Optimal Paternalism*

Manuel Amador, MIT Iván Werning, MIT and NBER

Preliminary and Incomplete

Abstract

If one accepts some paternalistic objectives, how should this shape policy? This paper addresses this issue by considering the trade-off that results when agents have superior information regarding their own situation and tastes but a paternalistic principal views agents' tastes as as being biased. Applications include: savings behavior by hyperbolic discounting consumers, schooling choice by teenagers and many situations with externalities. We show that under certain conditions the optimal mechanism takes a simple threshold form. Particular attention is given to the savings case and the implication for forced minimum savings policies.

1 Introduction

If people suffer from self-control problems, what should be done to help them? Most analysis lead to a simple conclusion: the optimality of taking over individual's choices and running their lives for them. For example, in models with quasi-hyperbolic agents it is generally desirable to impose a particular savings plan on individuals. To obtain less extreme prescriptions we require some meaningful trade-off to the benefits of commitment. In this paper we model a trade-off between commitment and flexibility.

It is convenient to reformulate the same issue using a "paternalistic" interpretation. Thus, consider the maximizing a "paternalist's" utility function

*We'd like to thank Andy Atkeson and George-Marios Angeletos for helpful comments and discussions.

over two goods consumed by an "agent", subject to a standard budget constraint. We assume the paternalist's utility function differs from the agent's and, in particular, that the agent's preferences are biased – relative to the paternalist's preferences – towards one particular good. The question we ask is: What is the optimal mechanism in this situation?

If this were the whole story then clearly the optimal thing to do would be to force the agent to consume a certain allocation: that which maximizes the paternalist's utility function. However, it is likely that the individual's has superior and relevant information about their own situation. If the paternalist finds this information relevant then imposing a particular allocation – thus disregarding the individual's information completely – is likely to be undesirable.

Our main example modifies the intertemporal taste-shock preference specification introduced by Atkeson and Lucas (1995). Each period agents receive an i.i.d. taste shock that affects their desire for current consumption. To this preference specification we incorporate quasi-hyperbolic discounting [Laibson (1997)]. Quasi-hyperolic discounting implies that preferences over consumption plans change over time and generate a desire for commitment. In this setup we characterize the optimal allocation from the point of view of the time 0-self. One can interpret this solution as providing the optimal commitment device, constrained only by the incentive problems generated by the asymmetric information.

Although we focus on the hyperbolic discounting model as our main application, the crucial feature of our analysis is a disagreement in preferences between a 'paternalist' and an 'agent'. Hyperbolic models provide one rational for disagreement in preferences between the different time selves. We discuss some other rationales for disagreement about preferences.

The rest of the paper is organized as follows. After a brief discussion of related literature, Section 2 lays out the basic intertemporal model. Section 3 then briefly detours to discuss some alternative interpretations. Section 4 begins the analysis of the intertemporal model by solving the case with two taste shocks. Section 5 shows the kind of problems encountered with more than two shocks. Sections 6 and 7 analyze the case with a continuum of shocks – these sections contain the main results of the paper. Section 8 extends the analysis to arbitrary time horizons. Section 9 concludes. An appendix collects some proofs.

1.1 Related Literature

Several strands of literature are related in different ways to this paper and we review briefly those most related.

O'Donahue and Rabin (2003) advocate studying paternalism normatively by modeling the errors or biases agents make explicitly and applying standard public finance analysis to these environments. We subscribe to this suggestion and take a further step by studying the mechanism design problem associated with our environment instead of optimizing over a particular set of policies [1].

Economists have long been interested in the implications of and justifications for discounting the future at a lower rate than individuals. Phelan (2002) studies the implications for long-run inequality from insurance in a taste shock framework where the planner does not discount the future while agents discount the future at a positive rates. Caplin and Leahy (2001) discuss another justification for a welfare functional that discounts the future at a lower rate than individuals. Note that in both papers the planner and agents discount the future exponentially.

There is a large literature on social security policies that attempts to take into account the possible "undersaving" by individuals. Diamond (1977) discussed the case where agents may undersave due to mistakes. Feldstein (1985) examines the case where agents discount the future at a higher rate than the planner to study the optimal pay-as-you-go in an OLG model. Using hyperbolic-discounting preferences Imrohoroglu, Imrohoroglu and Joines (2000) perform a quantitative exercise measuring the welfare effects of pay-as-you-go systems. Diamond and Koszegi (2002) use a model with hyperbolic discounting agents to study the policy effects of endogenous retirement choices.

Athey, Atkeson and Kehoe (2002) study a problem of optimal monetary policy with a trade-off between comittment and flexibility. In this regard their paper is closely related to ours. In their model the central banker is benevolent but suffers from a time-consistency problem that could lead to high average inflation. Each period the central banker receives a shock that affects its objective function. They study the optimal policy design problem when this shock is private information.

[1] DellaVigna and Mermandier (2003) study a mechanism design problem that results from the maximization of profits for a firm facing hyperbolic, and possibly naive, costomers.

2 The Model

We first consider the case with two 'non-trivial' periods, $t = 1, 2$, and an initial period $t = 0$. Section 3 extends the analysis to arbitrary N periods.

Each period agents receive an i.i.d. taste shock θ, normalized so that $E\theta = 1$. The taste shock affects the marginal utility of current consumption making consumption more valuable for higher θ.

The utility for self-1 from periods $t = 1, 2$ with taste shock θ is

$$\theta u(c_1) + \beta U(c_2).$$

where $u(\cdot)$ and $U(\cdot)$ are increasing, concave and continuously differentiable[2] and $\beta \leq 1$. The notation allows $U(\cdot) \neq u(\cdot)$ because this greater generality later facilitates the extension to N periods in section 7.

The utility for self-0 from periods $t = 1, 2$ is

$$\theta u(c_1) + U(c_2).$$

Agents have quasi-hyperbolic discounting: *self-t* discounts the entire future at rate $\beta \leq 1$ and in this respect, there is *disagreement* among the different *t-selves* and $1 - \beta$ is a measure of this disagreement. On the other hand, there is *agreement* regarding taste shocks: everyone values the effect of θ in the same way.

To simplify notation we are assuming no exponential discounting for self-0. Likewise, we will assume that the interest rate between periods is zero. Both assumptions can be relaxed without affecting the analysis.

