Nonlinear Capital Taxation

Iván Werning, MIT

June 8, 2010 (9:20am)
Preliminary and Incomplete
do not circulate

Abstract

This paper studies tax design in a dynamic Mirrleesian economy, where agents face labor productivity shocks that are private information. I characerize a system of nonlinear taxes on savings that implement any incentive compatible allocation. I restrict the savings tax to be independent of the current state. The tax schedule is differentiable under quite general conditions and its derivative, the marginal tax, coincides with the wedge in the agent's intertemporal Euler equation. Although I allow for nonlinear schedules, a linear tax often suffices. Otherwise, the tax can always be made convex—in this sense, progressive taxation is always feasible. Finally, I show how the savings tax can be made independent of the history of shocks.

1 Introduction

When insurance is imperfect, individuals may wish to smooth their consumption by saving and dissaving a riskless asset. However, if insurance is limited because of asymmetric information, it may be best to restrict free access to savings, since optimal incentive compatible allocations may not be attainable otherwise [Rogerson, 1985]. Some distortion must be introduced in savings choices to achieve constrained-efficient allocations. In dynamic Mirrleesian models, featuring privately observed labor productivity shocks, this leads to the possibility that capital taxation may be desirable (Diamond and Mirrlees, 1978; Golosov et al., 2003; Kocherlakota, 2005). The question I address here is the precise form of such a distortionary tax system.

Of course, one possibility is to ban individuals from saving altogether. Such a tax system would essentially replicate the direct mechanism used in the revelation principle, which controls consumption directly. Less extreme alternatives seem more palatable. At a practical level, they are more comparable to actual or potential tax policy. On a theoretical level, they can be more revealing, as when marginal tax rates make explicit the distortions that are implicit in the allocation.

Kocherlakota [2005] and Albanesi and Sleet [2006] provide alternative tax systems that are less extreme than direct mechanisms. Kocherlakota studied a tax system where wealth is taxed linearly, with a tax rate that varies with the past history of shocks as well as with the current shock. The dependence on the current shock is crucial. It ensures that saving is never optimal for the agent regardless of the reporting strategy. Albanesi and Sleet study a model with i.i.d. shocks and provide an implementation where wealth is taxed jointly with current labor income in a nonlinear way. In both these implementations, the dependence of the wealth tax on the current shock makes the after-tax rate of return faced by agents risky.

The goal of this paper is to offer a new implementation with some desirable features. I study tax systems that allow the savings tax to be nonlinear but require it to be independent of the current shock. Thus, the after-tax rate of return faced by agents is riskfree. I show that the tax schedule can be made differentiable quite generally. The marginal tax rate then coincides with the intertemporal wedge in the standard consumption Euler equation.[1] Interestingly, although the tax schedule is allowed to be nonlinear, I show that a linear tax suffices in some cases. Otherwise, the tax schedule can always be made convex. Thus, progressive taxation, in the sense of increasing marginal tax rates, is always

[1]The intertemporal wedge is defined, for any given allocation, as the proportional adjustment in the rate of return on capital that is required for the standard consumption Euler equation, i.e. letting $U_{c,t}$ denote marginal utility at time t the wedge τ_t is such that $U_{c,t} = \beta(1+r_t)(1-\tau_t)\mathbb{E}_t[U_{c,t+1}]$.

feasible. Finally, I show two ways of removing history dependence in the savings tax. These results do not require utility to be separable in consumption and labor. All of these features contrast quite sharply with previous implementations.

Some well-known examples in the literature seem directly at odds with my results. These examples have motivated the need for state-dependent savings taxes, that is, for conditioning on the current shock or labor income. For instance, Kocherlakota [2004, 2005] shows that if the savings tax is differentiable and state-independent, then the agent may plan on a "double deviation": saving some positive amount and reporting a lower shock in the next period. In the example, productivity types are discrete and the optimal allocation has the agent indifferent between reporting truthfully and misreporting to be the productivity type immediately below. When the agent is allowed to save, any small amount of positive savings breaks this indifference in favor of misreporting. Similarly, although a savings tax may discourage savings under the truth-telling strategy, saving some positive amount is strictly optimal under a misreporting strategy. As a result, the double deviation is always desirable and the intended implementation fails. These examples allow one to conclude that optimal allocations, discrete shocks, differentiable tax functions and state-independent taxes are incompatible. Kocherlakota relaxes the latter assumption and proceeds to characterize state-dependent taxes.

This paper takes a different route. I show that a nonlinear tax schedule allows for an implementation without conditioning on the current shock. I do not rule out kinks a priori, only to show that kinks are mostly unnecessary. Indeed, the tax schedule can be made differentiable quite generally.

Before moving on, it is important to understand that kinks are inherently natural with finite shocks to productivity. Even in the static Mirrlees [1971] model, kinks in the labor-income tax are optimal in this case (see Section 3).[2] It is only natural to expect the same for the savings tax in dynamic extensions. In that case, it seems inconsistent to insist on the differentiability of the savings tax, but allow for kinks in the labor-income tax. At a more fundamental level, it seems onerous to expect a model to yield smooth solutions, such as differentiable tax schedules, when primitives, such as the distribution of productivity, are not assumed continuous.

In any case, I show that kinks are unnecessary under plausible conditions. In particular, the savings tax schedule can be made differentiable when productivity shocks are continuously distributed. The "double deviation" example, discussed above, had a finite

[2] In Mirrlees (1971), kinks may be required in the income-tax schedule even when the distribution of productivity is continuous if bunching is optimal over some interval of productivities. Unlike those due to the assumption of finite shocks, these kinks have no counterpart in the savings tax schedule I characterize.

number of shocks. I show that this assumption is not innocuous: when shocks are continuously distributed, double deviations cease to be a problem. To understand this, recall that indifference was crucial for the desirability of double deviations. But with a continuous distribution of shocks, indifference is a knife-edge phenomena that occurs with probability zero.

Moreover, even with finite productivity shocks, kinks are unimportant. First of all, for a finite but large number of productivity shocks, I show that the kinks required to implement the optimal allocation are minimal. That is, they shrink and vanish (with left and right derivatives converge to each other) as the number of shocks is expanded. More importantly, even when the number of productivity types is small, differentiable tax schedules can implement allocations that come arbitrarily close to optimal ones. In fact, the closure of the set of allocations achievable by differentiable tax schedules is the full set of incentive compatible allocations. Thus, essentially no welfare is lost by insisting on differentiable tax schedules.

Although I allow the savings tax to be nonlinear, I show that in some cases a linear one suffices. The logic for this is as follows. Suppose a tax schedule induces some agent to choose a certain level of savings. Now, replace this schedule with one that is tangent to it at that savings level and above it everywhere else. The agent will continue to find the same level of savings optimal, so this new tax schedule implements the same allocation. In this sense, the tax schedule is not unique, only the marginal tax rate is pinned down. In particular, any concave tax schedule can be replaced by its linear tangent. Thus, the question is whether the desired allocation can be implemented by a concave tax schedule. I show that this is the case whenever the agent's value function is concave in wealth, or, equivalently, when consumption rises with wealth. Numerical explorations for a set of standard distributions of skills and preferences always find that this is the case. I conclude that although a nonlinear tax is generally required to remove the dependence of the tax on the current state, a linear tax will work in important cases.

The same tangency principle explains why the tax schedule can always be made strictly convex. In this sense, progressive taxation of savings, with rising marginal tax rates, is always feasible.

