfr., e.g., Johnson, Souleles and Parker (2006).

¹⁵For a discussion on the Gaussian assumption in rational inattention models see Lewis (2007).

horizon. Not observing consumption in turn implies that once the stream of probabilities is chosen at the beginning of period, the update of the beliefs is deterministic in the choice of the signal. While Lewis (2008)'s framework does deliver upward-sloping age profiles as average consumption over a fixed time length, it does not allow to study unconditional moments of consumption nor conditional response of consumption to shocks as in my framework.

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

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

4 Solution Methodology

4.1 Discretizing the Framework

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

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

Let $|S|$ be the dimension of the *belief simplex* which approximates the distribution $g(w)$ and let $\Gamma \equiv \left\{ g \in \mathbb{R}^{|S|} : g(i) \geq 0 \text{ for all } i \sum_{i=1}^{|S|} g(i) = 1 \right\}$ denote the set of all prob-

ability distribution on Π . The initial condition for the problem is $g(w_0)$.

The consumer enters each period choosing the joint distribution of consumption and financial possibilities. From the previous section, the control variable for the discretized set up as the probability mass function $\Pr(w, c)$ where $c \in \Omega_c$ and $w \in \Omega_w$, constrained to belong to the set of distributions. Given $g(w_0)$ and $\Pr(c_t, w_t)$ and the observation of c_t consumed in period t , the belief state is updated using Bayesian conditioning:

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

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

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

The interpretation of (15) is akin to its continuous counterpart. The capacity of the agents to process information is constrained by a number, $\bar{\kappa}$, which denotes the upper bound on the rate of information flow between the random variables C and W ¹⁶ in time t . Finally, the objective function (8) in the discrete world amounts to

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

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, 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 curvature of the utility function and the fact that the belief space has a bounded support in $[0, 1]$. The 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 $Q : (\Omega_w \times Y \times \mathcal{B} \rightarrow [0, 1])$ and (14) has the Feller property. Then, the conditions for applying the Theorem of the Maximum are fulfilled which guarantees the existence of a solution. In the next section, I provide sufficient conditions to guarantee uniqueness.

Casting the problem of the consumer in a recursive Bellman equation formulation,

¹⁶Recall from the argument in *Section 2.1* that both W and C are random variables before the household has acquired and processed any information.

the full discrete-time Markov program amounts to:

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

subject to:

(θ :)

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

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

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

$$1 \geq \Pr(c_t, w_t) \geq 0 \quad \forall (c_t, w_t) \in B, \forall t \quad (21)$$

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

The Bellman equation in (17) takes up as its argument the marginal distribution of wealth $g(w_t)$ and uses as the control variable the joint distribution of wealth and consumption, $\Pr(c_t, w_t)$. The latter links the behavior of the agent with respect to consumption (c), on one hand, and income (w) on the other, hence specifying the actions over time. The first term on the right hand side of (17) is the utility function $u(\cdot)$. The second term, $\sum_{w_t \in \Omega_w} \sum_{c_t \in \Omega_c} V(g'_{c_t}(w_{t+1})) \Pr(c_t, w_t)$, represents the expected continuation value of being in state $g(\cdot)$ discounted by the factor $\beta = 1/R = 0.9881$. This corresponds to interest rate $R = 1.012$ which gives an annualized gross real rate of investment $R^4 = 1.0489$, with a quarterly frequency of the data. The expectation is taken with respect to the endogenously chosen distribution $\Pr(c_t, w_t)$. I have discussed the relations in (18)-(21) earlier. Moreover, I appended the equation in (20) which constrains the choice of the distribution to be consistent with the initial prior $g(w_t)$.

Next, I analyze the main properties of the Bellman recursion (17) and derive conditions under which it is a contraction mapping and show that the mapping is isotone.

4.3 Properties of the Bellman Recursion

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

First, note that the 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). The 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 the convexity of the value function and the fact that the value iteration is a contraction mapping. All the proofs are in Appendix A.

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

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

with $0 \leq \beta < 1$ and $\|\cdot\|$ 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 the value functions together with the supreme norm form a *Banach space*; 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*

$$V \leq U \implies HV \leq HU$$

that is the value recursion H is an isotonic mapping.

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

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

5 Numerical Technique and its Predictions

I solve the model by transforming the underlying partially observable Markov decision process into an equivalent, fully observable Markov decision process with a state space that consists of all probability distributions over the core¹⁷ state of the model (wealth).

For a model with n core states, w_1, \dots, 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 tetrahedon in a single, two, three or four-dimensional space, respectively. Formally, a belief simplex is defined as the *convex hull*¹⁸ of belief

¹⁷The state of the model is a probability distribution of wealth, i.e., $g(w)$. For lack of a better alternative, I call *core* state the random variable w whose distribution is the state of the model. This nomenclature is borrowed from information theory and AI literature. cfr. Puterman (1994) .

¹⁸A *convex hull* of a set of points is defined as the closure of the set under convex combination.

states from an *affinely independent*¹⁹ set B . The points of B are the vertices of the belief simplex. The convex hull formed by any subset of B is a face of the belief simplex. To address the issue of dimensionality in the state space of my model, I use a grid-based approximation approach. The idea of a grid based approach is to use a finite grid to discretize the uncountably infinite continuous state space. The implementation has the following steps: I place a finite grid over the simplex point, I compute the values for points in the grid, and I use a kernel regression to interpolate solution points that fall outside the grid.

5.1 Belief Simplex and Dynamic Programming

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

To overcome this issue, Astrom (1965) proposed an *information state* approach. It is based upon the idea that all the information needed to act optimally can be summarized by a vector of probabilities over the system, the *belief* state. Let $g(w)$ denote the probability that the wealth is in state $w \in \Omega_w$ where Ω_w is assumed to be a finite set. Probability distributions such as $g(w)$ that are defined on finite sets are in fact *simplices*. Let n be the possible values that w can assume. The discretization of the core state is an equi-spaced grid with $n = 20$ values of w ranging from 1 to 10. The points in the simplex Δ are n distinct values for the marginal pdf $g(w)$ in the interval $I \equiv [0, 1]$. The simplex is constructed using uniform random samples from the unit simplex. The reason why I use this methodology is that it is computationally faster than non-uniform grid and it is able to handle higher dimensional space.²⁰ In my model, each point in the simplex is an n -array whose column contains m random values in the $[0, 1]$ range and whose sum per row is 1. To span the simplex I use $m = (n - 1)!$.²¹ The distribution of values within the simplex is uniform in the sense that it has the conditional probability of a uniform distribution over the whole m -cube, given that the sum per row is 1. The algorithm calls three types of random processes that determine the placement of random points

¹⁹A set of belief states $\{g_i\}$, $1 \leq i \leq z$ is called *affinely independent* when the vectors $\{g_i - g_z\}$ are linearly independent for $1 \leq i \leq z$.

²⁰At least compared to the ndgrid library functions in Matlab. This is because the algorithm creates the simplex directly while when using ndgrid it is necessary to define a uniform grid over the whole $n - 1$ space and then sectioning the resulting grid so that each simplex point sum to one.

²¹With $n = 20$, the proposed sampling produces the same results for sample size of $m = (n - k)!$, for $k = 1, \dots, 5$. I have not tried cases with $k < 5$. When $k > 1$, even if the algorithm produces the same results it takes longer to converge (about 3 minutes more per iteration).

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

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

Let core states and behavior states be sorted in descending order. I impose the constraint $c < w$,²². Then, given the symmetry in the dimensionality of Ω_c and Ω_w , the joint distribution of consumption and wealth for a given multidimensional grid point is square matrix with rows corresponding to levels of consumption. Summing the matrix per row results in 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, $g(w)$. Given the initial belief simplex, its successor belief states can be determined by Bayesian conditioning at each multidimensional point of the simplex and gives the expression:

$$g(w'|c) = \sum_i T(w'; w_i, c) \Pr(w_i|c) = \Pr(w'|c). \quad (23)$$

Let \mathcal{V} be the set of all bounded real-valued function V on G . Then, the Bellman optimality equation of the household is described by (17)-(21).

Without loss of generality, I restrict the columns of the matrix $\Pr(c, w)$ to sum to the marginal pdf of wealth in the main diagonal. Moreover, because some of the values of the marginal $g(w)$ per simplex-point are exactly zero given the definition of the envelope for the simplex, I constrain the choices of the joint distributions corresponding to those values to be zero. This handling of the zeroes 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 $\frac{n*(n-1)}{2}$.²³ Once the *belief simplex* is set up, I initialize the joint probability distribution of consumption and wealth per belief point and solve the program of the household by backward induction iterating on the value function $V(g(w))$. To map the

²²The constraint $c < w$ makes economic sense since there is no borrowing in this economy. To encode this constraint without complicating the model, one may assume that κ_t in (18) is the capacity left after the consumer has processed his spending limits.

Note also that this constraint is computationally convenient reducing the number of choice variables from $n^2 = 400$ to $\frac{n*(n+1)}{2} = 210$ per iteration.

²³To illustrate this point, two examples 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., $g \equiv \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$, the consumer has to choose $\frac{3*(2)}{2} = 3$ points on the joint distribution, $\{p_1, p_2, p_3\}$ placed as:

finer state space into Matlab possibilities, I interpolate the value function with the new values of (23) using a kernel regression of $V(\cdot)$ into $g'(w'|_a)$. I use an Epanechnikov kernel with smoothing parameter $h = 2.7$.²⁴ A kernel regression approximates the exact non linear value function in (17) with a piece-wise linear function. The following propositions illustrate this point.

Proposition 2. *If the utility is CRRA with a parameter of risk aversion $\gamma \in (0, +\infty)$ and if $\Pr(c_j, w_i)$ satisfies (18)-(21), then the optimal n -step value function $V_n(g)$ 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 α -vectors, $\alpha : \Omega_w \rightarrow R$, are $|W|$ -dimensional hyperplanes.