Confining ourselves to periods 1 and 2 an alternative interpretation to hyperbolic discounting is available of the two preferences above. One can simply assume that the correct welfare criterion does not discount future utility at the same rate as agents do, although both do so exponentially[3]. Although this alternative interpretation is available for two-periods, as we shall see, it does not allow as straightforward an extension to more periods.

We investigate the optimal allocation from the point of view of *self-0* subject to the constraint that θ is private information of *self-1*. The essential tension is between tailoring consumption to the taste shock and the self-1's

[2] Note that a taste shock for period $t = 2$ is not included in this expression. However the absence of the shock is only apparent since c_2 cannot depend on θ_2 and $E\theta_2 = 1$.

[3] Phelan (2002) studies such a model focusing on the long-run implications for the distribution of consumption.

constant higher desire for present consumption. This generates the trade-off between flexibility and commitment for *self-0*.

To solve for the allocation preferred by *self-0* let the amount of income available to be spent in periods 1 and 2 is y. We now set up the optimal direct truth telling mechanism given y. For simplicity we first consider the case where the shock take on two values: high, θ_h, with probability p and low, θ_l, with probability $1-p$. Sections 5 and 6 extend the analysis to allow for more types.

Mechanism Design Problem 2x2

$$v_2(y) \equiv \max_{\substack{c_{1h},c_{2h}\\ c_{1l},c_{2l}}} \{p[\theta_h u(c_{1h}) + U(c_{2h})] + (1-p)[\theta_l u(c_{1l}) + U(c_{2l})]\}$$

$$\theta_h u(c_{1h}) + \beta U(c_{2h}) \geq \theta_h u(c_{1l}) + \beta U(c_{2l}) \tag{1}$$

$$\theta_l u(c_{1l}) + \beta U(c_{2l}) \geq \theta_l u(c_{1h}) + \beta U(c_{2h}) \tag{2}$$

$$c_{1h} + c_{2h} \leq y$$

$$c_{1l} + c_{2l} \leq y$$

This problem maximizes the expected utility from consuming in $t = 1, 2$ total resources y from the point of view of the $t = 0$ self, subject to the constraint that θ is private information of self-1. In the budget constraints the interest rate is normalized to zero. The incentive constraints (1) and (2) reflect the fact that it must be in agent-θ's self interest to report their true type, thus obtaining the allocation that is intended for them.

The problem imposes a budget constraint for each state: there is no possibility for insurance across θ's. We later study some aspects of the problem that does allow insurance and discuss some reasons for focusing on the case without insurance.

Once we have solved for $v_2(\cdot)$ the optimal allocation for self-0 is

$$\max_{c_0} \{\theta_0 u(c_0) + \beta v_2(y_0 - c_0)\}$$

where y_0, c_0 and θ_0 represents the initial $t = 0$, income, consumption and taste shock, respectively. In what follows we concentrate attention on the mechanism design problem for the non-trivial periods.

In the next section we offer two alternative interpretations of the above model.

3 Alternative Interpretations

3.1 Schooling

One interesting reinterpretation of the model is for the choice between schooling and leisure choice. In many cases the relevant decision maker is not yet an adult so that we can interpret paternalism literally as a struggle between the preferences of parent and child. Alternatively, other adults may be altruistically concerned about children without parents and support paternalistic legislation.

The preference from the point of view of the child are given by the utility function:

$$\theta u\left(l\right)+\beta U\left(s\right)$$

where s represents schooling time and l represents other valuable uses of time. The taste parameter θ affects the relative valuation between schooling and other activities. The paternalist has preferences given by $\theta u\left(s\right)+U\left(s\right)$ so that more weight is given to schooling time.

The allocation of time is constrained by the endowment of 1 (normalized) unit of time:

$$s+l\leq 1$$

Note that in this example it is especially natural not to consider insurance: time cannot be transferred across agents.

3.2 Externalities

One interpretation for a divergence of preferences between the planner and the agents is when consumption of a good, x^2 say, by any agent generates positive externalities for the rest of the population. Agents do not internalize the effects of their consumption on other agents but the planner does.

To make this precise, suppose that the agent with taste shock θ obtains the following utility when the allocation is $(x,z)\equiv\left(x\left(\theta\right),z\left(\theta\right)\right)$

$$V\left(\theta,(x,z)\right)\equiv\theta u\left(x\left(\theta\right)\right)+\beta U\left(z\left(\theta\right)\right)+(1-\beta)\int U\left(x\left(\theta\right)\right)dF\left(\theta\right) \tag{3}$$

it follows that our welfare criterion will be:

$$W=\int V\left(\theta,(x,z)\right)dF\left(\theta\right)=\int\left[\theta u\left(x\left(\theta\right)\right)+U\left(z\left(\theta\right)\right)\right]dF\left(\theta\right).$$

This shows that we can represents $\theta u(x) + U(z)$ as the relevant utility function for agent-θ from the planner's point of view[4]. That is, such a representation leads to the same welfare functional.

4 Optimal Allocation

If $\beta = 1$ then we can implement the ex-ante first-best allocation given by the solution to:

$$\begin{aligned} \theta^s \frac{u'(c_{1s})}{U'(c_{2s})} &= 1 \\ c_{1s} + c_{2s} &= y \end{aligned}$$

for $s = h, l$. Obviously, with $\beta = 1$ this allocation can be implemented by giving self-1 full reign of choices over the budget constraint. We now show that with β below but sufficiently close to 1 we can also implement the ex-ante first-best allocation[5].

Lemma 1. There exists a $\beta^* < 1$ such that for $\beta \in [\beta^*, 1]$ the first-best allocation is implementable.

Proof. At $\beta = 1$ the incentive constraints are slack at the ex-ante first-best allocation. Thus, there is an interval around $\beta = 1$ for which the incentive constraints continue to hold. Indeed, one can define β^* to be the value of β for which the incentive constraint (2) holds with equality at the first best allocation. ■

For $\beta < \beta^*$, if we attempt to implement the (ex-ante) first-best allocation the incentive constraint (2) will be violated: it is attractive for the low type to claim being a high type to increase present consumption. The next proposition characterizes the optimum allocation in such cases.

Proposition 1. *The optimum can be attained with the budget constraint holding with equality:* $c_{1h} + c_{2h} = y$.