Finally, I explore the possibility of making the savings tax history independent. It is important to understand that most allocations, including optimal ones, will feature intertemporal wedges that are history dependent. Thus, the challenge is to create a savings tax that is history independent yet manages to deliver the appropriate history-dependent wedges.[3]

[3]Studying the subset of allocations and welfare that obtain using relatively simple and history-

I show that this can be done in two ways. The first implementation has agents saving in a history independent manner. The history independent schedule is set as the upper envelope of the history dependent ones. As long as the intertemporal wedge varies across histories this upper envelope schedule will feature a kink at the equilibrium savings point. Indeed, the left and right derivatives are precisely the smallest and largest intertemporal wedge, respectively.

The second way of making the savings tax history independent avoids generating kinks. In the model, individual savings and tax transfers are indeterminate due to standard Ricardian equivalence arguments. Unlike previous implementations, this one relies on agents saving in a history dependent manner. The idea is to have wealth be a sufficient statistic for the desired intertemporal wedge. Different wealth positions then place agents on different points of the history independent schedule, precisely where the marginal tax rate equals the desired intertemporal wedge.

A simple and attractive feature of my implementation is that, if an incentive-compatible allocation is obtained by allowing agents to save freely, then this arrangement belongs to the set studied here. In other words, the tax systems I consider would include the one that generated the allocation in the first place. In contrast, the implementation in Kocherlakota [2005] would yield non-zero and stochastic taxes for such an allocation.[4]

It is virtually impossible to judge implementations objectively. They are all just as good within the logic of the model—including the direct mechanism. Unspecified considerations may guide us, but until those are made explicit it remains a subjective exercise. In the meantime, it seems useful to understand many possible tax systems, studying their properties, the choices they offer agents, the informational requirements they place on the tax agency, and the like. The main contribution of this paper should be seen as an effort in this direction.

The rest of the paper is organized as follows. I begin by deriving the implementation for a two-period economy in Section 2. I then take up the differentiability of the savings tax schedule in Section 3, considering both the case with a continuous distribution and that with finite shocks. In Section 4 I extend the implementation to longer horizons. Section 5 explores how to make the savings tax history independent.

---

independent tax systems (including the tax on savings and income) is an interesting different question that is not attempted here. The goal here, as stated at the outset of the paper, remains to implement *any* incentive compatible allocation, not a subset. The goal now is to attempt this with a history-independent savings tax, but allowing the labor-income tax to remain history dependent.

[4]One cannot ask the same question for the implementation in Albanesi and Sleet [2006], since their results only apply to optimal allocations (which feature savings distortions).

2 Two Period Horizon

It is useful to begin with a two-period horizon. Section 4 extends the implementation to longer horizons.

Time periods are denoted $t = 0, 1$. Productivity is given by θ_t. Initial productivity, θ_0, is known, while θ_1 is uncertain. It is realized privately to the agent at $t = 1$. An allocation specifies c_0, y_0, $c_1(\theta_1)$, $y_1(\theta_1)$ and delivers utility

$$v_0^* = U^0(c_0, y_0, \theta_0) + \beta \mathbb{E}\left[U^1\left(c_1(\theta_1), y_1(\theta_1), \theta_1\right)\right],$$

where $U^t(c, y, \theta)$ represents the utility in period $t = 0, 1$, assumed to be increasing in c and θ, concave in (c, y) and satisfy the standard single-crossing condition that the marginal rates of substitution $U_c^t(c, y, \theta)/U_y^t(c, y, \theta)$ is increasing in θ, for any given (c, y).

Unlike Kocherlakota [2005] and Albanesi and Sleet [2006], the implementation I develop does not require additively separable utility of the form $U(c, y, \theta) = u(c) - h(y, \theta)$. Theoretically, the additively separable case is an important benchmark since it is required for the Inverse Euler characterization of constrained optimal allocations. Empirically, Aguiar and Hurst [2005] provide evidence supporting departures from separability in the direction of assuming that consumption (expenditures) c and labor y are Hicksian complements, so that $U_{cy} > 0$.

Incentive compatibility requires truth telling to be optimal

$$U^1\left(c_1(\theta_1), y_1(\theta_1), \theta_1\right) \geq U^1\left(c_1(r_1), y_1(r_1), \theta_1\right)$$

for all $\theta_1 \in \Theta$ and all $r_1 \in \Theta$; here, r_1 represents the report made by the agent regarding the true shock θ_1.

This concludes the description of the model's environment. The main goal of this paper is to extend a given mechanism so that it allows for saving and attains the same equilibrium allocation. For that purpose, it is unnecessary to introduce technology, or to set up a planning problem and solve for optimal allocations.

2.1 Nonlinear Tax Implementation

For any given incentive compatible allocation $(c_0, y_0, c_1(\theta_1), y_1(\theta_1))$, consider an implementation that confronts the agent with the following budget constraints:

$$\begin{aligned} \tilde{c}_0 + a_1 &\leq c_0, \\ \tilde{c}_1 &\leq R(a_1) + c_1(r_1). \end{aligned}$$

Here where $R(\cdot)$ represents the retention function, with $R(0) = 0$. If the net interest rate is r then $R(a) - (1+r)a$ represents a nonlinear tax on wealth at $t = 1$.

Without loss of generality, one can take R to be increasing. Equivalently, consider the budget constraints

$$\begin{aligned} \tilde{c}_0 + M(x_1) &\leq c_0, \\ \tilde{c}_1 &\leq x_1 + c_1(r_1), \end{aligned}$$

for $M = R^{-1}$, with $M(0) = 0$. If the net interest rate is r then $M(x) - x/(1+r)$ represents a nonlinear savings tax at $t = 0$. Clearly, one can go back and forth between R and M. I shall use the second formulation and refer to M as the savings tax function.

These budget constraints strictly augment the original direct mechanism by adding a saving choice x_1; the constraints effectively specialize to the direct mechanism when $x_1 = 0$. The agent takes M as given and optimizes by choosing a saving and reporting strategy. The tax function M is said to implement the proposed allocation $(c_0, y_0, c_1(\theta_1), y_1(\theta_1))$ if the agent finds it optimal to save zero $x_1 = 0$. The original incentive compatibility of the allocation then ensures that truth telling is optimal.

Consider the agent's problem in two stages. In the second period, the agent enters holding wealth x_1, then productivity θ_1 is realized and the agent makes a report r_1. The utility obtained is

$$V_1(x_1, \theta_1) \equiv \max_{r_1} U^1\big(c_1(r_1) + x_1, y_1(r_1), \theta_1\big).$$

Define the set of maximizers to be

$$R^*(x_1, \theta_1) \equiv \arg\max_{r_1} U^1\big(c_1(r_1) + x_1, y_1(r_1), \theta_1\big)$$

Given a choice of x_1, expected utility is

$$W_0(x_1; M) \equiv U^0\big(c_0 - M(x_1), y_0, \theta_0\big) + \beta \mathbb{E}[V_1(x_1, \theta_1)].$$

Suppose the allocation is incentive compatible so that $W_0(0; M) = v_0^*$ if $M(0) = 0$. Now define $M^*(x)$ so that

$$W_0(x_1; M) = v^* \qquad \forall x_1.$$

Solving for the tax schedule M gives

$$M^*(x_1) \equiv c_0 - \Psi^0\big(v^* - \beta\mathbb{E}[V_1(x_1, \theta_1)], y_0, \theta_0\big), \tag{1}$$

where $\Psi^t(\cdot, y, \theta)$ denotes the inverse of $U^t(\cdot, y, \theta)$. Of course, since this construction implies indifference, if taxes are set higher than M^* for $x_1 \neq 0$, then $x_1 = 0$ remains optimal for the agent. Conversely, any tax schedule below M^* offers a better expected welfare opportunity than v^* and cannot implement the desired allocation. I summarize these arguments in the following proposition.

