

# The Dynamics of Optimal Taxation when Human Capital is Endogenous

Marek Kapička\*

Department of Economics

University of California, Santa Barbara

August 21, 2009

## Abstract

This paper characterizes the dynamics of optimal income taxes in a multi-period Mirrleesian economy where agents have private information about both their ability and their human capital. I provide theoretical results showing conditions under which marginal income taxes are decreasing over time. The main reason is that schooling and labor supply are substitutes in the short run, while in the long run both become complements.

I calibrate the economy to U.S. data and solve numerically for the dynamics of efficient allocations and optimal taxes in two reforms: A reform where the government minimizes the costs of delivering a given distribution of lifetime utilities, and a Rawlsian reform, where the government only cares about the utility of the least able agent. In both reforms the marginal income taxes

---

\*E-mail: mkapicka@econ.ucsb.edu. I would like to thank Henning Bohn and seminar participants at Georgetown University, University of Texas at Austin and various conferences for their helpful comments and suggestions.

decrease significantly over time, on average by 25% in the first reform and by 20% in the second reform.

J.E.L Codes: E6, H2

Keywords: optimal taxation, income taxation, human capital

## INTRODUCTION

I study efficient allocations and optimal taxes in a dynamic economy where agents are heterogeneous in their ability to produce output, and can invest in human capital to augment their productivity. Neither their abilities nor human capital are observable by the government. I calibrate the economy to US data and solve numerically for the dynamics of efficient allocations and optimal taxes. I also provide analytical results for several special cases.

I provide quantitative results for two types of reforms. In the first one, the government takes the existing distribution of utilities as given and minimizes the present value of costs of delivering that distribution. In the second reform, the government is Rawlsian and minimizes the present value of costs of delivering a given lifetime utility to the least able agent. In both reforms, the average marginal income taxes decrease significantly over time: under the first reform from 41% to 16%, and under the Rawlsian reform from 65% to 45%. However, the cross sectional profile of marginal income taxes is very different. Under the first reform the marginal income tax schedule is very similar to the pre-reform tax schedule, with lower ability agents facing lower marginal income tax rates. Under the Rawlsian reform the marginal income tax schedules are U-shaped: low ability agents face extremely high tax rates, agents in the middle of the ability distribution face lower income taxes, and the income taxes increase again for high ability agents. In both reforms the shape of the marginal income tax schedule

does not vary much over time.

While under the first reform the government saves only 3.31% of resources, under the Rawlsian reform the government saves 115.5% of resources. It is not surprising that the Rawlsian reform saves more resources because it places much less constraints on the problem. But the difference in magnitudes and small savings under the first reform indicate that most of the potential gains can be found from reallocation of resources across individuals rather than across time.

Since human capital is unobservable, the efficient allocation can be implemented by using only one tool: income tax. The incentives to accumulate human capital are therefore provided only indirectly, through variations of income tax rates over time and across agents. The analysis of the efficient allocations shows that the optimal tax reform starts with relatively high marginal income tax schedules, and decreases them over time. There are several effects that contribute to this result. On one hand, current labor supply is complementary to past human capital investment: agents with higher schooling in the past supply more labor because they are more productive. On the other hand, since both labor and current human capital investments require time, current labor supply is a substitute to current human capital investment. In the short run, past human capital investments are already made, and only the substitution effect remains. Thus, marginal income taxes should be relatively high to encourage human capital investment and future labor supply. In a two period economy, only the complementarity effect applies in the second period, and the marginal income tax rates decrease under a very general set of conditions. In an infinite horizon economy both effects operate in the long run. Numerical simulations show that the complementarity effect ultimately dominates, and marginal income tax rates decrease over time.

## Related Literature

Recent research on optimal taxation with private information followed the seminal contributions of [Mirrlees \(1971\)](#), [Mirrlees \(1976\)](#), and [Mirrlees \(1986\)](#), and extended them to dynamic economies. It has focused primarily on cases when the dynamics of efficient allocations is driven by gradual revelation of private information over time ([Albanesi and Sleet \(2006\)](#), [Battaglini and Coate \(2008\)](#), [Farhi and Werning \(2005\)](#)), when the dynamic is driven by an aggregate state variable ([Werning \(2007\)](#)) or both ([Golosov, Kocherlakota, and Tsyvinski \(2003\)](#), [Kocherlakota \(2005\)](#)). In contrast, this paper focuses on a case when the driving force behind the dynamics is an unobservable individual specific state variable, namely human capital. Such environments have been recently studied in a two period setting, namely by [Grochulski and Piskorski \(2005\)](#) who study unobservable risky human capital and by [Albanesi \(2007\)](#), who studies observable risky physical capital. This paper abstracts from riskiness of human capital, but allows for fully dynamic environment with infinitely many periods.

Unlike most of this research ([Werning \(2007\)](#) is an exception), this paper assumes that individual abilities are permanent. The main role of income taxation is therefore to redistribute resources across time and people, rather than to provide insurance against dynamically evolving shocks. While the assumption of permanent shocks may seem restrictive, it is worth noting that heterogeneity in fixed effects is estimated to contribute something between 47% ([Storesletten, Telmer, and Yaron \(2004\)](#)) and 90% ([Keane and Wolpin \(1997\)](#)) to the variance of lifetime utility and thus can be considered at least as important as idiosyncratic shocks.

Dynamic private information environments where the dynamics is driven by an unobserved individual state variable have not been studied much so far because such problems have been relatively hard to solve. [Boháček and Kapička \(2008\)](#) show in a

somewhat different setting that these problems can be conveniently partitioned them into a static and a dynamic subproblem, with both of them being relatively easy to solve on their own. This paper uses similar partitioning to solve the social planner's problem.

This paper is also closely related to [Kapička \(2006\)](#) where I analyze the optimal steady state allocations when human capital is unobservable, the government is restricted to use current income taxes and agents cannot borrow or save.<sup>1</sup> This paper extends these results in two ways. First, no exogenous restrictions on the tax system are imposed. Second, I now solve for the whole transitional dynamics of efficient allocations. These additional results come at a cost, however. First, I assume that the government can freely borrow and save at an exogenously given interest rate rather than assuming that the resource constraint must clear in each period. Second, I assume that preferences take a very particular form: there are no income effects on leisure. Such preferences were recently used by [Diamond \(1998\)](#) to show that the analysis of the static optimal taxation problem can be much simplified. This specification brings even more benefits in a dynamic setting, because it simplifies computation of the efficient allocations.<sup>2</sup>

## THE MODEL

The agents live for  $T$  periods, where  $T$  can be infinite. Each period agents can choose their working time  $l_t$ , and time spent by accumulating human capital  $s_t$ . The

---

<sup>1</sup>See also [Diamond and Mirrlees \(2002\)](#) who analyze unobservable human capital investments in a static framework. While their framework cannot capture all the aspects of human capital accumulation, it allows them to give exact conditions when endogeneity of human capital will cause the optimal marginal income taxes to decrease.

<sup>2</sup>In addition, [Saez \(2001\)](#) shows that, at least in a static setting, the optimal tax system is not so different from the optimal tax system when income effects are nonzero.

preferences are given by

$$\sum_{t=0}^T \beta^t [c_t - V(l_t, s_t)], \quad 0 < \beta < 1,$$

where  $V : \mathbb{R}_+^2 \rightarrow \mathbb{R}$  is disutility from working and schooling. It is twice continuously differentiable on  $\mathbb{R}_{++}^2$  in both arguments, strictly increasing and strictly convex. It is also assumed that  $V_{ls} \geq 0$  so that current schooling and current labor supply are substitutes. Note that the period utility is linear in consumption: there are no income effects on leisure.

Each agent is assigned some minimal lifetime utility entitlement  $\underline{U}(\theta) < \infty$ . The function  $\underline{U} : \Theta \rightarrow \mathbb{R}$  is thus the lower bound on the utility distribution a social planner must deliver.

The agent's human capital at the beginning of period  $t + 1$ ,  $h_{t+1}$ , depends on time spent accumulating it in last period  $s_t$ , previous human capital  $h_t$ , and is given by a function  $G : \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}_+$ :

$$h_{t+1} = G(h_t, s_t).$$

The human capital accumulation function  $G$  is continuously differentiable on  $\mathbb{R}_{++} \times (0, 1]$ , strictly increasing and strictly concave. Schooling time required to have next period human capital  $h_{t+1}$  when current human capital is  $h_t$  is obtained by inverting the production function and will be denoted by  $S(h_t, h_{t+1})$ .

Each individual is associated with an ability level  $\theta \in [\underline{\theta}, \bar{\theta}] = \Theta$ , which is constant over time. Agent's ability and human capital determine her skills  $\theta h_t$  and, together with labor supply, her output  $y_t = \theta h_t l_t$ . The ability is drawn from a distribution function  $F$ , which is differentiable and has density  $f(\theta)$ .

Agent's abilities, as well as her human capital and labor supply, are private infor-

mation of the agent, and are not observed by the social planner. The only exception is the initial human capital  $h_0$ , which is identical for all agents and is assumed to be publicly observable. Consumption and output are both publicly observable as well.