Define $\beta^{\#} \equiv \theta_l/\theta_h < \beta^*$ then:

[4]Note that this is not the utility actually attained by agent-θ which is given by the expression in (3).

[5]This result no longer holds with a continuum of types, see section 6.

(a) *if* $\beta > \theta_l/\theta_h$ *separation is optimal, i.e.* $c_{1h} > c_{1l}$ and $c_{2h} < c_{2l}$
(b) *if* $\beta < \theta_l/\theta_h$ *pooling is optimal, i.e.* $c_{1l} = c_{1h}$ and $c_{2l} = c_{2h}$
(c) *if* $\beta = \theta_l/\theta_h$ *both separating and pooling are optimal*

Proof. First, $\beta^* > \beta^\#$ because,

$$\begin{aligned}\beta^* &\equiv \theta_l \frac{u(c_h) - u(c_l)}{U(y - c_l) - U(y - c_h)} \\ &> \theta_l \frac{u'(c_h)(c_h - c_l)}{U'(y - c_h)(c_h - c_l)} \\ &= \theta_l \frac{u'(c_h)}{U'(y - c_h)} = \frac{\theta_l}{\theta_h} \equiv \beta^\#\end{aligned}$$

Now, consider the case where $\beta > \beta^\#$ and suppose that $c_{1h} + c_{2h} < y$. Then an increase in c_{1h} and a decrease in c_{2h} that holds $\theta_l u(c_{1h})/\beta + u(c_{2h})$ unchanged increases $c_{1h} + c_{2h}$ and the objective function. Such a change is incentive compatible because it strictly relaxes (1) and leaves (2) unchanged. It follows that we must have $c_{1h} + c_{2h} = y$ at an optimum. This also proves that separating is optimal in this case, proving part (a). Analogous arguments prove parts (b) and (c). ■

Note that for $\beta = \beta^\#$ there are optimal allocations with $c_{1h} + c_{2h} < y$; nevertheless, the proposition shows that in such cases there also exist (two) optima with $c_{1h} + c_{2h} = y$. For the case with $u(\cdot) = U(\cdot)$, if pooling is optimal then $c_{1l} = c_{1h} = y/2$.

Figure 1 below we show a typical situation. We set $u = U$ with $u(c) = c^{1-\sigma}/(1-\sigma)$ and use the following parameters: $\sigma = 2$, $\theta_h = 1.2$, $\theta_l = .8$, $p = .5$ and $y = 1$. The figure shows consumption in the first period as a function of β. We display the optimal c_1 for the high and low type. For comparison we also plot the optimal ex-post c_1, for both types. These are always higher than what the optimal allocation allows.

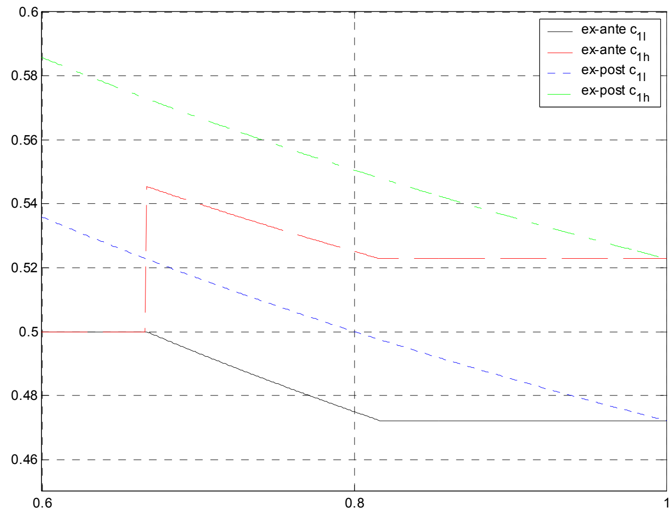

Figure 1: optimal c_1 with two shocks as a function of β.

5 Three Types: Money Burning

With two types we were able to characterize the optimal allocation. Proposition 1 showed that it enjoys certain nice properties. In particular, the budget constraint is binding for both types and we found simple necessary and sufficient conditions for a pooling or separating outcome to be optimal. In this section we illustrate the types of difficulties one encounters when trying to extend these results to situations with more than two types by studying the case with three types.

Suppose there are three possible taste shocks: $\theta_h > \theta_m > \theta_l$. It is easy to see that if pooling is not optimal then $c_{1l} < c_{1m} < c_{1h}$ and two incentive constraints bind: (i) the low-type considering the medium-type's allocation and (ii) medium-type considering the high-type's allocation. By similar arguments to Proposition 1 one can show that the budget constraint holds with equality for θ_l and θ_h. However, no similar argument applies to θ_m.

Indeed, it is simple to construct robust examples where it is optimal for the budget constraint for θ_m to hold with strict inequality. In this sense, 'money burning' may be optimal. The figures below illustrate one such case. This figure uses $\theta_l = .8$, $\theta_m = .85$ and $p_l = .59$, $p_m = .01$ and $p_h = .41$ (θ_h is set so that $E\theta = 1$). Figure 2 shows the optimal allocation as a function of β for the range of β's for which complete separation, $c_{1l} < c_{1m} < c_{1h}$, of the three types is optimal. Figure 3 plots the optimal level of $c_{1m} + c_{2m}$ as a function of β for the same range.

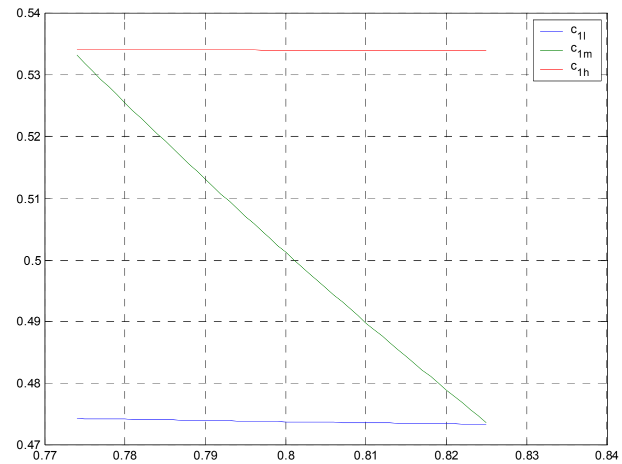

Figure 2: optimal c_1 as a function of β in the separating region with 3 shock types.