Proposition 1. *Any incentive compatible allocation is implemented for any $M(x_1) \geq M^*(x_1)$ with $M(0) = M^*(0) = 0$, where $M^*(x_1)$ is defined by equation (1).*

The tax schedule M^* is the lowest tax that implements any given allocation. It does so by making the agent indifferent to saving any quanitity. Tax schedules are not unique, since higher taxes for $x \neq 0$ help discourage saving $x_1 \neq 0$ further.

In the next two sections I study the shape of tax functions M satisfying Proposition 1. I start with the local property of differentiability. Since M^* is a lower envelope, it contains important local information about any feasible schedule M. In particular, if M is differentiable at $x = 0$ then its derivative, the marginal tax, must coincide with that of M^*. Indeed, it turns out that, M can be differentiable at $x = 0$ only if the same is true for M^*. I turn to investigate conditions for the latter.

3 Differentiability: Marginal Taxes and Wedges

For any allocation, define m^* to be the marginal rate of substitution by

$$U_c^0(c_0, y_0, \theta_0) m^* \equiv \beta\mathbb{E}\big[U_c^1\big(c_1(\theta_1), y_1(\theta_1), \theta_1\big)\big].$$

So that m^* corresponds to the shadow price that makes the agent's standard intertemporal Euler equation hold. If the interest rate is r, then

$$m^* - \frac{1}{1+r}$$

is a measure of the distortion implicit in the allocation. This is sometimes termed the intertemporal wedge or implict marginal tax rate.

When $M^*(x_1)$ is differentiable at $x_1 = 0$. The first order condition for the agent implies that the marginal tax should equal the $M^{*\prime}(0) = m^*$. Thus, in my implementation the marginal tax equals the intertemporal wedge. That is, implicit and explicit marginal tax rates coincide. I now investigate conditions for differentiability.

3.1 Left and Right Derivatives

The key to establishing differentiability properties for M^* is to verify the differentiability of $V_1(\cdot, \theta_1)$. Envelope theorems can be used for this, since V_1 is defined by a maximization. In particular, Corollary 4 in Milgrom and Segal [2002] applies in our case and implies that $V_1(\cdot, \theta_1)$ is left- and right-differentiable and that are given by

$$\frac{\partial}{\partial x_1} V_1(x_1+, \theta_1) = \max_{r \in R^*(x_1, \theta_1)} U_c^1(c_1(r) + \tilde{x}_1, y_1(r), \theta_1)$$
$$\frac{\partial}{\partial x_1} V_1(x_1-, \theta_1) = \min_{r \in R^*(x_1, \theta_1)} U_c^1(c_1(r) + \tilde{x}_1, y_1(r), \theta_1)$$

respectively. Thus, kinks can only occur where $R^*(x_1, \theta_1)$ is not unique. These represent situations where the agent is indifferent to two or more reports r_1.

Note that

$$\frac{\partial}{\partial x_1} V_1(x_1-, \theta_1) \leq \frac{\partial}{\partial x_1} V_1(x_1+, \theta_1)$$

so that only convex, but not concave, kinks are possible. It also follows that

$$M^{*\prime}(x_1-) \leq m^* \leq M^{*\prime}(x_1+).$$

Next I investigate when these inequalities can be assured to hold with equality.

3.2 Continuous Distribution of Shocks

I now show that the tax schedule $M^*(x_1)$ is differentiable if the distribution of productivity is continuous. The idea is that indifference of the agent with respect to the report, is a "knife edge" phenomena. From an ex-ante perspective, when productivity is continuously distributed, indifference occurs with probability zero. It follows that V_1 is almost everywhere differentiable and that $\mathbb{E}[V_1(\cdot, \theta_1)]$ is differentiable.

Proposition 2. *If θ_1 is continuously distributed then $M^*(x_1)$ defined by equation (1) is differen-*

tiable and $M^{*\prime}(0) = m^*$.

Proof. Incentive compatibility implies that $c_1(\theta_1)$ and $y_1(\theta_1)$ are non-decreasing functions of θ_1. It follows that there are at most countably many points of discontinuity. Define $\hat{\Theta}$ to be the set of points of discontinuity in $c_1(\cdot)$. By Theorem 3 in Milgrom and Segal (2002), the function $V_1(x_1, \theta_1)$ is differentiable with respect to x_1 at $x_1 = 0$ and $\theta_1 \notin \hat{\Theta}$, with derivative given by the envelope formula: $V_x(0, \theta_1) = U^1_c(c_1(\theta_1), y_1(\theta_1), \theta_1)$ (since $r_1 = \theta_1$ is optimal by incentive compatibility). The function of interest to us is

$$V_1(x_1) \equiv \mathbb{E}[V(x_1, \theta_1)],$$

the integral of a function that is differentiable at $x_1 = 0$ except on a set of countable points. With a continuous distribution for θ_1 the set of points of non-differentiability is of probability zero. It then follows that $V(x_1)$ is differentiable at $x_1 = 0$ with derivative

$$V_1'(0) = \mathbb{E}\left[V_x(0, \theta_1)\right] = \mathbb{E}\left[U^1_c\left(c_1(\theta_1), y_1(\theta_1), \theta_1\right)\right]. \qquad \square$$

In other words, if the net interest rate is r, the marginal tax rate $M'(0) - 1/(1+r)$ exists and coincides with the intertemporal wedge in the Euler equation.

3.3 Finite Shocks

If productivity types are finite, so that $\Theta = \{\theta^1, \theta^2, \ldots, \theta^N\}$, then indifference may occur with positive probability, produces a convex kink: $M^{*\prime}(0_-) < M^{*\prime}(0_+)$. Indeed, this will typically be the case at optimal allocations, because downward incentive constraints bind.

I first show that, with finite types, kinks occur in the labor income tax schedule. Thus, they are not limited to the savings tax, but inherent to the Mirrlees [1971] model.

With finite types, incentive constraints bind downward at an optimal allocation, so that

$$v_1(\theta^n) \equiv U^1\left(c_1(\theta^n), y_1(\theta^n), \theta^n\right) = U^1\left(c_1(\theta^{n-1}), y_1(\theta^{n-1}), \theta^n\right) \qquad \forall n = 2, \ldots, N.$$

In addition, output is non-decreasing, so that $y_1(\theta^{n-1}) \leq y_1(\theta^n)$ for $n = 2, \ldots, N$. Following Mirrlees [1971], one seeks an incodme tax schedule $T(Y)$, defined for all $Y \geq 0$, so that agents face the budget constraint

$$c_1 \leq y_1 - T(y_1)$$

in the second period. Ensuring that agents choose the original allocation requires

$$U^1(Y - T(Y), Y, \theta^n) \leq v_1(\theta^n)$$

for all $Y \geq 0$, with equality at $Y = y(\theta^n)$ for $n = 1, \ldots, N$. This is equivalent to

$$\begin{aligned} T(Y) &\geq T^*(Y) \qquad && \forall Y \\ T(Y) &= T^*(Y) \qquad && \forall Y = y(\theta^n) \text{ for } n = 1, \ldots, N. \end{aligned}$$

where

$$T^*(Y) \equiv \max_n \{Y - \Psi^1[v_1(\theta^n), Y, \theta^n)]\}.$$

Note that T^* is differentiable whenever the maximization over n that defines it is unique. Conversely, whenever the maximum is attained by more than one n then the function T^* has a convex kink. This is precisely what happens, at $y(\theta^n)$ for $n = 1, \ldots N-1$ when downward incentive constraint are binding. I summarize this observation in a proposition.