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

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

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, the $\langle \alpha_n^i, g \rangle$ operator is linear in the belief space. Therefore, the convex property is given by the fact that V_n is defined as the maximum of a set of convex (linear) functions and, thus, obtains a convex function as a result. The optimal value function V^* is the limit for $n \rightarrow \infty$ and, because all the V_n are convex function, so is V^* .

Proposition 3. *Assuming the CRRA utility function and 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).*

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

²⁴Epanechnikov kernel is an optimum choice for smoothing because it minimizes asymptotic mean integrated squared error (cfr. Marron, J. S. and Nolan, D. (1988)). I use the algorithm proposed in Beresteanu, A. and C. F. Manski (2000) and experiment with smoothing parameter $h \in [0.3 : 0.3 : 4.2]$. For the characteristics of the problem, and the optimization routine used (csmmwel), for different specification of utility functions and Lagrange multiplier θ , the parameter $h = 2.7$ performs best in terms of computational time and convergence of the value function.

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

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

Benchmark Values

| | Discretization |
|---|--------------------|
| Wealth Space W | [1 : 0.4737 : 10] |
| Consumption Space C | [0.8 : 0.1158 : 3] |
| Mean of Income, \bar{y} | 1.1 |
| Joint Distribution per simplex point, $p(c, w)$ | 20×20 |
| Marginal C | 20×1 |
| Marginal W | 20×1 |
| Coeff. risk aversion, γ | 1 |
| Interest rate, R | 1.012 |
| Discount Factor, β | 0.9881 |

Table 1

6 Results

In this section, I investigate the dynamic interplay of information flow and degree of risk aversion. In particular, I study different specifications of the baseline model changing degrees of risk aversion, $\gamma \in \{0.5, 1, 2, 5, 7\}$, and different Lagrange multipliers, $\theta \in [0.2, 4]$, representing the shadow costs of processing information in (18). Time path for each individual are average across simplex points. For the time series of the aggregate economy, I perform 10,000 Monte Carlo runs and simulate the model for each path for $T = 80$ periods. Then, I compute average across runs and simplex-points. Sample statistics are calculated after I compute these averages. I choose this way of calculating average to compare my model, tailored for individual behavior, to aggregate data. I divide the results into three parts: (1) interaction of information flow and risk aversion; (2) implications of information constraint on lifetime consumption; and (3) consumption

reaction to temporary income shocks.

| <i>Statistics</i> | | | | |
|-------------------|-------------------|-------------------|-------------|--------------------|
| $\theta=0.2$ | | | | |
| | CRRA $\gamma = 7$ | CRRA $\gamma = 5$ | Log Utility | CRRA $\gamma = .5$ |
| $E(C)$ | 1.14 | 1.09 | 1.08 | 1.02 |
| $std(C)$ | 0.08 | 0.09 | 0.11 | 0.14 |
| κ | 2.03 | 1.99 | 1.87 | 1.72 |

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

Table 2

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

This finding is documented in Table 2 and in Figures 5-6. Figure 5a plots the difference between the mean of the time series of consumption between $\theta = 0$ and $\theta > 0$. After deriving the time path of consumption as described above, I calculate the mean and standard deviation of the average of the time path and subtract from it the mean of the time path for the full information equivalent ($\theta = 0$).²⁵ Figure 5a shows how this difference changes as θ varies and when utility is logarithmic. Figure 5b plots the corresponding difference in standard deviation of consumption as a function of θ .

²⁵For the parameter of the model, when $\theta = 0$ a full information solution $c_t^f = \beta w_t + (1 - \beta)\bar{y}$ has mean $E(c_t^f) = 1.124$ and standard deviation $std(c_t^f) = 0.0713$.

Loss in consumption due to increasing processing effort

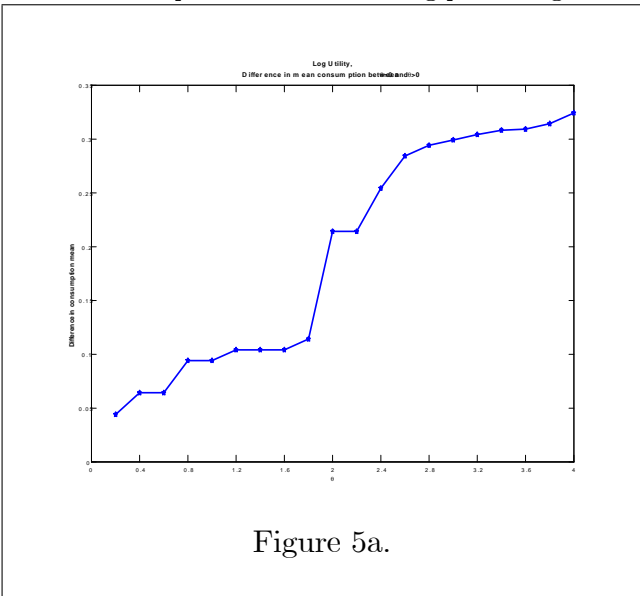


Figure 5a.

Difference in std. of consumption due to processing effort.

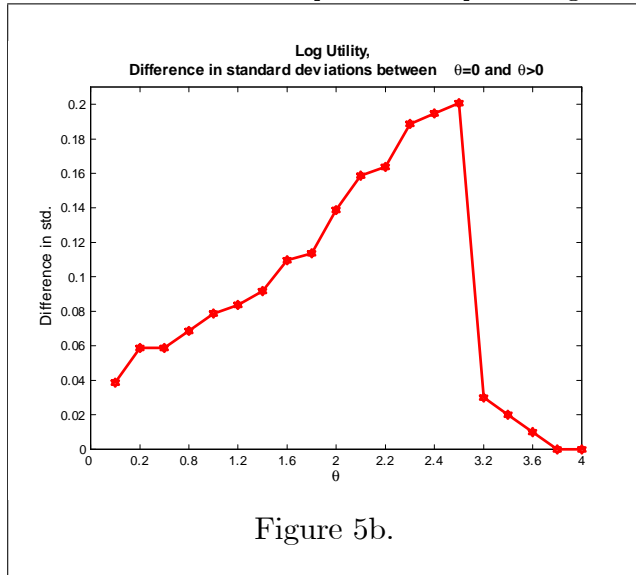


Figure 5b.

To understand this result, consider what happens in the full information ($\theta = 0$) case. With $R\beta = 1$, the agent smooths consumption regardless of his utility. To appreciate how preferences towards risk play out with processing limits ($\theta > 0$), consider Figure 6c. It plots the optimal distribution of consumption for two individuals ($\gamma \rightarrow 1$ and $\gamma = 5$) when information is very costly to process ($\theta = 3$). In this case, a rational agent consumes a fixed amount every period in the limits of his net worth. This requires very little bits of information. In Figure 6c note how a person with log-utility puts probability mass mostly on the lower values of consumption while a more risk averse agent sacrifices smoothing consumption to allocate some probability on higher values of consumption. Assuming the same θ , the resulting effect is solely due to consumer preference. Now consider Table 2 and Figures 6a-6b.

Marginal Distribution of Consumption, Log Utility

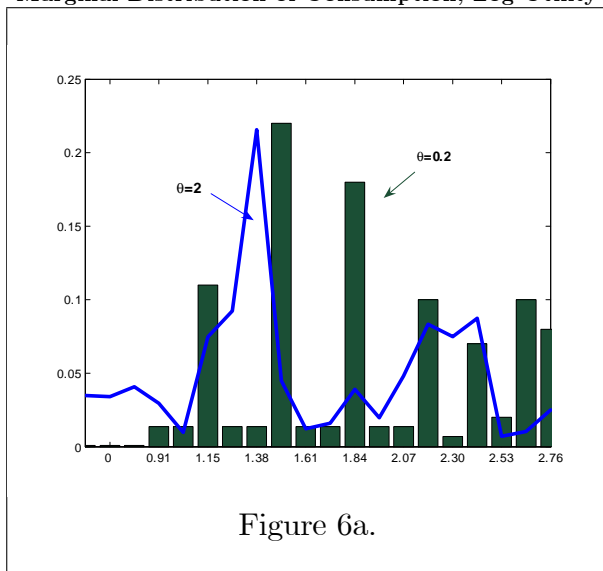


Figure 6a.

Marginal Distribution of Consumption, CRRA $\gamma=5$

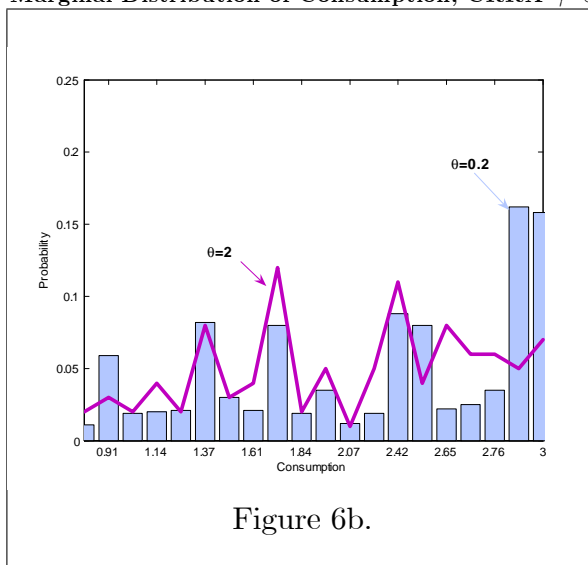


Figure 6b.

