The Life-Cycle Effects of Information-Processing Constraints [1]

Kurt F. Lewis *

Board of Governors of the Federal Reserve System, Washington, DC 20551, USA

Abstract

This paper extends the rational inattention framework of Sims (2006) to a finite-horizon dynamic setting. This is accomplished by creating a structure in which an agent faced with information-processing constraints necessarily views states as distributions and makes approximate, distributional choices for controls. In the model, limited information processing capacity is used optimally, and agents have the opportunity to trade processing capacity for higher expected future income. The framework is applied to the canonical life-cycle model of consumption and saving, and an analysis is conducted of the impact of preference parameters on optimal attention allocation is conducted. The model produces a distinct hump-shaped profile in aggregate consumption.

Key words: Rational Inattention, Information-Processing Constraints, Life-Cycle, Consumption Expenditures Path.

JEL Codes: E21, D83, D91

* Board of Governors of the Federal Reserve System, 20th Street and Constitution Avenue, NW, Washington, DC 20551. *Ph.* (202) 452-3424. *Fax.* (202) 452-2301.

Email address: `kurt.f.lewis@frb.gov` (Kurt F. Lewis).

[1] I would like to thank Kurt Anstreicher, John Geweke, Ryan Haley, Beth Ingram, Ricardo Reis, Gene Savin, Chris Sims, Antonella Tutino, Todd Walker, anonymous referees, and the faculty and seminar participants at the Board of Governors of the Federal Reserve, the University of Iowa, Kansas State University and Southern Methodist University. Special thanks to Charles Whiteman for guidance and many helpful discussions. All remaining errors are my own. The views expressed in this paper are solely those of the author and do not necessarily represent the views of the Federal Reserve Board, its members or its staff.

Date and Version Information *27 May 2010, Ver. 5.1*

1 Introduction

Mounting evidence suggests that the hump-shaped profile of aggregate consumption over the life-cycle is a consequence of individual choice, and not an artifact of other factors such as growth in family size or accumulation of durable goods or even general economic growth during an agent's life [Aguiar and Hurst (2005, 2007), Fernández-Villaverde and Krueger (2004), Gourinchas and Parker (2002)]. The hump-shape runs counter to intuition regarding the smoothing behavior of economic agents: one would expect consumers to save and dissave so as to keep consumption relatively constant from year to year. Recent work in Aguiar and Hurst (2005, 2007) points out that actual consumption levels may not fluctuate as much as consumption expenditure, thus explaining the hump-shaped profile in aggregate consumption expenditure as a shape aided by shopping, rather than a pure consumption pattern alone. This paper proposes a complimentary hypothesis that a weakness in the relationship between wealth and consumption expenditure could help to create such a profile in the aggregate consumption patterns of US consumers. Credit-based shopping allows for the potential of much more loosely accounted-for shopping practices and lends itself to approximate (from an expenditure standpoint) shopping behavior. In particular, the hump shape arises in a framework that includes uncertainty and information-processing constraints popularized recently in the Rational Inattention (RI) paradigm of Sims (1998, 2003, 2006). Moreover, circumstances arise under which there is a nontrivial probability that the representative agent fails to consume all his wealth prior to death despite knowing its timing precisely.[2]

The simple life-cycle setting developed here uses much of the same technology of the Sims (2006) two-period model, but with the addition of a trade-off allowing agents to give up processing power for something of economic value. This paper focuses on the dynamics of the information problem and the value of information-processing capacity, and demonstrates anew that simple applications of uncertainty and information-processing constraints can produce substantial changes in otherwise standard models.

Many papers in the life-cycle literature share the presumption that uncertainty regarding the agent's environment likely plays a role in the profile of consumption over time [e.g. Caballero (1990), Carroll (1992, 2004), Deaton (1992), Luo (2004), Sims (2003)][3]. How the agent confronts this uncertainty is at the center of several attempts to resolve the consumption puzzle. Most of these (excluding those of Sims and Luo) assume that the agent is aware of

[2] This model studies aggregate consumption profiles, the agents discussed throughout the paper are the representative agent in the model. An area of future research includes heterogeneous agents (that is, agents with differing information-processing abilities or other characteristics that would yield different optimal uses of information capacity). Here, the terms agent and representative agent will be used interchangeably.

[3] A recent addition to the literature, Tutino (2008), examines lifetime consumption and savings within the context of a recursive RI model. This paper, and the device of finite time periods combined with the aggregation across individual representative households generates the closest analog to a RI version of the canonical life-cycle problem presented in section 3 because the solution procedure can be one of solving the whole problem at the same time.

all information in the system at all times, even in cases where "all the information" includes knowing complicated distributions over many variables. This represents a large abstraction from reality that is accepted in the name of model tractability. On the other hand, as demonstrated in Sims (2003), information-processing constraints produce results which look more like observed data; a similar result characterizes this extension.

Another contribution of this paper is that it is the first to offer the representative agent an opportunity exercise choice regarding the stock of his information-processing capacity. This capacity decision is separate from how to optimally allocate the processing capacity that is being utilized. We see that the agent chooses to forgo some of his processing capacity in exchange for potential future income, and that the choice is not binary. That is, he chooses to forgo some, but not all, in particular situations suggesting that there is a value to information processing that could potentially be used to price certain information services in future research.

The remainder of the paper is structured as follows: Section 2 contains a discussion of the rational inattention problem applied specifically in this model. Section 3 describes the life-cycle consumption problem and a brief description of its history. Section 4 lays out the specifics of the model, while section 5 provides an analysis of the model results and discusses the optimal allocation of information processing that leads to the hump-shaped behavior. Section 6 concludes.

2 Rational Inattention and Approximate Decision-Making

The idea behind rational inattention can traced back at least as far as a 1978 address to the AEA meetings by Herbert Simon, who titled a portion of his talk "Attention As The Scarce Resource" (Simon, 1978, p. 13), and to work that Sims was conducting much earlier than Sims (1998, 2003, 2006). Moreover, the idea is intuitive: as Sims points out, "...modeling agents as finite-capacity channels...fits well with intuition; most people every day encounter, or could very easily encounter, much more information that is in principle relevant to their economic behavior than they actually respond to" (Sims, 2006, p. 2). What rational inattention provides is a comprehensive framework for relaxing the assumption of unlimited information-processing ability without abandoning the assumption that resource are used optimally, and without introducing arbitrary frictions.

Rational inattention is not the limiting the information used by an agent in his or her decision-making. Neither does it correspond to delaying or disguising the information either, as with most information frictions. The central theme of rational inattention is one of optimal choice regarding how to reduce uncertainty. Agents make decisions that affect how the uncertainty in their world is reduced by choosing what to pay attention to: the agent has *all the current information* at his or her disposal, but chooses optimally to not pay attention some if it—thus the monicker.[4]

[4] An example of a recent, fully developed paradigm for information frictions is the "inattentiveness"

2.1 A Generalization

The amount of information-processing required to solve economic problems depends on the complexity of the problem, but even the simplest problems have processing requirements. To see how information-processing constrained models differ from their unconstrained counterparts, consider the two-period model of Sims (2006). This is an undiscounted "cake-eating" problem in which the agent takes a given amount of wealth, w, and divides it optimally between consuming c in period one and $w-c$ in period two. That is, for CRRA preferences,

$$\max_{c \leq w} \frac{c^{1-\gamma} + (w-c)^{1-\gamma}}{1-\gamma},$$

for a given w. The solution to this problem is an optimal decision rule, denoted f, that describes the optimal plan for the choice variable, c, given a value for the state variable, w. That is, the solution is a one-to-one mapping from the state-space to the choice-space, described by $c^* = f(w)$. The solution to the agent's maximization problem here is given by:

$$c^* = f(w) = \frac{w}{2},$$

that is, the agent should consume half his wealth in each of the two periods. For a given value of w, this describes a corresponding value for c. Even if wealth is characterized by a probability distribution, the optimal f describes a mapping from each potential value of w to a single corresponding value for c.

To set the stage for the information-constrained problem to come, consider a generalization of the cake eating problem in which the cake (wealth) and bites of the cake (consumption) only come in a finite set of discrete values $c_1, c_2, \ldots, c_{N_c}$ and $w_1, w_2, \ldots, w_{N_w}$. Suppose further that wealth is characterized by a probability distribution $g(w)$. The decision rule, $c^* = f(w) = w/2$, becomes the method for generating a set of conditional distributions – one for each wealth value. Each of these conditional distributions for consumption is degenerate, that is, the joint distribution $f(c, w)$ describes the same thing as the $c^* = f(w) = w/2$: a one-to-one mapping from state-space to choice-space.

Under this "generalization," the agent's optimization problem is to choose the joint distribution $f(c, w)$ to:

$$\max_{\{f(c_i,w_j)\}} \sum_{i=1}^{N_c} \sum_{j=1}^{N_w} \frac{c_i^{1-\gamma} + (w_j - c_i)^{1-\gamma}}{1-\gamma} f(c_i, w_j) \tag{1}$$

of Reis (2006a,b), wherein price-setting producers and consumers update their information only occasionally, but *completely*. Thus while inattention is the optimal response to a finite capacity to process, inattentiveness results from an inability to acquire information frequently.

subject to:

$$\sum_{i=1}^{N_c} f(c_i, w_j) = g(w_j) \qquad \forall \ j = 1, \ldots, N_w \tag{2}$$

$$f(c_i, w_j) \in [0, 1] \qquad \forall \ i, j \tag{3}$$

$$f(c_i, w_j) = 0 \qquad \text{for } c_i > w_j. \tag{4}$$

The properties of the problem and the optimum are qualitatively unchanged under this generalization. Suppose that the marginal distribution of wealth is triangular, with higher levels of wealth more probable. The optimal decision rule is the joint distribution $f(c, w)$ that describes the same one-to-one mapping that divides wealth into two halves and consumes one in each time period. Under the generalization, however, this is accomplished by assigning probability to specific (c_i, w_j) pairs. That is, given a distribution for wealth, the agent disperses the probability weight $g(w_j)$ across the possible values $\{c_i\}_{i=1}^{N_c}$ such that weight is only allowed where $c_i \leq w_j$. The optimal choice, as seen in figure 1, is to place all of the probability of being at wealth node w_j on the pair $(c_i = w_j/2, w_j)$, that is, $f(c_i = w_j/2, w_j) = g(w_j)$.

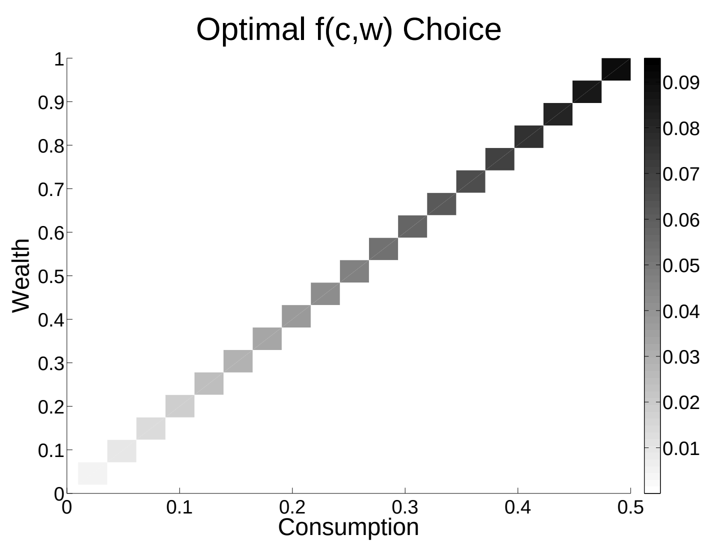


Fig. 1. A One-to-One Mapping via the Joint Distribution $f(c, w)$.

Figure 1 represents the joint distribution of c and w over the $[0, 1]$ interval when the (c, w) space is discretized. The darkness of the boxes indicates the weight of probability on that specific (c, w) pair. The darker the box, the higher the probability weight, as indicated by the legend on the right-hand side of the figure. The boxes get darker as they progress "northeast" because the (exogenous) marginal distribution of wealth, $g(w)$, is triangular. The solution, $f(c_i = w_j/2, w_j) = g(w_j)$, demonstrates that within this generalization the one-to-one mapping takes the form of creating a set of conditional distributions of c given w that are degenerate at $c_i = w_j/2$.