Proposition 3. *Suppose the set of productivity types is finite. Any income tax schedule $T : \mathbb{R}_+ \to \mathbb{R}$ that implements the optimal allocation cannot be differentiable at points $Y = y(\theta^n)$ for $n = 1, \ldots, N-1$. Moreover, if T has left and right derivatives at any of these points then there must be a convex kink: $T'(Y_-) < T'(Y_+)$.*

The situation is illustrated in Figure 1 for the case with two shocks. In conclusion, with finite types, kinks are pervasive at the optimum. They occur both for labor-income taxes as well as the savings tax.

Nevertheless, two ideas suggest that kinks due to finite shocks should be of little concern.

First, if incentive constraints bind, one can perturb the allocation so that they are slack. In this way, one can come arbitrarily close to any desired allocation, and its welfare, with a differentiable tax function.

Proposition 4. *Suppose shocks are finite $\Theta = \{\theta^1, \theta^2, \ldots, \theta^N\}$. Take any incentive compatible allocation c_0, y_0, $c_1(\theta_1)$, $y_1(\theta_1)$. Then for any $\varepsilon > 0$ there exists differentiable tax functions M_ε and T_ε that implement allocations (c_0^ε, y_0^ε, $c_1^\varepsilon(\theta_1)$, $y_1^\varepsilon(\theta_1)$) so that $|c_0^\varepsilon - c_0| < \varepsilon$, $|y_0^\varepsilon - y_0| < \varepsilon$, $|c_1^\varepsilon(\theta_1) - c_1(\theta_1)| < \varepsilon$ and $|y_1^\varepsilon(\theta_1) - y_1(\theta_1)| < \varepsilon$.*

It is also possible to add a technological constraint for the allocations (c_0^ε, y_0^ε, $c_1^\varepsilon(\theta_1)$, $y_1^\varepsilon(\theta_1)$) to be feasible.

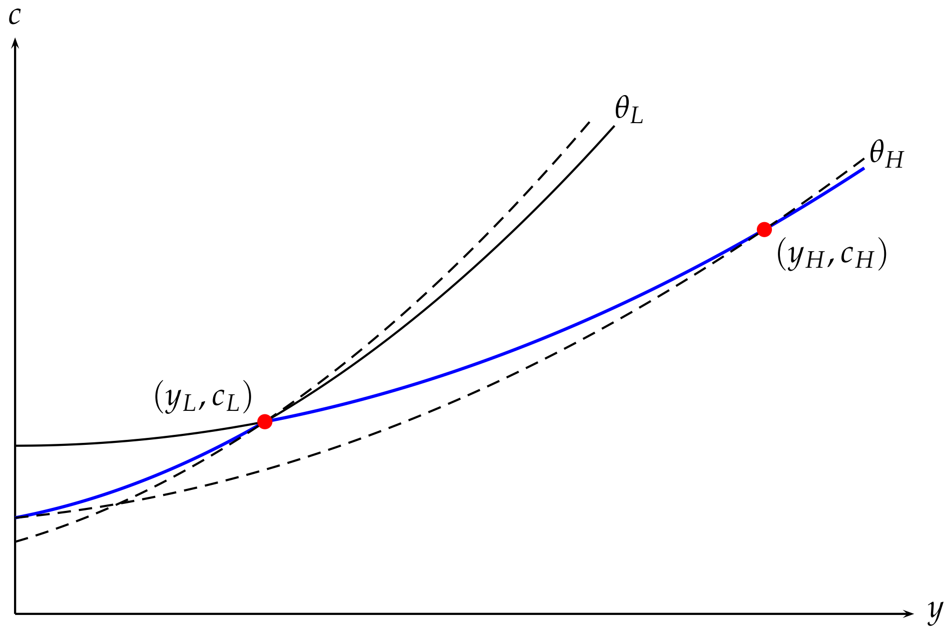

Figure 1: An example with two types $\Theta = \{\theta_L, \theta_H\}$ where the incentive constraint for θ_H initially binds. A small increase in savings $x > 0$ makes the indifference curves steeper, and reporting L optimal for both agent types, θ_L and θ_H, becomes optimal.

Second, I conjecture that the size of this kink disappears as the number of shocks is increased $N \to \infty$ so that Θ becomes dense in some interval $[\underline{\theta}, \bar{\theta}]$. As the grid points are expanded the distance between any two types shrinks and the probability at any one type vanishes. As a result, for large N most agents will be indifferent to a consumption-labor bundle very close to their own; only a vanishing fraction may be indifferent to a more distant point. For small but positive savings $x_1 > 0$ agents underreport, but the deviation in consumption and labor is small and vanishing. As a result, the left and right derivative in the envelope formula approach each other.

4 Linear and Progressive Capital Taxation

Interestingly, although I allow for nonlinear taxation, in some cases a linear tax will do. To see this, recall that Proposition 1 states that any function $M(x_1) \geq M^*(x_1)$ will implement the desired allocation. It follows that, whenever $M^*(x_1)$ is concave that one can set $M(x_1)$ to be a linear function that acts as a supporting tangent at $x_1 = 0$. Inspection of equation

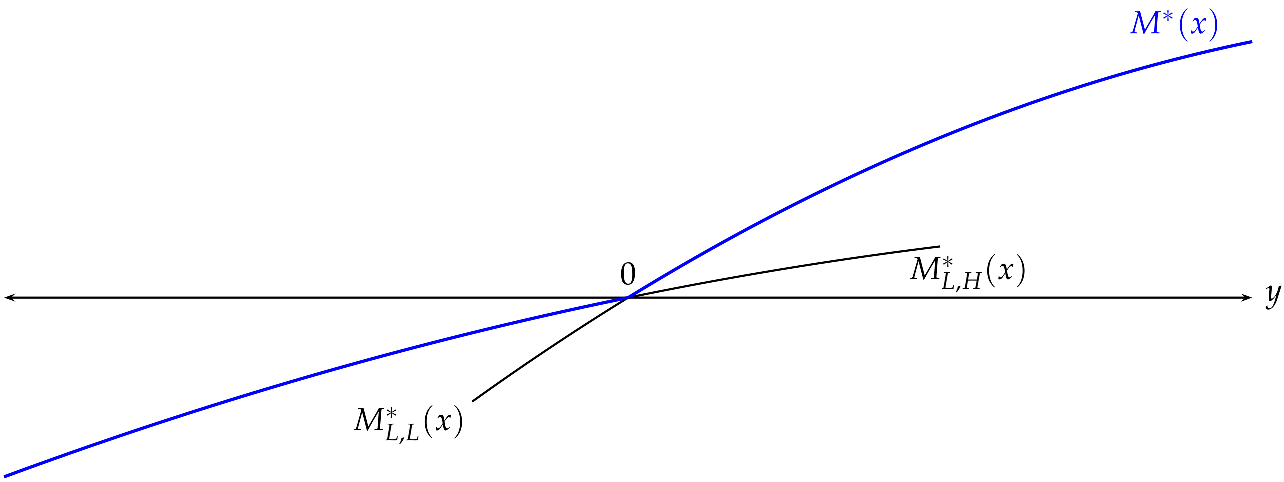

Figure 2: An example with two types Here $M^*_{L,H}(x)$ and $M^*_{L,L}(x)$ respectively represent the tax functions required for a truth telling strategy and a strategy that always reports the low shock. That is, $M^*_{L,H}(x_1) = c_0 - \Psi^0(v^* - \beta\mathbb{E}[U^1(c_1(\theta_1) + x_1, y_1(\theta_1), \theta_1)], y_0, \theta_0)$ and $M^*_{L,L}(x) = c_0 - \Psi^0(v^* - \beta\mathbb{E}[U^1(c_1(\theta_L) + x_1, y_1(\theta_L), \theta_1)], y_0, \theta_0)$.