Marginal Distribution of Consumption, $\theta=3$

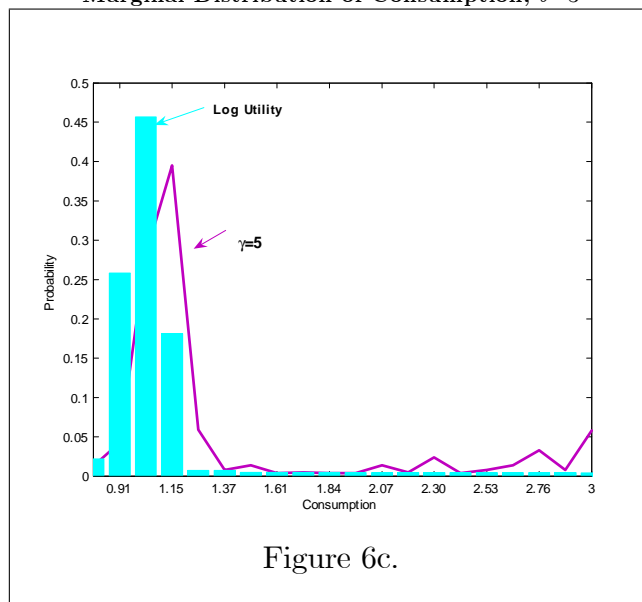


Figure 6c.

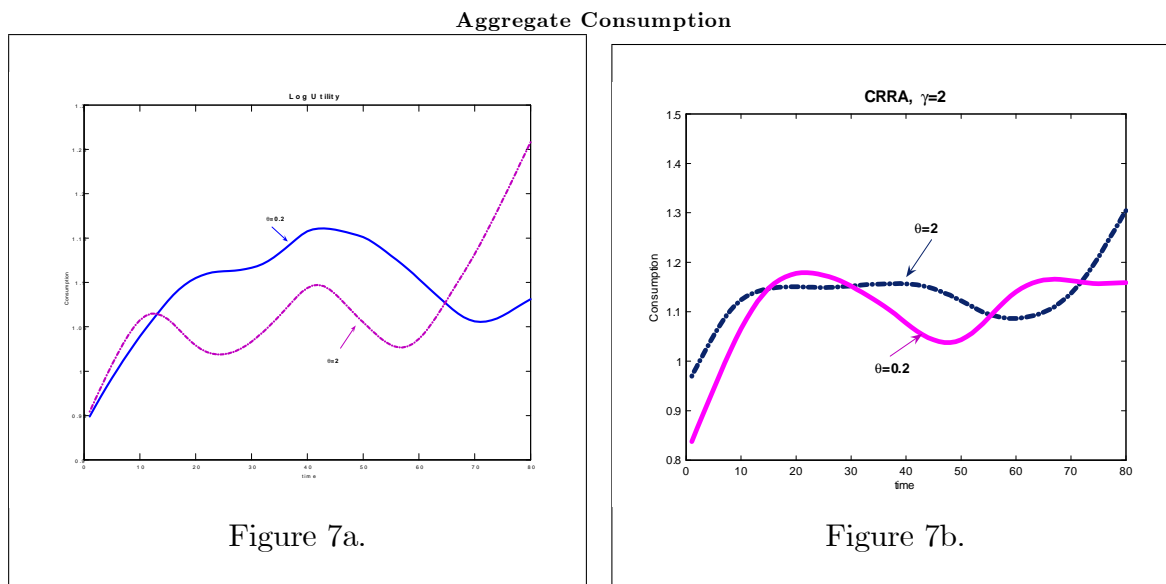
When $\theta = 2$, people select how much information they want to process and which values of wealth to be better informed about according to their utility. Also in this case, the higher the degree of risk aversion, the higher the quest for information (κ). This is exactly what Table 2 shows. In the table, the higher the coefficient of risk aversion, γ , the higher the information collected by the agent, κ , and the higher the mean of consumption. The same story can be told in terms of probability distribution as in 6a-6c. For a given level of θ , a person with log utility would be better informed on extreme values of wealth to avoid such values. This knowledge makes it possible to assign high

probability to the middle value of consumption, as his utility commands. By contrast, a consumer with CRRA, $\gamma = 5$, wants to avoid low values of consumption for high values of wealth. Processing information about these events decreases the likelihood of their occurrence and makes it possible to place high probability on high value of consumption. This mechanism makes consumption more persistent for people with a higher degree of risk aversion (cfr. Figure 6a-6c).

Processing capacity (κ) strengthens this effect. This is because high information flow allows consumers to enjoy high and smooth consumption throughout their life time. If information flows at very low rate, households update their knowledge slowly over time and wait to modify their behavior until they have sufficient knowledge of their financial possibilities. Inertial behavior of consumption due to low information flow induces sharp changes in consumption after the consumer accumulates information. This mechanism makes consumption more volatile for people with lower information flow.

Figure 5b plots the standard deviation of consumption for several values of θ . As pointed out, for very high shadow cost of processing information $\theta > 3$, consumption does not vary over time. For $0 < \theta < 3$, the volatility of consumption increases with θ . This result makes sense. To see why, consider again the full information version of the model. People's will to smooth consumption in full information is limited by the finite flow of information available. When deciding on the precision of their signals, risk averse people trade off lower volatility in consumption for better knowledge of low value of wealth.

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



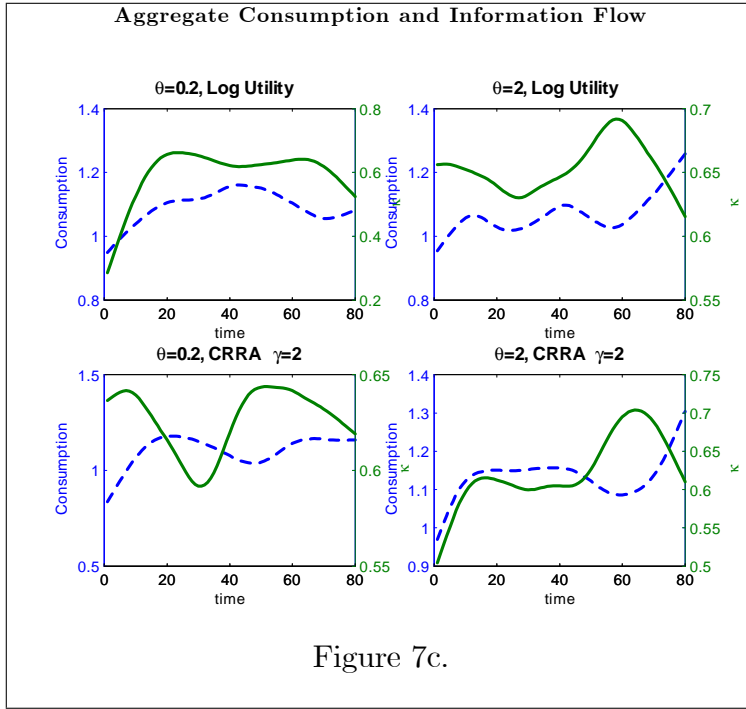


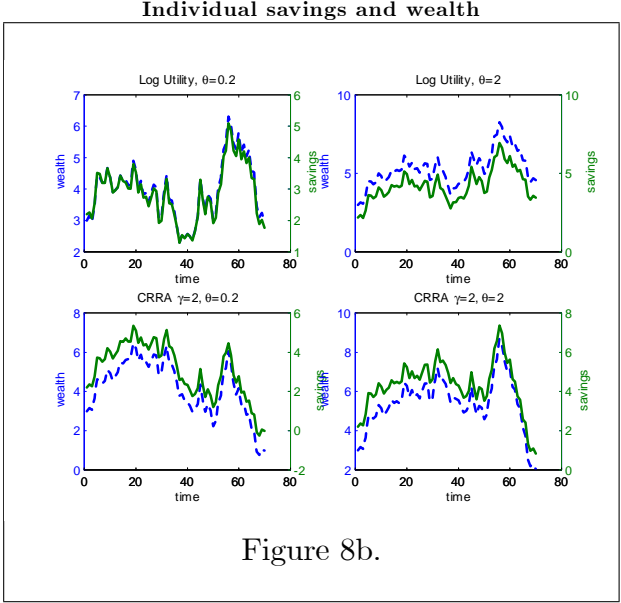
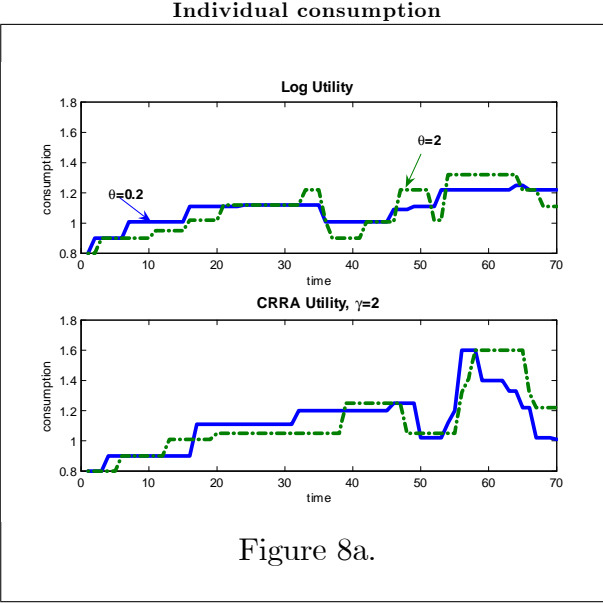
Figure 7c.