2.1.1 *States, Decisions, and Constrained Mutual Information*

The joint distribution of c and w in figure 1 shows a strong relationship between the variables; a one-to-one relationship to be specific. The tools of the rational inattention framework can be used to create an approximate (one-to-many) decision-making paradigm by limiting the strength of the relationship between state and decision variables through the concept of mutual information. The key to this is the constrained *mutual information*; a measurement of the amount of uncertainty about one variable that is reduced by observation of another. In each period we will constrain the amount of mutual information agents can process between consumption and wealth. As this paper will be working in discrete distributions, we demonstrate the mutual information of two discrete random variables x and y having joint distribution f, identical N-point support, and marginal distributions p and g by

$$MI(x,y) = \sum_{j=1}^{N}\sum_{i=1}^{N} f(x_i,y_j)\cdot\log[f(x_i,y_j)] - \sum_{i=1}^{N} p(x_i)\cdot\log[p(x_i)] - \sum_{j=1}^{N} g(y_j)\cdot\log[g(y_j)]. \quad (5)$$

The value of $MI(x,y)$ is equal to the sum of three components: beginning at the far right, the entropy of the marginal distribution of x, the entropy of the marginal distribution of y, and the entropy of the joint distribution of x and y.[5] Mutual information is a generalization of correlation in that as the conditional distributions of x given y (or y given x) tighten around a single x (or y), the amount of mutual information increases. Note that if x and y were independent, then $f(x_i,y_j) = p(x_i)g(y_j)$, and it is easy to see that $MI(x,y) = 0$. That is, observing a random variable x independent of y provides no information about y and vice-versa. Indeed, as the set of conditional distributions becomes degenerate, the amount of mutual information increases toward a maximum. Mutual information is as large as it can be in figure 1 because for each draw of w, a *specific* c value is determined, and vice versa.

2.1.2 *Mutual Information and Information-Processing Capacity*

The heart of the rational inattention framework is the idea that agents are incapable of processing all the information related to their economic decisions. In models in which wealth is the state variable, if the agent does not *exactly* know his wealth in each period, he is modeled as having a distribution over wealth which stems from a noisy signal he has received. In the rational inattention framework, the agent is also modeled as choosing the joint distribution of states and decisions, in this case wealth and consumption.[6] This can be thought of as stemming from the amount of attention that must be paid to prices in order to purchase a specific value of consumption, rather than a specific bundle of goods. An exact level of

[5] It should be noted that while $\log(0)$ is undefined, $x\log(x)$ is a continuous function on $x \in [0,\infty)$ and it can be demonstrated via L'Hôpital's Rule that $\lim_{x\to 0} x\log(x) = 0$.

[6] Because the wealth distribution is fixed for a given time period, what is being chosen is actually a set of conditional distributions of consumption given wealth.

consumption spending *can* be met, but in order to do so a large investment of information-processing capacity must be made: prices must be compared and calculated, expenditures within the time period must be accounted for precisely and updated continually. All of this is possible, but it requires processing many small pieces of information. Processing this information tightens the conditional distributions of consumption given wealth, moving the agent closer to a one-to-one mapping between consumption and wealth. [7]

The mutual information constraint is literally limiting the amount of information that can be known about wealth by observing consumption (and vice-versa). This does not need to limit the interpretation of the mutual information constraint. Rather than thinking about this strictly from the point of view that an observation about consumption informs the agent about his wealth, think about mutual information as being an element in a cohesive model of "close-enough" decision-making. The agent is aware of his wealth, but only approximately, and he is aware of his consumption, also approximately. The "tightness" of the mapping between states and decisions is determined by the mutual information constraint. That is, the mutual information constraint controls how approximately a decision is made in this model, in addition to literally describing how much wealth-information can be discovered from consumption.

The rational inattention framework uses the metric of mutual information to quantify the amount of information-processing capacity the agent is using to solve his optimization problem. By placing a constraint on mutual information, the framework limits the strength of the relationship between c and w by limiting the precision with which either variable can be understood by the agent. As the amount of information the agent can process is reduced from the amount required to produce the one-to-one relationship described in figure 1, the agent must decide how best to allocate the finite resource of processing capacity across the space of his choice variable. An important result of Shannon (1948) should be noted: the information-processing capacity constraint will always bind when the amount of mutual information allowed is less than or equal to the amount of mutual information required for an optimal unconstrained mapping. The agent will use all available capacity to process the information in his environment, thus capacity equals the amount of information processed. [8]

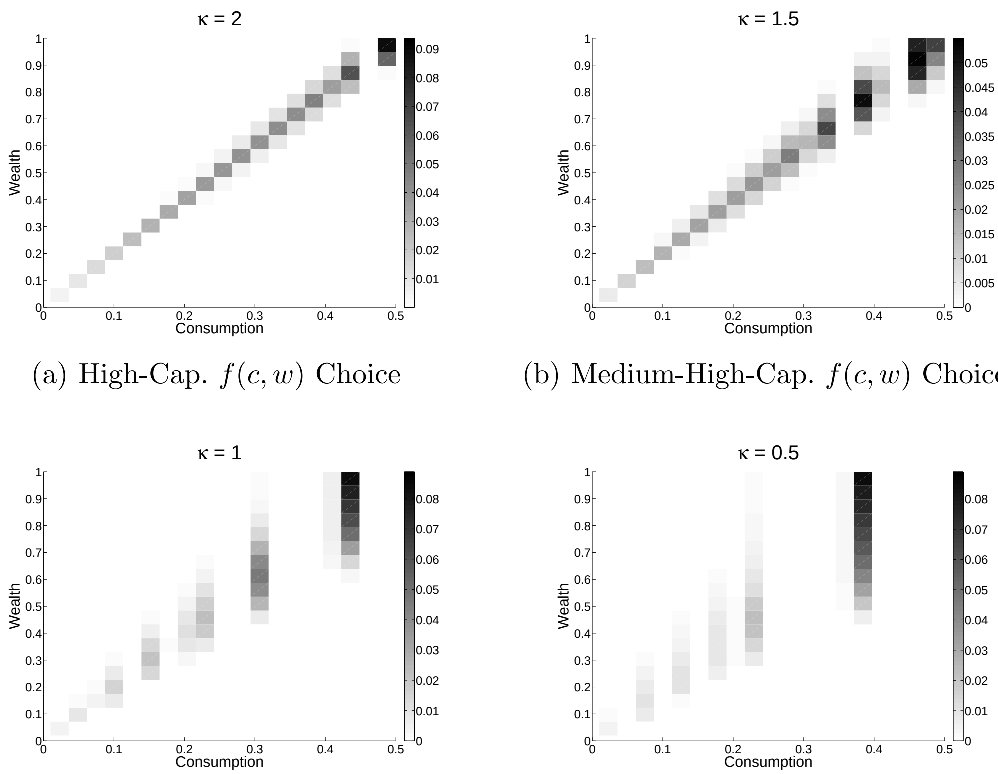
[7] The information-processing requirement of such small decisions as an aggregate consumption-savings decision could feel to the reader as an overly-complicated apparatus for such a small, well-understood problem. Part of the parsimonious nature of the RI paradigm is not displayed in the current literature because the tools are still being developed. Using concepts such as total correlation (a multi-variate mutual information analog), the author hopes in the future to examine much more complicated problems including multiple dimensions of decision-making for an agent, without changing the basic setup of the model. The agent will still draw upon the single pool of attention-resources to solve all his problems.

[8] The problem faced by agents in an RI model is a close analog to what information theorists call the "optimal coding" problem. This problem is one of fitting a message of size X through a channel with capacity $Y < X$. It is shown in the information theory literature that there is an optimal way to code any message sent through such a channel such that it will fill the channel's capacity and represent the message with the minimum loss. The agent's task is to choose the optimal way to filter the larger amount of information he or she has been presented with into a small amount of

The agent's optimization problem in the information-processing constrained universe is the same as the one detailed in equations (1) through (4) above, with the addition of the following constraint on the amount of mutual information in the model:

$$MI(c,w) \leq \kappa \tag{6}$$

As the amount of information-processing capacity (κ) decreases, the effect on the agent is similar to that of increasing the noise in a signal-extraction version of the same problem. In the past, economic models have tried to explain the difference between theory and empirical observation in many models by assuming the existence of an exogenous noise that complicates the understanding of the state of the model. The rational inattention framework does something similar to this by describing an environment in which the "noise" is endogenously determined rather than exogenously given: it arises from the agent's inability to accurately assess the state of the model because he does not have the information-processing resources to do so.



(a) High-Cap. $f(c,w)$ Choice

(b) Medium-High-Cap. $f(c,w)$ Choice

(c) Medium-Low-Cap. $f(c,w)$ Choice

(d) Low-Cap. $f(c,w)$ Choice

Fig. 2. Optimal Choice of Joint Distribution $f(c,w)$ Given Triangular $g(w)$ for Various Levels of Information-Processing Capacity.

"encoded" information, that gets across the *general* message of the data without every last detail. This gives a sense of how rational inattention can be leveraged to create a framework for modeling agents who are making decisions under a "that's-close-enough" style of thinking.

Figure 2 demonstrates the effect of lowering the information-processing capacity: a discretization of the consumption choices by the agent. In panel 2(c), the agent chooses to place no probability weight on consuming at the level of, for example, 0.275. The agent's noisy signal for wealth generates consumption behavior that does not as closely track potential differences in wealth as the information-processing *unconstrained* model would. That is, there is a range of possible wealth levels that could result in a level of consumption, and also a range of possible consumption levels that correspond to a given potential wealth level. This is a many-to-many mapping rather than a one-to-one mapping, but still describes how the agent chooses consumption based on both his information regarding wealth and his preferences.[9]

3 The Life-Cycle Model

The canonical life-cycle model has no growth in income or household size, no borrowing, no shocks to the income stream or preferences, and a certain ending period. The agent has a constant income, no initial savings, discounts the future geometrically, and has no uncertainty regarding the future. That is, a consumer with period utility of consumption, c_t, given by $U(c_t)$, discount rate β, initial wealth W_0, period income of y_t, savings of S_t, and facing a constant interest rate $(1+r)$, chooses $c_1, c_2, \dots$ to

$$\max_{\{c_t\}_{t=1}^{T}} \quad \sum_{t=1}^{T} \beta^t U(c_t)$$

subject to:

$$\begin{aligned} W_{t+1} &= (1+r)W_t + y_{t+1} - c_{t+1}, \qquad && t = 1, \dots, T-1 \\ S_t &\geq 0, && t = 1, \dots, T \end{aligned}$$

This "stripped-down" life-cycle model [demonstrated in Modigliani (1986), and dating back to the mid-50's] is a textbook staple and produces the well-known result in figure 3,[10] of

[9] Recall that the optimality of the choice of $f(c, w)$ is a function of the preferences of the agent. The MI constraint would bind for many different specifications of f. For example, as κ decreased, the agent could simply "blur" the diagonal line to weaken the relationship between consumption and wealth until $MI \leq \kappa$. But, the shape of the distribution is the result of the agent's preference for a "lumpy" marginal distribution of consumption. Note that the bottom-right panel of figure 2 indicates that the marginal distribution of consumption places most of its weight on a single value a little below the optimal consumption value that would correspond with the mean of the triangular wealth distribution in the unconstrained model.

[10] This is a picture of the result for $\beta(1+r) = 1$. Additionally, U needs to be monotone increasing and concave, in this case, it is CRRA. In general, the path of consumption will increase or decrease

consuming a fixed fraction of life-time wealth in each period.

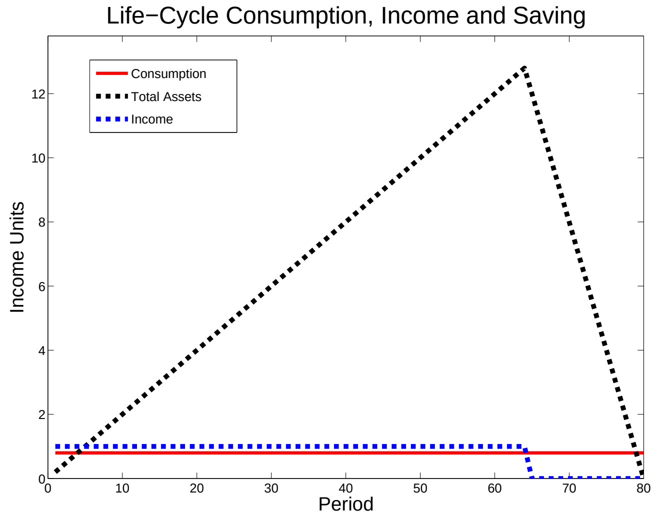


Fig. 3. The Stripped Down Life-Cycle Model

What is observed in the data is clearly different from figure 3; as one recent study put it: "...total consumption expenditures, as well as expenditures for non-durables and durables display a significant hump over the life cycle, even after accounting for changes in family size," [(Fernández-Villaverde and Krueger, 2004, pg. 2)].

The framework presented here can produce the observed hump-shape without the use of income growth, changes in household size, non-geometric discounting, adjustment costs, or many of the other common mechanisms imposed to generate these results. With very few free parameters, the uncertainty induced through approximation constrained by mutual information can qualitatively replicate the observed hump. The only changes to the canonical model are the introduction of uncertainty to income and wealth, and the approximate mechanism by which agents can choose how and, in this model: *if and when*, they would like to work to resolve this uncertainty.

4 A Life-Cycle Model with Information-Processing Constraints

What is the value of a unit of information-processing capacity? I address this question in two ways.

First, it is known from Shannon (1948) that an information channel of any capacity can be optimally used by any signal.[11] Thus, no matter the amount of channel capacity an agent has, the agent will fill that amount regardless of how "complex" their decision is. Thus, an

depending on the value of βR; see e.g. Yaari (1964).