(1) shows that this will be the case if $\mathbb{E}[V_1(\cdot, \theta_1)]$ is concave.[5] A sufficient condition for this is for each $V_1(\cdot, \theta_1)$ function to be concave, for all $\theta_1 \in \Theta$. Of course, this condition is not necessary: $\mathbb{E}[V_1(\cdot, \theta_1)]$ may be if $V(\cdot, \theta_1)$ is only concave for a subset of values of θ_1.

Concavity of $V_1(x_1)$ is not a foregone conclusion. A sufficient condition for concavity is that the agent faces as a convex tax schedule in the second period, so that the the loci of points (y, c) available to him are in a convex relation. This will be the case if labor-income taxation is progressive.

Proposition 5. *(a) If $V_1(x_1)$ is concave a linear schedule $M(x_1) = (1+\tau)x_1$ implements the optimal allocation with $x_1 = 0$. (b) If labor income taxation is progressive in period $t = 1$, then $V_1(x_1)$ is concave.*

[5]This follows by the following fact. Suppose $f : \mathbb{R} \to \mathbb{R}$ is decreasing and concave and that $g : \mathbb{R} \to \mathbb{R}$ is decreasing and convex Then $f(g(\cdot))$ is concave. Start with the fact that f is concave to deduce:

$$\alpha f(g(x_1)) + (1-\alpha) f(g(x_2)) \leq f(\alpha g(x_1) + (1-\alpha) g(x_2)).$$

Now, by convexity of g,

$$\alpha g(x_1) + (1-\alpha) g(x_2) \geq g(\alpha x_1 + (1-\alpha) x_2).$$

Since f is decreasing it follows that

$$f(\alpha g(x_1) + (1-\alpha) g(x_2)) \leq f(g(\alpha x_1 + (1-\alpha) x_2)).$$

Putting the first and last inequality together yields the desired result.

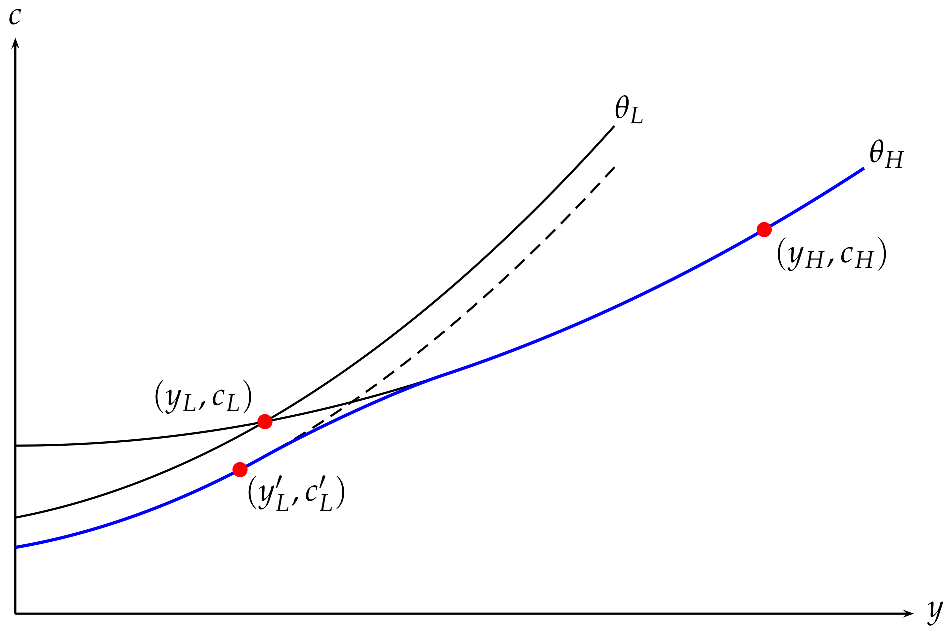

Figure 3: A smooth tax function can be created by shifting the allocation for θ_L so that the incentive constraints are slack. The lower envelope (in blue) represents the retention function $Y - T(Y)$.

5 Arbitrary Finite Horizon

Now suppose there are periods $t = 0, 1, \ldots, T$. The agent has utility function

$$\sum_{t=0}^{T} \beta^t \mathbb{E}[U^t(c_t, y_t, \theta_t)]. \tag{2}$$

Assume $\{\theta_t\}$ is a Markov process with $\theta_t \in \Theta$.

Consider any incentive compatible allocation $(c(\theta^t), y(\theta^t))$ with equilibrium values $v(\theta^t)$ satisfying

$$v(\theta^t) = U^t\big(c(\theta^t), y(\theta^t), \theta_t\big) + \beta \mathbb{E}[v(\theta^{t+1})|\theta_t]. \tag{3}$$

The agent's optimization problem has the Bellman equation

$$w(r^{t-1}, \theta_t) = \max_{r_t} \big\{U^t\big(c(r^t), y(r^t), \theta_t\big) + \beta \mathbb{E}[w(r^t, \theta_{t+1})|\theta_t]\big\}. \tag{4}$$

Here the agent enters a period having reported r^{t-1} in the past and learns his true shock θ_t; thus, the state variable is (r^{t-1}, θ_t). The agent then decides what report to make r_t.[6]

[6]The Markov assumption allows us to ignore the history of shocks θ^{t-1}. Also, for an incentive compatible allocation, truth telling is optimal even if the agent has made false reports in the past.

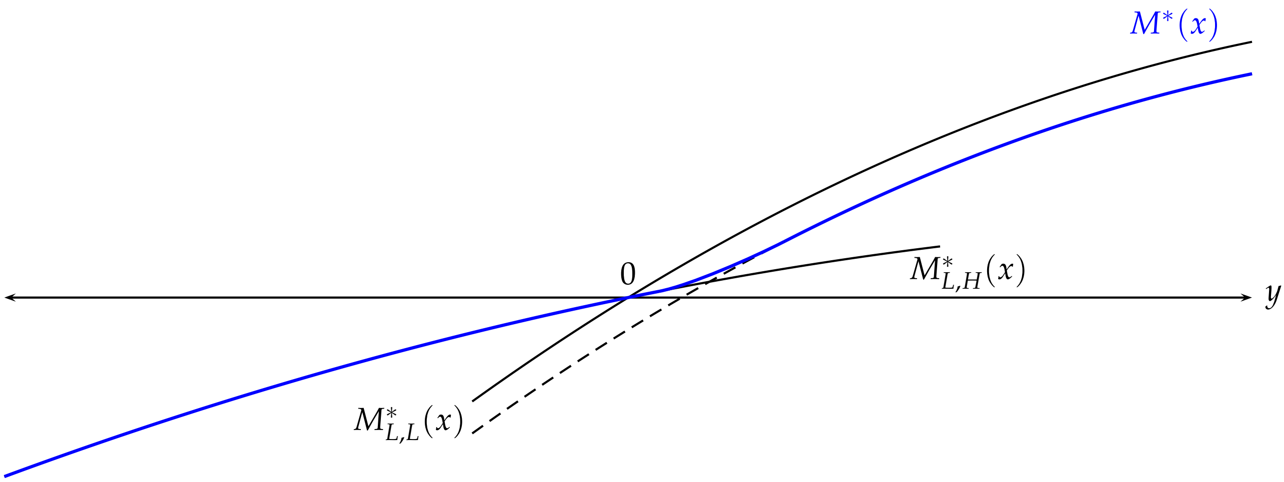

Figure 4: ...smoothing out the savings tax.