### Efficient Allocations

An *allocation* is defined as a collection of functions  $\sigma = (u, y, h')$  where  $u : \Theta \rightarrow \mathbb{R}_+$  specifies lifetime utility,  $y = \{y_t\}_{t=0}^T$  with  $y_t : \Theta \rightarrow \mathbb{R}_+$  specifies output in period  $t$ , and  $h' = \{h_{t+1}\}_{t=0}^T$  with  $h_{t+1} : \Theta \rightarrow \mathbb{R}_+$  specifies recommended amount of human capital at the beginning of period  $t + 1$ .<sup>3</sup>

Since the agent's ability is her private information, the social planner needs to motivate the agent to provide a truthful information about her. At the beginning of period 0 the agent reports her type to the social planner. She chooses a human capital sequence  $\hat{h}' = \{\hat{h}_{t+1}\}_{t=0}^T$ . The utility of a  $\theta$ -type agent who reports  $\hat{\theta}$  is

$$W_\sigma(h_0, \theta; \hat{\theta}) = \max_{\hat{h}'} u(\hat{\theta}) - \sum_{t=0}^T \beta^t [V(\frac{y_t(\hat{\theta})}{\theta \hat{h}_t}, S(\hat{h}_t, \hat{h}_{t+1})) - V(\frac{y_t(\hat{\theta})}{\theta h_t(\hat{\theta})}, S(h_t(\hat{\theta}), h_{t+1}(\hat{\theta})))]$$

The incentive compatibility constraint requires the allocation to be such that  $\theta$ -type agent prefers to report her own type to any other report and chooses the recommended human capital sequence. Thus, an allocation  $\sigma$  is *incentive compatible* if, for all  $\theta \in \Theta$ ,

$$\begin{aligned} \{h_{t+1}(\theta)\}_{t=0}^T &\in \arg \max W_\sigma(h_0, \theta; \theta) \\ W_\sigma(h_0, \theta; \theta) &\geq W_\sigma(h_0, \theta; \hat{\theta}) \quad \forall \hat{\theta} \in \Theta. \end{aligned} \tag{1}$$

Note that  $W_\sigma(h_0, \theta; \theta) = u(\theta)$  by construction and so any incentive compatible allo-

---

<sup>3</sup>Schooling  $s_t$  and labor supply  $l_t$  can be recovered by inverting the human capital production function and output production function respectively.

cation delivers the distribution of utilities  $u$  by construction. Let  $\Sigma^{IC}$  be the set of allocations that are incentive compatible.

The social planner minimizes the costs of delivering the distribution of lifetime utilities  $\underline{U}$  or, equivalently, maximizing the present value of revenues, subject to the constraint that the allocation must be incentive compatible. It is assumed that the interest rate equals to the discount rate.

Since  $\sum_{t=0}^T \beta^t c_t = \sum_{t=0}^T \beta^t (u_t + V(l_t, s_t)) = u + \sum_{t=0}^T \beta^t V(l_t, s_t)$ , the revenue maximization problem can be written as

$$\Psi(\underline{U}, h_0) = \min_{\sigma \in \Sigma^{IC}} \int_{\Theta} \left\{ \sum_{t=0}^T \beta^t [y_t(\theta) - V(\frac{y_t(\theta)}{\theta h_t(\theta)}, S(h_t(\theta), h_{t+1}(\theta)))] - u(\theta) \right\} f(\theta) d\theta,$$

subject to the promise keeping constraint

$$u(\theta) \geq \underline{U}(\theta) \quad \forall \theta \in \Theta. \quad (2)$$

The *efficient allocation* is the allocation that attains the minimum of the social planner's problem. Note that the efficient allocation does not determine the sequence of period utilities, or, equivalently, the sequence of consumption allocations, but only their present values.<sup>4</sup>

The social planner's revenue maximization problem has the advantage that it does not directly depend on any particular social welfare function. But it in fact encompasses one particular form of government's preferences toward inequality. If  $\underline{U}(\theta) = -\infty$  for all  $\theta > \underline{\theta}$  then the social planner's preferences are Rawlsian: the social planner cares only about the lifetime utility of the lowest ability agent  $\underline{\theta}$ .

---

<sup>4</sup>The nonnegativity constraint on consumption is not imposed explicitly, but it is checked numerically in the quantitative simulations.

## FIRST ORDER APPROACH

In this section I will characterize the incentive compatible allocations by using the first order approach. The incentive compatible allocations will be easier to characterize in terms of labor supply  $l_t(\theta) = \frac{y_t(\theta)}{\theta h_t(\theta)}$  rather than in terms of output. Thus, from now on, an allocation  $\sigma$  is given by  $\sigma = (u, l, h')$ .

Necessary conditions for incentive compatibility are given by the envelope condition in  $\theta$  and the first-order condition in  $h_{t+1}$ :<sup>5</sup>

**Proposition 1** *If an allocation is incentive compatible then*

$$u(\theta) = \int_{\underline{\theta}}^{\theta} \sum_{t=0}^T \beta^t V_{l,t}(\varepsilon) l_t(\varepsilon) \frac{d\varepsilon}{\varepsilon} + u(\underline{\theta}), \quad (3)$$

and

$$\frac{V_{s,t}}{G_{s,t}} \geq \beta(V_{l,t+1} \frac{l_{t+1}}{h_{t+1}} + V_{s,t+1} \frac{G_{h,t+1}}{G_{s,t+1}}) \quad (4)$$

with equality only if  $S(h_t, h_{t+1}) > 0$ , where  $V_{x,t}(\theta) = V_x[l_t(\theta), S(h_t(\theta), h_{t+1}(\theta))]$ , and  $G_{x,t}(\theta) = G_x[h_t(\theta), S(h_t(\theta), h_{t+1}(\theta))]$ .

Equation (3) shows how the agent's lifetime utility varies with her type. The variation in lifetime utility is proportional to the informational rent the agent obtains from having a given ability level. The utility of the lowest type agent, who gets no informational rent, is given by  $u(\underline{\theta})$ . Equation (4) is the Euler equation in human capital investment. The left hand side equals to the marginal costs of increasing human capital next period by one unit. The right hand side shows that there are two types of private benefits from such an investment. First, higher investment in human capital increases agent's income next period. This effect is equal to  $V_{l,t+1} \frac{l_{t+1}}{h_{t+1}}$ . Second, higher investment in human capital today decreases the amount of next pe-

---

<sup>5</sup>All proofs can be found in Appendix 1.

riod schooling time required to achieve a given level of human capital two periods ahead. This effect is captured by the term  $V_{s,t+1} \frac{G_{h,t+1}}{G_{s,t+1}}$ . The equation (4) thus says that as long as the agent is not limited by the requirement that schooling is negative, marginal costs of investing in human capital equal to the marginal benefits of doing so.

Constraints (3) and (4) limit social planner's choices in a conceptually different way. In the absence of the Euler equation (4) the social planner would be able to provide incentives for schooling and labor supply separately. Conditional on labor supply, schooling can be chosen without any distortions, i.e. marginal costs of schooling equated to social marginal benefits of schooling. As shown by Boháček and Kapička (2008), the social planner would typically want to increase schooling above the level the agents would privately choose. If, on the other hand, the constraint (3) is missing, then the social planner can eliminate distortions in labor supply completely by setting labor supply to be such that  $V_{l,t} = \theta h_t$ . Schooling is given by (4), and the social planner uses individual specific lump-sum transfers to satisfy the promise keeping constraint (2).

Let  $\Sigma^{FOC}$  be the set of allocations that satisfy the first order conditions (3) and (4). The constraints (3) and (4) are necessary for an allocation to be incentive compatible, but they are not sufficient. Hence  $\Sigma^{IC} \subseteq \Sigma^{FOC}$  and the solution to the relaxed social planner's problem may not be identical to the solution of the social planner's problem. There are two reasons why this may be true. First, an individual could find it profitable to deviate jointly in her choice of human capital and in her choice of report.<sup>6</sup> Second, even if joint deviations were not profitable, the envelope condition (3) might still not be enough to prevent a deviation in the report itself. This problem is well known from the static optimal taxation literature.

---

<sup>6</sup>Kocherlakota (2005) argues that similar joint deviations are in some cases profitable in a model where savings are hidden.

I will therefore compute the efficient allocations *assuming* that the first order approach is valid. After that, I will provide a numerical ex-post verification procedure that will justify this procedure.

## The Relaxed Social Planner's Problem

A *relaxed* social planner's problem is defined as a problem of maximizing the present value of revenue of delivering  $(\underline{U})$  by choosing an allocation that belongs to  $\Sigma^{FOC}$ :

$$\hat{\Psi}(\underline{U}, h_0) = \min_{\sigma \in \Sigma^{FOC}} \int_{\Theta} \left\{ \sum_{t=0}^T \beta^t [y_t(\theta) - V(l_t(\theta), S(h_t(\theta), h_{t+1}(\theta)))] - u(\theta) \right\} f(\theta) d\theta, \quad (5)$$

subject to the promise keeping constraint (2).

Let  $\mu(\theta)$  be the Lagrange multiplier on the Envelope condition (3). Define a *cumulative distortion function*  $X_\mu$  by

$$X_\mu(\theta) = \frac{1 - F(\theta)}{\theta f(\theta)} \frac{\int_{\theta}^{\bar{\theta}} \mu(\varepsilon) f(\varepsilon) d\varepsilon}{1 - F(\theta)}.$$

The cumulative distortion function captures the redistributive aspects of the social planner's problem. It consists of two terms. The first one is the ratio of the mass of people with abilities above any given ability level to a density of people with that ability. The second term reflects the social planner's preferences toward redistribution implicit in the promise keeping constraint (2).<sup>7</sup>

The cumulative distortion plays an important role in the determination of the efficient allocations. The reason is that knowledge of the cumulative distortion for a given agent allows one to determine the whole sequence of efficient labor supply and human capital allocations:

---

<sup>7</sup>Note that in case of Rawlsian preferences  $\mu(\theta) = 1$  for all  $\theta > \underline{\theta}$ , and so  $X_\mu(\theta) = \frac{1 - F(\theta)}{\theta f(\theta)}$ .