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

1. *Consumption is hump-shaped. It gets to its peak later for individual that have low information flow. The effect is stronger as the degree of risk aversion increases..*
2. *Individuals with high information flow, by having sharper signals on their wealth, have savings behavior that follows closely their wealth. Furthermore, the lower the degree of risk aversion, the higher the fluctuations of savings per period.*
3. *Individuals with low information flow, tend to consume a constant amount every period. They increase their consumption only if the information they process points them towards a significant increase in wealth. The higher the degree of risk aversion, the less volatile the time path of consumption for these types.*

Figures 7-8 illustrate these points for aggregate and individual time series behavior, respectively. The simulations are derived by drawing the time path of consumption and wealth from $p^*(c, w)$, after the value iteration has converged. Figures 7a-7c plot the average across the Monte Carlo runs and simplex points (i.e., initial beliefs about wealth). Individual time series (Figures 8a-8b) are an average of initial beliefs. To have some interesting transitional dynamics, I begin the simulation with an initial condition for wealth far from the steady state²⁶.

²⁶For the grid in the model, the steady state value of wealth is $\cong 5.65$ and I initialize the simulation with $w_0 = 3$.



To appreciate the results, consider what would happen with full information. In such a case, consumption smoothing ($R\beta = 1$) implies an immediate ($T = 1$) adjustment of consumption to its long-run optimal values and no transient behavior. Thus, in that case from $T = 2$ onwards, the simulations lead to a constant time path. Now consider Figures 7(a-c)-8(a,b). The hump in consumption comes from Result 1 and a simple intuition: information-constrained people are cautious (degree of risk aversion $\gamma \geq 1$), consume a little and collect information about wealth before they change consumption. For a fixed θ , the more risk averse they are (cfr. Figure 7a with log utility and Figure 7b with CRRA, $\gamma = 2$), the longer they wait before increasing their consumption. This inertial behavior in consumption leads to an increase in savings and, as a result, in wealth (cfr. Figure 8a-8b). Processed information keeps signaling the increase in wealth until households realize that they are wealthy enough to increase their consumption. Thus, the hump in consumption is the mirrored image of the rise (until people know they are rich) and fall (once people know they are rich) in wealth. Note that, depending on the history of income shocks, consumption can have more than one hump in its path. To see why, consider a high realization of income occurring after a hump in consumption. Over time, signals about wealth convey such information, consumers start saving and history as well as humps repeat themselves. These effects are enhanced by the shadow cost of processing information, θ , with higher costs forcing long periods of inertia in consumption followed by sizeable changes. Note also the relationship between consumption and information flow (Figure 7c): risk averse agents would rather push forward consumption in times in which they are processing information about wealth. Finally, note from 7(a-b)-8(a,b) how the peak in consumption occurs later for an individual with higher degree of risk aversion and lower information flow. The rationale for this result is that more cautious people wait to be better informed about their wealth before modifying their consumption

behavior. In particular, since a consumer with CRRA utility ($\gamma = 2$) chooses to be better informed about low values of wealth than a log utility consumer (cfr. Figures 7a and 7b), he processes news about high value of wealth slower than his log counterpart. The resulting additional savings for precautionary motives are triggered by both the curvature of the utility function and the bound on information-processing constraint.

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

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

The logic behind this result is easily understood by considering the interdependence of information flow and coefficient of risk aversion. A risk averse person is more likely to be affected by negative events than positive ones. As soon as he receives signals that his wealth is lower than what he thought, he reacts by decreasing his consumption. The change in behavior and its persistence are more consistent the more risk averse and uninformed the consumer is. This occurs because consumers wait to gather more information before changing their behavior and, in the meanwhile, build up a savings buffer. Thus, the temporary change in income propagates slowly over time. A positive temporary income shock triggers the opposite behavior in a risk averse uninformed person. The intuition is that this type of consumer is concerned about negative wealth fluctuations and allocates most of his information capacity to prevent this event. A signal that indicates positive wealth may be ignored, generating extra savings in the meanwhile. Once this is acknowledged, a prudent consumer distributes the additional consumption driven by the income shock plus savings throughout his lifetime. 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.²⁷

²⁷*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) refers to the observation that aggregate consumption is smoother than permanent income in that it reacts with a less than one-to-one ratio to shocks to permanent income.

IRF to a temporary increase in income

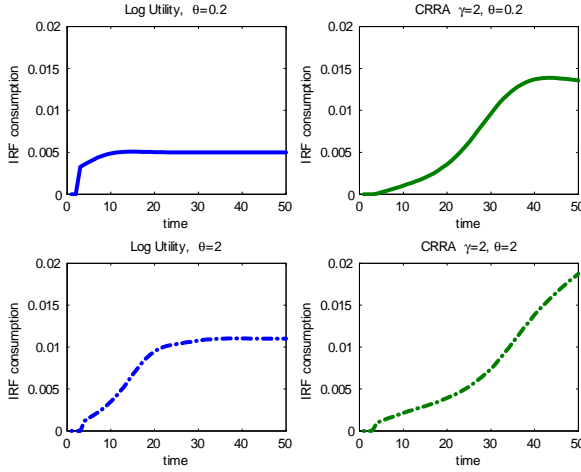


Figure 9a.

IRF to a temporary decrease in income

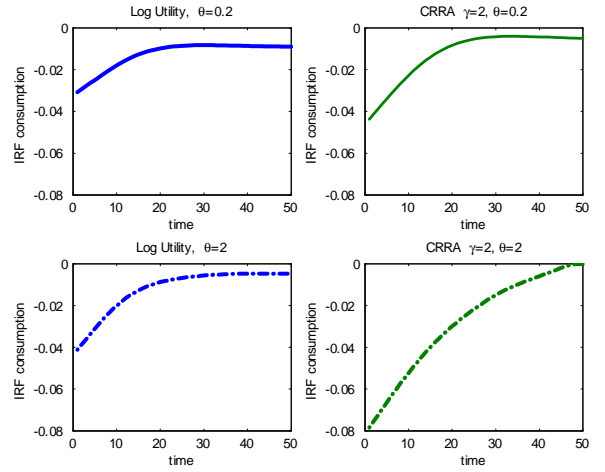


Figure 9b.

Since the model is non-linear, let me first explain how the impulse responses are generated and then focus on the intuition that the graphs suggest. I simulate the model drawing 10,000 times from the same optimal policy distribution under two scenarios. In the first, I draw from a distribution with constant mean of the shock to income. In the second, I assume that the mean of the shocks increase/decrease in the very first period (one-time shocks) and then revert back to its original distribution. Impulse responses of consumption are the difference between the two income paths averaged over simplex-points and 10,000 Monte Carlo draws of income. The impulse response functions are plotted in Figures 9a-9b. Consider Figures 9a first. They display a positive (Figure 9a) and a negative (Figure 9b) shock to income respectively. Note that for both log and CRRA $\gamma = 2$ and for different value of the shadow cost ($\theta = 0.2 \vee \theta = 2$) the reaction to a negative shocks ($\Delta y = |1|$) starts from the very first period. However, the extent of the reaction varies across utilities and information costs. When $\theta = 0.2$, a log utility-type consumer reacts on impact by increasing savings to an extent lower than the shock. He then adjust savings and consumption so to distribute the averse shock throughout time. The same log-type but with $\theta = 1$ decreases more consumption on impact than his $\theta = 0.2$ counterpart. He increases consumption slowly over time until it reaches its new long-run value. Likewise, a consumer with risk aversion $\gamma = 2$ varies his saving when the shock hits to an extent that depends on his information flow. In particular, note that for $\theta = 2$ the decrease in consumption on impact and in the following periods is so significant that consumers can use the accumulated savings to restore their original consumption plan. The endogenous asymmetric response to shocks, even in this very simple setting, makes rational inattention models observationally distinct from any other standard macroeconomic model. In those frameworks, either there is no asymmetric

reaction (as in LQG) or the asymmetric response is due to the asymmetric magnitude of the shocks (as in models a la Lucas). These implications make the theory appealing from an empirical standpoint (e.g., think about consumers' reactions to a tax break vs. being fired from the job). Moreover, they make the theory suitable to study the impact of policy changes on private sectors decision. The 2008 Tax Rebate provides one such example.

6.1 Sensitivity analysis and policy implications

A feature of the model worth exploring is how consumption's reaction to shocks depends on the initial value of wealth.

Drawing a time series from the probability distribution that solves the model, it is natural that the farther away wealth is from its steady state, the more consumption reacts to a shock to wealth. The interesting prediction of the model with an information-processing constraint is that for either the log or the CRRA, $\gamma > 1$ utility, it does matter for the impulse response whether we start from a value of wealth above or below the steady state. In both cases, the reactions are faster in case of a negative shock than a positive one. However, extent and timing are different with wealthier people reacting faster and with sharper decrease in consumption to a negative shock than poorer people do when facing the same kind of shock. This is due to the fact that poorer people already consume small amount so that when a negative shock hits, even if they receive immediately signal of the news, they only gradually reduce their consumption. Savings slowly accumulate over time until the shock is absorbed. For a given processing capacity, wealthier people can afford to reduce their consumption as soon as they acknowledge the negative shock. The jump start in savings makes it possible for them to absorb the shock faster. By contrast, a positive shock has a stronger effect on poorer people than wealthier one. To see why, consider a tax rebate. Taking two individuals with the same characteristics in terms of risk aversion and information-processing constraints but different initial net worth, the wealthier person takes longer to change his consumption behavior. When the change does occur, the magnitude is smaller than the one for a poorer person. The intuition for this result is that an increase in disposable income for a less wealthy person implies a more sizeable financial break than the same amount does to a wealthy person. Risk aversion prevents both types of consumers from immediately disposing of the additional credit but it has a bigger effect on impact for the more constrained consumer.