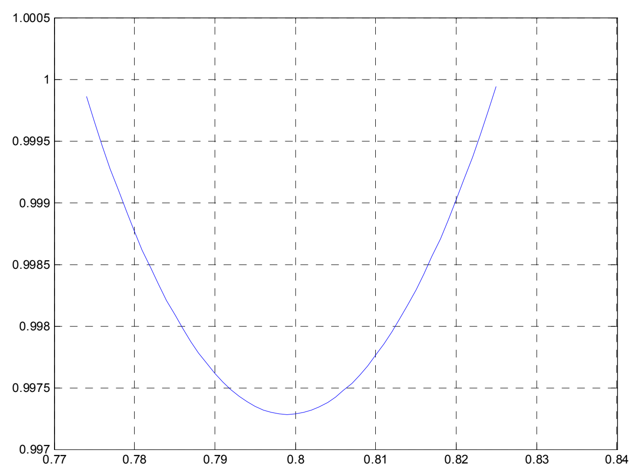

Figure 3: Optimal $c_{1,m} + c_{2,m}$ as a function of β.

A similar argument to part (b) of Proposition 1 can be used to establish that pooling must occur between m and h whenever $\beta < \theta_m/\theta_h$. However, there are cases with $\beta > \theta_m/\theta_h$ where pooling m and h is optimal. As the figure illustrates, this occurs in the present example for β just below $.775 > \theta_m/\theta_h = .66$.

6 Continuum of Types

Although one is able to exhaustively characterize the case with two types, we saw above that the case with three type case imposes new difficulties. It is tempting to conclude that not much can be said for situations with situations with more than two types. Fortunately, we are able to find a situation for which progress can be made.

In this section we study the case with a continuum of types and find an assumption on the distribution of θ under which we can solve the optimal allocation. Indeed, when this assumption is satisfied the optimal mechanism takes an extremely very simple form. We also show that when this assumption on the distribution does not hold this simple mechanism is not optimal.

Suppose we have a continuous distribution of types that can be represented by a density $f(\theta)$ over the interval $\left[\underline{\theta}, \overline{\theta}\right]$. The agent faces a truth telling mechanism $(c(\theta), C(\theta))$ or equivalently $(u(\theta), U(\theta))$ and solves,

$$v(\theta) \equiv \max_{\hat{\theta} \in \left[\underline{\theta}, \overline{\theta}\right]} \left\{ \frac{\theta}{\beta} u\left(\hat{\theta}\right) + U\left(\hat{\theta}\right) \right\}.$$

It follows that we can represent $v(\theta)$ as

$$v(\theta) = \int_{\underline{\theta}}^{\theta} \frac{1}{\beta} u(\theta)\, d\theta + v(\underline{\theta}) \tag{4}$$

[see Milgrom and Segal (2002)]. One can also show that incentive compatibility implies that u must be non-decreasing. Thus, incentive compatibility implies (4) and monotonicity of u. We can now state the mechanism design problem for the continuous type case.

Continuous-type Mechanism Design Problem:

$$\max \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta u(\theta) + U(\theta)\right] f(\theta)\, d\theta$$

$$\begin{aligned}\frac{\underline{\theta}}{\beta}u\left(\underline{\theta}\right)+U\left(\underline{\theta}\right)+\int_{\underline{\theta}}^{\theta}\frac{1}{\beta}u\left(\theta\right)d\theta &= \frac{\theta}{\beta}u\left(\theta\right)+U\left(\theta\right)\\ c\left(u\left(\theta\right)\right)+C\left(U\left(\theta\right)\right) &= y\\ u\left(\theta'\right) &\geq u\left(\theta\right)\text{ for }\theta'\geq\theta\end{aligned}$$

Note that this problem is convex: the objective function is linear in u, U and the constraint set is convex (they are linear except for the resource constraint which is strictly convex). Below we apply duality theory to find the maximum.

For any density function f and support $\left[\underline{\theta},\overline{\theta}\right]$ define the function

$$g\left(\hat{\theta}\right)\equiv\frac{E\left[\theta|\theta\geq\hat{\theta}\right]}{\hat{\theta}}$$

Note that $g\left(\theta\right)$ is continuos and that $g\left(\overline{\theta}\right)=1$. Thus, there is always an interval $\left[\theta_p,\overline{\theta}\right]$ for which $g\left(\theta\right)\leq 1/\beta$ for $\theta\in\left[\theta_p,\overline{\theta}\right]$.

The following lemma shows that pooling always occurs in the upper tail.

Lemma. An optimal contract pools all the agents with $\theta\in\left[\theta_p,\overline{\theta}\right]$.

The next proposition characterizes the rest of the optimal allocation under the following assumption on the density f and β.

A1 The density f is differentiable and satisfies

$$\theta\frac{f'\left(\theta\right)}{f\left(\theta\right)}\geq-\frac{2-\beta}{1-\beta}$$

for all $\theta\leq\theta_p$ where where θ_p is the highest solution θ_p to $g\left(\hat{\theta}\right)\leq 1/\beta$.

Assumption A1 bounds the elasticity of f from below by a negative number that depends on β. Note that we do not need to impose the bound on the whole support of f, only on $\theta\leq\theta_p$. The lower bound is continuous and decreasing in β. The highest lower bound of -2 is attained for $\beta=0$ and as $\beta\rightarrow 1$ the lower bound goes to $-\infty$[6].

[6] This is consistent with the fact that for $\beta=1$ no restriction on f is required for the allocation we propose below to be optimal.

Note that for any density with $\theta f'(\theta)/f(\theta)$ bounded from below there exists an interval $\left[0, \beta^f\right]$ for which A1 is satisfied.

Many densities satisfy A1 trivially for all β, for example A1 is trivially satisfied for all density functions that are non-decreasing. No doubt this in itself provides a large class of interesting cases. However, this class does not include any distributions with unbounded support. It turns out that in many interesting cases one can show that A1 is satisfied. For example, A1 holds for the exponential distribution and for the log-normal, Pareto and gamma distributions with some restriction on the parameter space.