**Theorem 2** *An allocation  $\sigma^*$ , together with the Lagrange multipliers on the envelope condition  $\mu^*$ , solves the relaxed social planner's problem if and only if satisfies the envelope condition (3), the promise keeping constraint (2), and for all  $\theta \in \Theta$ ,*

$$[l^*(\theta, h_0), h'^*(\theta, h_0)] \in \arg \max_{l, h'} \sum_{t=0}^T \beta^t \{ \theta h_t l_t - V_t - V_{l,t} l_t X_{\mu^*}(\theta) \}, \quad (6)$$

*s.t. Euler Equation (4),*

$$\mu^*(\theta [u(\theta) - \underline{U}(\theta)]) = 0. \quad (7)$$

The relaxed social planner's problem can thus be broken into a continuum of separate maximization problems (6), connected only through the cumulative distortion function. This is very convenient both analytically and computationally. Computing the cumulative distortion function may be complicated, but it is an essentially a static problem. On the other hand, all the dynamics in the model is captured in the maximization problem (6), where only  $X_{\mu}(\theta)$  needs to be known.

## IMPLEMENTATION IN A MARKET ECONOMY

Consider now a market economy with income taxes. The agent's problem in such an economy is to maximize utility subject to a present value budget constraint, taking the tax system as given. In this section I define the efficient tax system that implements the efficient allocation, and show that it depends only on the past history of incomes.

To define such a tax system, I first construct a single tax function  $\Phi(\{y_t\}_{t=0}^T)$  that depends on the infinite sequence of incomes and implements the efficient allocation. I will then show that this tax function is equivalent to a sequence of tax functions  $\{\Phi_t\}_{t=0}^T$ ,  $\Phi_t : \mathbb{R}_+^{t+1} \rightarrow \mathbb{R}$  where taxes are paid each period and depend only on the past history of incomes.

Define a function  $\varphi(\{y_t\}_{t=0}^T, \theta)$  that solves the following problem:

$$\max_{\{h_{t+1}\}_{t=0}^T} \sum_{t=0}^T \beta^t [y_t - V(\frac{y_t}{\theta h_t}, S(h_t, h_{t+1}))] - \varphi(\{y_t\}_{t=0}^T, \theta) = W_{\sigma^*}(h_0, \theta; \theta),$$

where  $\sigma^* = \{c_t^*(\theta), y_t^*(\theta), h_{t+1}^*(\theta)\}_{t=0}^T$  is the efficient allocation. The function  $\varphi(\{y_t\}_{t=0}^T, \theta)$  is the present value of tax payments that will keep a  $\theta$ -type agent indifferent between choosing a sequence  $\{y_t\}_{t=0}^T$  and the efficient allocation. The tax function  $\Phi$  is defined as a supremum of functions  $\varphi$ :

$$\Phi(\{y_t\}_{t=0}^T) = \sup_{\theta \in \Theta} \varphi(\{y_t\}_{t=0}^T, \theta). \quad (8)$$

The function  $\Phi$  satisfies the following property:

**Lemma 3**  $\Phi(\{y_t^*(\theta)\}_{t=0}^T) = \sum_{t=0}^T \beta^t [y_t^*(\theta) - c_t^*(\theta)].$

By construction of the tax function  $\Phi$  and by Lemma 3, when an agent with ability  $\theta$  faces the tax function  $\Phi$ , she prefers a sequence  $\{y_t^*(\theta)\}_{t=0}^T$  to any other sequence  $\{y_t\}_{t=0}^T$ . In other words, the tax function  $\Phi$  implements the efficient allocation.

The tax policy  $\Phi$  defined in (8) has an undesirable feature that it depends on all future income and human capital levels. Fortunately,  $\Phi$  is not the only tax policy that implements the efficient allocation. There are other tax policies, that are equivalent to the tax policy  $\Phi$ , but do not depend on future variables. Following [Werning \(2007\)](#), I consider the following sequential representation of agent's tax liabilities. The agent pays taxes each period and tax payments in period  $t$ ,  $\Phi_t(y^t)$ , are given by

$$\Phi_t(y^t) = \beta^{-t} [\Phi(\{y_i\}_{i=0}^t) - \Phi(\{y_i\}_{i=0}^{t-1})],$$

where  $\Phi(\{y_i\}_{i=0}^t)$  is a tax paid by someone with an income sequence  $(y_0, y_1, \dots, y_t, 0, 0, \dots)$ .  $\Phi_t(y^t)$  represent additional tax liabilities of an agent that, instead of not working from

period  $t$  on decides to work for one additional period  $t$  and chooses income  $y_t$ . The present value of tax payments under the sequential representation is given by

$$\sum_{t=0}^T \beta^t \Phi_t(y^t) = \sum_{t=0}^T \beta^t \beta^{-t} [\Phi(\{y_i\}_{i=0}^t) - \Phi(\{y_i\}_{i=0}^{t-1})] = \Phi(\{y_i\}_{i=0}^T),$$

and so is, for any sequence  $\{y_i\}_{i=0}^T$ , exactly identical to the tax liabilities under the original tax system  $\Phi$ . Hence the sequence of tax functions  $\{\Phi_t\}_{t=0}^T$  implements the efficient allocation as well.

### SOME EXAMPLES SOLVED

Analytical results for the maximization problem (6) are not available with the current level of generality. This section will therefore analyze its the solution in several special cases. Each of them will illustrate some aspects of the general model. In all the examples below, marginal income tax rate refers to the marginal rate of substitution between consumption and output,  $\tau_t(\theta) = 1 - \frac{V_{l,t}(\theta)}{\theta h_t(\theta)}$ . Marginal income tax rates will be analyzed in terms of their dynamics, and compared to the marginal income tax rate  $\tau^s$  one would obtain in a static economy without human capital for the same cumulative distortion  $X$ .<sup>8</sup>

**Separable Utility, Full Depreciation.** Start by assuming that the disutility of labor and schooling is additively separable and human capital fully depreciates each period. One then obtains the following result:

**Proposition 4** *If  $V(l, s) = v(l) + e(s)$  and  $G(h, s) = s$  then the optimal marginal income tax rates are given by*

---

<sup>8</sup>The static marginal income tax rate  $\tau^s$  is given by  $\tau^s = \frac{(1+\nu^{-1})X}{1+(1+\nu^{-1})X}$ , where  $\nu = \frac{V_l}{lV_{ll}}$  (see [Diamond \(1998\)](#)).

$$\tau_0 = \tau^s$$

$$\frac{\tau_t}{1 - \tau_t} = \frac{1 + \epsilon^{-1}}{2 + \nu^{-1} + \epsilon^{-1}} \frac{\tau^s}{1 - \tau^s}, \quad t = 1 \dots T,$$

where  $\nu = \frac{v'(l)}{lv''(l)}$  and  $\epsilon = \frac{e'(s)}{se''(s)}$  are the elasticities of labor supply and schooling.

The marginal income tax rate in period 0 is thus identical to the static marginal income tax rate  $\tau^s$ . As follows from the second part of the Proposition, from period 1 on the marginal income tax rates are lower than  $\tau^s$ , and are constant over time. The reason why the marginal tax rate is higher in period 1 is that labor supply in all periods except for period 0 is complementary to investment in human capital: higher labor supply increases incentives to accumulate human capital in the previous period. On the other hand, the initial human capital  $h_0$  is fixed at the time of the reform, and will not respond to variations in the marginal tax rates in period 0.

Unobservability of human capital is essential in obtaining this time profile of marginal income tax rates. One can easily show that if human capital were observable, the social planner would like to set the marginal income tax rates to be equal to  $\tau^s$  in all periods. At the same time, he would equalize marginal cost of schooling to social marginal benefits by setting  $e'(h_{t+1}) = \beta\theta l_{t+1}$ . This violates the Euler equation (4), because individuals only equate marginal costs of schooling to private marginal benefits, which are lower because of the income tax.<sup>9</sup> Schooling subsidies are needed to increase incentives to invest in human capital.

In the absence of the envelope condition (3) the cumulative distortion  $X$  is equal to zero. The social planner would set both labor supply and human capital to their first best levels. A closed form solution for the first best levels can be obtained, and is given by  $l_0 = h_0^\nu \theta^\nu$ , and for  $t > 0$ ,  $l_t = k_l \theta^\nu \frac{1+\epsilon}{1-\nu\epsilon}$  and  $h_t = k_h \theta^\epsilon \frac{1+\nu}{1-\nu\epsilon}$  for some constants  $k_l$  and  $k_h$ . Those values are in general incompatible with both the promise keeping

---

<sup>9</sup>It is easy to show that while the Euler equation (4) is given by  $e'(h_{t+1})h_{t+1} = \beta v'(l_{t+1})l_{t+1}$ , the social planner would like to set  $e'(h_{t+1})h_{t+1} = \beta[1 + (1 + \nu^{-1})X]v'(l_{t+1})l_{t+1}$ .

constraint (2) and the envelope condition (3).