Incentive compatibility is equivalent to the requirement that

$$w(r^{t-1}, \theta_t) = v(r^{t-1}, \theta_t),$$

so that v defined by the recursion (3) solves the Bellman equation (4).

5.1 Nonlinear Tax Implementation

For any given incentive compatible allocation $(c(\theta^t), y(\theta^t))$, consider the budget constraints

$$\tilde{c}_t + M(x_{t+1}, r^t) \leq x_t + c(r^t), \tag{5}$$

with initial condition $x_0 = 0$ and terminal condition $x_{T+1} \geq 0$. If the interest rate in period t is r_t, then $M(x_{t+1}, r^t) - x_{t+1}/(1 + r_t)$ represents a tax on savings.

The goal of the implementation is to find a sequence of functions $\{M(x_{t+1}, r^t)\}$ to ensure that the agent finds it optimal not save and set $x_{t+1} = 0$ for all $t = 0, 1, \ldots, T$. To simplify, I refer to $\{M(x_{t+1}, r^t)\}$ as tax functions. The agent's dynamic programming problem has state variables x_t, r^{t-1} and θ_t with Bellman equation

$$V(x_t, r^{t-1}, \theta_t) = \max_{r_t, x'} \left\{ U^t\big(c(r^t) + x_t - M(x_{t+1}, r^t), y(r^t), \theta_t\big) + \beta \mathbb{E}[V(x_{t+1}, r^t, \theta_{t+1}) \mid \theta^t] \right\} \tag{6}$$

with $V(x_{T+1}, r^T, \theta_{T+1}) = 0$ so that

$$V(x_T, r^{T-1}, \theta_T) = \max_{r_T} U^T\big(c(r^T) + x_T, y(r^T), \theta_T\big).$$

It is useful to define $W(x_t, x_{t+1}, r^t, \theta_t)$ as the right hand side of the Bellman equation

$$W(x_t, x_{t+1}, r^t, \theta_t) \equiv U^t\big(c(r^t) + x_t - M(x_{t+1}, r^t), y(r^t), \theta_t\big) + \beta \mathbb{E}[V(x_{t+1}, r^t, \theta_{t+1}) \mid \theta^t] \quad (7)$$

I proceed recursively, starting in the last period and working backwards.

5.2 Last Two Periods

I start at $T-1$ when only two periods remain. The argument is similar to the two period case studied previously, but now there are shocks and reporting in both periods, the argument is a bit different.

At $t = T-1$ impose

$$W(0, x_T, r^{T-1}, \theta_{T-1}) \leq v(r^{T-2}, \theta_{T-1}) \qquad \forall x_T, r^{T-1}, \theta_{T-1} \quad (8)$$

$$W(0, 0, (r^{T-2}, \theta_{T-1}), \theta_{T-1}) = v(r^{T-2}, \theta_{T-1}) \qquad \forall r^{T-2}, \theta_{T-1} \quad (9)$$

The first inequality ensures that if the agent arrives in period $t = T-1$ without any assets $x_{T-1} = 0$, then the agent will find it optimal to report the current shock truthfully, $r_{T-1} = \theta_{T-1}$ and not save, $x_T = 0$. The second equality ensures that if the agent does this he attains the same utility as with the original direct mechanism. This last condition is equivalent to ensuring that the original allocation is available and simply fixes $M(0, r^{T-1}) = 0$.

These conditions are not only sufficient to implement the desired allocation, they are also necessary. If any of the first inequalities were violated, then the agent can do better by deviating to a another savings level $x_T \neq 0$. Similarly, if any of the second set of equalities were violated, then even agents that do not save would be taxed, $M(0, r^{T-1}) > 0$, and the original allocation would not be attainable for the agent budget constraints. As a result the agent would be attain strictly less utility than intended.

Two more related remarks are in order. First, these conditions do not rule out that if the agent were to deviate and save $x_T \neq 0$ that he will not then prefer to deviate in his reporting r_T at $t = T$. Second, if the agent misreports in the current period $r_{T-1} \neq \theta_{T-1}$ then $x_T \neq 0$ may be optimal. That is, in both senses, double deviations may be better than a single deviation. There is no need to impose that double deviations are not strictly better than single deviations. What the conditions ensure is that no deviations are better than any set of deviations.

The inequalities are equivalent to imposing that

$$M(x_T, r^{T-1}) \geq M^*(x_T, r^{T-1}) \tag{10}$$

where M^* is the upper envelope over θ_{T-1},

$$M^*(x_T, r^{T-1}) \equiv \max_{\theta_{T-1}} \tilde{M}^*(x_T, r^{T-1}, \theta_{T-1}), \tag{11}$$

of the function

$$\begin{aligned}\tilde{M}^*(x_T, r^{T-1}, \theta_{T-1}) &\equiv c(r^{T-1}) \\ &- \Psi^{T-1}\left[v(r^{T-2}, \theta_{T-1}) - \beta\mathbb{E}[V(x_T, r^{T-1}, \theta_T) \mid \theta_{T-1}], y(r^{T-1}), \theta_{T-1}\right].\end{aligned} \tag{12}$$

Intuitively, $\tilde{M}^*$ represents a fictitious tax function that would ensure that agents are indifferent to any savings and report, so that inequality (8) holds with equality for all x_T, r^{T-1} and θ_{T-1}. Such a tax function is not feasible, since it must depend on the true type θ_{T-1} in addition to the reports r^{T-1}. Taking the upper envelope over true types θ_{T-1} to define M^* yields a tax function that depends only on reports r^{T-1}, while ensuring that the inequality (8) continues to hold. Because one takes the upper envelope over θ_{T-1}, agents confronted with M^* will not typically be indifferent among several saving and reporting strategies.

Conversely, M^* is the lowest possible tax function that prevents a deviation from $x_{T-1} = 0$, since any lower tax function would necessarily violates inequality (8) for some type θ_{T-1}. Thus, the set of all feasible tax functions M are those that lie above M^*, satisfying inequality (10).

Finally, note that for each r^{T-1} and θ_{T-1} the fictitious tax $\tilde{M}^*$ is exactly akin to the tax function defined in equation (1) for the previous two-period horizon case, where it was assumed that $\theta_{T-1} = \theta_0$ was known. The innovation here, when this last assumption is dropped, is to take the upper envelope over θ_{T-1} and to condition on the history r^{T-1}. The new construction rules out misreporting in period $T-1$, as well as in period T.

5.3 Other Periods

I now work backward inductively defining a tax schedule for all periods $t = 0, 1, \ldots T-1$. Suppose tax schedules for periods $s = t+1, t+2, \ldots, T-1$ have already been constructed. Associated with these tax schedules $\{M(x_{s+1}, r^s)\}_{s=t+1}^{T-1}$ are value functions $\{V(x_s, r^{s-1}, \theta_s)\}_{s=t+1}^{T-1}$. I seek to construct a tax schedule $M(x_{t+1}, r^t)$ and value function $V(x_t, r^{t-1}, \theta_t)$ for period t.