[11] See footnote 8 on page 7 concerning information theory's "optimal coding problem."

additional tool must be used to evaluate how much it is worth to the agent to give up some of that capacity.

Second, a large number of services exist to collate, cull, and organize information for the purposes of presenting the clearest picture of any topic to an end-user and customers pay for this service. How much is it worth? Here, within a consumption-expenditure model, there is some strong similarity with the Aguiar and Hurst (2005, 2007) shopping-time literature. Information-processing can be thought of as encompassing a broad range of activities or thought processes related to the consumption expenditure decision, rather than a specific measure like shopping time. For example, it is possible to purchase groceries over the internet and have them delivered to a home. While Aguiar and Hurst (2007) demonstrate that there is a tangible return to time spent shopping for goods in grocery stores, internet services such as this reduce the amount of time shopping without limiting the amount of information gained, because the information has been pre-processed and organized in a more efficient manner. One can compare across stores without leaving their desk, compare final costs of bundles without visiting multiple stores, etc., *if they have paid to access this information-processing service.*[12] One possible way to do this is to consider an environment in which the agent has the opportunity to divide his time in the current period between two activities: time devoted to processing information related to the current consumption-wealth decision, and time devoted to increasing future income.

The division of time spent in the current period on the two activities is represented by the parameter, $\alpha_t \in [0, 1]$, which is chosen by the agent. Current-period information processing activities can be thought of as time spent balancing one's checkbook or checking one's debit card balance, as well as clipping coupons, checking internet sites for sales, comparison shopping within and across similar stores, determining where the lowest gas prices are locally, etc. The maximum fraction of time that can be spent in period t is $\alpha_t = 1$. The formula for period t processing capacity is written:

$$\kappa_t = \alpha_t \kappa^M, \quad t = 1, \ldots, T \tag{7}$$

where κ^M is the maximum information-processing capacity of the agent. As α_t increases, the agent spends more time processing information.

The other side of the tradeoff is the expected income of the agent in the following period. One can think about the intuition of this side of the tradeoff in the following way: this is time spent working overtime, or time spent reorganizing an investment portfolio, or time spent playing golf with the boss, or anything that is likely to increase the expected value of next period's income. Here, income is modeled in a fairly simplistic way: period income arrives as a distribution over K nodes that are fixed for the entire lifetime. There is no

[12] Purchasing the service described here boils down to having internet access in most cases, but there exist levels and speeds of that access. The point is simply that time, effort, and money can be interchanged here on a more parsimonious level than shopping time because the returns to information-processing can be more widely realized.

growth in income in this model. By spending time to improve income prospects, the agent shifts probability from lower income states to higher income states for the next period. In each period, income takes one of K values, e_r, $r = 1, \ldots, K$. If in the previous period all effort was devoted to information processing ($\alpha_{t-1} = 1$), the income distribution at time t is uniform. As α_{t-1} decreases, probability is shifted toward values as depicted in figure 4, according to the following formula for the distribution of per-period income:

$$b_t(e_r|\alpha_{t-1}) = \frac{r^{2(1-\alpha_{t-1})}}{\sum_{s=1}^{K} s^{2(1-\alpha_{t-1})}}. \tag{8}$$

where r is a node index and the summation over s guarantees that the income distribution sums to one. This distribution has the property that lower-α distributions stochastically dominate (to first order) higher-α distributions.[13]

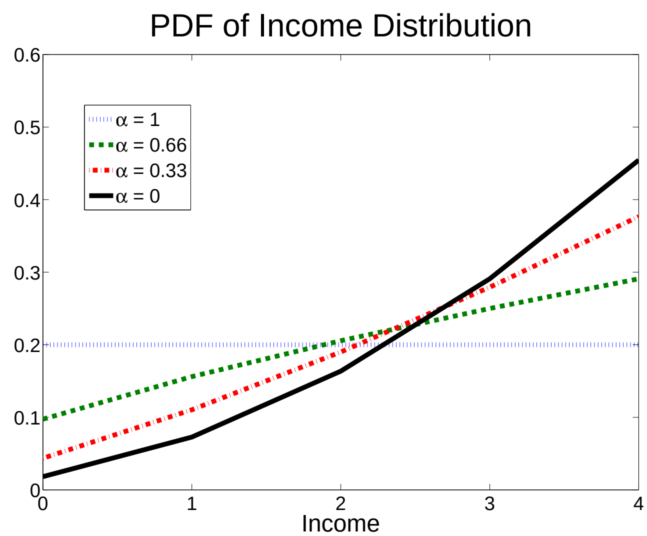


Fig. 4. An Example of the Income Process When $e_1 = 0$, $e_2 = 1$, etc.

In each time period, the agent will have the opportunity to vary his total pool of attention, as well as how it is used. This is important because to date, work in rational inattention assumes only an exogenous information processing constraint that binds. The intuition behind the binding constraint is obvious: no one can process all the information available to them. However, one could consider devoting more time to information processing than he or she does currently. The flexibility must be capped such that no one can process all information, thus the κ^M parameter; but an important addition to the RI literature is the concept of "paying" for additional processing capacity, which is the central idea behind the model components described in equations (7) and (8). By beginning to analyze the value of a unit of information-processing capacity, we can move toward models that consider the ways in which the market helps agents solve their optimal attention-allocation problem.

[13] Technical Appendix A discusses the relative merits of this income process in response to questions regarding some potential alternatives. It is included for the benefit of the referee.

4.1 The Agent's Problem

The agent has the same period utility function as in the canonical model:

$$U(c) = \frac{c^{1-\gamma}}{1-\gamma}.$$

The objective function, however, is the generalization described in section 2.1: the choice variable is the joint distribution of consumption and wealth in each time period and the objective is to maximize lifetime *expected* utility. The choice variable, the joint distribution over consumption and wealth in each period, is a choice of probability weights on a fixed domain of (c, w) pairs. That is, the agent is choosing a set of probabilities for c_i and w_j, $i, j = 1, \ldots N$: $f(c_i, w_j)$'s, which means that agents are placing weight on points, not choosing the points themselves. *The grids of support for consumption and wealth are identical, and the points are evenly spaced.* The problem is to choose α and the probabilities at each date to

$$\max_{\{f_t(c_i,w_j),\alpha_t\}_{t=1}^{T}} \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \beta^t U(c_i) f_t(c_i, w_j). \tag{9}$$

In choosing the weights, the agent has a standard budget/borrowing constraint in that he is unable to consume more than his available wealth. That is, in placing weight on consumption-wealth pairs, he is unable to assign positive probability to situations where consumption exceeds wealth:

$$f_t(c_i, w_j) = 0 \text{ for } c_i > w_j, \quad t = 1, \ldots, T. \tag{10}$$

Also, the standard constraint regarding probability weights must hold

$$0 \leq f_t(c_i, w_j) \leq 1\,, \quad i = 1, \ldots, N;\; j = 1, \ldots, N;\; t = 1, \ldots, T. \tag{11}$$

Additionally, recall that the model is one of having a distribution over wealth and making choices that are sets of conditional distributions of consumption given wealth. This is modeled as the choice of a joint distribution that is restricted to agree with the marginal distribution of wealth. Therefore,

$$\sum_{i=1}^{N} f_t(c_i, w_j) = g_t(w_j)\,, \quad j = 1, \ldots, N;\; t = 1, \ldots, T. \tag{12}$$

4.2 The Information-Processing Constraint

The information-processing constraint is a per-period constraint. That is, the mutual information between consumption and wealth is restricted to be less than the total capacity κ_t in each time period. The mutual information calculation in this discrete-distribution case is identical in form to that of (5):

$$\sum_{j=1}^{N}\sum_{i=1}^{N}\log[f_t(c_i,w_j)]\cdot f_t(c_i,w_j)-\sum_{i=1}^{N}\left(\log\left(\sum_{j=1}^{N}f_t(c_i,w_j)\right)\cdot\sum_{j=1}^{N}f_t(c_i,w_j)\right) \\ -\sum_{j=1}^{N}\log(g_t(w_j))\cdot g_t(w_j)\leq\kappa_t\,,\quad t=1,\ldots,T \tag{13}$$

where the second term is the entropy of the marginal distribution of consumption that results from choices in the joint.

4.3 The Wealth Transition

The challenge in the dynamic RI framework is the transition of the state variable. Because the state variable in the general framework is a distribution, as is the choice variable $f_t(c_i, w_j)$, determining the next period's distribution of wealth is a matter of determining the probability of being at each potential wealth node.

The wealth distribution is fully determined by the joint distribution of consumption and wealth in the past period and the distribution of per-period income in the current period. Current income is independent of prior wealth and is received prior to any consumption decision. That is, the timing works as follows: The choice variable in period $t-1$, $f_{t-1}(c_i, w_j)$, is combined with the current period's income distribution, $b_t(e)$ during the working portion of the agent's life, to determine his current marginal distribution of wealth. In the retired portion of the agent's lifetime, the consumption-wealth joint distribution in the previous period fully determines the wealth marginal distribution in the current period. The equation for the transition of the marginal distribution of wealth during the employed portion of the lifetime is

$$g_t(w_j)=\sum_{r=1}^{\min(K,j)}\left[b_t(e_r)\cdot\sum_{p=j-r+1}^{D_{t-1}}f_{t-1}(c_{p-j+r},w_p)\right],\quad t=2,\ldots,R-1;\ j=1,\ldots,N, \tag{14}$$

with the transition during retirement being given by

$$g_t(w_j) = \sum_{p=j}^{D_{t-1}} f_{t-1}(c_{p-j+1}, w_p), \quad t = R, \ldots, T\,; \quad j = 1, \ldots, N, \tag{15}$$

where K is the number of nodes in the income distribution, R is the first period in which the agent is retired, N is the number of grid-points in the support of both the consumption and wealth distributions and D_{t-1} guarantees that the correct diagonals of the f distribution are being used in the calculations, see figure 5. The wealth transition represented in equations (14) and (15) are facilitated by the choice of a specific "gridding" of the supports for the discrete distributions involved in the model.

The support for the wealth and consumption distributions in this model is a pair of identical, N-point grids that begin at zero and increase in equally spaced increments. The support for the income distribution is the first K nodes of the support of the wealth and consumption distributions. By using the same support for all the distributions, we keep the state space as small as possible for a given set of distributions. The transition from the current period wealth distribution ($g_t(w)$) to next period's wealth distribution ($g_{t+1}(w)$), given the choice of the joint distribution $f_t(c, w)$ and per-period income distribution $b_{t+1}(e|\alpha_t)$ is outlined in figure 5. The process is as follows: for a given node within the future wealth distribution (for example $w = 1$, in figure 5's example), we find the probability by looking for all the possible combinations of current wealth, consumption, and income that could bring us to that point, and sum the probabilities of those events.

This paper is interested in examining the aggregate phenomena regarding the consumption hump observed in the cross-sectional data and therefore uses each period's simple average across potential consumption values for the representative agent. Each period, the representative agent makes his $f(c, w)$ choice and receives a draw from the resulting marginal distribution of c. A simple average of these values is obtained to examined the aggregate life-cycle phenomena seen in cross-sectional data. The aggregated consumption point is used in the transition for the representative agent because it simplifies the calculations while allowing examination of the aggregate result.