**Inseparable Utility, Full Depreciation.** Now assume, more realistically, that labor supply and schooling are substitutes in a sense that higher labor supply increases marginal disutility from schooling and vice versa. Assume again that human capital fully depreciates. If the economy lasts for only three periods 0, 1, 2 then the marginal income tax rates are characterized as follows:

**Proposition 5** *If  $V(l, s) = \frac{(l+s)^{1+\nu^{-1}}}{1+\nu^{-1}}$ ,  $G(h, s) = s$  and  $T = 2$  then the optimal marginal tax rates satisfy  $\tau_0 > \tau_1 > \tau_2$  and  $\tau_0 < \tau^s$ .*

The first part of the proposition shows that the marginal income tax rates are now strictly decreasing in all periods. The reason is that, in addition to complementarity between future labor supply and current schooling, current labor supply is a substitute to current schooling. In period 0 the complementarity effect is zero and a lower labor supply has an extra benefit of increasing incentives to invest in human capital at time 0. On the other hand, in period 2 the substitution effect is zero since no schooling takes place in period 2. Higher labor supply has an extra benefit of increasing schooling in period 1. In period 1, both the complementarity and substitution effects are in place and the marginal tax rate lies somewhere in between.

The level of the marginal income tax rates in the first period is determined by two effects. First, as mentioned in previous paragraph, the fact that labor supply and schooling in period 0 are substitutes tends to increase the marginal income tax rate. Second, the elasticity of labor supply increases because labor supply constitutes only a part of nonleisure time. The second part of the proposition shows that with CRRA utility the second effect dominates.<sup>10</sup>

**Separable Utility, Incomplete Depreciation.** Next proposition returns to

---

<sup>10</sup>However, one can show that with CARA utility it is the first effect that dominates. Human capital considerations can therefore increase marginal income tax rates in the first period.

the case when utility is separable in labor and schooling, but allows for less than full depreciation of human capital.

**Proposition 6** *If  $V(l, s) = v(l) + e(s)$ ,  $G(h, s) = h^\alpha s^{1-\alpha}$  and  $T = 2$  then the optimal marginal tax rates satisfy  $\tau_0 > \tau_1 > \tau_2$  and  $\tau_0 = \tau^s$ .*

If human capital does not fully depreciate then the marginal income tax rates again decrease over time. The reason is now that higher labor supply in period 2 increases incentives to invest in human capital in both period 0 and period 1, while higher labor supply in period 1 increases incentives to invest in human capital only in period 0. Thus, labor supply in period 2 should be encouraged more, i.e. marginal tax rates should be lower. Marginal income tax rates in period 0 again coincide with marginal tax rates in a static economy, because the utility is additively separable in labor and schooling.

**Inseparable Utility, 2 periods.** If the economy has only two periods then analytical results are available even without imposing any restrictions on human capital production function, and for utility function that is not separable in labor supply and schooling:

**Proposition 7** *If  $V(l, s) = \frac{(l+s)^{1+\nu-1}}{1+\nu-1}$  and  $T = 1$  then the optimal marginal tax rates satisfy  $\tau_0 > \tau_1$  and  $\tau_1 < \tau^s$ .*

The marginal income tax rates are again decreasing, and, at least in the second period, are lower than the marginal income tax rates in a static economy.

Previous examples show several principles that most likely drive the intertemporal profile of marginal income taxes in a more general infinitely lived economy. First, the intertemporal profile will be driven by the strength of substitutability between labor supply and current schooling, and complementarity between labor supply and past

schooling. Whenever the role of complementarity grows over time, as it does in the second and fourth example, marginal income taxes will decrease with time. Second, the intertemporal profile will depend on how fast human capital depreciate over time. One can, however, determine the magnitude of those effects only through numerical simulations.<sup>11</sup>

Finally, note that in all four examples marginal income tax rates decrease over time regardless of the initial human capital level  $h_0$ . What matters is how much schooling people choose relative to the labor supply, and not the absolute level of schooling (or human capital). In all four examples schooling is inefficiently low relative to labor supply, because people do not take into account all social benefits from schooling.

## QUANTITATIVE RESULTS

In this section I will numerically compute the optimal allocations and show how the welfare maximizing tax reform looks like. Set  $T = \infty$ . The disutility of nonleisure time and the human capital production function take the form

$$V(l, s) = \frac{(l + s)^{1+\nu^{-1}}}{1 + \nu^{-1}}$$

$$G(h, s) = (1 - \delta)h + \delta(hs)^\alpha.$$

The parameters to be chosen are the utility parameters  $k$  and  $\beta$ , the human capital production function parameters  $\alpha$  and  $\delta$ , and the ability density  $f(\theta)$ .

**Utility Function Parameters.** The time period is one year and so  $\beta = 0.96$ , implying the annual interest rate of 4%. The parameter  $\nu$  is closely linked to the Frisch elasticity of labor supply. Indeed, in a static model with no human capital,

---

<sup>11</sup> Kapička (2006) shows that if the human capital production is Cobb-Douglas then the complementarity effect dominates the substitutability effect in the steady state of an infinite period economy. The long run marginal income tax rate is then strictly lower than  $\tau^s$ , as in Example 4.

the elasticity of labor supply would be equal to  $\nu$ . Endogenous human capital accumulation complicates this link for three reasons. First, short-run labor elasticity is not the same as long-run labor elasticity. Second, labor supply responses to temporary changes in wages or taxes are different to labor supply responses to permanent changes in wages or taxes. And, third, the link depends whether schooling (or part of thereof) takes place on the job and is included in the measures of reported hours worked.<sup>12</sup>

To obtain a fraction of schooling attributed to reported hours worked I proceed as follows. According to [Francis and Ramey \(2009\)](#), people nowadays spend on average about 2.6 hours per week at school, and about 23 hours at work. The ratio of formal schooling to reported hours of work is  $\psi = \frac{23}{2.6}$ . One can show that the model predicts that steady state labor supply to schooling ratio is  $\frac{l}{s} = p = \frac{1-\beta(1-\delta(1-\alpha))}{\beta\delta(1-\alpha)}$ , identical for all agents. For those two facts to be consistent with each other, a fraction  $\frac{\psi-p}{1+\psi}$  of schooling must take the form of on the job training. This number turns out to be 0.67.

Empirical studies typically find that the short-run Frisch labor supply elasticity is between 0.1 and 0.3.<sup>13</sup> For the benchmark calibration I therefore choose  $k$  to be such that the short run elasticity of labor supply with respect to permanent wage changes is 0.2, assuming that 67% of schooling takes place on the job and is therefore included in the measure of labor supply. The resulting value of  $\nu$  is 0.302.

**Human Capital Production Function Parameters.** The evidence regarding the parameters of the human capital production function is very diverse. The estimates of  $\alpha$  range from about 0.5 to over 0.9. The evidence regarding  $\delta$  ranges from

---

<sup>12</sup>[Imai and Keane \(2004\)](#) show that omitting human capital accumulation biases the estimates of the labor supply elasticity downwards. Their revised estimates are not directly applicable to this model, however, because in their environment human capital accumulation takes a very different form of learning by doing.

<sup>13</sup> See [Pencavel \(1986\)](#), Table 1.22. For females, the labor supply elasticity is typically higher.

0.0016 to 0.089.<sup>14</sup> For the benchmark simulation I choose  $\alpha = 0.625$ , as estimated by [Huggett, Ventura, and Yaron \(2009\)](#), and  $\delta = 0.04$ .

**Distribution of Abilities.** The distribution of abilities is calibrated in the following way. I first parameterize the actual U.S. tax code. Then I compute the steady state distribution of incomes under this tax code and set the distribution of shocks to be such that the distribution of earnings in the model is the same as in the data: If the c.d.f. of earnings in the data is  $\hat{F}(Y)$  and the model predicts steady state earnings of a  $\theta$ -type agent to be  $\hat{Y}(\theta)$  then the c.d.f of the ability distribution is given by  $F(\theta) = \hat{F}(\hat{Y}(\theta))$ . The tax return data for 1992, smoothed by a 7-point moving average, are used for the empirical distribution of earnings. The ability distribution is discretized by a vector of  $N = 300$  ability levels.

The U.S. income tax code is approximated by the following function:

$$T(y) = \kappa_0 \left( y - (y^{-\kappa_1} + \kappa_2)^{-\frac{1}{\kappa_1}} \right) + (\tau_c + \tau_{ss})y.$$

This functional form has been recently used frequently in the quantitative public finance literature.<sup>15</sup> The consumption tax rate  $\tau_c$  is set at 5.2%, following [Mendoza, Razin, and Tesar \(1994\)](#). The payroll tax rate  $\tau_{ss}$  is set at 12.4%. [Gouveia and Strauss \(1994\)](#) show that parameters  $\kappa_0 = 0.258$  and  $\kappa_1 = 0.768$  best approximate the U.S. income tax. The remaining parameter  $\kappa_2$  is chosen in order to match the average marginal income tax rate of 17.4%, as reported by [Gouveia and Strauss \(1994\)](#). The resulting value is  $\kappa_2 = 0.000226$ .

The calibration yields the following results. The average income per person in the calibrated allocation is \$45321 and the Gini coefficient of income is 0.397. Leisure,

---

<sup>14</sup> See [Browning, Hansen, and Heckman \(1999\)](#), Table 2.3 and 2.4, [Huggett, Ventura, and Yaron \(2009\)](#) and [Trostel \(1993\)](#).