Impulse Response function as a function of wealth, $\Delta y=0.02$

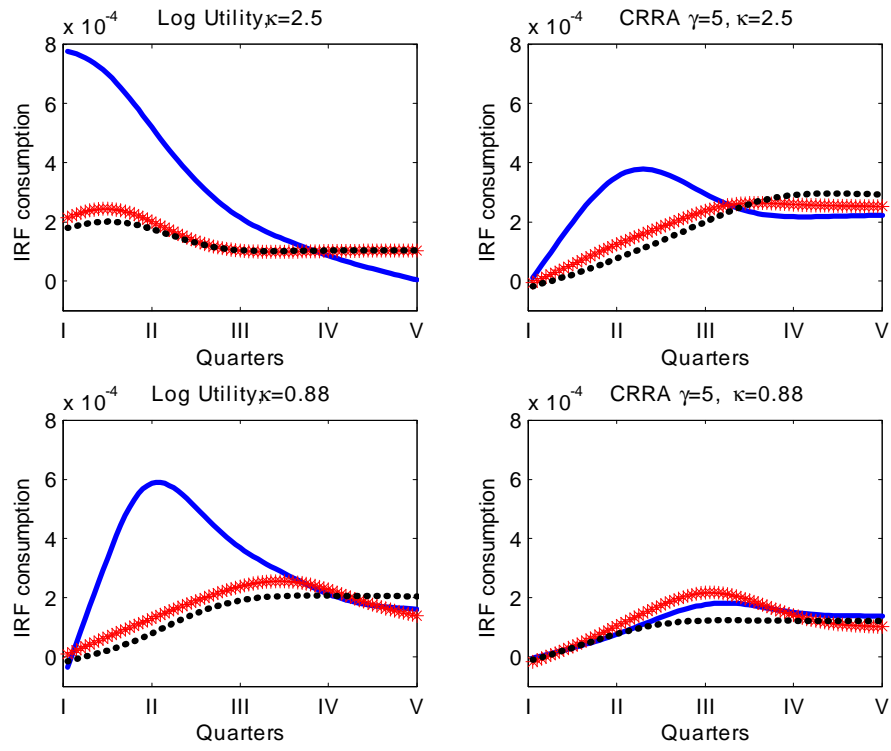


Figure 10. Solid: $w_0 = 1.94$; Star Dashed: $w_0 = 3.3$; Dotted: $w_0 = 5.2$

Even in its simplicity, the model can be used to address important policy questions. In particular, it can be used to analyze the effectiveness of tax policy reforms on individual consumption and savings decisions. Figure 10 displays the impulse response function of consumption to a stimulus payment which increases income of 2% with respect to its (constant) long run level. The discretized solutions are generated using equi-spaced grid of consumption and wealth, with 50 points each. Consumption takes up value in $[0.5, 3]$ while wealth ranges from 1 to 10. I use the same parameters ($R = 1.012$ and $\beta = 1/R$) of the baseline model and a simplex of size $(50!) * (49)$ and two specifications of utility functions. In both cases I choose θ so that the capacities corresponds to $\cong 2.5$ bits and 0.88 bits²⁸. Once the value iteration converged, I generate the impulse response function by simulating a time series path of consumption and wealth with 10,000 Monte Carlo runs for each initial condition on wealth. I consider three initial values of wealth as a proxy of population with low, middle and low net worth. I then average the time series per quarters and simplex points. Figure 10 gives interesting insights on the effect of the stimulus on consumer spending. For the degrees of risk aversion considered and information capacity, the reaction of the stimulus is higher the lower the initial wealth. This is not surprising, as the stimulus payments have bigger impact on the disposable income of credit constrained consumers than richer people. For a given amount of information capacity and wealth, the higher the risk aversion, the lower the spending in the first quarter. This result also makes sense. If a consumer is risk averse and have no credit frictions, he allocates more attention in processing information about low values of wealth. This leads to processing information slower and, in turn, reacting slower to positive news to income (Result 3). Finally for a given wealth and degree of risk aversion, the lower the information processing capacity, the lower the response of consumption spending to the rebate. The findings in Figure 10 can be summarized as:

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

1. *is stronger the lower the initial net worth.*
2. *is more delayed the higher the degree of risk aversion.*
3. *is more persistent but less effective the lower the information-processing capacity.*

The insights one can gather from the model have strong policy implications on the effectiveness of tax reform on people's behavior. The 2008 tax rebate provides one such example. The model predicts that such a policy has greater response on impact for individual with low net worth. Figure 10 also suggests that the effect will be mild and spread out through several quarters for middle-high income households. These findings are consistent with the empirical evidence on consumers spending of 2001 tax rebates (cfr., Johnson, Parker and Soules (2006)).

²⁸The constraint $\kappa = 2.5$ corresponds to $\theta_{\log} = 0.01$ and $\theta_{crra} = 0.05$ for the log case and the crra, $\gamma = 2$ case respectively, while $\kappa = 0.88$ is given by $\theta_{\log} = 0.1$ and $\theta_{crra} = 0.9$.

7 Conclusions

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

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

Furthermore, the model predicts that consumption path has humps. Under information-processing constraints, a hump occurs when people consume a little and save a lot while collecting information about wealth. When consumers realize that they are rich, they increase consumption and decumulate savings. This increase stops when they acknowledge that their wealth is low again: they start to save and process more information. Thus, consumption decreases. Consistent with the previous two results, I find that the peak in consumption is delayed the more the individual becomes risk averse.

Differing from other life-cycle models, in my model there could be more than one hump in the consumption path. Depending on the history of the income shocks, a very low or very high realization of income affects consumers' signal through its effect on wealth. Consumers react to the news by varying savings and information over time, thereby generating another hump.

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

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

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. Stokey, (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] Astrom, K. (1965), *Optimal Control of Markov decision process with incomplete state estimation*. Journal of Mathematical Analysis and Applications 10, pp.174-205
- [8] Broda, C. and J. Parker (2008), *A preliminary analysis of how household spending changed in response to the receipt of a 2008 economic stimulus payment*. Mimeo.
- [9] 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.
- [10] Caballero, R. J. (1995), *Near-rationality, Heterogeneity and Aggregate Consumption*, Journal of Money, Credit and Banking, 27 (1), pp29-48.
- [11] Campbell, J.Y. (1987), *Does Savings Anticipate Declining Labor Income? An alternative Test of the Permanent Income Hypothesis*, Econometrica, 55, pp.1249-1273.
- [12] Campbell, J.Y. and A. Deaton (1989), *Why is Consumption so Smooth?* Review of Economic Studies, 56, pp.357-374.
- [13] 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.
- [14] 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.
- [15] Carroll, C. D. (2003) *Macroeconomic Expectations of Households and Professional Forecasters*, Quarterly Journal of Economics, 118 (1), pp. 269-298.

- [16] 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.
- [17] Cover, T.M. and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991
- [18] 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.
- [19] Deaton, A. (1992), *Understanding Consumption*, Oxford, Oxford University Press.
- [20] Dynan, Karen (2000), *Habit Formation in Consumer Preferences: Evidence from Panel Data*, American Economic Review, 90, pp. 391-406.
- [21] Flavin, M. A. (1981), *The Adjustment of Consumption to Changing Expectations about Future Income*, Journal of Political Economy, 89, pp. 974-1009.
- [22] Friedman, M. (1957), *A Theory of the Consumption Function*, Princeton, Princeton University Press.
- [23] Goodfriend, Marvin (1992), *Information-Aggregation Bias*, American Economic Review, 82, pp. 508-519.
- [24] Gourinchas, P-O, and J.A. Parker (2002), *Consumption over the Life-Cycle*, *Econometrica*, vol. 70 (1), pp.47-89.
- [25] Hall, R. (1978), *Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence*, Journal of Political Economy, 86, pp.971-987.
- [26] Hellwig, C. (2004), *Heterogeneous Information and the Benefits of Transparency*, Discussion paper, UCLA.
- [27] Hellwig, C. and L. Veldkamp, (2006) *Knowing What Others Know: Coordination Motives in Information Acquisition*, NYU, mimeo.
- [28] Johnson, D., Parker, J. and N. Soules (2006) *Household expenditure and the income tax rebates of 2001*. American Economic Review, Vol. 96, No. 5.
- [29] Kim, J., Kim, S., Schaumburg, E. and C. Sims (2005). *Calculating and Using Second Order Accurate Solutions of Discrete Time Dynamic Equilibrium Models*. mimeo.
- [30] Lewis, K. (2007), *The life-cycle effects of information-processing constraints*. Working Paper, University of Iowa.
- [31] Lewis, K. (2008), *The Two-Period Rational Inattention Model: Accelerations and Analyses*", *Computational Economics*, Forthcoming. Currently available as Federal Reserve Financial & Economics Discussion Series Paper, No. 2008-22, Board of Governors .

- [32] Lorenzoni, G. (2006), *Demand Shocks and Monetary Policy*, MIT, mimeo.
- [33] Lucas, R. E., J. (1973), *Some International Evidence on Output-Inflation Tradeoffs*, *American Economic Review*, 63(3), 326–334.
- [34] Luo, Y. (2007), *Consumption Dynamics, Asset Pricing, and Welfare Effects under Information Processing Constraints*, forthcoming in *Review of Dynamics*.
- [35] 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.
- [36] Lusardi, A. (2003), *Planning and Savings for Retirement*, Dartmouth College, unpublished.
- [37] MacKay, David J. C (2003). *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003
- [38] Macrowiack, B., and M. Wiederholt (2007), *Optimal Sticky Prices under Rational Inattention*, Discussion paper, Northwestern/ ECB mimeo.
- [39] 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.
- [40] Mankiw, N. G., Reis, R. and J. Wolfers (2004) *Disagreement in Inflation Expectations*, NBER Macroeconomics Annual 2003, vol. 18, pp. 209-147.
- [41] Mondria, J. (2006), *Financial Contagion and Attention Allocation*, Working paper, University of Toronto.
- [42] Morris, S., and H. S. Shin (2002), *The Social Value of Public Information*, *American Economic Review*, 92, 1521–1534.
- [43] Moscarini, G. (2004), *Limited Information Capacity as a Source of Inertia*, *Journal of Economic Dynamics and Control*, pp. 2003–2035.
- [44] Mullainathan, S. (2002), *A Memory-Based Model of Bounded Rationality*, *Quarterly Journal of Economics*, 117 (3), pp. 735-774.
- [45] Orphanides, A. (2003), *Monetary Policy Evaluation with Noisy Information*, *Journal of Monetary Economics* 50 (3), pp. 605-631.
- [46] Parker, J. (1999) *The Reaction of Household Consumption to Predictable Changes in Social Security Taxes*, *American Economic Review*, 89 (4), pp. 959-973.
- [47] Peng, L. (2005), *Learning with Information Capacity Constraints*, *Journal of Financial and Quantitative Analysis*, 40(2), 307–329.