Recall that these functions $W(x_t, x_{t+1}, r^t, \theta_t)$ are given by (7) using next period's value function $V(x_{t+1}, r^t, \theta_{t+1})$ and the current tax schedule $M(x_{t+1}, r^t)$, which must be constructed. I impose that, whatever the value function $V(x_{t+1}, r^t, \theta_{t+1})$, that the tax function $M(x_{t+1}, r^t)$ be such that the implied W satisfy

$$W(0, x_{t+1}, r^t, \theta_t) \leq v(r^{t-1}, \theta_t) \qquad \forall x_{t+1}, r^t, \theta_t \tag{13}$$

$$W(0, 0, (r^{t-1}, \theta_t), \theta_t) = v(r^{t-1}, \theta_t) \qquad \forall r^{t-1}, \theta_t \tag{14}$$

These conditions are exactly analogous to the ones presented above. The previous conditions (8)–(9) are the special case of conditions (13)–(14) for $t = T-1$. Thus, the induction argument could begin at $T-1$ directly with the general arguments laid out in this subsection; however, the previous subsection was included for presentation purposes as warm up to the more general ideas.

The reasons for imposing these conditions are the same as before. They ensure that in period t the agent will find truth telling and no savings, $r_t = \theta_t$ and $x_{t+1} = 0$ optimal, obtaining the same allocation and utility as the original allocation. They are both necessary and sufficient for this.

The same arguments lead us to define

$$M^*(x_{t+1}, r^t, \theta_t) \equiv c(r^t) - \Psi^t \left[v(r^{t-1}, \theta_t) - \beta \mathbb{E}[V(x_{t+1}, r^t, \theta_{t+1}) \mid \theta_t], y(r^t), \theta_t \right]$$

and

$$M^*(x_{t+1}, r^t) \equiv \max_{\theta_t} M^*(x_{t+1}, r^t, \theta_t), \tag{15}$$

and set $M(x_{t+1}, r^t)$ to be any function above this:

$$M(x_{t+1}, r^t) \geq M^*(x_{t+1}, r^t). \tag{16}$$

Given such a choice for $M(x_{t+1}, r^t)$ this then defines a value function $V(x_t, r^{t-1}, \theta_t)$ using the Bellman equation (6).

Continuing this way gives a tax schedule $M(x_{t+1}, r^t)$ and value function $V(x_t, r^{t-1}, \theta_t)$

for $t = 0, 1, \ldots, T-1$. This construction ensures that the Bellman equation (6) holds with the maximum is attained by truth-telling and no savings, for all $t = 0, 1, \ldots, T-1$. By the principle of optimality this implies that truth telling and no savings solves the agent's problem of maximizing (2) among all possible reporting and savings strategies satisfying the budget constraint (5). This proves the following.

Proposition 6. *Any incentive-compatible allocation $\{c(\theta^t), y(\theta^t)\}_{t=0}^{T}$ can be implemented using the budget constraints (5) by any sequence tax functions on savings $\{M(x_{t+1}, r^t)\}_{t=0}^{T-1}$ satisfying the inequalities (16), where $M^*(x_{t+1}, r^t)$ defined implicitly from $\{M(x_{t+1}, r^t)\}_{t=0}^{T-1}$ using conditions (15) and the Bellman equation (6). Conversely, if a sequence of tax functions $\{M(x_{t+1}, r^t)\}_{t=0}^{T-1}$ implements the incentive-compatible allocation $\{c(\theta^t), y(\theta^t)\}_{t=0}^{T}$ it must satisfy inequalities (16).*

This construction creates all feasible tax schedules that implement the desired allocation. Note that the tax schedules $\{M(x_{s+1}, r^t)\}_{s=t+1}^{T-1}$ chosen to satisfy (15)–(16) affect the lowest tax schedule possible, $M^*(x_{t+1}, r^t)$, in period t given by (15) and, thus, affect the available tax schedule choices for period t satisfying (16). Indeed, higher choices for the functions $\{M(x_{s+1}, r^t)\}_{s=t+1}^{T-1}$ lead to lower values of $V(x_{t+1}, r^t, \theta_{t+1})$ and hence lower values for $M^*(x_{t+1}, r^t)$, i.e. more choices for $M(x_{t+1}, r^t)$. In particular, because of this decreasing relation between future and current tax schedules, it would be incorrect to interpret the sequence generated by always setting $M(x_{t+1}, r^t) = M^*(x_{t+1}, r^t)$ for $t = 0, \ldots, T-1$ as the lowest possible tax schedules (this is only the case for period $T-1$). Relatedly, seeking the lowest tax functions $\{M(x_{t+1}, r^t)\}_{t=0}^{T-1}$ is not a well-posed problem, outside the two-period horizon. Nevertheless, the important point is that the characterization provided by (15)–(16) is exhaustive, providing all the tax schedules that implement the allocation.

It is interesting to note how the tax schedules are dependent on each other. In other words, constructing the tax schedules for one period cannot be done independently from other periods. Similarly, the set of feasible tax schedule for period t depends on the entire allocation for periods $t, t+1, \ldots, T$. This is unlike the implementation in Kocherlakota [2005], which employs state-dependent linear taxes on capital; there, only the consumption allocation in periods t and $t+1$ is needed to compute tax rates at $t+1$. In the present case, however, the entire allocation is only relevant for the global nonlinear properties of the tax schedules. As I show next, the marginal tax rates always equal the intertemporal wedge, which can also be computed from the consumption allocation at t and $t+1$.

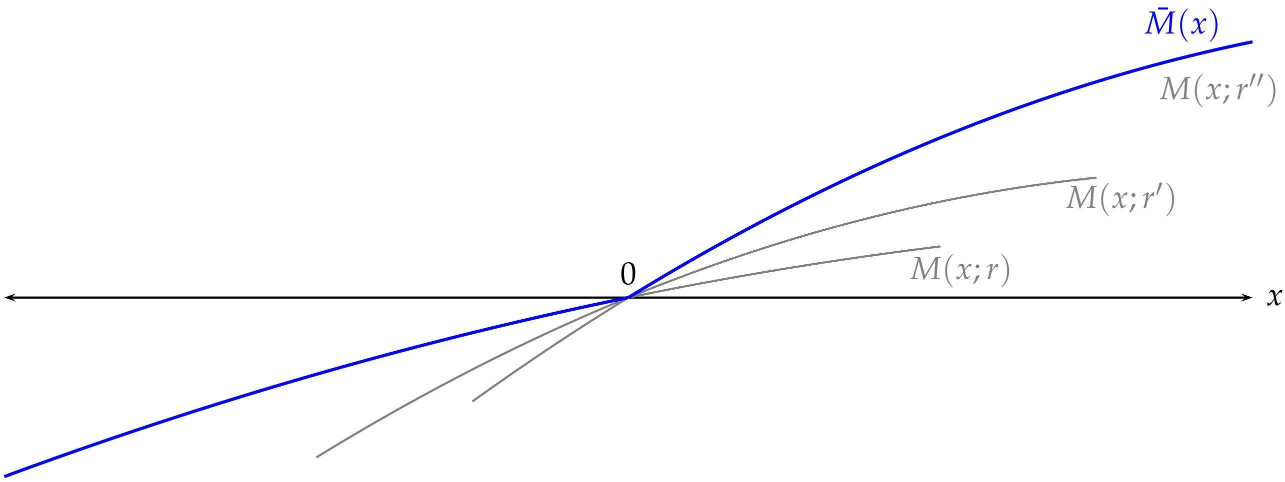

Figure 5: A kinked upper envelope $\bar{M}(x)$ function constructed from $M(x;r)$ functions that have not been transposed.