We now show that under assumption A1 agents with $\theta \geq \theta_p$ are pooled and the rest of the agents are offered their unconstrained optimum. That is, the mechanism offers the whole budget line to the left of some point (c_1^*, c_2^*), given by the unconstrained optimum of the (ex-post self) θ_p-agent[7].

Proposition 2. The following allocation is optimal if and only if A1 holds:

$$c(\theta) = \begin{cases} c^*(\theta) & \text{for } \theta < \theta_p \\ c^*(\theta_p) & \text{for } \theta \geq \theta_p \end{cases}$$

where θ_p is θ_p is the highest solution θ_p to $g\left(\hat{\theta}\right) \leq 1/\beta$ and $c^*(\theta)$ is the best ex-post allocation for agent-θ : $c^*(\theta) \equiv \arg\max_{c_1+c_2 \leq y} \{u(c_1) + \beta U(c_2)\}$.

Proof. We make use of saddle-point results for linear spaces [see Luenberger (1969, Chapter 8)]. The associated Lagrangian becomes, upon integrating by parts,

$$\begin{aligned} & \mathcal{L}(u, U, \mu, \lambda) \\ \equiv & \int_{\underline{\theta}}^{\overline{\theta}} \left\{ [\theta u(\theta) + U(\theta)] f(\theta) - \left[\frac{\theta}{\beta} u(\theta) + U(\theta)\right] \mu'(\theta) \right. \\ & \left. -\mu(\theta) \frac{1}{\beta} u(\theta) - \lambda(\theta) [c(u(\theta)) + C(U(\theta)) - y] f(\theta) - \gamma'(\theta) u(\theta) \right\} d\theta \\ & +\mu\left(\overline{\theta}\right) \left[\frac{\overline{\theta}}{\beta} u\left(\overline{\theta}\right) + U\left(\overline{\theta}\right)\right] - \mu(\underline{\theta}) \left[\frac{\underline{\theta}}{\beta} u(\underline{\theta}) + U(\underline{\theta})\right] \\ & +\gamma\left(\overline{\theta}\right) u\left(\overline{\theta}\right) - \gamma(\underline{\theta}) u(\underline{\theta}) \end{aligned}$$

[7] We can offer the whole budget set $c_1+c_2 \leq y$ (instead of just the budget line $c_1+c_2 = y$) to the left of (c_1^*, c_2^*) without any loss since agents will obviously find it optimal to locate themselves on the budget line.

The partials w.r.t. $U\left(\overline{\theta}\right)$ yields,

$$\frac{\partial \mathcal{L}}{\partial U\left(\overline{\theta}\right)} = \mu\left(\overline{\theta}\right)$$

we will show below that $\mu\left(\overline{\theta}\right)$ is non-positive, then since $U\left(\overline{\theta}\right)$ can be lowered this implies that $\mu\left(\overline{\theta}\right) = 0$. Likewise,

$$\frac{\partial \mathcal{L}}{\partial u\left(\overline{\theta}\right)} = \mu\left(\overline{\theta}\right)\frac{\overline{\theta}}{\beta} + \gamma\left(\overline{\theta}\right) = 0$$

implies that $\gamma\left(\overline{\theta}\right) = 0$.

The f.o.c.'s for u and U are,

$$\left(f\left(\theta\right) - \frac{1}{\beta}\mu'\left(\theta\right)\right)\theta - \frac{1}{\beta}\mu\left(\theta\right) = \lambda\left(\theta\right)c'\left(u\left(\theta\right)\right)f\left(\theta\right) + \gamma'\left(\theta\right) \quad (5)$$

$$f\left(\theta\right) - \mu'\left(\theta\right) = \lambda\left(\theta\right)C'\left(U\left(\theta\right)\right)f\left(\theta\right) \quad (6)$$

Separating region. We first focus on $\theta < \theta_p$. We take $\gamma\left(\theta\right) = \gamma'\left(\theta\right) = 0$ for this region. At the proposed allocation we have that:

$$c'\left(u\left(\theta\right)\right) = \frac{\theta}{\beta}C'\left(U\left(\theta\right)\right)$$

since $\theta u'/\beta U' = 1$, $c' = 1/u'$ and $C' = 1/U'$ for $\theta < \theta_p$. Substituting this into (5) and (6)

$$\mu\left(\theta\right) = -\left(1-\beta\right)\theta f\left(\theta\right)$$

$$\mu'\left(\theta\right) = -\left(1-\beta\right)f\left(\theta\right) - \left(1-\beta\right)\theta f'\left(\theta\right) < 0$$

if and only if

$$\theta\frac{f'\left(\theta\right)}{f\left(\theta\right)} > -1$$

The sign of $\lambda\left(\theta\right)$ is equal to the sign of $f - \mu'$ thus

$$\theta\frac{f'\left(\theta\right)}{f\left(\theta\right)} \geq -\frac{2-\beta}{1-\beta}$$

Pooling region. Next, we focus on $\theta \geq \theta_p$. At the proposed allocation we have that,

$$c'\left(u\left(\theta\right)\right) = \frac{\theta_p}{\beta}C'\left(U\left(\theta\right)\right),$$

setting $\lambda(\theta) = 0$ in (5) and (6) we get obtain

$$\gamma'(\theta) = \left(f(\theta) - \frac{1}{\beta}\mu'(\theta)\right)\theta - \frac{1}{\beta}\mu(\theta) \tag{7}$$

$$\mu'(\theta) = f(\theta) \tag{8}$$

integrating (8) using the boundary condition $\mu\left(\overline{\theta}\right) = 0$ we obtain:

$$\mu(\theta) = -\int_{\theta}^{\overline{\theta}} f(\theta)\, d\theta = -(1 - F(\theta))$$

substituting this into into (7) we obtain,

$$\gamma'(\theta) = -\left(\frac{1}{\beta} - 1\right) f(\theta)\,\theta + \frac{1}{\beta}(1 - F(\theta)),$$

which can be integrated to yield,

$$\gamma\left(\overline{\theta}\right) - \gamma(\theta) = \int_{\theta}^{\overline{\theta}} \gamma'(\theta) = -\left(\frac{1}{\beta} - 1\right)\int_{\theta}^{\overline{\theta}} f(\theta)\,\theta + \frac{1}{\beta}\int_{\theta}^{\overline{\theta}} (1 - F(\theta)).$$