- [48] Peng, L., and W. Xiong (2005), *Investor Attention, Overconfidence and Category Learning*, Discussion paper, Princeton University.
- [49] Pischke, Jorn-Steffen (1995) *Individual Income, Incomplete Information, and Aggregate Consumption*, *Econometrica*, 63 (4), pp. 805-840.
- [50] Puterman, M.L. (1994), *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, Inc.
- [51] Reis, R. A. (2006) *Inattentive Consumers*, *Journal of Monetary Economics*, 53 (8), 1761-1800.
- [52] 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.
- [53] Rotemberg, Julio J., and Michael Woodford, 1999. "The Cyclical Behavior of Prices and Costs." in John B. Taylor, and Michael Woodford (ed.), *Handbook of Macroeconomics*.
- [54] Shimer, Robert, 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*. 95 (1): 25–49.
- [55] Smets, Frank, and Rafael Wouters, 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association*. 1 (5): 1123–1175.
- [56] Smets, Frank, and Rafael Wouters, 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review*. 97 (3): 586–606.
- [57] Sims, C. A. (1998), *Stickiness*, *Carnegie-rochester Conference Series On Public Policy*, 49(1), 317–356.
- [58] Sims, C. A. (2003), *Implications of Rational Inattention*, *Journal of Monetary Economics*, 50(3), 665–690.
- [59] Sims, C. A.(2005), *Rational Inattention: a Research Agenda*, Princeton University mimeo.
- [60] Van Nieuwerburgh, S., and L. Veldkamp (2004a), *Information Acquisition and Portfolio Under-Diversification*, Discussion paper, Stern School of Business, NYU.
- [61] Stokey, N. L. and R. E. Lucas Jr., with Edward C. Prescott (1989) *Recursive Methods in Economic Dynamics*, Cambridge, Harvard University Press.
- [62] Van Nieuwerburgh, S., and L. Veldkamp (2004b), *Information Immobility and the Home Bias Puzzle*, Discussion paper, Stern School of Business, NYU.

- [63] Van Nieuwerburgh, S., and L. Veldkamp, 2006, *Learning Asymmetries in Real Business Cycles*, Journal of Monetary Economics 53 (4), pp.753-772.
- [64] Wilson, Andrea (2003) "Bounded Memory and Biases in Information Processing," unpublished, Princeton University.
- [65] 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.
- [66] Woodford, M. (2003), *Interest and Prices*, Princeton. Princeton University Press
- [67] Zeldes, S. (1989) *Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence*, Quarterly Journal of Economics, May 1989, 275-98..

8 Appendix A

8.1 Proof of Proposition 1.

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

Proof. The H mapping displays:

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

with

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

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

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

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,

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

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

8.2 Proof of Corollary.

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

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

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

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

$$V \leq U \implies$$

$\forall g(w), c,$

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

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

8.3 Proof of Proposition 2.

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

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

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

and therefore if I define the vectors

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

I have the desired

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

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

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

by the induction hypothesis

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

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

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

With the above:

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

At this point, it is possible to define

$$\alpha_{p,c}^j(w) = \sum_{w' \in \Omega_w} \alpha_{n-1}^i(w') T(\cdot : w, c) \cdot p. \quad (31)$$

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

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

and define:

$$\alpha_{p,c,g} = \arg \max_{\{\alpha_{p,c}^j\}_j} \langle \alpha_{p,c}^j, g \rangle. \quad (33)$$

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

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

Now

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

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

8.4 Proof of Proposition 3.

Proof. The first task is to prove that $\{\alpha_n^i\}_i$ 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 (25), 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 $M = |\{\alpha_{n-1}^j\}|$, the sets $\{\alpha_{p,c}^j\}$ 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. ■

9 Appendix B

9.1 Concavity of Mutual information in the Belief State.

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

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

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

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

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

Q.E.D. ■

10 Appendix C

Pseudocode

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

- Step 1: Build the simplex. Construct an 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 $V(g'_{c_j}(\cdot)) = 0$.
- Step 3: For each simplex point, find $p^*(c, w)$ s.t.

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

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

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

→ 180-320 iterations each taking 8min-20min

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

11 Appendix D

11.1 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 (20) in the main diagonal of the joint distribution to be chosen, $\Pr_t(c_j, w_i)$. Note that such a restriction is *WLOG*. I then show the analytical details of the derivatives with respect to control and states.

11.2 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 $\Delta_k v(l) \equiv v(l_{k_1}) - v(l_{k_2})$ and θ as the shadow cost of processing information. Then, the optimal control for the program (17)-(21) amounts to:

$$\begin{aligned} & \partial p^*(c_{k_1}, w_{k_2}) : \\ & \Delta_k u(c) + \beta \Delta_k V(g'_c(\cdot)) = p^*(c_{k_1}, w_{k_2}) \left(-\Delta_k u'(c) \theta p^*(w_{k_2}|c_{k_1}) - \beta \Delta_k V'_{p^*}(g'_c(\cdot)) \right) \quad (36) \end{aligned}$$

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 (36) is that the optimal probability of consumption and wealth depends on both levels of current and intertemporal utility and marginal utility. In

particular, there is an intertemporal trade-off by consuming the maximum value of wealth allowed by the signal, c_{k_2} and a lower consumption c_{k_1} . 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, $\beta \Delta_k V(g'_c(\cdot))$. 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 $\Delta_k V'_{p^*}(g'_c(\cdot))$. To this end, define the ratio between differential in utilities (current and discounted future) and differential in marginal current utility as $\Psi^\kappa \equiv \frac{\Delta_k u(c(\kappa)) + \beta \Delta_k V(g'_c(\cdot))}{-\theta[u'(c_{k_1}(\kappa)) - u'(c_{k_2}(\kappa))]}$. 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 $\Delta_k V(g'_c(\cdot)) = 1$ or, $\Phi^\kappa \equiv \frac{\beta}{-\theta[u'(c_{k_1}(\kappa)) - u'(c_{k_2}(\kappa))]}$. Then, an application of Chain rule and point-wise differentiation leads to

$$p^*(c_{k_1}, w_{k_2}) = \Lambda(k_1, k_2) p^*(c_{k_1}) \quad (37)$$

where

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

Let me focus on the explanation for the terms $\Lambda(k_1, k_2)$ which characterize the optimal solution of the conditional distribution $p^*(w_{k_2}|c_{k_1})$.

The first term $\Lambda_1(\Psi^\kappa, p^*(c_{k_1}, w_{k_2})) \equiv \exp(\Psi^\kappa / (p^*(c_{k_1}, w_{k_2})))$ 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 $\exp(\Psi^\kappa)$ is a likelihood ratio between utilities in the two states of the world (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 $p^*(c_{k_1}, w_{k_2})$. This matches the intuition because 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 $g'_{c_{k_i}}(\cdot)$, $i = 1, 2$. 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 $\tilde{\Delta} T_j \equiv T(\cdot; w_{k_2}, c_{k_j}) - \left(\sum_i T(\cdot; w_i, c_{k_1}) p^*(w_i|c_{k_j}) \right)$ for $j = 1, 2$. Evaluating the derivative with respect to the state *almost surely* reveals

that $\Lambda_2 \equiv \exp\left(-\left(\Phi^\kappa \Lambda'_1 \tilde{\Delta} T_1\right)/p(c_{k_1})\right)$ while $\Lambda_3 \equiv \exp\left(-\left(\Phi^\kappa \Lambda'_1 \tilde{\Delta} T_2/p(c_{k_2})\right)\right)$. The terms Λ_2 and Λ_3 reveal that in setting the optimal distribution $p^*(c_{k_1}, w_{k_2})$ 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 closed form solution is when the agent is risk neutral. Consider the framework in Section (3.2) and let utility take up the form $u(c) = c_t$, 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 $u(c) = w$, which implies $c|w \propto U(w_{\min}, w_{\max})$. 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 the continuation value, exploiting risk neutrality, the optimal policy function amounts to:

$$p^*(w_{k_2}|c_{k_1}) = \frac{e\left(\frac{[(c_{k_1}-c_{k_2})+\beta\Delta_k\bar{V}(g'_c(\cdot))]}{\theta}\right)}{\sum_j \tilde{\Delta} T_j} \quad (38)$$

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 because 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, the consumption profiles in the region $c < w$ are independent on the value of wealth.