<sup>15</sup>See [Castañeda, Díaz-Giménez, and Ríos-Rull \(1999\)](#), [Conesa and Krueger \(2006\)](#), or [Conesa, Kitao, and Krueger \(2009\)](#).

schooling and human capital all increase with abilities. Since the human capital production function exhibit increasing returns, human capital distribution is more unequal than the distribution of labor supply or schooling. The labor-schooling ratio  $p$  is 2.27 is independent of the abilities, and so all people spend 69.4% of their nonleisure time by working, and the remaining fraction by schooling.

## Benchmark Results

The calibration procedure generates a distribution of lifetime utilities  $\underline{u}$ , and a distribution of initial human capital levels  $h_0$ . In the benchmark reform I numerically compute the relaxed social planner’s cost minimization problem (2), where the social planner needs to deliver a lifetime utility of at least  $\underline{u}$ , taking the initial distribution of human capital as given.<sup>16</sup> I verify that the resulting allocation is incentive compatible. (see Appendix 2 for details).

The first panel of Figure 1 plots the average optimal marginal income tax rates for all periods. They decrease by about 25 percentage points, from 41% to 16%. The decline in average marginal tax rates is consistent with the intuition conveyed in the examples: initially, labor supply and schooling are substitutes, and it is optimal to set the marginal income tax rates high to encourage schooling and future labor supply. In the long run, the complementarity effect takes over and marginal income tax rates

---

<sup>16</sup>Calibrating the ability distribution to the data creates a theoretical problem, because it implies that there is a history of incomes for each agent, and that initial human capital stock varies across people. The social planner could in principle use this information to infer the ability of each agent. The differentiated lump-sum taxes could then be used. See [Berliant and Ledyard \(2004\)](#). To circumvent this problem, I will assume that the government is prohibited by a constitution to use the initial human capital and incomes before period zero as a direct argument in the tax function.

decline.<sup>17</sup>

[INSERT FIGURE 1 HERE]

The second panel of Figure 1 plots the optimal marginal tax rates in periods 0, 50, and 250, and compares them with the pre-reform tax schedule. The shape of the marginal income tax schedule is approximately the same over time. Thus, although the marginal income tax schedule shifts over time, taxes do not become any more (or less) progressive. The shape of the marginal income tax schedule itself doesn't change much from its pre-reform shape. This is mostly driven by the restriction that the reform cannot change the distribution of utilities.

The optimal tax reform provides modest cost savings of 3.32% of government's present value of resources. This is equal to approximately \$300 per person per year. The first panel of Figure 2 shows that the aggregate output increases significantly in the long run by 60% (to \$71826 per person), but declines in the short run (by 11% to \$39817 per person immediately after the tax reform is implemented). The reason is that a decreasing pattern of marginal income taxes motivates people to increase schooling. The third panel of Figure 2 shows that, in the short run, schooling increases by 25% on average, while labor supply decreases by 12%. In the long run, both schooling and labor supply increase by 18% on average. Human capital increases by 31% on average in steady state.

[INSERT FIGURE 2 HERE]

The distribution of a change in income, labor supply and schooling is plotted in the second and fourth panel of Figure 2. All agents are qualitatively affected in the same

---

<sup>17</sup> If schooling were observable then the marginal income taxes would be approximately constant over time. See [Boháček and Kapička \(2008\)](#) for numerical simulations. Similarly to Example 1, unobservability of human capital is therefore critical in obtaining the decreasing marginal income taxes over time.

way: income and labor supply decrease in the short run and decrease in the long run, while schooling increases both in the short and long run. The Figure shows that long run increases in income schooling and labor supply are the largest for the high ability agents. On the other hand, the short run decline in income and labor supply is U-shaped: the decline is the smallest for the low ability and for the high ability agents.

### Rawlsian Reform

A Rawlsian social planner is only constrained by the requirement that the utility of the lowest able agent must be  $\underline{u}(\theta)$ . He is free to choose the utility of all the other agents freely, subject to the constraints (3) and (4). Figure 3 shows the distribution of lifetime utilities before and after the reform. The lifetime utility of a majority of agents decreases. Interestingly, the lifetime utility of the ablest agents increases, despite the fact that the social planner does not care about their utility directly. The only reason why this happens is that it helps to increase the social planner's revenue.

[INSERT FIGURE 3 HERE]

Figure 4 shows that, on average, the marginal income tax rates are now much higher than in the previous reform. Marginal income taxes now decrease from 65% to 45% on average. Initially, almost all people face higher marginal income taxes, while in the long run people earning more than 50000 USD a year face lower tax rates. The pattern of the marginal income tax schedule is now drastically different from its pre-reform shape, with tax rates first decreasing with income, and then increasing again. It's shape is caused primarily by the shape of the cumulative distortion  $X = \frac{1-F}{\theta f}$ .<sup>18</sup>

---

<sup>18</sup> See Saez (2001) for the optimal marginal income tax schedule in a static setting, which has a similar shape.

Marginal income tax rates decrease over time for all ability levels. This is true despite the fact that lower ability agents decrease their human capital over the transition. The decline in marginal income taxes is, however, significantly smaller for those agents.

[INSERT FIGURE 4 HERE]

The Rawlsian reform provides savings of 115.5% of the present value of resources, equal to \$10471 per person and year. That is, the savings are 35 times higher than in the previous reform. It is, however, not surprising that the difference is large, since the Rawlsian problem provides significantly less constraints on the social planner's problem. In the long run, the aggregate income increases by only 30%, while average schooling and labor supply actually slightly decreases. Therefore, almost all the welfare gains are obtained not from an overall increase of schooling or labor supply, but from a better reallocation of those across people. In other words, the increase in welfare comes mainly from an improvement in the correlation between ability and human capital or labor supply.

### **Alternative Assumptions**

This section investigates the intertemporal profile of the marginal income taxes under alternative assumptions about the utility function parameter  $\eta$  and the human capital production function parameters  $\delta$  and  $\alpha$ . I focus on the Rawlsian reform because, due to exogeneity of the cumulative distortion, the comparison is easier to make.

The first panel of Figure 5 shows the average marginal income tax rates when  $\eta$  is set to a low and high value of 0.20 and 0.39, and when  $\alpha$  is set to a low and high value of 0.5 and 0.7. Higher parameter values decrease the marginal income tax rates, while lower parameter values increase them. This is intuitive, because higher  $\eta$

corresponds to higher labor supply elasticity, and higher  $\alpha$  implies that human capital is more sensitive to schooling. In both cases, marginal income tax rates create larger distortions. Hence, in the optimum, marginal income tax rates decrease. The second panel of Figure 5 shows that with higher depreciation of human capital the decline in average marginal income taxes is steeper and faster. This is also intuitive, because the convergence to steady state is now faster. However, the magnitude of the decrease is not affected by the depreciation rate very much.

[INSERT FIGURE 5 HERE]

## CONCLUSIONS

This paper analyzes efficient allocations in a dynamic economy where private information ability shocks are permanent and human capital is endogenous and unobservable by the government. The main contribution is that it does in a framework that is tractable enough so that the efficient allocations can be analyzed quantitatively. The paper also provides several theoretical results regarding the nature of optimal income taxes.

These results rely on several assumptions: There are no income effects, the government can freely borrow and save at an exogenously given interest rate, and the ability shocks are permanent. What happens if these assumptions are relaxed?

When the assumption of no income effects is relaxed, the maximization problem (6) will also depend on individual's consumption. Thus, the dimension of the problem increases and, at least for the infinitely lived economy, the problem becomes hard to compute numerically. The assumption of permanent shocks is critical because it ensures that the cumulative distortion is constant over time. If private information about abilities is revealed only gradually, this result will no longer hold. Similar complications arise when the assumption that the government has access to perfect credit

markets is relaxed. On the other hand, one could extend the model by incorporating life-cycle elements quite easily, as long as the life-cycle profile of productivity is nonstochastic. This important extension is left for future research.

## References

- Albanesi, S. (2007). Optimal taxation of entrepreneurial capital with private information. Working paper, Columbia University. 4
- Albanesi, S. and C. Sleet (2006). Dynamic optimal taxation with private information. *Review of Economic Studies* 73(1), 1–30. 4
- Battaglini, M. and S. Coate (2008). Pareto efficient income taxation with stochastic abilities. *Journal of Public Economics* 92(3-4), 844–868. 4
- Berliant, M. and J. Ledyard (2004). Optimal dynamic nonlinear income taxes with no commitment. working paper, Washington University. 21
- Boháček, R. and M. Kapička (2008). Optimal human capital policies. *Journal of Monetary Economics* 55, 1–16. 4, 10, 22
- Browning, M., L. P. Hansen, and J. J. Heckman (1999). Micro data and general equilibrium models. In M. Woodford and J. Taylor (Eds.), *Handbook of Macroeconomics, Volume 1, Chapter 8*. North Holland. 20
- Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (1999). Earnings and wealth inequality and income taxation: Quantifying the trade-offs of switching to a proportional income tax in the u.s. Working paper, University of Pennsylvania. 20
- Conesa, J. C., S. Kitao, and D. Krueger (2009). Taxing capital? not a bad idea after all! *American Economic Review* forthcoming. 20
- Conesa, J. C. and D. Krueger (2006). On the optimal progressivity of the tax code.