Using the fact that:

$$\int_{\theta}^{\overline{\theta}} \left(1 - F\left(\tilde{\theta}\right)\right) d\tilde{\theta} = (1 - F(\theta))\left\{E\left[\tilde{\theta}|\tilde{\theta} \geq \theta\right] - \theta\right\}$$

(which is obtained by integration by parts) and setting $\gamma\left(\overline{\theta}\right) = 0$ we find:

$$\gamma(\theta) = (1 - F(\theta))\,\theta\left\{\frac{1}{\beta} - \frac{E\left[\tilde{\theta}|\tilde{\theta} \geq \theta\right]}{\hat{\theta}}\right\}.$$

Thus $\gamma(\theta) \geq 0$ if and only if,

$$E\frac{\left[\theta|\theta \geq \hat{\theta}\right]}{\hat{\theta}} \leq \frac{1}{\beta},$$

for all $\theta \geq \theta_p$.

To summarize, we have constructed positive Lagrange (λ, μ, γ) multipliers. Since the problem is convex verifying the first order conditions establishes that c and (λ, μ, γ) are a saddle-point of the Lagrangian function. Thus c is optimal. Conversely we have seen that if A1 does not hold then the allocation cannot generate a positive multiplier for μ, since the problem is convex this implies that c cannot be optimal if A1 fails. ■

7 Extension to N Periods

We now extend the previous analysis to N "non-trivial" periods. We write the problem recursively with $k \geq 3$ remaining periods as follows.

N period Mechanism Design Problem

$$v_k(y) = \max_{c,\, y} \int \left[\theta u(c(\theta)) + v_{k-1}(y'(\theta))\right] dF(\theta)$$

$$\theta u(c(\theta)) + \beta v_{k-1}(y'(\theta)) \geq \theta_h u\left(c\left(\hat{\theta}\right)\right) + \beta v_{k-1}\left(y'\left(\hat{\theta}\right)\right) \text{ for all } \theta, \hat{\theta} \in \Theta$$

$$c(\theta) + y'(\theta) \leq y \text{ for all } \theta \in \Theta$$

This is exactly the same general structure as the problem analyzed previously, the only difference is that v_{k-1} has substituted for the role of U. Since we did not require any special structure on U for our analysis, except concavity and monotonicity, all the previous results go through.

Proposition 3. If A1 holds then the optimal mechanism in the N-period problem can be characterized as imposing a minimum amount of saving $S_t(y_t)$ for period t (i.e. agents can consume in period t at most $y_t - S_t(y_t)$) and this minimum is a function of resources y_t.

Note that it is crucial for the simple recursive structure of this extension that the agent's preferences are hyperbolic. Indeed the alternative setup where the planner and the agent both discount exponentially at different rates does not offer have such a simple representation.

The ease with which this multi-period problem can be studied should be contrasted with the difficulties that arise in studying equilibria of the simplest setups with hyperbolic discounting (e.g. Laibson, Krusell and Smith). The vantage point of designing the game hyperbolic discounters play cannot be discounted.

We next consider the case with CRRA preferences. It is easy to see that in this case the optimum mechanism is homothetic so that c_t and y_{t+1} at each stage evolves proportionally to y_t. If in addition A1 holds it follows that the mechanism imposes a *minimum saving rate* for each period.

Proposition 4. If A1 holds and $u(c) = c^{1-\sigma}/(1-\sigma)$ the optimal mechanism in the N-period problem can be characterized as imposing a minimum saving rate s_t for each period t. The saving rate is independent of period t resources.

Another property of this solution under A1 is worth mentioning. Suppose that agents can save behind the planner's back as in Cole and Kocherlakota (2001). Such hidden saving obviously reduce the set of allocations that are incentive compatible. A relevant question is whether our mechanism remains with this smaller set. It turns out that it does.

Under assumption A1 the mechanism imposes only a minimum on savings in each period. Thus, the agent is implicitly allowed to save any additional amount with the principal. However, the agent choses not to do so. By saving behind the principal's back the agents can do no better: the principal is already maximizing the subsequent utility given the resources. That is, hidden savings are dominated by savings with the principal. It follows that the allocation remains incentive compatible when we add that agents can deviate with hidden savings. We summarize this discussion in the following proposition.

Proposition 5. If A1 holds then optimal mechanism in the N-period problem without hidden savings is also implementable, and thus optimal, if we allow agents hidden savings.

8 Optimal Allocations Assuming no Money Burning

In this section we study cases where A1 may fail. Specifically we show that without A1 the allocation described in proposition 1 can be improved upon by drilling holes into the separating section. That is, removing intervals in regions where A1 is not satisfied improves the planners welfare. This helps understand why condition $A1$ is required for the result.

When an interval is removed agents that previously found their tangency in this region will move either to one edge or the other. The critical issue is the relative fraction of agents moving to the left versus the right. The planner gains from agents that move to the left but loses from those that move to the right.

How many agents move to the right versus the left depends positively on f'. This is why f' plays a critical role in A1. In particular, if $f' > 0$ more agents move to the right and such a change is clearly undesirable.

To make this idea precise suppose that we are offering a segment of the budget line between the tangency point for θ_L and that of θ_H, with associated allocation c_L and c_H. Define the θ^* that is indifferent from the allocation c_L and c_H then $\theta^* \in (\theta_L, \theta_H)$ for $\theta_H > \theta_L$. We know then that the $\theta \in (\theta^*, \theta_H)$ types will pick the c_H allocation and the $\theta \in (\theta_L, \theta^*)$ types will pick the c_L allocation.