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

$$\begin{aligned} \partial p^*(c_{k_1}, w_{k_2}) : \\ \Delta_k u(c(\kappa)) + \beta \Delta_k V(g'_c(\cdot)) = p^*(c_{k_1}, w_{k_2}) (-\theta \Delta_k u'(c(\kappa)) + \beta \Delta_k V_{p^*}(g'_c(\cdot))) \end{aligned} \quad (39)$$

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

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

where

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

Note that by Chain rule $\frac{\partial V'(g'_{c_{k_j}}(\cdot))}{\partial \Pr(c_{k_j}, w_{k_2})} = \frac{\partial V'(g'_{c_{k_j}}(\cdot))}{\partial (g'_{c_{k_j}}(\cdot))} \frac{\partial (g'_{c_{k_j}}(\cdot))}{\partial \Pr(c_{k_j}, w_{k_2})}$, for $j = 1, 2$. Plug (36)

in the second term of the above expression and evaluating point-wise the derivatives delivers

In $c_j = c_{k_1}$,

$$\begin{aligned} \Rightarrow \frac{\partial g(\cdot | c_{k_1})}{\partial \Pr(c_{k_1}, w_{k_2})} &= \frac{\partial \left[\frac{1}{p(c_{k_1})} \left(\sum_i T(\cdot; w_i, c_{k_1}) \Pr(w_i, c_{k_1}) \right) \right]}{\partial \Pr(c_{k_1}, w_{k_2})} = \\ &= \frac{1}{p(c_{k_1})} \left(T(\cdot; w_{k_2}, c_{k_1}) - \frac{\left(\sum_i T(\cdot; w_i, c_{k_1}) \Pr(w_i, c_{k_1}) \right)}{p(c_{k_1})} \right) \end{aligned}$$

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

$$\Pr(c_{k_1}, w_{k_2}) = \Lambda(k_1, k_2) \Pr(c_{k_1}) \quad (40)$$

where

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

while

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

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:

$$\begin{aligned} \frac{\partial V(g(w_k))}{\partial g(w_k)} &\stackrel{a.s}{=} \\ &\left(u(c_k(\kappa)) + \beta (V(g'_{c_k}(\cdot))) - \left(u(c_{l_{\max}}(\kappa)) + \beta V(g'_{c_{l_{\max}}}(\cdot)) \right) \right) + \\ &-\theta \left(\log \left(\frac{\Pr(c_k, w_k)}{p(c_k)g(w_k)} \right) u'(c_k(\kappa)) \Pr(c_k, w_k) - \log \left(\frac{\Pr(c_{l_{\max}}, w_k)}{p(c_{l_{\max}})g(w_k)} \right) u'(c_{l_{\max}}(\kappa)) \Pr(c_{l_{\max}}, w_k) \right) + \\ &+\beta \sum_j \left[\frac{\partial V(g'_{c_{k_j}}(\cdot))}{\partial (g'_{c_{k_j}}(\cdot))} \left(\frac{\partial (g'_{c_{k_j}}(\cdot))}{\partial g(w_k)} \right) \Pr(c_j, w_k) \right]. \end{aligned}$$

Combining first order conditions and the envelope condition after some algebra leads to the result in (40).

12 Appendix E.

12.1 A simple example

To illustrate how a consumer with information constraints differs from a consumer with full information and a consumer with no information, consider the following 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 cannot borrow, $c \leq w$ and, if his check bounces he suffers $c = 0$. He derives utility from consumption 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 $c = w$, $\forall c \in C, \forall w \in W$.

In contrast, if he cannot gather any information about wealth besides that provided by the prior, the consumer will avoid unpleasant surprises 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 . The ex-ante uncertainty embedded in the prior for w is calculated 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 given by C ($\mathcal{H}(W|C)$). In formulae, the mutual information or capacity of the channel amounts to:

$$I(C; W) = \sum_{w \in W} \sum_{c \in C} p(c, w) \log\left(\frac{p(c, w)}{p(c)p(w)}\right)$$

To see what this formula implies, consider first the situation in which information can flow at infinite rate. In this case ex-post uncertainty is 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 $\mathcal{H}(W|C) = 0$, so the mutual information in this case will be $I(C; W) = \mathcal{H}(W) = 1.5 \text{ bits}$.

Instead, if the consumer has zero information flow or, equivalently, if processing information is 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 $I(C; W) = 0$ and no reduction in the uncertainty about wealth occurs upon observing consumption. The intuition is that 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. Let $\bar{\kappa} = 0.3$ be the maximum amount of information flow that the

consumer can process. Let the probability matrix of the consumer be:

| $c \backslash w$ | $P(w = 2)$ | $P(w = 5)$ | $P(w = 8)$ |
|------------------|------------|-------------|-------------------|
| $P(c = 2)$ | 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 zeroes on the lower left corner of the matrix encode a non-borrowing constraint $c \leq w$.²⁹ The program of the consumer is:³⁰

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

s.t.

$$\bar{\kappa} \geq I(C; W).$$

Given $\bar{\kappa} = 0.3$,³¹ the optimal policy sets $p_1^* = 0.125$, $p_2^* = 0.125$, $p_3 = 0.125$, which corresponds to $\Pr(C = 2) = 0.75$, $\Pr(C = 4) = 0.25$, $\Pr(C = 6) = 0$. This leads to an expected utility of $E^\kappa(u(c)) = 0.87$. 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 information for the highest value to gain a more precise knowledge of the lower value of wealth is driven by the functional form of utility. For

²⁹I append a non-borrowing constraint $c \leq w$ to reduce the number of the probabilities to be calculated in Figure (b), thereby keeping the example easy. Figure (b) can be rationalized by assuming that the consumer acquires a signal on wealth $w^s = w + \varepsilon$ and chooses the distribution of ε always such that the support of ε is in $(0, -\infty]$. One can think that $\bar{\kappa}$ is net of the bits used to set the desired support of ε .

³⁰In details:

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

and

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

³¹Note that such a bound of information flow is unrealistically low. However I decided to trade off realism for simplicity in this example.

instance, a consumer with the same bound on processing capacity but CRRA utility with coefficient of risk aversion, say, $\gamma = 5$, would have chosen a probability $\Pr(C = 2)$ lower than his log-utility counterpart. This is because higher degrees of risk aversion induce the consumers to be better informed about low values of wealth to avoid such occurrences. The intuition is that because the attention of the consumer within the limits of the Shannon capacity is allocated according to his utility, the degree of risk aversion plays an important role in determining what events receive the consumer's attention. A log-utility consumer wants to be well informed about the middle values of his wealth, while a high risk averse consumer selects a signal which provides sharper information about the lower values of wealth, so that he can avoid high disutility. The opposite direction is taken by the less risk-averse agent.

12.2 Analytical Results for a three-point distribution

In this section I will focus on 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.³²

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:

| | | | |
|-----------------|-------|-------|-------|
| $C \setminus W$ | 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 |

³²A 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.

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:

$$\begin{aligned} x_1 &= g_1 \\ x_2 + x_4 &= g_2 \\ x_3 + x_5 + x_6 &= 1 - g_1 - g_2 \end{aligned}$$

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:

$\Pr(c_j, w_i) :$

| | | | | |
|-----------------|-------|-------------|---------------------------------|------|
| $C \setminus W$ | w_1 | w_2 | w_3 | |
| c_1 | g_1 | p_1 | p_2 | (41) |
| c_2 | 0 | $g_2 - p_1$ | p_3 | |
| c_3 | 0 | 0 | $1 - (g_1 + g_2) - (p_2 + p_3)$ | |

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

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

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

$$g'(\cdot|c_j) = \sum_i T(\cdot; w_i, c_j) \Pr(w_i|c_j). \quad (42)$$

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:

$$EW' = R(w_t - c_t) + \bar{Y}. \quad (43)$$

The approximation is necessary since (43) 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 (42) together with the specifications of $T(\cdot; w_i, c_j)$ and $\Pr(w_i, c_j)$, I can explicitly evaluate $g'(\cdot|_{c_j})$ *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 $c_j = c_1$,

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

In $c_j = c_2$

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

In $c_j = c_3$

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

Then, the first order conditions and envelope conditions amount to

∂p_1 :

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

Note that $\frac{\partial g'_{c_j}(\cdot)}{\partial p_j} = 0$ for $j \in \{1, 2, 3\}$.³³ This result is not driven by the specification chosen for the transition function $T(\cdot; w_i, c_j)$, but it is a feature of the three point

³³To see this, plug (42) in $\frac{\partial g'_{c_j}(\cdot)}{\partial p_j}$ for $j \in \{1, 2\}$ and evaluating pointwise the derivatives delivers

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

$$\frac{1}{(g_1 + p_1 + p_2)^2} \begin{bmatrix} 0.81p_2 - 0.15g_1 \\ -(0.56p_2 - 0.15g_1) \\ -0.25p_2 \end{bmatrix} = 0$$

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

$$\frac{p_3}{(g_2 - p_1 + p_3)^2} \begin{bmatrix} -0.15 \\ 0.15 \\ 0 \end{bmatrix} = 0$$

distribution. Indeed, because 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 $V' \left(g'_{c_j}(\cdot) \right)$ through the transition function and, as a result, the update for the marginal $g'_{c_j}(\cdot)$ according to (42). That is, the marginal probability on wealth $g'_{c_j}(\cdot)$ in this case tends to its ergodic value $\bar{g}_{c_j}(\cdot)$. It follows that $V' \left(\bar{g}_{c_j}(\cdot) \right) \xrightarrow{a.s.} \bar{V}^* \left(\bar{g}_{c_j}(\cdot) \right)$ 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:

$\partial p_1 :$

$$\begin{aligned} & [u(c_1(\kappa)) - u(c_2(\kappa)) + \beta (\bar{V}(\bar{g}_{c_1}(\cdot)) - \bar{V}(\bar{g}_{c_2}(\cdot)))] \\ & = p_1 \left(\theta ([u'(c_1(\kappa)) - u'(c_2(\kappa))]) \ln \left(\frac{p_1}{(g_1 + p_1 + p_2)} \right) \right) \end{aligned} \quad (44)$$

Similarly, for the second control

$\partial p_2 :$

$$\begin{aligned} & [u(c_1(\kappa)) - u(c_3(\kappa)) + \beta (\bar{V}(\bar{g}_{c_1}(\cdot)) - \bar{V}(\bar{g}_{c_3}(\cdot)))] \\ & = p_2 \left(\theta ([u'(c_1(\kappa)) - u'(c_3(\kappa))]) \ln \left(\frac{p_2}{(g_1 + p_1 + p_2)} \right) \right) \end{aligned} \quad (45)$$