Figure 5 outlines the interaction between each of the three distributions involved in forming the distribution of wealth in a given time period: the state variable $g_t(w)$, the current income variable $b_{t+1}(e)$, and the choice variable $f_{t+1}(c, w)$. Note first how the period t distribution of wealth, $g_t(w)$, restricts possible forms of the period t joint distribution. The columns in the f matrix composed entirely of zeroes and surrounded by dashed boxes represent the restriction in the joint distribution due to the fact that the marginal distribution of wealth has no probability weight on those wealth values. Therefore, the choices available to the agent regarding $f_t(c, w)$ are as follows: for each wealth level w_j, the conditional distribution for consumption divides the weight from $g_t(w_j)$ among the feasible elements of $f_t(c_i, w_j)$. This is done in period t under the processing constraint that the mutual information of c and w not exceed κ_t, while also choosing α_t in order to balance the benefits of increased processing capacity with increases in expected future income. The choice of α_t fully determines the weights in $b_{t+1}(e)$ as well. Once $f_t(c, w)$ and $b_{t+1}(e)$ have been determined, we have

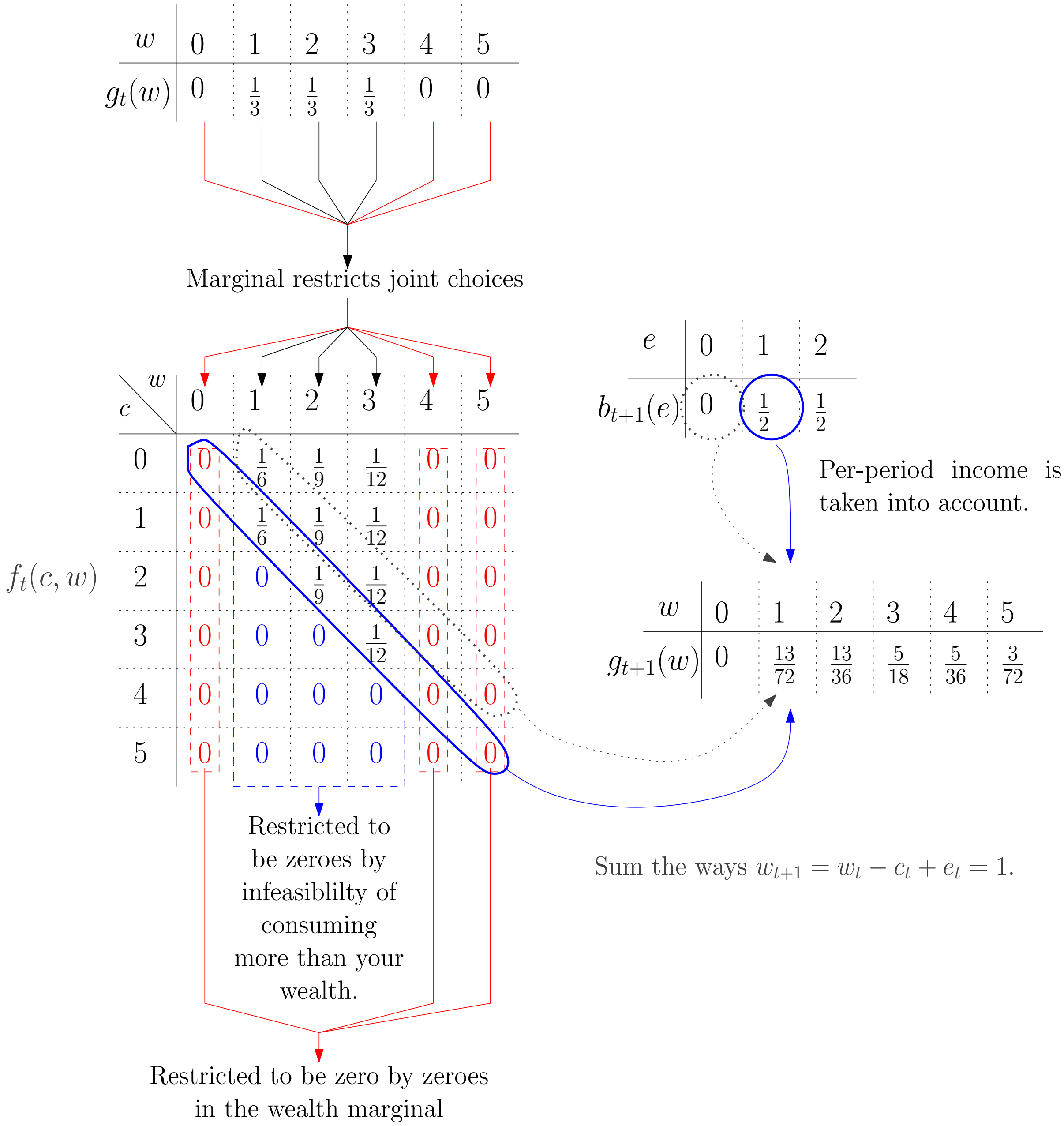


Fig. 5. An Example of the Transition of Wealth From One Period to the Next.

determined the marginal distribution of wealth in period $t + 1$. The probability weights of period-t based decisions are combined thusly in order to determine the weights in $g_{t+1}(w)$: Take, for example, $g_{t+1}(w = 1)$. There are two and only two ways to get to $w_{t+1} = 1$ in a borrowing-constrained model: The first is to consume everything in period t and receive an income of one at the beginning of period $t + 1$. The second is to consume all but one unit in period t and receive no income at the beginning of period $t + 1$. Figure 5 focuses on the first of these two possibilities by summing the probability of all the consumption-wealth combinations that leave behind zero post-consumption wealth in period t (the solid outline around the main diagonal), and multiplying that by the probability of receiving one unit of income in period $t + 1$. A similar calculation characterizes post-consumption wealth of one and income of zero (the dotted outline around the first off-diagonal). These two calculations are then summed to arrive at the total probability of being at $w_{t+1} = 1$. It should be noted that the specific values in the f_t matrix of figure 5 in no way represent an optimal choice and are simply for illustrative purposes.

Regarding information concerning the next period's wealth, note that "...the agent must allow some noise to affect the choice of c in the current period, but can use the noisy observation that entered determination of c to update beliefs about next period's w" [(Sims, 2006, p. 18)]. This is accounted for in equations (14) and (15) and clearly reflected in figure 5. That is, the restriction on information processing in the current period constrains decisions regarding consumption, but also effects the information known about the following period's wealth. Where attention is focused by the agent (in the wealth-consumption space) in the current period will have an impact on the precision of the agent's wealth distribution in the following period(s).

4.4 Parameters and Initial Conditions

In the analysis that follows, the life-cycle is divided into $T = 8$ periods, where the agent is employed for six periods and retires in period $R = 7$. The life-cycle is assumed to begin during the working portion of life and we assume that it takes place during ages 25-80, meaning that a model period is just under seven years and $\beta = 0.96^{\frac{55}{8}} \approx 0.76$. The initial state, $g_1(w)$, is assumed to be a flat one-period income distribution, with the exception that there is no weight placed on the $w_1 = 0$ node. We want to be able to analyze preferences with higher risk aversion, and therefore do not want to force agents to absorb zero consumption in the initial state. In future periods, there can be probability on a per-period income of zero, but agents with higher-γ values choose never to allow this to become a problem.

The value for N (the number of nodes in the wealth, and therefore also consumption, grid) is determined by the value for K (the number of nodes in the income grid) and the retirement age R in this model. In each period, the income received adds to potential existing wealth and the maximum possible value of wealth increases. Therefore, the total size of the wealth and consumption grids is given by $N = (R - 2)(K - 1) + K$, and in the figures that follow, $K = 9$, making $N = 49$.

As in Lewis (2009), this problem is handed over to a numerical optimizer. The optimization is performed using a combination of programs known as AMPL and KNITRO. AMPL is a front end for many powerful optimizers, one of which being KNITRO. By front end we mean the following: Problems are entered into AMPL via a very explicit system which essentially requires nothing more than copying the objective function in equation (9) as well as the constraints in equations (7), (8), (10)-(15) into a file exactly as they appear above. Once the problem has been described to AMPL it performs what is called pre-solve, which looks at the problem and does what it can to eliminate complexity from a hill-climbing perspective by performing basic exercises such as solving for equality-constrained variables and so forth. Finally, AMPL performs what is known as *automatic* or *algorithmic* differentiation. The speed and accuracy of any optimizer depend on the information available about the hill being climbed. Automatic differentiation (AD) provides the gradients without the truncation errors of a procedure like divided differencing or the excessive memory usage of symbolic differentiation. AD is best thought of as a close cousin of symbolic differentiation in that both are the result of systematic application of the chain rule. However, in the case of AD, the chain rule is applied not to symbolic expressions but to actual numerical values.[14]

Given the specifications for K, R, and N above, the number of free parameters (5882, after accounting for adding-up and zero-restriction constraints) seems *very* large. Sims had 345 free parameters and needed 11 minutes. However, using AMPL/KNITRO on his problem [See section 2.1 above; see Lewis (2009) for more detail regarding the numerical optimization issues.] required 1-2 seconds. The 5882-parameter problem of this section requires about 2 minutes. The problem itself is straightforward from a numerical optimization standpoint except for the pre-retirement transition of wealth probabilities, specifically elements pertaining to per-period income. With the exception of α_t, the problem is a very well-posed optimization problem. The objective is linear, and the constraints would be convex if not for the effect of α on per-period income (that is, the trade-off variable accounts for eight of the 5800+ choice variables). Thus, the problem does not appear to be badly behaved. Several specifications of the model have been tested with dozens of random starting points for both $f_t(c, w)$ and α_t, and always within each specification, optima were identical across starting values.

5 Analysis and Results

The addition of uncertainty and information-processing constraints to the canonical model results in a clear hump-shape in the aggregate consumption path, as indicated in figure 6. The initial slope of the consumption hump is the result of "buffer-stock" style savings early in the life-cycle designed to protect again low wealth states in the future. The downward

[14] For a discussion on this and further exposition of AD, see Griewank (2000) and Rall (1981). For a discussion specific to its application within AMPL, see Gay (1991).

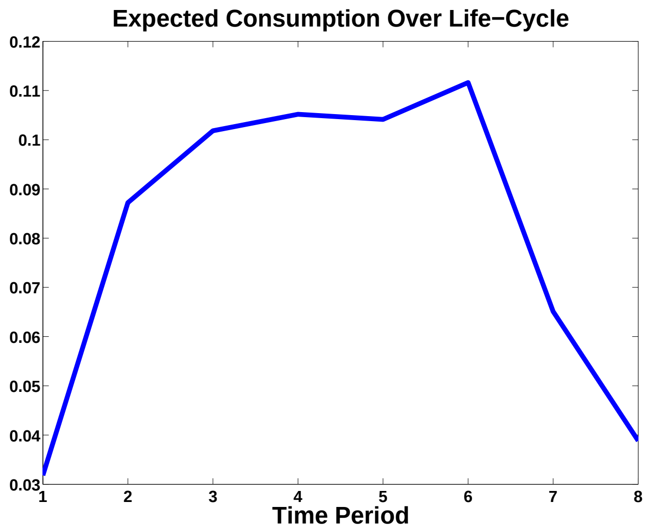


Fig. 6. The Path of Aggregate Consumption Over the Life-Cycle.

slope at the end is the result of a struggle between the desire to consume as much as possible in each period and the desire to avoid a large probability of being left with zero wealth in the final period. This behavior is the result of unresolvable uncertainty on the part of the agent.

To understand where this hump-shape comes from, we examine the impact of decreasing the maximum information-processing capacity level, κ^M. A note about discounting and returns: in this model the gross rate of return is one making the value of $\beta R \approx 0.76$. As has been documented in the life-cycle literature [see, e.g. Yaari (1964)], the path of consumption in the canonical model will only be flat for $\beta R = 1$, while $\beta R > 1$ leads to growth and $\beta R < 1$ leads to decline. Therefore, when the α_t trade-off is eliminated and the infomation-processing capacity is made very large, we would expect the consumption path to decay, as indicated in figure 7.

As the information-processing constraint is tightened, we see a clear hump-shape emerge in the aggregate consumption path, as seen in figure 8.

It is important to remember that the agent never has exact knowledge of his wealth or income. This unresolvable uncertainty due to information-processing constraints is what gives rise to the hump. Careful examination of the differences in the decay of consumption in closing time periods shows that while the unconstrained model (figure 7) has a decrease in consumption, the information-processing constraint clearly plays a role in the profile of the decrease (note the change in the profile of the decrease in panel 8(a) compared to figure 7).

The downside of the hump is partially created by the fact that the no-processing-constraint model in which $\beta R < 1$ has a natural downward slope. However, we can see that the both sides of the hump clearly are produced by the information processing constraints even in the more standard universe of $\beta R = 1$, in figure 9.

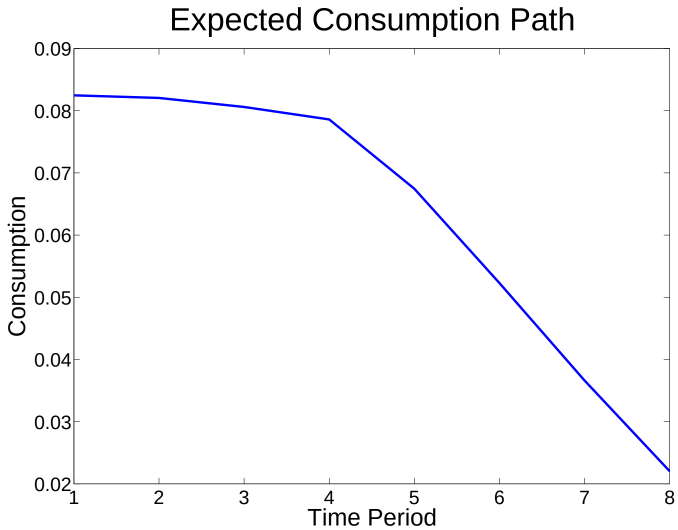


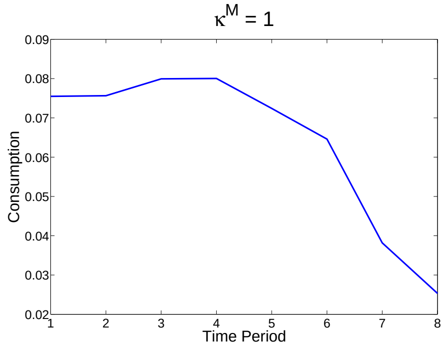
Fig. 7. The Path of Consumption for $\alpha_t = 1$, $\forall\, t$, and $\kappa^M = 10$.