- Journal of Monetary Economics* 53(7), 1425–1450. 20
- Diamond, P. (1998). Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review* 88, 83–95. 5, 14
- Diamond, P. and J. Mirrlees (2002). Optimal taxation and the le chatelier principle. Working paper, MIT. 5
- Farhi, E. and I. Werning (2005). Inequality and social discounting. *Journal of Political Economy* 115(1), 365–402. 4
- Francis, N. and V. A. Ramey (2009). A century of work and leisure. *American Economic Review: Macroeconomics forthcoming*. 19
- Golosov, M., N. R. Kocherlakota, and A. Tsyvinski (2003). Optimal indirect and capital taxation. *Review of Economic Studies* 70, 569–587. 4
- Gouveia, M. and R. P. Strauss (1994). Effective federal individual income tax functions: An exploratory empirical analysis. *National Tax Journal* 47(2), 317–339. 20
- Grochulski, B. and T. Piskorski (2005). Optimal wealth taxes with risky human capital. Working paper, Federal Reserve Bank of Richmond. 4
- Huggett, M., G. Ventura, and A. Yaron (2009). Sources of lifetime inequality. Working paper, Georgetown University. 20
- Imai, S. and M. P. Keane (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review* 45(2), 601–641. 19
- Kapička, M. (2006). Optimal taxation and human capital accumulation. *red* 9, 612–639. 5, 18, 29
- Keane, M. and K. Wolpin (1997). The career decisions of young men. *jpe* 105, 473–522. 4

- Kocherlakota, N. R. (2005). Zero expected wealth taxes: A mirrless approach to dynamic optimal taxation. *Econometrica* 73, 1587–1621. 4, 10
- Mendoza, E. G., A. Razin, and L. L. Tesar (1994). Effective tax rates in macroeconomics: Cross-country estimates of tax rates on factor incomes and consumption. *Journal of Monetary Economics* 34(3), 297–323. 20
- Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38, 175–208. 4
- Mirrlees, J. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics* 6, 327–358. 4
- Mirrlees, J. (1986). The theory of optimal taxation. In K. J. Arrow and M. D. Intriligator (Eds.), *Handbook of Mathematical Economics, vol. III, Chapter 24*. Elsevier. 4, 29
- Pencavel, J. (1986). Labor supply of men: A survey. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics, Volume 1, Chapter 1*. North Holland. 19
- Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies* 68, 205–229. 5, 23
- Storesletten, K., C. Telmer, and A. Yaron (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics* 51, 609–633. 4
- Trostel, P. A. (1993). The effect of taxation on human capital. *Journal of Political Economy* 101, 327–350. 20
- Werning, I. (2007). Optimal fiscal policy with redistribution. *Quarterly Journal of Economics* 22(3), 925–967. 4, 13

## APPENDIX 1: PROOFS

**Proof of Proposition 1.** The proof of the envelope condition (3) is a fairly straightforward modification of the proof that can be found in (Kapička 2006) or in (Mirrlees 1986). To show necessity of (4), differentiate  $W(h_0, \theta, \theta)$  w.r.t.  $h_{t+1}$ . If  $h_{t+1} > G(h_t, 0)$  then

$$\begin{aligned} 0 &= \frac{\partial}{\partial h_{t+1}} \sum_{t=0}^T \beta^t [c_t(\theta) - V(\frac{y_t(\theta)}{\theta h_t}, S(h_t, h_{t+1}))] \\ &= -\frac{V_{s,t}}{G_{st}} - \beta(V_{l,t+1} \frac{y_{t+1}}{\theta h_{t+1}^2} + V_{s,t+1} \frac{G_{h_{t+1}}}{G_{st+1}}). \end{aligned}$$

Equation (4) follows from the fact that  $\frac{y_{t+1}}{\theta h_{t+1}^2} = \frac{l_{t+1}}{h_{t+1}}$ . ■

**Proof of Theorem 2.** Define  $\Delta$  to be the set of all labor supply and human capital sequences  $(l, h')$  that satisfy the Euler equation (4) for all  $t \geq 0$ .<sup>19</sup>

Necessity of the conditions is obvious. For sufficiency, let  $\sigma^* = (u^*, l^*, h'^*)$ , together with  $\mu^*$  be an allocation that satisfies the conditions of the theorem. Let  $\sigma = (u, l, h') \in \Delta$  be an alternative allocation that satisfies the envelope condition (3) and the promise keeping constraint (2). Write the social planner's objective function as follows:

$$\begin{aligned} \hat{\Psi}(\underline{U}, h_0) &= \int_{\Theta} \left\{ \sum_{t=0}^T \beta^t [\theta h_t^*(\theta) l_t^*(\theta) - V_t^*(\theta)] - u^*(\theta) \right\} f(\theta) d\theta \\ &= \int_{\Theta} \sum_{t=0}^T \beta^t \left[ \theta h_t^*(\theta) l_t^*(\theta) - V_t^*(\theta) - \mu^*(\theta) \int_{\underline{\theta}}^{\theta} V_{l,t}^* l_t^* \frac{d\varepsilon}{\varepsilon} \right] f(\theta) d\theta + \xi_{\mu^*, u^*} \\ &= \int_{\Theta} \sum_{t=0}^T \beta^t \left[ \theta h_t^*(\theta) l_t^*(\theta) - V_t^*(\theta) - \frac{V_{l,t}^*(\theta) l_t^*(\theta)}{f(\theta)\theta} \int_{\theta}^{\bar{\theta}} \mu^*(\varepsilon) f(\varepsilon) d\varepsilon \right] f(\theta) d\theta + \xi_{\mu^*, u^*}, \end{aligned}$$

where  $\xi_{\mu, u} = \int_{\Theta} \{(\mu(\theta) - 1)u(\theta) - \mu(\theta)u(\underline{\theta})\} f(\theta) d\theta$ . The second equality follows from

<sup>19</sup>Writing  $(l, h') \in \Delta$  means that this property holds for all  $\theta \in \Theta$ .

the fact that the envelope condition holds with equality for  $(u^*, l^*, h^*)$ . The third equality follows from integration by parts.

It follows from (7) that  $u^*$  maximizes  $\xi_{\mu^*, u}$ . To see this note first that  $u^*(\underline{\theta}) = \underline{U}(\underline{\theta})$  since the social planner can always uniformly decrease utility without affecting the constraints (3) and (4). Second,  $\mu(\theta) \leq 1$  for all  $\theta > \underline{\theta}$ , with  $\mu(\theta) = 1$  if the promise keeping constraint doesn't bind. This holds since if  $\mu(\theta) > 1$  then  $u(\theta) = \infty > \underline{U}(\theta)$  which in turn implies that  $\mu(\theta) = 1$ , a contradiction. But if  $\mu(\theta) \leq 1$  then  $(\mu(\theta) - 1)u^*(\theta) = (\mu(\theta) - 1)\underline{U}(\theta) \geq (\mu(\theta) - 1)u(\theta)$  for any  $u(\theta) \geq \underline{U}(\theta)$ .

Since  $(l^*, h^*)$  solves the problem by (6) and  $u^*$  maximizes  $\xi_{\mu^*, u}$ ,

$$\begin{aligned} \hat{\Psi}(\underline{U}, h_0) &\geq \int_{\Theta} \sum_{t=0}^T \beta^t \left[ \theta h_t(\theta) l_t(\theta) - V_t(\theta) - \frac{V_{l,t}(\theta) l_t(\theta)}{f(\theta)\theta} \int_{\theta}^{\bar{\theta}} \mu^*(\varepsilon) f(\varepsilon) d\varepsilon \right] f(\theta) d\theta + \xi_{\mu^*, u} \\ &= \int_{\Theta} \left\{ \sum_{t=0}^T \beta^t [\theta h_t(\theta) l_t(\theta) - V_t(\theta)] - u(\theta) \right\} f(\theta) d\theta. \end{aligned}$$

The last equality follows from reverse integration by parts and from the fact that the envelope condition holds with equality for  $(u, l, h')$ . ■

**Proof of Proposition 4.** The Euler equation can now be written as

$$e'(h_{t+1})h_{t+1} = \beta v'(l_{t+1})l_{t+1}.$$

Let  $\lambda_{t+1}$  be the Lagrange multiplier on this constraint. The Lagrangean is

$$\max_{l, h} \sum_{t=0}^T \beta^t \{ \theta h_t l_t - v(l_t) - e(h_{t+1}) - v'(l_t) l_t X \} + \sum_{t=0}^{T-1} \beta^t \lambda_{t+1} \{ \beta v'(l_{t+1}) l_{t+1} - e'(h_{t+1}) h_{t+1} \}$$

The first order conditions in  $l_0$  is

$$\frac{\theta h_0}{v'(l_0)} - 1 = (1 + \nu^{-1})X.$$

The first order conditions in  $l_t$  and  $h_t$  for  $t = 1 \dots T$  are

$$\begin{aligned}\frac{\theta h_t}{v'(l_t)} - 1 &= (1 + \nu^{-1})(X - \lambda_t), \\ \beta \frac{\theta l_t}{e'(h_t)} - 1 &= (1 + \epsilon^{-1})\lambda_t.\end{aligned}$$

Using the Euler equation, eliminating  $\lambda$ , and using the definition of  $\tau$ , the desired result obtains. ■

**Proof of Proposition 5.** The Euler equation can now be written as

$$v'(l_t + h_{t+1})h_{t+1} = \beta v'(l_{t+1} + h_{t+2})l_{t+1}.$$

Let  $\lambda_{t+1}$  be the Lagrange multiplier on this constraint. The Lagrangean is

$$\begin{aligned}\max_{l,h} \sum_{t=0}^2 \beta^t \{ \theta h_t l_t - v(l_t + h_{t+1}) - v'(l_t + h_{t+1})l_t X \} \\ + \sum_{t=0}^1 \beta^t \lambda_{t+1} \{ \beta v'(l_{t+1} + h_{t+2})l_{t+1} - v'(l_t + h_{t+1})h_{t+1} \}\end{aligned}$$