And finally:

$\partial p_3 :$

$$\begin{aligned} & [u(c_2(\kappa)) - u(c_3(\kappa)) + \beta (\bar{V}(\bar{g}_{c_2}(\cdot)) - \bar{V}(\bar{g}_{c_3}(\cdot)))] \\ & = p_3 \left(\theta (u'(c_2(\kappa)) - u'(c_3(\kappa))) \ln \left(\frac{p_3}{(g_2 - p_1 + p_3)} \right) \right) \end{aligned} \quad (46)$$

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

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

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:

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

where $\{\alpha'_j\}_j$ are a set of vectors each of them generated for a particular observation of previous values of consumption c_j and $\langle \cdot, \cdot \rangle$ denotes the inner product $\langle \alpha'_j, \bar{g}'_{c_j}(\cdot) \rangle \equiv \sum_{w' \in \Omega_w} \alpha'_j(w') T(\cdot : w, c_j) \cdot p(c_j|w)$. 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

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

where $\tilde{\delta}(w - w_i) = \text{Dirac}(w - w_i)$ is the *Dirac* delta function with the 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:

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

The central issue in the particle filter approach is how to obtain a set of particles to approximate $\bar{g}'(\cdot | c_j)$ from the set of particles approximating $g(w)$. The usual Sampling Importance Re-sampling (SIR) approach (Dellaert et al., 1999; Isard and Blake, 1998) samples particles using the motion model $T(\cdot; w_i, c_j)$, 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 $\Pr(c_j | \cdot)$ is too peaked or when there is a small overlap between prior and posterior likelihood. The 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 addressed by inserting the likelihood inside the mixture

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

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

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

with μ_ω^i any likely value associated with the i^{th} component of the transition density $T(\cdot; w_i, c_j)$, e.g., its mean. In this case, we have that $\mu_\omega^i = w_i + \Delta(c_j)$. Then, $\bar{g}'(\cdot | c_j)$ can be regarded as a mixture of N transition components $T(\cdot; w_i, c_j)$ with weights $\varpi_i \Pr(c_j | \mu_\omega^i)$. Therefore, sampling a new particle w'_j to approximate $\bar{g}'(\cdot | c_j)$ can be carried out by

selecting one of the N components, say i_m , with probability $\varpi_i \cdot \Pr(c_j|\mu_\omega^i)$ and then sampling w'_i from the corresponding component $T(\cdot; w_{i_m}, c_j)$. 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

$$\varpi'_m \propto \frac{\Pr(c_j|w'_m)}{\Pr(c_j|\mu_\omega^{i_m})}.$$

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

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

where $\tau(\cdot)$ 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 $\sigma_\omega^2 \equiv \delta^2(\pi_3 - \pi_1) - (\mu_\omega - \omega_2)^2$ 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 $V'(\vec{g}'_{c_j}(\cdot)) = \arg \max_{\{\alpha'_j\}_j} \langle \alpha'_j, \vec{g}'_{c_j}(\cdot) \rangle = \sum_{k,l} \tilde{\varpi}'_k \tilde{\varpi}'_l \tau(w_l|w_k, \Sigma_k)$, where $\tilde{\varpi}'_k \equiv \left(\frac{\Pr(c_j|w'_k)}{\Pr(c_j|\mu_\omega^k)} \right)$, $\tilde{\varpi}'_l \equiv \left(\frac{\Pr(c_j|w'_l)}{\Pr(c_j|\mu_\omega^l)} \right)$. Since the result of the arg max is just one of the member of the set $\{\alpha'_j\}_j$ 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:

$$p_1^*(\vec{g}, \theta) = \frac{g_1(\psi_1 - \theta\beta\nu_1)}{\theta g_1(\text{LambertW}(\chi_1) x_{12} - \text{LambertW}(\chi_{11}) x_{11}) + 2g_1(\psi_1 - \theta\beta\nu_1)}; \quad (47)$$

$$p_2^*(\vec{g}, \theta) = \frac{g_1(\psi_2 - \theta\beta\nu_2)}{\theta g_1(\text{LambertW}(\chi_2) x_{21} - \text{LambertW}(\chi_2) x_{22}) + 2\psi_2 g_1(\psi_2 - \theta\beta\nu_3)}; \quad (48)$$

$$p_3^*(\vec{g}) = \frac{\psi_3 - \theta\beta v_3}{\theta x_3 \text{LambertW}(\chi_3)} \quad (49)$$

where

- $\psi_1 \equiv \left(\frac{e^{-\gamma(c_2 - \theta\kappa)} e^{-\gamma(c_1 - \theta\kappa)}}{\gamma} \right); \psi_2 \equiv \left(\frac{e^{-\gamma(c_3 - \theta\kappa)} e^{-\gamma(c_1 - \theta\kappa)}}{\gamma} \right); \psi_3 \equiv \left(\frac{e^{-\gamma(c_3 - \theta\kappa)} e^{-\gamma(c_2 - \theta\kappa)}}{\gamma} \right);$
- $\nu_1 \equiv g_2 (\psi'_3) + (g_2 - g_1) (\psi'_2 - \psi'_3);$
- $\nu_2 \equiv g_2 (\psi'_1) + (g_2 - g_1) (\psi'_1 - \psi'_2);$
- $\nu_3 \equiv (1 - g_2 - g_1) (\psi'_2) + (g_2 - g_1) (\psi'_3 - \psi'_1)$
- $\chi_1 \equiv \frac{(\psi_1 - \theta\beta v_1)\psi_1}{\theta g_1 (e^{\gamma(c_2 - c_1)})}; x_{11} \equiv e^{-\gamma(c_1 - \theta\kappa)}, x_{12} \equiv e^{-\gamma(c_2 - \theta\kappa)};$
- $\chi_{21} \equiv \frac{(\psi_2 - \theta\beta v_2)\psi_2}{\theta g_1 (e^{\gamma(c_3 - c_1)})}, x_{21} \equiv e^{-\gamma(c_1 - \theta\kappa)}, x_{22} \equiv e^{-\gamma(c_3 - \theta\kappa)}$ and
- $\chi_3 \equiv \frac{\psi_3 - \theta\beta v_3}{\theta g_2 (e^{\gamma(c_3 - c_2)})}, x_3 \equiv e^{\gamma(c_3 - c_2)}.$

and $\text{LambertW}(\cdot)$ is the LambertW function that satisfies $\text{LambertW}(x) e^{\text{LambertW}(x)} = x$ ³⁴. 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 $\frac{\partial \text{LambertW}(x)}{\partial x} = \frac{\text{LambertW}(x)}{x(1 + \text{LambertW}(x))}$ 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 rationale 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 (18) 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

³⁴Formally, the LambertW function is the inverse of the function $f : \mathbb{C} \rightarrow \mathbb{C}$ given by $f(x) \equiv xe^x$. Hence $\text{LambertW}(x)$ is the complex function that satisfies

$$\text{LambertW}(x) e^{\text{LambertW}(x)} = x$$

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

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

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

$p_1 = \Pr(c_1, w_2)$, $p_2 = \Pr(c_1, w_3)$ 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 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 .

13 Appendix F:

13.1 The Mathematics of Rational Inattention

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

Formally, the starting point is a set of possible events whose probabilities of occurrence are p_1, p_2, \dots, 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, \dots, p_n . If x is a chance variable, then $H(x)$ indicates its entropy; thus x is not an argument of a function but a label for a number, to differentiate it from $H(y)$ say, the entropy of the chance variable y .

Quantities of the form $H = -\sum_i p_i \log p_i$ play a central role in Information Theory as measures of information, choice and uncertainty. The quantity H goes by the name of entropy³⁵ and p_i is the probability of a system being in cell i of its phase space.

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

³⁵See, for example, R. C. Tolman, *Principles of Statistical Mechanics*, Oxford, Clarendon, 1938.

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 to account for the change in unit of measurement.

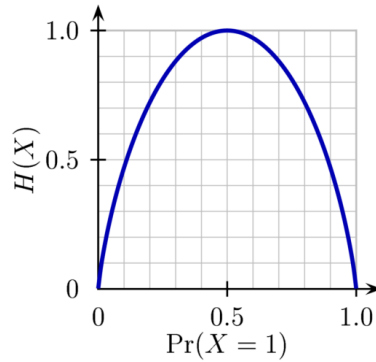


Figure a: Entropy of two choices with probability p and $q=1-p$ as function of p .

Remark 1. $\mathcal{H} = 0$ if and only if all the p_i but one are zero, with the one remaining 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, \dots, 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} increases³⁶.

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

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

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

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

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

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

or

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

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

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

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

³⁶The only case in which \mathcal{H} remains unchanged is when the transformation results in just one permutation of p_j .

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, \dots, 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 sequences will be possible signals for the channel. Given a channel, one may be interested in measuring its capacity to transmit information. In general, with different lengths of symbols and constraints on the allowed sequences, the capacity of the channel is defined as:

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

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

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

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

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

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

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

In formulae:

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

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

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

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

where the maximum is with respect to all possible information sources used as input to the channel (i.e., the probability distribution of Y , $p(Y)$). If the channel is noiseless, $E(\mathcal{H}_y(x)) = E(\mathcal{H}(X(|Y))) = 0$. For example, think about a newspaper editor who wants to maximize his sales. To do that, he has to choose the allocation of space for his articles in such a way that it is attractive for the consumers. In this example, Y is the random variable *space*, X the random variable *sales*, the channel's capacity is the maximum number of pages in the newspaper and the channel itself is the best articles' allocation of space which signals that the journal is worth buying.