6 History Independent Taxation

Finally, I explore the possibility of making the savings tax history independent. It is important to understand that most allocations, including optimal ones, will feature intertemporal wedges that are history dependent. Thus, the challenge is to create a savings tax that is history independent yet manages to deliver the appropriate history-dependent wedges.

I show that this can be done in two ways. The first implementation has agents saving in a history independent manner. The history independent schedule is set as the upper envelope of the history dependent ones. As long as the intertemporal wedge varies across histories this upper envelope schedule will feature a kink at the equilibrium savings point. Indeed, the left and right derivatives are precisely the smallest and largest intertemporal wedge, respectively.

The second way of making the savings tax history independent avoids generating kinks. In the model, individual savings and tax transfers are indeterminate due to standard Ricardian equivalence arguments. Unlike previous implementations, this one relies on agents saving in a history dependent manner. The idea is to have wealth be a sufficient statistic for the desired intertemporal wedge. Different wealth positions then place agents on different points of the history independent schedule, precisely where the marginal tax

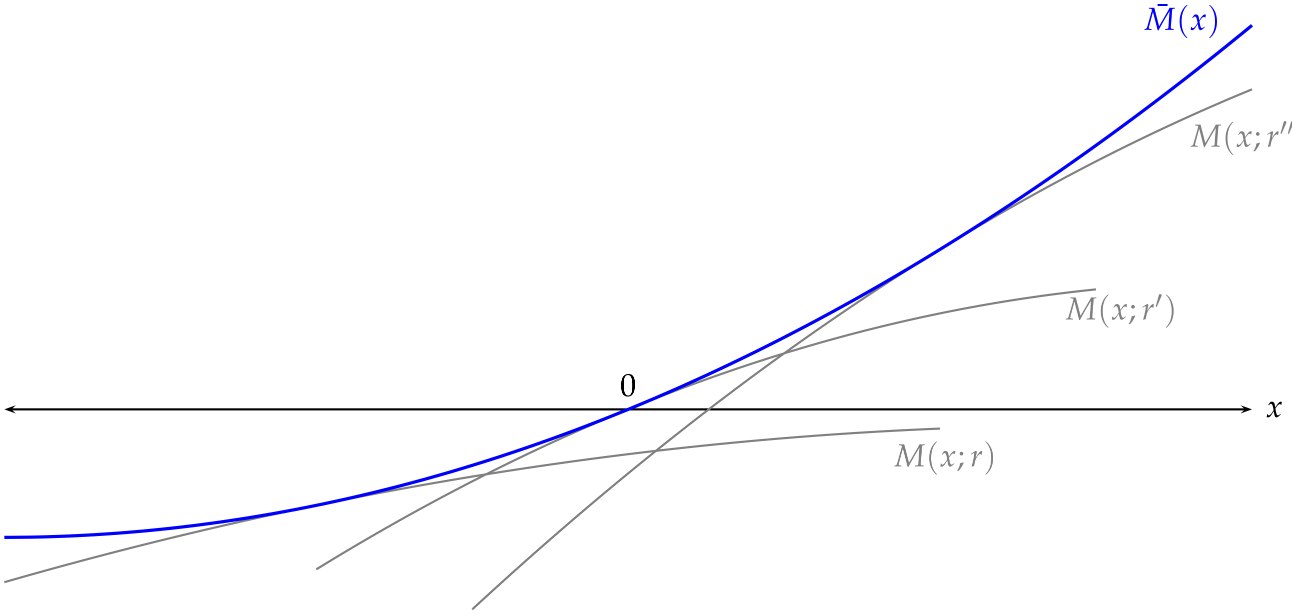

Figure 6: A smooth upper envelope $\bar{M}(x)$ function constructed from transposed $M(x;r)$ functions.

rate equals the desired intertemporal wedge.

[to be completed]

References

Mark Aguiar and Erik Hurst. Consumption versus expenditure. *Journal of Political Economy*, 113(5):919–948, October 2005.

Stefania Albanesi and Christoper Sleet. Dynamic optimal taxation with private information. *Review of Economic Studies*, 2006. forthcoming.

Peter A. Diamond and James A. Mirrlees. A model of social insurance with variable retirement. *Journal of Public Economics*, 10(3), 1978.

Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski. Optimal indirect and capital taxation. *Review of Economic Studies*, 70(3):569–587, 2003.

Narayana R. Kocherlakota. Figuring out the impact of hidden savings on optimal unemployment insurance. *Review of Economic Dynamics*, 7:541–554, 2004.

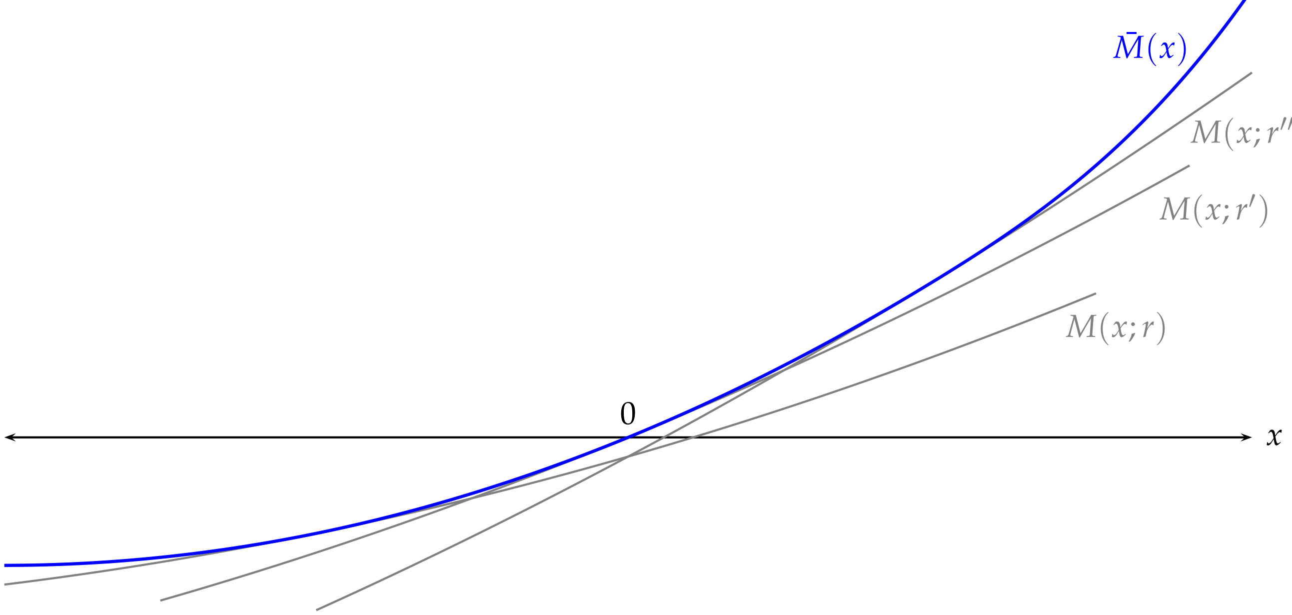

Figure 7: A smooth upper envelope $\bar{M}(x)$ function constructed from transposed $M(x;r)$ functions.

Narayana R. Kocherlakota. Zero expected wealth taxes: A mirrlees approach to dynamic optimal taxation. *Econometrica*, 73(5):1587–1621, 2005.

Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, March 2002.

James A. Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38(2):175–208, 1971.

William P. Rogerson. Repeated moral hazard. *Econometrica*, 53(1):69–76, 1985.