Let  $v_t = v(l_t + h_{t+1})$  and  $\vartheta_t = \frac{\theta h_t}{v'_t} - 1$ . The first order conditions in  $l_0, l_1, l_2$  are

$$\vartheta_0 = (1 + \frac{v''_0}{v'_0} l_0)X + \lambda_1 \frac{v''_0}{v'_0} h_1 \tag{9}$$

$$\vartheta_1 = (1 + \frac{v''_1}{v'_1} l_1)(X - \lambda_1) + \lambda_2 \frac{v''_1}{v'_1} h_2 \tag{10}$$

$$\vartheta_2 = (1 + \frac{v''_2}{v'_2} l_2)(X - \lambda_2) \tag{11}$$

The first order conditions in  $h_1, h_2$ , together with the Euler equation imply

$$\vartheta_1 = \frac{v_0''}{v_0'} l_0 X + \lambda_1 \left(1 + \frac{v_0''}{v_0'} h_1\right) \quad (12)$$

$$\vartheta_2 = \frac{v_1''}{v_1'} l_1 (X - \lambda_1) + \lambda_2 \left(1 + \frac{v_1''}{v_1'} h_2\right). \quad (13)$$

To show the first part of the proposition, note that  $\vartheta_1 = \vartheta_0 - (X - \lambda_1)$  and  $\vartheta_2 = \vartheta_1 - (X - \lambda_1 - \lambda_2)$ . Since  $X - \lambda_1 > X - \lambda_1 - \lambda_2$ , the Proposition will be proven if  $X - \lambda_1 - \lambda_2$  is positive. To obtain this, eliminate  $\vartheta_1$  by using (10) and (12), and  $\vartheta_3$  by using (11) and (13). Combining both equations and using  $\frac{v_2''}{v_2'} l_2 = \nu^{-1}$ , one gets

$$x - \lambda_2 - \lambda_1 = \frac{\frac{v_0''}{v_0'} l_0 \lambda_2 + \frac{v_2''}{v_2'} h_2 \lambda_1}{2 + \nu^{-1} - \frac{v_0''}{v_0'} l_0}.$$

Since  $\nu^{-1} = \frac{v_0''}{v_0'} (l_0 + h_1) > \frac{v_0''}{v_0'} l_0$ ,  $X - \lambda_1 - \lambda_2 > 0$ .

The second part of the Proposition follows from rewriting (9) as

$$\vartheta_0 = (1 + \nu^{-1})X - (X - \lambda_1) \frac{v_0''}{v_0'} h_1.$$

and the fact that  $X - \lambda_1 > 0$ . ■

**Proof of Proposition 6.** The Euler equation can now be written as

$$e'(s_t) s_t = \beta \alpha e'(s_{t+1}) s_{t+1} + \beta (1 - \alpha) v'(l_{t+1}) l_{t+1}.$$

Let  $\lambda_{t+1}$  be the Lagrange multiplier on this constraint. The Lagrangean is

$$\begin{aligned} \max_{l,h} \sum_{t=0}^2 \beta^t \{ \theta h_t l_t - v(l_t) - e(s_t) - v'(l_t) l_t X \} \\ + \sum_{t=0}^1 \beta^t \lambda_{t+1} \{ \beta \alpha e'(s_{t+1}) s_{t+1} + \beta(1-\alpha) v'(l_{t+1}) l_{t+1} - e'(s_t) s_t \} \end{aligned}$$

Let  $\vartheta_t = \frac{\theta h_t}{v'(l_t)} - 1$ . The first order conditions in  $l_0, l_1, l_2$  are

$$\vartheta_0 = (1 + \nu^{-1})X \quad (14)$$

$$\vartheta_1 = (1 + \nu^{-1})[X - (1 - \alpha)\lambda_1] \quad (15)$$

$$\vartheta_2 = (1 + \nu^{-1})[X - (1 - \alpha)\lambda_2] \quad (16)$$

The first order conditions in  $h_1, h_2$ , together with the Euler equations imply the following set of conditions:

$$\frac{1 + \nu^{-1}}{1 + \varepsilon^{-1}} [X - (1 - \alpha)\lambda_2] \beta(1 - \alpha) v'(l_2) l_2 = (\lambda_2 - \alpha \lambda_1) e'(s_1) s_1 \quad (17)$$

$$\frac{1 + \nu^{-1}}{1 + \varepsilon^{-1}} [X - (1 - \alpha)\lambda_1] \beta(1 - \alpha) v'(l_1) l_1 = \lambda_1 e'(s_0) s_0 - (\lambda_2 - \alpha \lambda_1) \beta \alpha e'(s_1) s_1 \quad (18)$$

where  $\nu^{-1} = \frac{v''(l_t)}{v'(l_t)} l_t$  and  $\varepsilon^{-1} = \frac{e''(s_t)}{e'(s_t)} s_t$ , both assumed constant.

To show the first part of the Proposition, assume by contradiction that  $\lambda_2 < \lambda_1$ .

The equation (17), together with the Euler equation implies that

$$\frac{1 + \nu^{-1}}{1 + \varepsilon^{-1}} [X - (1 - \alpha)\lambda_2] = \lambda_2 - \alpha \lambda_1. \quad (19)$$

Using equality (19) one can show that if  $\lambda_2 < \lambda_1$  then it must be true that

$$(1 - \alpha)\lambda_2 > \frac{1 + \nu^{-1}}{2 + \nu^{-1} + \varepsilon^{-1}}X \quad (20)$$

$$(1 - \alpha)\lambda_1 > \frac{1 + \nu^{-1}}{2 + \nu^{-1} + \varepsilon^{-1}}X. \quad (21)$$

Equation (18) then implies that

$$\begin{aligned} \lambda_1 e'(s_0)s_0 &= \frac{1 + \nu^{-1}}{1 + \varepsilon^{-1}}[X - (1 - \alpha)\lambda_1]\beta(1 - \alpha)v'(l_1)l_1 + \frac{1 + \nu^{-1}}{1 + \varepsilon^{-1}}[X - (1 - \alpha)\lambda_2]\beta(\alpha)e'(s_1)s_1 \\ &< \frac{1 + \nu^{-1}}{2 + \nu^{-1} + \varepsilon^{-1}}X\beta(1 - \alpha)v'(l_1)l_1 + \frac{1 + \nu^{-1}}{2 + \nu^{-1} + \varepsilon^{-1}}X\beta(\alpha)e'(s_1)s_1 \\ &= \frac{1 + \nu^{-1}}{2 + \nu^{-1} + \varepsilon^{-1}}Xe'(s_0)s_0 \end{aligned}$$

where the first equality uses (19), the inequality uses (20) and (21), and the last equality uses the Euler equation. Cancelling  $e'(s_0)s_0$  one gets

$$\lambda_1 < \frac{1 + \nu^{-1}}{2 + \nu^{-1} + \varepsilon^{-1}}X,$$

a contradiction with (21). Therefore  $\lambda_2 \geq \lambda_1$ . One can easily rule out  $\lambda_2 = \lambda_1$  as long as  $\alpha > 0$ , and so  $\lambda_2 > \lambda_1$ . ■

**Proof of Proposition 7.** The social planner solves (6) for  $T = 1$ . The first order conditions in  $l_0$ ,  $l_1$  and  $h_1$  can be written as

$$\begin{aligned} \frac{\theta h_0}{V'_0} - 1 &= \left(1 + \frac{V''_0}{V'_0}l_0\right)X + \frac{\lambda}{G_{s,0}}\frac{V''_0}{V'_0} \\ \frac{\theta h_1}{V'_1} - 1 &= (1 + \nu^{-1})\left(X - \frac{\lambda}{h_1}\right) \\ \frac{\theta h_1}{V'_1} - 1 &= \frac{V''_0}{V'_0}l_0X + \frac{\lambda}{G_{s,0}}\left(\frac{V''_0}{V'_0} + \frac{1}{h_1} - \frac{G_{ss,0}}{G_{s,0}}\right) \end{aligned}$$

The Proposition is proven if  $\lambda > 0$ . This follows from the last two first order condi-

tions. ■

**Proof of Lemma 3.** It is immediate that

$$\varphi(\{y_t^*(\theta)\}_{t=0}^T, \theta) = \sum_{t=0}^T \beta^t [y_t^*(\theta) - c_t^*(\theta)]. \quad (22)$$