Let $\Delta(\theta_H, \theta_L)$ be the utility gain for the planner of such a move (normalizing income, $y = 1$)

$$\begin{aligned}
\Delta(\theta_H, \theta_L) \equiv & \int_{\theta^*(\theta_H,\theta_L)}^{\theta_H} \{\theta u(c^*(\theta_H)) + U(y - c^*(\theta_H))\} f(\theta)\, d\theta \\
& + \int_{\theta_L}^{\theta^*(\theta_H,\theta_L)} \{\theta u(c^*(\theta_L)) + U(y - c^*(\theta_L))\} f(\theta)\, d\theta \\
& - \int_{\theta_L}^{\theta_H} \{\theta u(c^*(\theta)) + U(y - c^*(\theta))\} f(\theta)\, d\theta
\end{aligned}$$

where the function $c^*(\theta)$ is defined implicitly by

$$\theta u'[c^*(\theta)] = \beta U'(y - c^*(\theta)) \tag{9}$$

and $\theta^*(\theta_H, \theta_L)$ is then defined by

$$\begin{aligned}
& \theta^*(\theta_H, \theta_L)\, u(c^*(\theta_H)) + \beta U(y - c^*(\theta_H)) \\
= & \theta^*(\theta_H, \theta_L)\, u(c^*(\theta_L)) + \beta U(y - c^*(\theta_L))
\end{aligned} \tag{10}$$

Notice that $\Delta(\theta_L, \theta_L) = 0$.

The following lemma regarding the partial derivative of $\Delta(\theta_H, \theta_L)$ is used below in the proof of the main proposition of this section.

Lemma. The partial of $\Delta(\theta_H, \theta_L)$ with respect to θ_H can be expressed as:

$$\frac{\partial \Delta}{\partial \theta_H}(\theta_H, \theta_L) = S(\theta_H; \theta^*) \frac{u'(c^*(\theta_H))}{\beta} \frac{\partial c^*(\theta_H)}{\partial \theta_H}$$

where $S(\theta; \theta^*)$ is defined by,

$$S(\theta,\theta^*) \equiv (y-\beta)(\theta-\theta^*)\theta^* f(\theta^*) - \int_{\theta^*}^{\theta}\left(\theta-\beta\tilde{\theta}\right) f\left(\tilde{\theta}\right) d\tilde{\theta}$$

Since $u'(c^*(\theta_H)) > 0$ and $\frac{\partial c^*(\theta_H)}{\partial \theta_H} > 0$, then $sign(\Delta_1) = sign(S(\theta_H,\theta^*))$.

The proof involves a lot of grinding and is included in the appendix.

Proposition 3. If A1 holds then $\theta_L \in \arg\max_{\theta_H \geq \theta_L} \Delta(\theta_H,\theta_L)$. In other words if A1 holds then punching holes into any offered interval is not optimal. Conversely, it is optimal to remove any interval where it doesn't hold.

Proof. From the lemma we only need to sign $S(\theta_H,\theta^*)$. Clearly, $S(\theta^*,\theta^*) = 0$. Taking derivatives we also get that

$$\frac{\partial S(\theta,\theta^*)}{\partial \theta} = [1-\beta]\theta^* f(\theta^*) - (1-\beta)\theta f(\theta) - \int_{\theta^*}^{\theta} f\left(\tilde{\theta}\right) d\tilde{\theta}$$

Notice that

$$\left.\frac{\partial S(\theta,\theta^*)}{\partial \theta}\right|_{\theta^*} = 0$$

$$\frac{\partial^2 S(\theta,\theta^*)}{(\partial\theta)^2} = -(2-\beta) f(\theta) - (1-\beta)\theta f'(\theta)$$

Note that $\partial^2 S(\theta,\theta^*)/(\partial\theta)^2$ does not depend on θ^*, just on θ. It follows that $sign\left(\frac{\partial^2 S(\theta,\theta^*)}{(\partial\theta)^2}\right) \leq 0$ if and only if

$$\frac{\theta f'(\theta)}{f(\theta)} \geq -\frac{2-\beta}{1-\beta} \tag{11}$$

That is, if A1 holds. Integrating $\partial^2 S(\theta,\theta^*)/(\partial\theta)^2$ twice:

$$S(\theta_H,\theta^*) = \int_{\theta^*}^{\theta_H}\int_{\theta^*}^{\theta} \frac{\partial^2 S\left(\tilde{\theta},\theta^*\right)}{\left(\partial\tilde{\theta}\right)^2} d\tilde{\theta} d\theta$$

Thus $S(\theta_H,\theta^*) \leq 0$ if $A1$ holds.

This implies then that $\Delta_1(\theta,\theta_L) \leq 0$ for all $\theta \geq \theta_L$ if A1 holds; and

$$\Delta(\theta_H,\theta_L) = \int_{\theta_L}^{\theta_H} \Delta_1(\theta;\theta_L) d\theta$$

so that

$$\frac{\theta f'(\theta)}{f(\theta)} \geq -\frac{2-\beta}{1-\beta} \Rightarrow \Delta(\theta_H, \theta_L) \leq 0 \ \ ; \text{for all } \theta_H \ \text{ and } \ \theta_L$$

and clearly $\theta_L \in \arg\max_{\theta_H \geq \theta_L} \Delta(\theta_H, \theta_L)$. In other words if A1 holds then punching holes into any offered interval is not optimal.

The converse is also true: if A1 does not hold for some open interval $\theta \in (\theta_1, \theta_2)$ then the previous calculations clearly show that it is optimal to remove the whole interval. In other words,

$$\begin{aligned}(\theta_1, \theta_2) \ \in \ & \arg\max_{\theta_L, \theta_H} \Delta(\theta_H, \theta_L) \\ & \text{s.t. } \theta_1 \leq \theta_L \leq \theta_H \leq \theta_2\end{aligned}$$

This concludes the proof. ■

Appendix

Proof of Lemma 2.

$$\begin{aligned}\Delta_1(\theta_H, \theta_L) \ = \ & [\theta_H u(c^*(\theta_H)) + U(y - c^*(\theta_H))] f(\theta_H) \\ & - [\theta^*(\theta_H, \theta_L) u(c^*(\theta_H)) + U(y - c^*(\theta_H))] f(\theta^*) \frac{\partial \theta^*}{\partial \theta_H} \\ & + \int_{\theta^*(\theta_H, \theta_L)}^{\theta_H} \{\theta u'(c^*(\theta_H)) - U'(y - c^*(\theta_H))\} f(\theta) \frac{\partial c^*(\theta_H)}{\partial \theta_H} d\theta \\ & + \{\theta^*(\theta_H, \theta_L) u(c^*(\theta_L)) + U(y - c^*(\theta_L))\} f(\theta^*) \frac{\partial \theta^*}{\partial \theta_H} \\ & - [\theta_H u(c^*(\theta_H)) + U(y - c^*(\theta_H)) f(\theta_H)]\end{aligned}$$