The feature of the model that is important is that information-processing constraints create both precautionary saving and dis-saving. The agent in the model is scaling back his or her consumption more dramatically than they would in the event that they knew their wealth *exactly* and could control their consumption *precisely*. Figure 9 demonstrates that this "precautionary dis-savings" generates a downside to the consumption hump without the aid of the normal consumption decay implied by $\beta R < 1$.

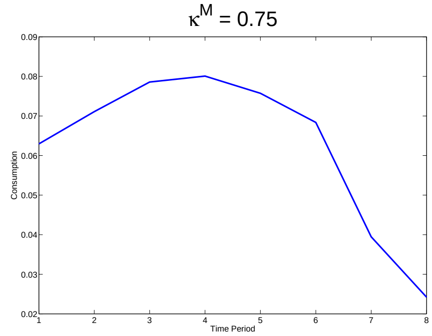
The particular consumption path in figure 6 is based on an agent who has fairly low risk aversion: CRRA utility with $\gamma = 0.5$. As would be expected, the consumption behavior of the agent changes as risk aversion rises, as demonstrated in figure 10: more risk averse agents are more frugal early in life and consume more in retirement as a result.

From figure 11 it is seen that risk aversion changes result in very different behavior during the "employed" portion of the life cycle. Careful examination of the first-period marginal distribution of consumption reveals that the more risk averse agent ($\gamma = 2$) chooses to place probability on only the lowest positive consumption point, regardless of the level of wealth. The other two parameterizations, $\gamma = 0.5$ and $\gamma = 1$, spread probability across multiple potential consumption levels. The three initial consumption strategies summarized in the first panel of figure 11 require different levels of information-processing capacity. While the consumption strategy of the agent when $\gamma = 0.5$ requires some processing capacity (he keeps roughly a third of his potential capacity), the strategy of comsuming the lowest level possible requires almost no information-processing capacity, and the $\gamma = 2$ agent chooses α_1 accordingly, (see Table 1).

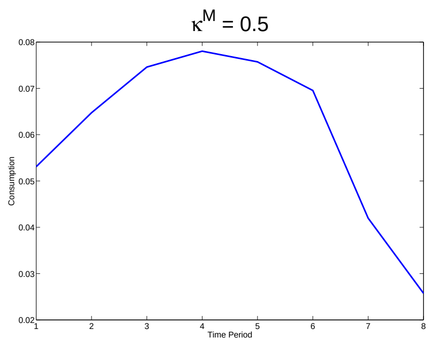
Figure 12 displays the joint distribution of consumption and wealth in the first period. Panel 12(a) is analogous to previous RI treatments in which the information-processing capacity of the agent is fixed exogenously. Panel 12(b) depicts the more general case in which the agent optimally chooses α_t. In each figure, as in the joint distribution figures of section 2, darker boxes indicate heavier probability weight, with the key to the value of the joint pdf given in



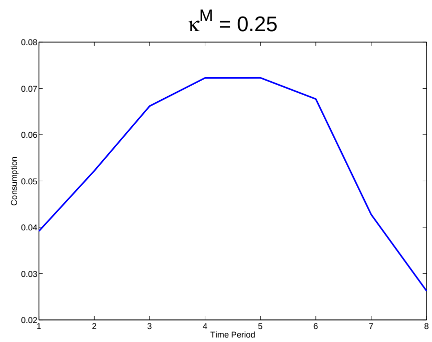
(a) [C_t Path, $\kappa^M = 1.0$



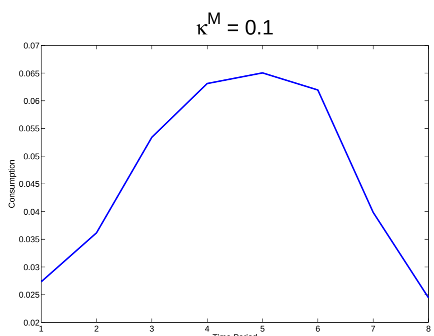
(b) C_t Path, $\kappa^M = 0.75$



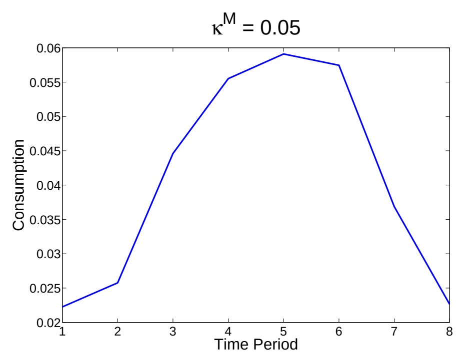
(c) [C_t Path, $\kappa^M = 0.5$



(d) [C_t Path, $\kappa^M = 0.25$



(e) [C_t Path, $\kappa^M = 0.1$

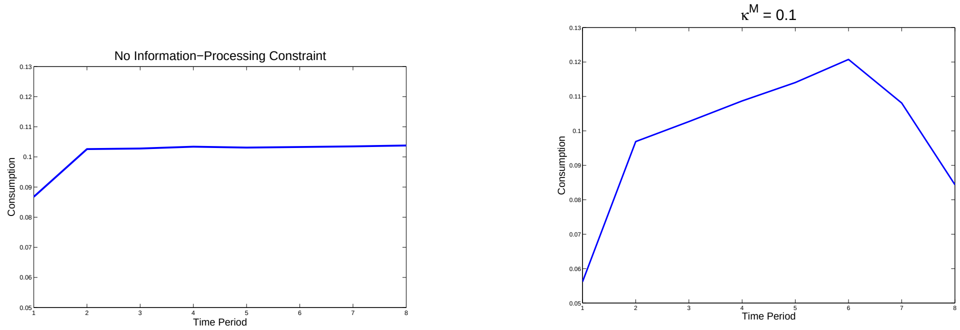


(f) [C_t Path, $\kappa^M = 0.05$

Fig. 8. Aggregate Consumption Path for Decreasing Levels of Information-Processing Capacity, Given a Fixed $\alpha_t = 1$ for All Time Periods.

	$\gamma = 0.5$	$\gamma = 2.0$
α_1	0.3	10^{-3}

Table 1
Values of Tradeoff Parameter, α, in the First Period.



(a) Aggregate Consumption Path, No Restrictions on Processing Capacity

(b) Consumption Path, $\kappa^M = 0.1$

Fig. 9. Aggregate Consumption Path for Decreasing Levels of Information-Processing Capacity, Given a Fixed $\alpha_t = 1$.

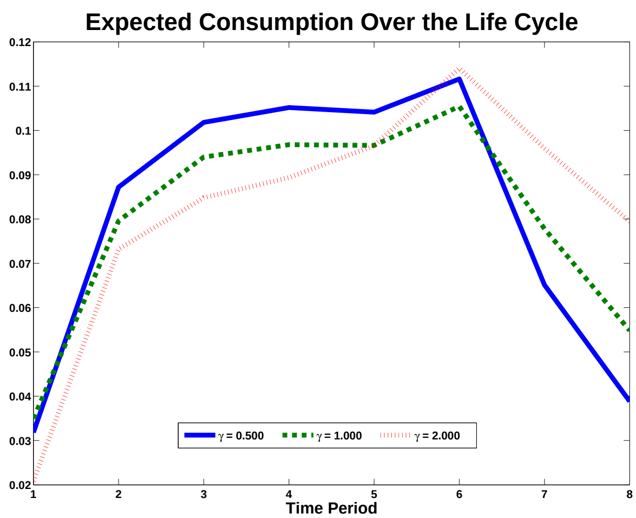


Fig. 10. The Path of Aggregate Consumption Over the Life Cycle for Different Risk Preferences.

the legend to the right of each distribution.

To understand the figure, consider the $\gamma = 2$ case in panel 12(a). The representative agent is choosing to save in the first period, regardless of wealth level. This is seen by the solid bar on the lowest positive consumption node, indicating that he is placing all his probability on consuming at the lowest non-zero level. With lower risk aversion ($\gamma = 0.5$), the agent in panel 12(a) chooses to place probability on higher levels of consumption for higher levels of wealth. For example, the consumption distribution conditional on the wealth level just above 0.1 places probability on five possible consumption points, with the majority of the weight being placed on 0.1. It can be seen in this panel that given lower risk aversion ($\gamma = 0.5$), the agent will assign positive probability to several consumption points that are on the feasibility boundary. This is because unlike the higher risk aversion parameterizations, consumption of zero is not penalized nearly as heavily.

The two panels in figure 12 allow analysis of the information-processing capacity and income tradeoff. By fixing $\alpha_1 = 1$ we can examine how the agent's behavior changes when he does not have the ability to trade information-processing capacity for expected income in period two. By looking at the $\gamma = 2$ panels of both figure 12(a) and 12(b), we see that the agent does not wish to use any of his available processing capacity. As a result, when offered the trade of information-processing capacity (something he does not need) for future income (something he wants very much), it is obvious that he will give up much of his processing capacity except for enough to know where the lowest positive consumption node is located. That is, the choice made by the risk averse agent when $\alpha_1 = 1$ is a choice that requires almost no information-processing capacity, so two cases are identical.[15]

The $\gamma = 0.5$ parameterization, on the other hand, requires the agent to balance current processing needs with future income desires. When α_1 is fixed, the agent places weight on consumption possibilities he would not consider when he is offered the choice of α_1. That is, when $\gamma = 0.5$, giving the agent the ability to reduce his processing capacity in exchange for increases in expected future income results in a change in the behavior of the agent. From Table 1, we see that when given the opportunity, the agent will give up 70% of his information-processing capacity in exchange for improvements in his future income. The optimal allocation of "precision" changes when the total amount of precision to be allocated changes. First, note that the levels of wealth and their associated probabilities are identical in panels 12(a) and 12(b). That is, for example, the probability of $w = 0.025$ is the same in both panels. Next, note that the agent assigns probability to consumption possibilities when α_1 is exogenously-fixed at 1 that he ignores when α_1 is endogenously set to 0.3. Additionally, note that the probability is placed more heavily on the feasible boundary when α_1 is fixed, and that consumption appears to be more strongly correlated with wealth. That is, when α_1 is determined endogenously, the probability mass of consumption at each of the three possibilities is spread more uniformly across wealth levels when compared to the exogenously-fixed α_1 choice. These differences are all a result of using less information-processing capacity. First, when the amount of information-processing capacity decreases, the agent can compensate by paying attention to fewer things. This is accomplished by considering a smaller set of consumption possibilities. By eliminating higher consumption levels, the agent is able to "spend more" of his information-processing capacity on the remaining three levels. In addition to this, the remaining consumption-wealth pairings will have less precise conditional distributions. The problem can be thought of as one of allocating a total pool of precision, or attention. First, the agent can reduce the number of (c, w) pairings over which he is trying to be precise, and then lower the precision of pairings still to be considered. The mathematics of the problem boil down to how much "correlation" (more precisely mutual information) can be represented in the optimally chosen joint distribution. As the agent reduces information-processing capacity, the optimal choice of the joint distribution must imply a weaker relationship between c and w. It is important to remember that the agent is choosing to give up the information processing capacity that represents the differ-

[15] Note that the only mutual information "connection" between c and w implied by the joint distribution in panel 12(a) concerns feasibility of the lowest positive consumption node. Beyond that, the distribution implies independence, thus the very low mutual information content and therefore information-processing requirement.

ence between panels 12(a) and 12(b). The information-processing capacity is being traded for future income prospects, meaning that the agent is acting optimally when he chooses to move from panel 12(a) to 12(b).

The optimal path of α is given in figure 13. Returning to the higher risk aversion specification ($\gamma = 2$), the amount of information-processing required for the first period is nearly zero, but the second period makes use of a significant amount. This is done for two reasons, the first of which is that it is no longer optimal to save nearly all income in the second period. The "buffer-stock" accumulation of the first period gives way to a lower marginal expected savings rate. Second, the space of potential wealth levels doubles from period 1 to period 2.[16] These two effects combine to make the optimal choice for the high risk aversion parameterization essentially a coin-flip over two higher levels of consumption than he consider in the first period (see figure 11). It should also be noted that period 2 represents the period in which the representative agent places probability on the highest consumption node he will consider during periods 1-5. This statement must be differentiated from describing the aggregate consumption path, which clearly continues to rise in subsequent periods for the high risk aversion agent. What is meant here is that the agent solves his attention allocation problem in such as way as to place weight on higher levels of consumption in period 2 than he did in period 1, and that he does not continue to consider even higher levels again in period 3. That is, the agent has reached the levels of consumption sustainable given his lifetime expected income by period 2. The next few working periods (3-5) are a process of fine-tuning the consumption choice, as seen by the fact that the $\gamma = 2$ parameterization places the majority of the weight on a single consumption level in periods 3 through 5.

The low risk aversion parameterization ($\gamma = 0.5$) also needs more processing power in the second time period due to the increase in the size of the wealth space. However, when $\gamma = 0.5$, the agent's consumption choices do not need to grow as much relative to the first period as the higher risk aversion parameterizations, so his need for information-processing capacity growth is much less sharp. Still, he continues to use more processing power than his high risk aversion counterpart because he wants to be able to consider more points than the high risk aversion parameterization.