Combining terms,

$\Delta_1(\theta_H, \theta_L) =$

$$\begin{aligned}& \left(\int_{\theta^*(\theta_H, \theta_L)}^{\theta_H} \{\theta u'(c^*(\theta_H)) - U'(y - c^*(\theta_H))\} f(\theta) d\theta\right) \frac{\partial c^*(\theta_H)}{\partial \theta_H} \\ & + \{\theta^*(\theta_H, \theta_L)[u(c^*(\theta_L)) - u(c^*(\theta_H))] + U(y - c^*(\theta_L)) - U(y - c^*(\theta_H))\} f(\theta^*) \frac{\partial \theta^*}{\partial \theta_H}\end{aligned}$$

Now, from (10) we have

$$\theta u'\left[c^*\left(\theta\right)\right]-U'\left(y-c^*\left(\theta\right)\right)=\left[\frac{\beta-1}{\beta}\right]\theta u'\left[c^*\left(\theta\right)\right]$$

Substituting above

$\Delta_1\left(\theta_H,\theta_L\right)=$

$$\left(\int_{\theta^*\left(\theta_H,\theta_L\right)}^{\theta_H}\left(\theta-\frac{1}{\beta}\theta_H\right)f\left(\theta\right)d\theta\right)u'\left(c^*\left(\theta_H\right)\right)\frac{\partial c^*\left(\theta_H\right)}{\partial\theta_H}$$
$$+\left\{\theta^*\left(\theta_H,\theta_L\right)\left[u\left(c^*\left(\theta_L\right)\right)-u\left(c^*\left(\theta_H\right)\right)\right]+U\left(y-c^*\left(\theta_L\right)\right)-U\left(y-c^*\left(\theta_H\right)\right)\right\}f\left(\theta^*\right)\frac{\partial\theta^*}{\partial\theta_H}$$

we also have that from (9)

$$-\frac{\theta^*\left(\theta_H,\theta_L\right)}{\beta}\left[u\left(c^*\left(\theta_L\right)\right)-u\left(c^*\left(\theta_H\right)\right)\right]=\left\{U\left(y-c^*\left(\theta_L\right)\right)-U\left(y-c^*\left(\theta_H\right)\right)\right\}$$

So,

$$\begin{aligned}\Delta_1\left(\theta_H,\theta_L\right) &= \left\{\left[\frac{1}{\beta}-1\right]\theta^* f\left(\theta^*\right)\right\}\left[u\left(c^*\left(\theta_H\right)\right)-u\left(c^*\left(\theta_L\right)\right)\right]\frac{\partial\theta^*}{\partial\theta_H}\\ &\quad-\left(\int_{\theta^*}^{\theta_H}\left(\frac{1}{\beta}\theta_H-\theta\right)f\left(\theta\right)d\theta\right)u'\left(c^*\left(\theta_H\right)\right)\frac{\partial c^*\left(\theta_H\right)}{\partial\theta_H}\end{aligned}$$

Differentiating (10) we obtain:

$$\frac{\partial\theta^*}{\partial\theta_H}\left[u\left(c^*\left(\theta_H\right)\right)-u\left(c^*\left(\theta_L\right)\right)\right]=-\left[\theta^* u'\left(c^*\left(\theta_H\right)\right)-\beta U'\left(y-c^*\left(\theta_H\right)\right)\right]\frac{\partial c^*\left(\theta_H\right)}{\partial\theta_H}$$

Using the fact that $\theta u'\left[c^*\left(\theta\right)\right]-\beta U'\left(1-c^*\left(\theta\right)\right)=0$ this implies

$$\frac{\partial\theta^*}{\partial\theta_H}\left[u\left(c^*\left(\theta_H\right)\right)-u\left(c^*\left(\theta_L\right)\right)\right]=\left[\theta_H-\theta^*\right]u'\left[c^*\left(\theta_H\right)\right]\frac{\partial c^*\left(\theta_H\right)}{\partial\theta_H}$$

Substituting back the result follows. ■

References

[1] Athey, Susan, Andy Atkeson and Patrick Kehoe (2002) "On the Optimality of Transparent Monetary Policy", Working Paper 613.

[2] Atkeson, Andrew and Robert E. Lucas, Jr. "On Efficient Distribution with Private Information." Review of Economic Studies, 59, no.3 (July 1992): 427-53.

[3] Bernheim, B. Douglas, Jonathan Skinner and Steven Weinberg (2001) "What Accounts for the Variation in Retirement Wealth Among US Households?", American Economic Review v91, n4 (September): 832-57

[4] Cole, Harold L. and Narayana R. Kocherlakota (2001) "Efficient Allocations with Hidden Income and Hidden Storage", Review of Economic Studies v68, n3 (July): 523-42.

[5] DellaVigna, Stefano and Ulrike Malmendier (2003) "Contract Design and Self-Control: Theory and Evidence", mimeo.

[6] Diamond, Peter (1977) "A Framework for Social Security Analysis", Journal of Public Economics, v8, n3: 275-98.

[7] Diamond, Peter and Botand Koszegi (2002) "Quasi-Hyperbolic Discounting and Retirement", mimeo, MIT.

[8] Feldstein, Martin S. "The Optimal Level of Social Security Benefits" Quarterly Journal of Economics v10, n2 (May 1985): 303-20

[9] Imrohoroglu, Ayse, Selahattin Imrohoroglu and Douglas H. Joines (2000) "Time Inconsistent Preferences and Social Security", Federal Reserve of Minneapolis Discussion Paper Number 136

[10] Krusell, Per and Anthony Smith (2002) "Consumption Saving Decisions with Quasi-Geometric Discounting", mimeo.

[11] Laibson, David (1996) "Hyperbolic Discount Functions, Undersaving, and Savings Policy", National Bureau of Economic Research Working Paper: 5635.

[12] Leahy, John and xxx "The Social Discount Rate", mimeo, New York University.

[13] Luenberger, David G. (1969) "Optimization by Vector Space Methods", John Wiley & Sons, Inc.

[14] O'Donaghue, Ted and Matthew Rabin (2003) "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes", mimeo, UC Berkeley.

[15] Phelan, Christopher (2002) "Opportunity and Social Mobility", mimeo, Federal Reserve Bank of Minneapolis.