To explain why the low risk aversion agent wishes to consider a broader range of consumption nodes than the higher risk aversion agent for a given wealth distribution, we examine figure 14, the joint distributions in period 6 – the period just before retirement. Before progressing, it must be clearly understood that the α-path for all parameterizations goes to one for periods 6-8 because of the tradeoff used in the model. Agents trade information processing capacity for benefits in their next period's income distribution. At retirement, the agent stops receiving income, so because there is no income in period 7 or 8, there is no reason to spend time trying to improve it in periods 6 and 7 (there is no "future" following period 8, so similarly there is no incentive to give up any processing capacity). Beginning in the

[16] The agent begins with an initial wealth level equal to a flat one-period income distribution with no weight on zero. When the agent moves to the second period he gets his K-node period income distribution whose lowest node is zero. As a result, the first period had K wealth levels while the second period has $K + (K - 1)$ wealth levels.

period directly before retirement, the agent no longer has anything to trade for his processing capacity, and as a result, the amount of information processed in period six ($\alpha_6 = 1$) is nearly three times that of previous periods. One aspect of this model is that the agent anticipates this increase in processing power and is waiting to increase consumption once it arrives.

To return to the question of how risk aversion impacts the range of consumption levels considered given a wealth distribution, figure 14 demonstrates that while the anticipated explosion of processing power does, in fact, cause agents with each parameterization to expand their attention considerably, a central difference between high and low risk aversion can be seen. Low-γ agents spread their attention more broadly, strategically spacing consumption allocation so that they can cover more of the consumption-wealth space with some accuracy, while high-γ agents focus more intently on a smaller number of points, paying special attention to the lower levels to guarantee precise behavior there. For example, in Sims's undiscounted two-period case discussed earlier, given a marginal distribution for wealth, it is clear that the optimal unconstrained choice would be to choose $c = w/2$ for every w. It was demonstrated that, in the information-processing constrained world (his figures 5 and 6), the less risk averse agents will spread their attention over a larger region of the (c, w) space, discretely, so as to generate *adequate* consumption over a broader range of wealth. That is, the consumption conditionals are centered around the $c = w/2$ optimum but give up tightness around the optimum and careful examination of lower consumption nodes for the ability to focus attention on consuming at higher levels when wealth is, in fact, high. The agent has a total amount of "preciseness" that can be used. He could, for example, be very precise around a few levels of wealth by forming very tight conditional distributions around $c_i = w_j/2$ for several w_j's. Or, he could be very imprecise around every w_j. Where and how the agents wish to be precise is a function of their preferences. More risk averse agents give up higher consumption in the high wealth state for the ability to consume more accurately at lower wealth levels, meaning tighter distributions around the optimum and fewer gaps in attention overall at lower consumption levels. They do this because they are concerned about consumption-wealth mismatches at lower wealth levels and want to be able to consume everything up to their boundary in these cases. Further, they are willing to pay the price of moderate consumption in high wealth states in order to do so.

This result, that lower risk aversion agents will spread their attention across more consumption possibilities, combined with the fact that they reach a sustainable level of consumption by period 2, accounts for the relatively flat nature of the α-path for $\gamma = 0.5$ in figure 13. Similarly, the high risk aversion agent has more spread-out consumption behavior in period 2 because by saving heavily in the first period, he has increased his potential wealth to a point where there is very little probability of a low wealth state, therefore it does not take a large amount of information-processing capacity to have tight conditionals at the few wealth levels he considers to be "low." (Note the tight conditional distributions for consumption given the lower three wealth levels in figure 15.) Because he is able to behave cautiously at low wealth levels at a fairly low "attention" cost, he is able to consider a consumption lottery that is essentially a coin flip over two higher consumption levels.

In exchange for giving up processing capacity, the agent receives the same thing in each time period: improvement in the next period's income distribution. Therefore, the variations in α must represent variations in the value of information-processing over time. Just as consumption-smoothing is the optimal result of the canonical model, there could be informational smoothing in inattention models as well. Future research will include an examination of differences in the marginal value of processing capacity that could cause a sort of information smoothing over time that could explain the type of "over-shooting" observed in the α-path of the log-preferences agent in figure 13.

Figure 16 indicates that the wealth distributions are similar across the three γ specifications. As wealth grows, the distributions spread out and as agents dissave, their wealth distributions collapse on a small number of points. When the representative agent reaches the point of retirement, he wants to eat from his savings as much as possible, but there is a struggle between not wanting to leave anything on the table and not wanting to have a high probability of zero consumption in the final period. As can be seen in the final pane of figure 16, each wealth distribution collapses on a point, though the point is different for each parameterization. The distance of that "collapsing point" from the origin and the shape around the point is a function of the agent's preferences regarding zero consumption: more highly-risk-averse agents bring their final wealth distribution to a sharper point than their less cautious counterparts, and that point is shifted to the right to ensure safety regarding zero-consumption. The "sharpness" of the final wealth distribution is increasing in γ because of the effect described earlier regarding how agents allocate their attention. Because of the focus of the higher risk aversion parameterization on a smaller number of points, these agents tend to eliminate probability weight from certain regions in the wealth distribution while leaving other regions untouched. This behavior is different from lower risk aversion parameterizations which tend to eliminate some probability from a large number of nodes rather than all probability from a small number of nodes. As a result, highest probability weight in the high risk aversion parameterization is higher than it's counterpart in the lower risk aversion parameterizations.

As a result of the inability to eliminate uncertainty, we observe in this model what have been called "accidental bequests" [see e.g. Hendricks (2002)]. These bequests result not from uncertainty regarding time of death, but from uncertainty arising from an inability to process all the information available. As is seen in figure 17, the expected bequest is increasing in γ. Note that according to figure 17, the representative agent (with CRRA preferences and $\gamma = 2$) has a 20% probability of leaving behind a bequest at least as large as a full period's consumption (just under seven years' worth).

6 Conclusions

Building on the ideas in the two-period model of Sims (2006), this paper presents a simple life-cycle framework for addressing the optimal allocation of attention to decisions over time. The framework is fully scalable in a finite-horizon model and could be used to study optimal behavior under processing constraints in more general economic environments. In the framework studied here, the value of information-processing capacity varies over time, and the agent's degree of risk aversion plays a significant role in determining that value. Life cycle agents with finite information-processing capacity display the hump-shape pattern of consumption observed so frequently in the data. Additionally, the struggle between wanting to consume as much as possible and wishing to avoid zero consumption can lead to a high probability that an agent will leave behind non-trivial wealth at death, thus generating an "accidental bequest" that is solely an artifact of imprecise knowledge.

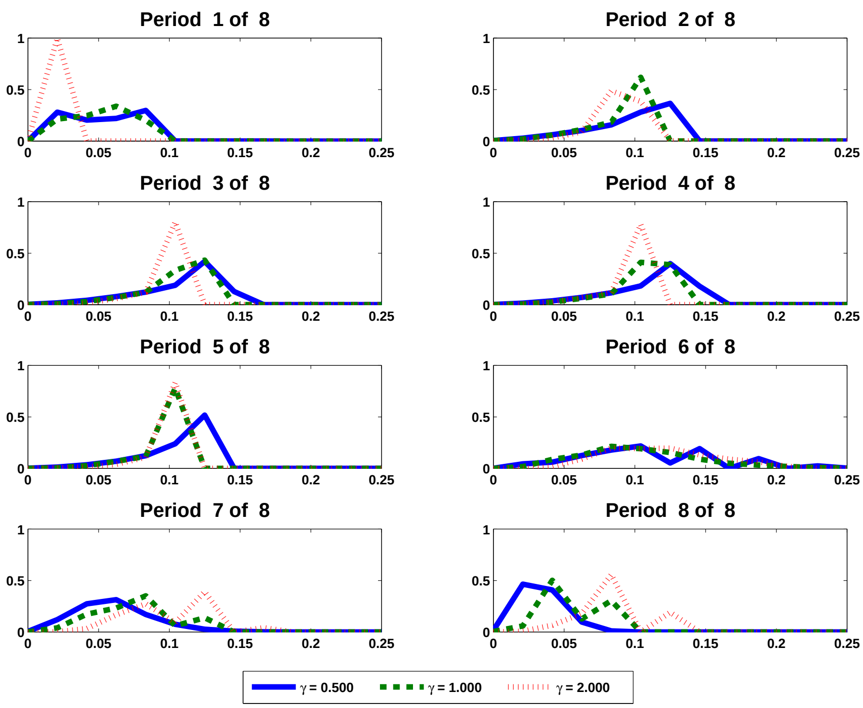
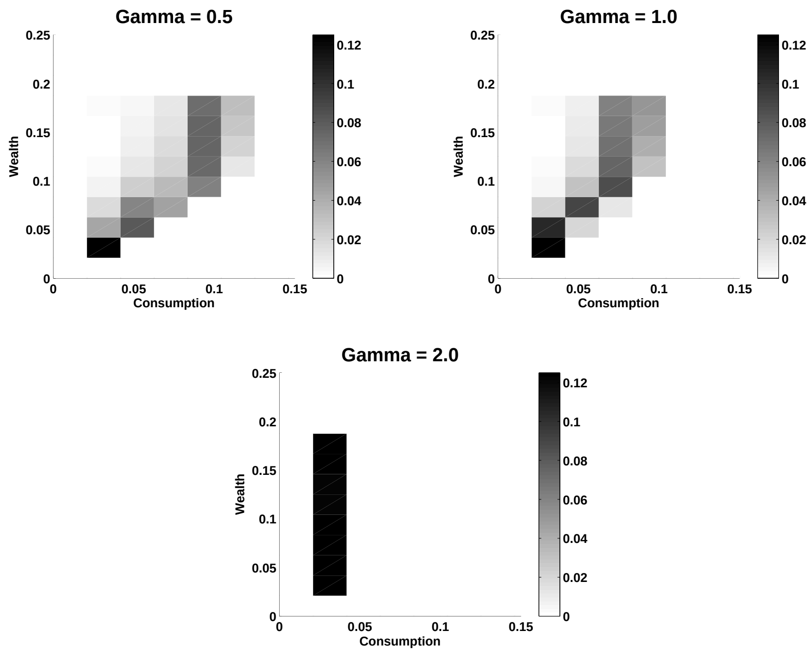
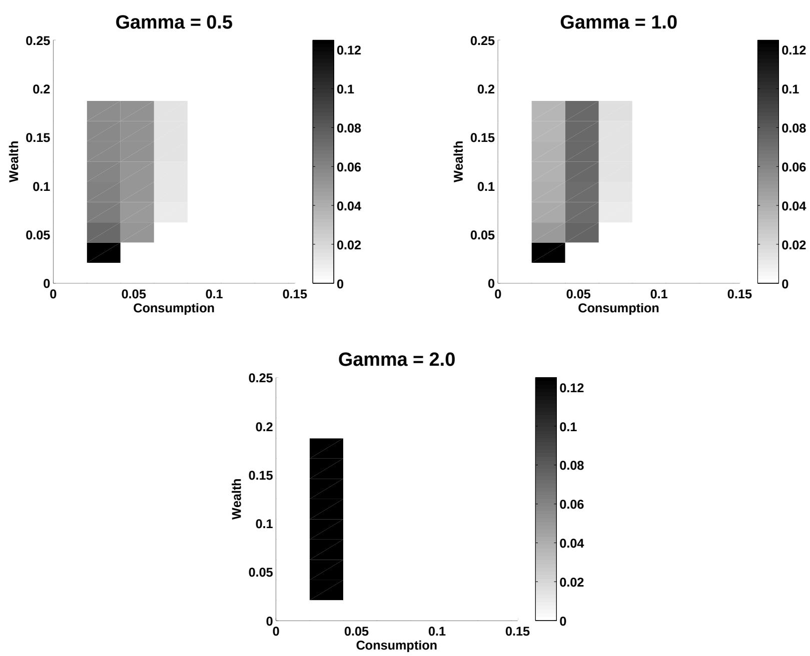


Fig. 11. The Marginal Distributions of Consumption Throughout the Life-Cycle for Different Levels of γ.



(a) The fixed-α choice



(b) The flexible-α choice

Fig. 12. The Choice of Joint Distribution in the First Period for the Fixed- and Flexible-α Cases.

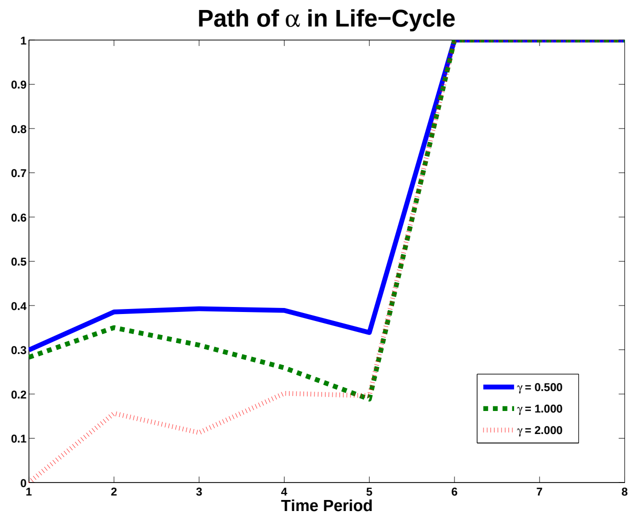


Fig. 13. The Path of the Choice Variable α_t Over Time for Different Risk Preferences.

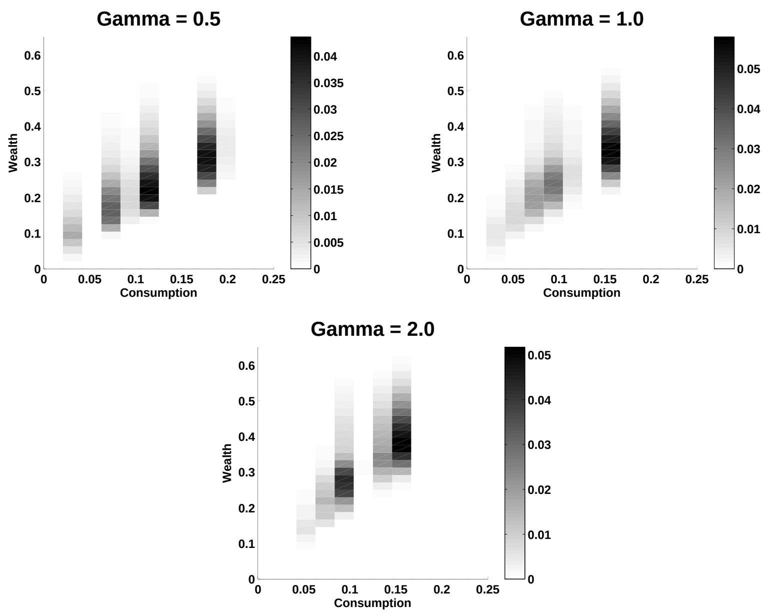


Fig. 14. The Choice of the Joint Distribution of Consumption and Wealth in the Sixth Time Period.

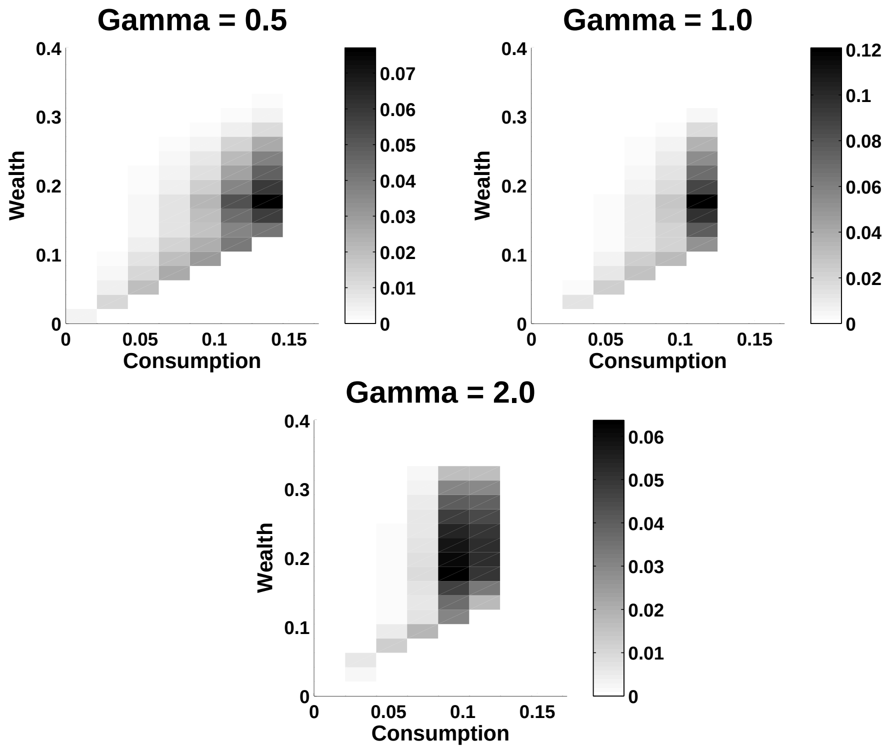


Fig. 15. The Choice of the Joint Distribution of Consumption and Wealth in the Second Time Period.

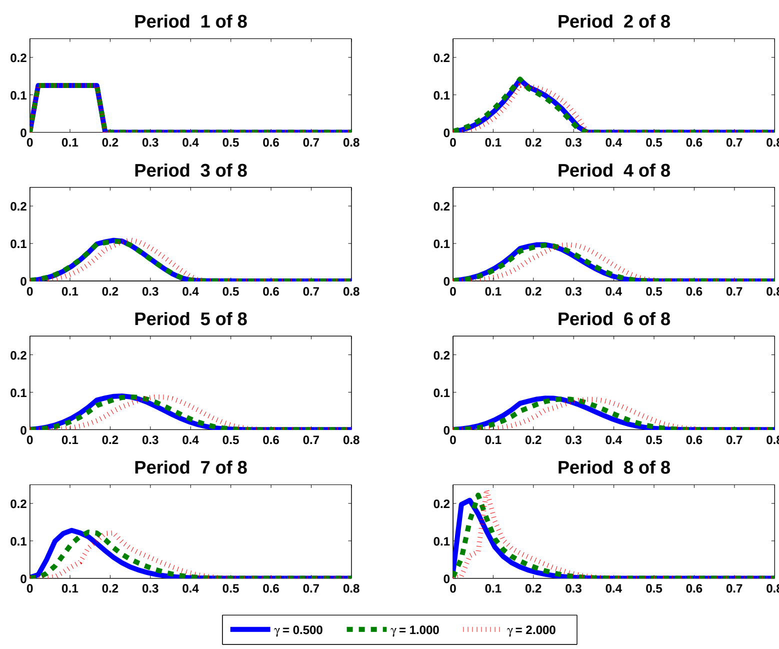


Fig. 16. The Marginal Distributions of Wealth Throughout the Life-Cycle for Different Levels of γ.

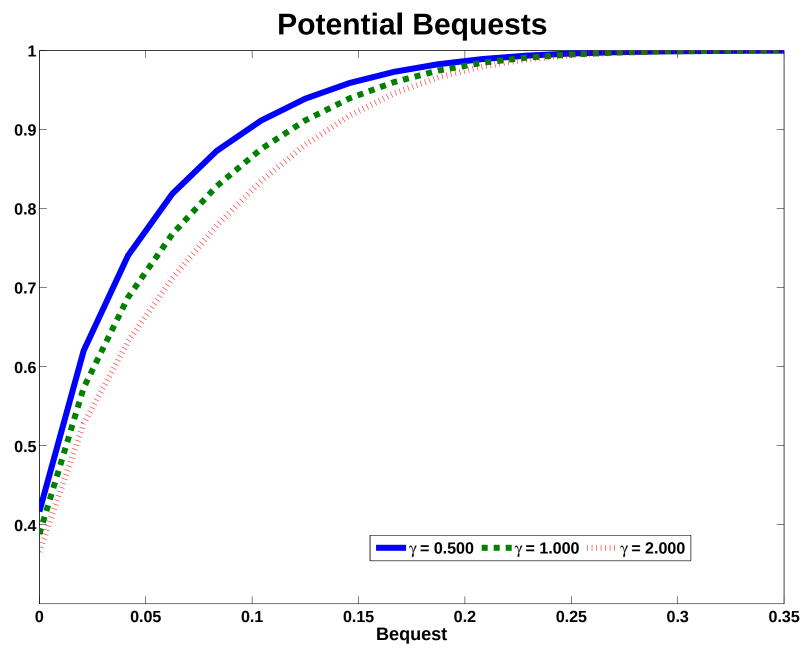


Fig. 17. The CDF of the Bequest Distribution for Different Levels of γ.

References

Aguiar, M., and E. Hurst (2005): "Consumption Versus Expenditure," *Journal of Political Economy*, 113(5), 919–948.

——— (2007): "Life-Cycle Prices and Production," *American Economic Review*, 97(5), 1533–1559.

Bussieck, M. R., and A. Pruessner (2003): "Mixed-Integer Nonlinear Programming," *SIAG/OPT Newsletter: Views & News*, 14(1), Available at: http://www.gamsworld.eu/minlp/siagopt.pdf.

Caballero, R. (1990): "Consumption Puzzles and Precautionary Savings," *Journal of Monetary Economics*, 25(1), 113–136.

Carroll, C. D. (1992): "The Buffer-Stock Theory of Savings: Some Macroeconomic Evidence," *Brookings Papers on Economic Activity*, 23, 61–156.

——— (2004): "Theoretical Foundations of Buffer-Stock Savings," Johns Hopkins University, Department of Economics Working Paper.

Deaton, A. S. (1992): *Understanding Consumption*. Clarendon / Oxford University Press, 1 edn.

Fernández-Villaverde, J., and D. Krueger (2004): "Consumption Over the Life-Cycle: Facts from Consumer Expenditure Survey Data," Working Paper, University of Pennsylvania.

Gay, D. M. (1991): "Automatic Differentiation of Nonlinear AMPL Models," *AT&T Bell Labratories Numerical Analysis Manuscript*, 91-05.

Gourinchas, P.-O., and J. A. Parker (2002): "Consumption Over the Life Cycle," *Econometrica*, 70(1), 47–89.

Griewank, A. (2000): *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, no. 19 in Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics.

Hendricks, L. (2002): "Intended and Accidental Bequests in a Life-Cycle Economy," Arizona State University Working Paper.

Lewis, K. F. (2009): "The Two-Period Rational Inattention Model: Accelerations and Analyses," *Computational Economics*, 33, 79–97.

Luo, Y. (2004): "Consumption Dynamics, Asset Pricing, and Welfare Effects under Information Processing Constraints," Discussion paper, Princeton University.

Modigliani, F. (1986): "Life Cycle, Individual Thrift, and the Wealth of Nations," *American Economic Review*, 76(3), 297–313.

Rall, L. B. (1981): *Automatic Differentiation: Techniques and Applications*, vol. 120 of *Lecture Notes in Computer Science*. Springer-Verlag.

Reis, R. (2006a): "Inattentive Consumers," *Journal of Monetary Economics*, 53(8), 1761–1800.

——— (2006b): "Inattentive Producers," *Review of Economic Studies*, 73(3), 793–821.

Shannon, C. E. (1948): "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 27, 379–423, 623–656.

Simon, H. A. (1978): "Rationality as Process and as Product of Thought," *The American Economic Review*, 68(2), 1–16, Papers and Proceedings of Ninetieth Annual Meeting of the American Economics Association.

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

——— (2003): "Implications of Rational Inattention," *Journal of Monetary Economics*, 50(3), 665–690.

——— (2006): "Rational Inattention: A Research Agenda," Working paper, Princeton University.

Tutino, A. (2008): "The Rigidity of Choice. Life cycle savings with information-processing limits," mimeo, Princeton University.

Yaari, M. E. (1964): "On the Consumer's Lifetime Allocation Process," *International Economic Review*, 5(3), 304–317.

A On the Entropy Effects of the Income Process

In equation (8), the mean of the time t income process is controlled by the value of α_{t-1}. What is seen in figure 4 is that in addition to the mean effect, there is also an effect on the entropy of the income distribution. As α_{t-1} gets closer to zero, more and more probability weight is taken from low income values and moved to high income values. This has the effect of reducing the entropy of the income process (concentrating more weight on fewer points reduces entropy). This, in turn, influences the entropy of the wealth distribution in the subsequent (and therefore every future) period. Thus, while the original intent of the α_t tradeoff was to represent the optimal mixing of two goals — increasing future income and increasing current processing capacity — the entropy-reducing side-effect of this particular income process potentially muddles the interpretation of the tradeoff represented by α_t.

The ideal solution to this problem would be to use an entropy-neutral (with respect to α_{t-1}) income process. This solution, however, is problematic. The existing income process removes probability from lower-income and places it on higher income values. This monotonically increases the mean and decreases the entropy of the income distribution. What is required is an income process whose shape does not change *at all* as the mean is increased by the α_t parameter. For example, a univariate Gaussian distribution with mean μ and standard deviation σ has a theoretical entropy value of $1/2\log(2\pi e) + \log(\sigma)$, which only depends on σ and therefore, any α_{t-1} scheme that shifts the distribution by making μ a simple function of α_{t-1} should achieve this entropy-neutral property.

However, employing a discretization of the Gaussian distribution on the wealth grid eliminates the tidiness of the theoretical entropy calculation. Consider the following income process: a discretized Gaussian distribution where $\mu = (1 - \alpha_{t-1})\mu_{HIGH} + \alpha_{t-1}\mu_{LOW}$. Thus, as $\alpha_{t-1} \to 0$, the agent is making the mean higher and higher, just as in the original process. The standard deviation of the discrete distribution is σ (fixed), which on a continuum, fixes the entropy of the income process. The discretization is accomplished by finding the value of the kernel of the $N(\mu, \sigma)$ distribution at each income node and then dividing by the sum to normalize the income process making it sum to one. Suppose that $\mu_{HIGH} = 14$, $\mu_{LOW} = 7$, $\sigma = 1$ and the income grid is the integers from 0 to 21. Figure A.1 shows the mean and entropy of the income distribution for $\alpha_{t-1} \in [0, 1]$.

What is seen in figure A.1 is the result of a continuous choice for α_{t-1} and its effect on a discrete grid. Note that the entropy values in this example (given on the right y-axis) differ in the seventh decimal place. This is a very small variation in the entropy, but from a numerical optimization point of view, catastrophic. The sine-wave pattern of entropy creates an effect that makes determination of the optimal choice of α_{t-1} extremely difficult, as indicated by the optimizer's inability to find an optimum under this income formulation, despite such a tiny change in entropy. While the change in entropy is very small, the derivatives are large and this problem becomes much like trying to numerically optimize a nonlinear problem whose constraint set can be thought of as a golf ball: many symmetric, equal-sized peaks and valleys. Having a numerical optimizer find an optimal tangency on that surface is theoretically *very*

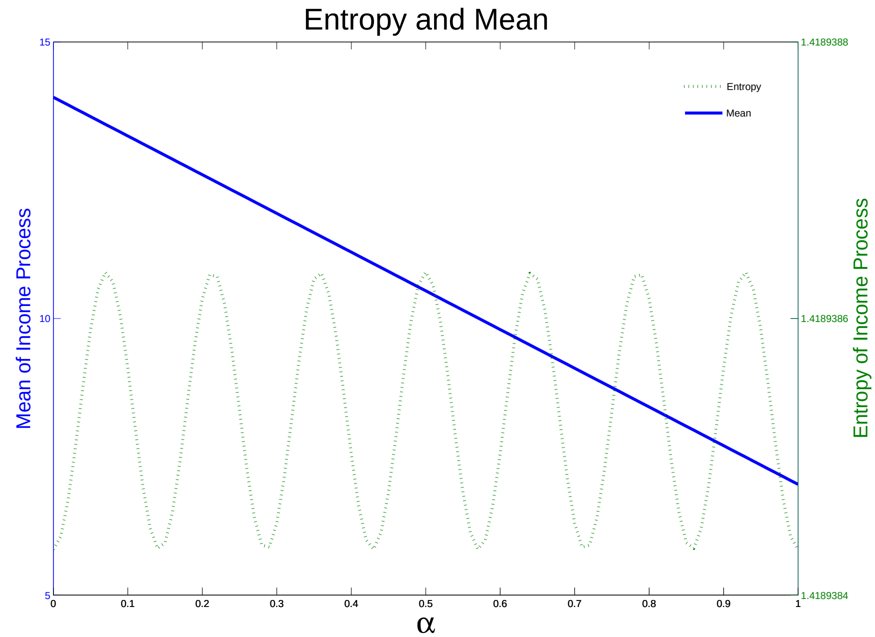


Fig. A.1. The Entropy and Mean of an Income Process for $\alpha_{t-1} \in [0, 1]$.

difficult, and in practice likely impossible. Therefore, while the original model has an income process in which entropy is monotonically increasing in α_{t-1}, it is at least optimizable. The peaks of figure A.1 are associated with the integer values between μ_{HIGH} and μ_{LOW}. *This sine-wave pattern exists for any continuous choice of α_{t-1} on a discrete grid.* The problem is a result of small changes in entropy due to moving tiny probability weights from lower portions of the support to higher portions of the support, as a result of changes to α_{t-1}.

The goal remains unsatisfied: a discrete distribution for income in which α_{t-1} has no effect on entropy while controlling the mean. One way of having two distributions over the same support p and q with the same entropy and different resulting means would be to have p be a simple re-ordering of q. That is, $p_i = q_j$ for some i and j. Imagine a scenario where q is a discretized Gaussian distribution with the conventional bell shape. It is possible to have p be simply the probabilities from q lined up in ascending order, over the same support. Thus, $H(p) = H(q)$ but $E(p) > E(q)$. Another option would be to have a small uniform distribution that shifts (e.g. a three-node pdf with 1/3 weight on each node), where α_{t-1} controlled *which* three nodes received probability. Either of the two potential solutions mentioned here require some kind of discretization of the choice of α_{t-1}, which *dramatically* complicates the problem. Allowing the choice of α_{t-1} to be from some discrete list (e.g. $\alpha_{t-1} \in \{0, 0.1, 0.2, \ldots, 1\}$), transforms the optimization problem into what is known as a mixed-integer nonlinear programming (MINLP) problem. These problems are the subject of the edge of optimization theory, are ill-tempered even under the best of circumstances and are not, at this time, a practical research avenue for this problem. [17]

[17] The method commonly used to solve these problems (known colloquially as "branch and cut" or "branch and bound") is unreliable to implement on a problem with this structure, given the high

A potential solution to the entropy-neutrality problem within the context of the above discussion using a continuous α_t choice would be to modify two probability nodes from the discretized Gaussian distribution generated by α_{t-1}. One could theoretically alter two of the probability nodes to achieve two simultaneous goals:

- The modified distribution has a specific entropy, fixed to eliminate the wave in figure A.1.
- The modified distribution sums to one.

This process could theoretically be used to zero out the effect seen in figure A.1, but the process would essentially involve finding the solution to the equation $p\log(p) = C$, and therefore mean an internal optimization designed to push the objective of the optimization $\min_p p\log(p) - C$ to zero, from the point of view of the computer, not just to an arbitrary tolerance chosen by the user. If the computer does not zero out the entropy difference completely, the problem illustrated in figure A.1 would persist. The internal optimization described here is *very* time-consuming (asking a computer to iterate on a problem, using smaller and smaller steps until "machine zero" is reached), if not computationally infeasible. Before this is pursued, it is reasonable to ask if the effect we are attempting to eliminate is important to the model results. The "side effect" we are trying to eliminate would allow the agent to get both higher expected income and lower uncertainty about that income. Is the agent interested in this side effect, or is the increase in the expected income the sole reason for the agent's choice of α_t?

A.1 Is the entropy feedback effect being used by the agent?

While it is potentially infeasible to use a continuous α_{t-1} to change the mean of the discrete income process in an entropy-neutral way, it *is* possible to use a continuous choice of α_{t-1} to change the entropy of the income process while leaving the mean of the distribution intact. Consider the following income process:

$$b'_t(e_r|\alpha_{t-1}) = \frac{(K^2 - (r-M)^2)^{(1-\alpha_{t-1})Z}}{\sum_{s=1}^{K} (K^2 - (s-M)^2)^{(1-\alpha_{t-1})Z}} \tag{A.1}$$

where M is the middle node of the income grid. This process is uniform when $\alpha_{t-1} = 1$, and as $\alpha_{t-1} \to 0$, weight is moved from the tails of the distribution to the center. Thus, the mean of the income distribution never changes, but the entropy monotonically increases in α_{t-1}.

Optimizing using this process will let us examine the tradeoff purely between information-processing capacity and the entropy of the future income process. The agent will have the ability to process more information "today", or have a more certain income "tomorrow."

dimensionality and complexity of the problem that remains after stipulating a vector of α_t's. For a quick overview of MINLP's, see Bussieck and Pruessner (2003) and the references therein.

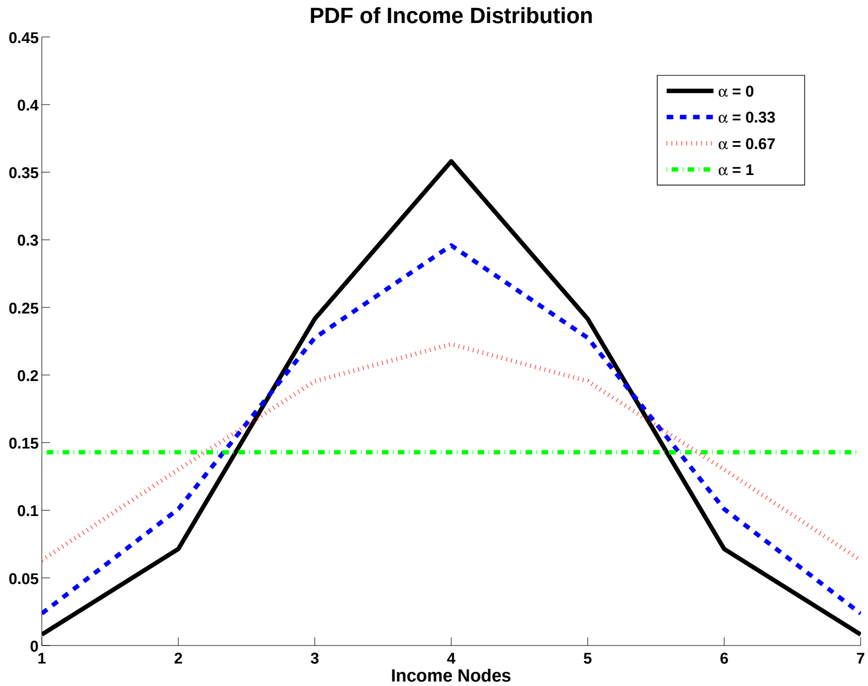


Fig. A.2. Mean-Neutral Income Process for Several α_t Choices, $K = 7$, $M = 4$, $Z = 20$

The expected income for "tomorrow" will not change, regardless of the agent's choice of α. What we see in figure A.3 is that the choice is clear: the agent prefers processing power to lowering the entropy of the income process, regardless of time period. That is, the agent will devote all his "energy" (as evidenced in α_t being pushed to the boundary $\alpha_t = 1$, $\forall t$) to one activity: information-processing; and completely ignore the alternative activity of reducing the entropy of the income process. This tradeoff is of no interest to the agent.

Figure A.3 demonstrates that the agent is unwilling to give up any processing power currently to reduce the entropy of the future problem. Later, we will look at the optimal choices for α_{t-1} given the original income process (the one specified in equation (8)). The role of Z is to control how entropy is affected by changes in α_{t-1}. The value for Z used to create figures A.2 and A.3 changes entropy at 10 times the rate of the entropy change in the original income process (the process in equation 8 which is the mean-shifting process), and values up to 50 times the rate of change in the original income process were tested with no effect on the results seen in figure A.3. The exercise performed here for much more extreme parameterizations indicate that we can look at the optimal choices for α_{t-1} under the original specification with reasonable expectation that the agent is making use of the feedback effect in order to get future and current benefits out of the income/processing-capacity tradeoff.

Therefore, while the current income process includes a feature that could be seen as a potential wrinkle in the model results, the important feature of monotonicity compensates for the inconvenience of the change in entropy, and the entropy effect appears to be inconsequential from the perspective of maximizing expected utility.

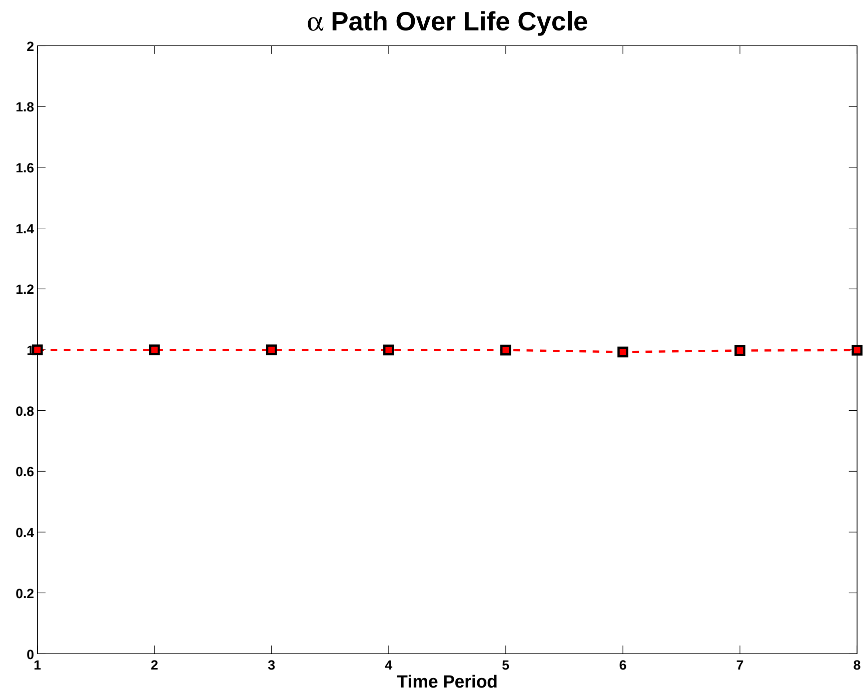


Fig. A.3. Choice of α_t, Mean-Neutral Income Process