The Lemma will thus be proven if

$$\varphi(\{y_t^*(\theta)\}_{t=0}^T, \theta) \geq \varphi(\{y_t^*(\theta)\}_{t=0}^T, \hat{\theta})$$

for any  $\hat{\theta} \in \Theta$ . Suppose by contradiction that there is some  $\hat{\theta} \in \Theta$  such that  $\varphi(\{y_t^*(\theta)\}_{t=0}^T, \theta) < \varphi(\{y_t^*(\theta)\}_{t=0}^T, \hat{\theta})$ . Then

$$\begin{aligned} W_{\sigma^*}(h_0, \hat{\theta}; \hat{\theta}) &= \max_{\{h_{t+1}\}_{t=0}^T} \sum_{t=0}^T \beta^t [y_t^*(\theta) - V(\frac{y_t^*(\theta)}{\hat{\theta}h_t}, S(h_t, h_{t+1}))] - \varphi(\{y_t^*(\theta)\}_{t=0}^T, \hat{\theta}) \\ &< \max_{\{h_{t+1}\}_{t=0}^T} \sum_{t=0}^T \beta^t [y_t^*(\theta) - V(\frac{y_t^*(\theta)}{\hat{\theta}h_t}, S(h_t, h_{t+1}))] - \varphi(\{y_t^*(\theta)\}_{t=0}^T, \theta) \\ &= \max_{\{h_{t+1}\}_{t=0}^T} \sum_{t=0}^T \beta^t [c_t^*(\theta) - V(\frac{y_t^*(\theta)}{\hat{\theta}h_t}, S(h_t, h_{t+1}))] \\ &= W_{\sigma^*}(h_0, \hat{\theta}; \theta), \end{aligned}$$

where the first equality follows from the definition of  $\varphi$ , the inequality follows from the proof hypothesis, the second equality follows from (22), and the last equality follows from the definition of  $W_{\sigma^*}$ . Thus,  $W_{\sigma^*}(h_0, \hat{\theta}; \hat{\theta}) < W_{\sigma^*}(h_0, \hat{\theta}; \theta)$ , a contradiction because  $\sigma^*$  is incentive compatible. ■

## APPENDIX 2: NUMERICAL SOLUTION

To verify that an allocation is incentive compatible, the inequality (1) needs to hold. To do so, I compute the following Bellman equation for  $t = 0 \dots \bar{t} - 1$ :

$$\Omega_t(h, \theta_i, \theta_j) = \max_{h'} \{c_t(\theta_j) - V\left(\frac{y_t(\theta_j)}{\theta_i h}, S(h, h')\right) + \beta \Omega_{t+1}(h', \theta_i, \theta_j)\}, \quad (23)$$

where  $\Omega_{\bar{t}}$  is a fixed point of the Bellman operator defined appropriately by the right hand side of (23),  $i = 1 \dots N$  denotes agent's productivity index, and  $j = 1 \dots N$  denotes agent's report index. Since  $\Omega_0 = W_\sigma$ ,<sup>20</sup> the incentive compatibility constraint is verified if

$$\Omega_0(h_0, \theta_i, \theta_i) \geq \Omega_0(h_0, \theta_i, \theta_j) \quad \forall i, j = 1 \dots N. \quad (24)$$

Computing (23) for all possible values of productivity shocks involves computation of  $N^2 = 90000$  fixed points and, in addition,  $\bar{t}N^2 = 22500000$  iterations toward period zero. I therefore only verify that (24) holds for all  $i = 1 \dots N$  and  $j = i - 5 \dots i + 5$ . In addition, I verify that (24) holds for every tenth  $i$  and all  $j = 1 \dots N$ . In all cases, I use an error tolerance equal to  $\frac{\varepsilon}{100}$ . I find that (24) is satisfied in all the computed cases.

---

<sup>20</sup>Strictly speaking,  $\Omega_0 = W_\sigma$  only if  $\bar{t}$  is infinite. For  $\bar{t}$  high enough,  $\Omega_0$  is an appropriate numerical approximation of  $W_\sigma$

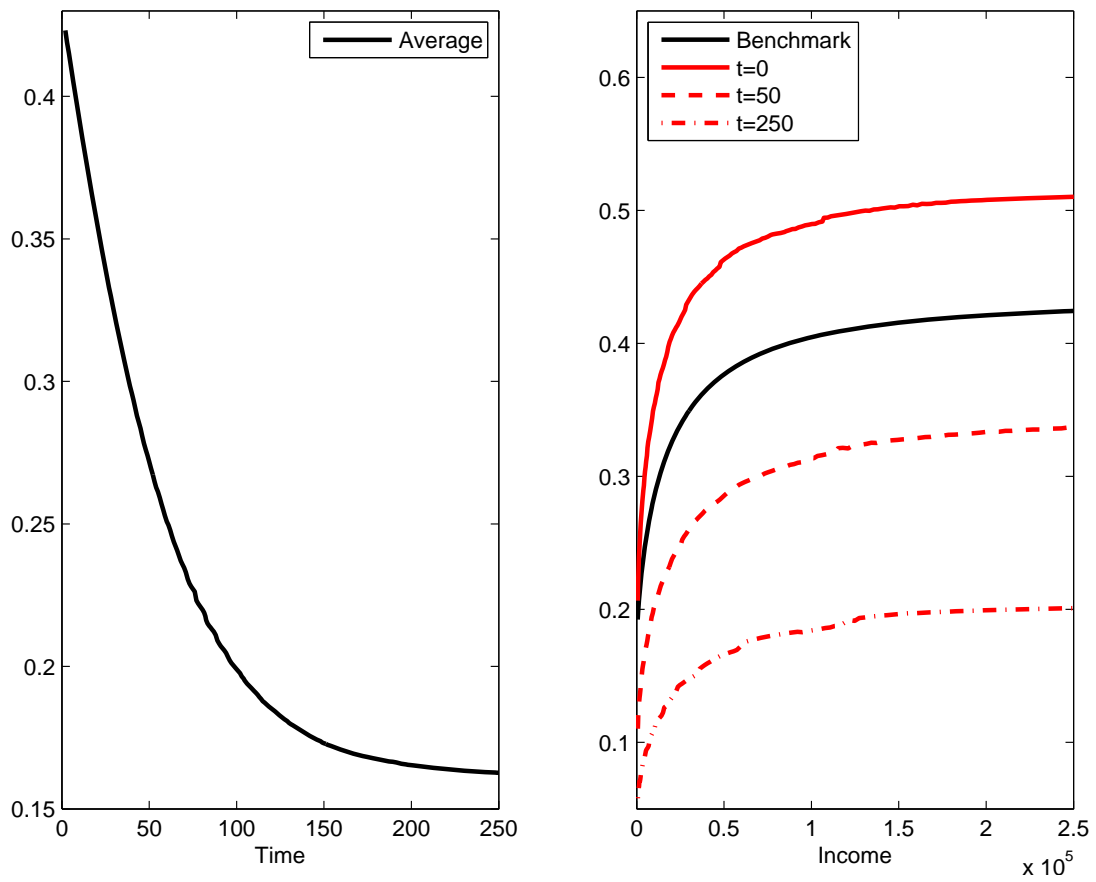


FIG. 1. Marginal Income Tax Rates, Benchmark Reform

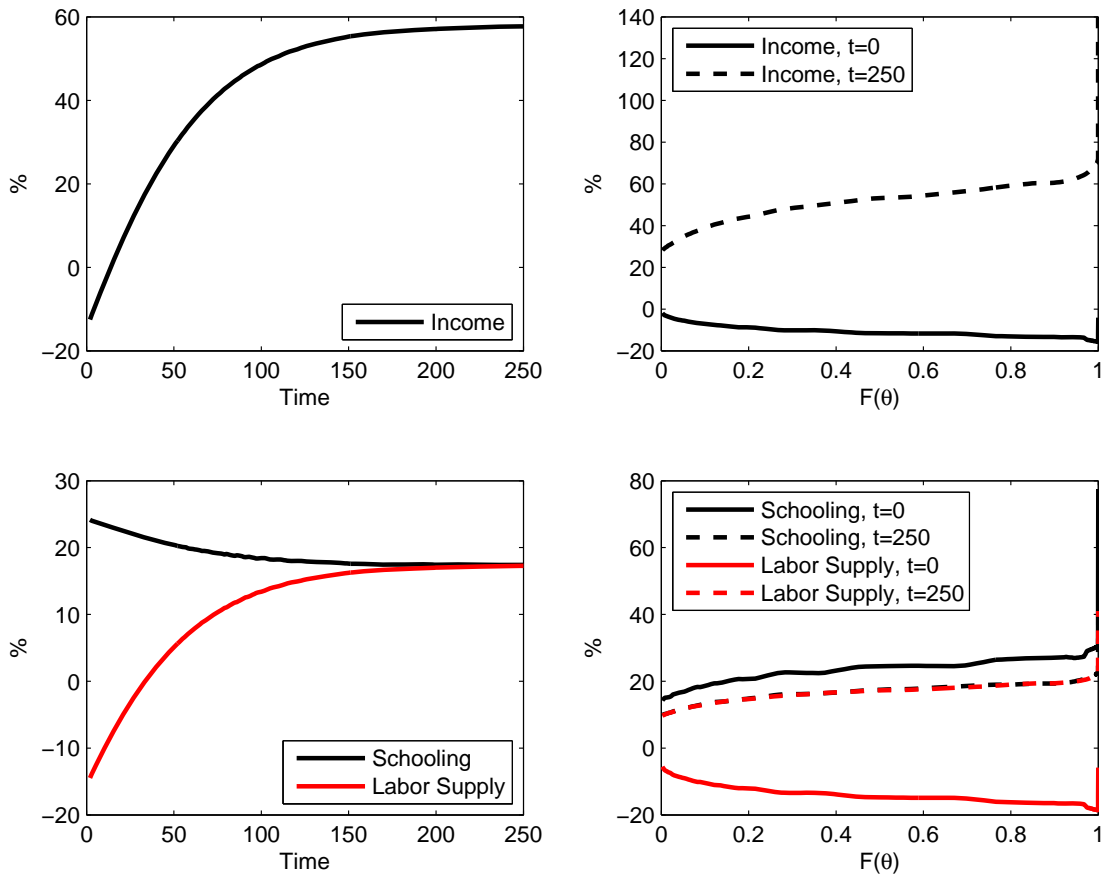


FIG. 2. Income, Schooling and Labor Supply, Change from Pre-Reform

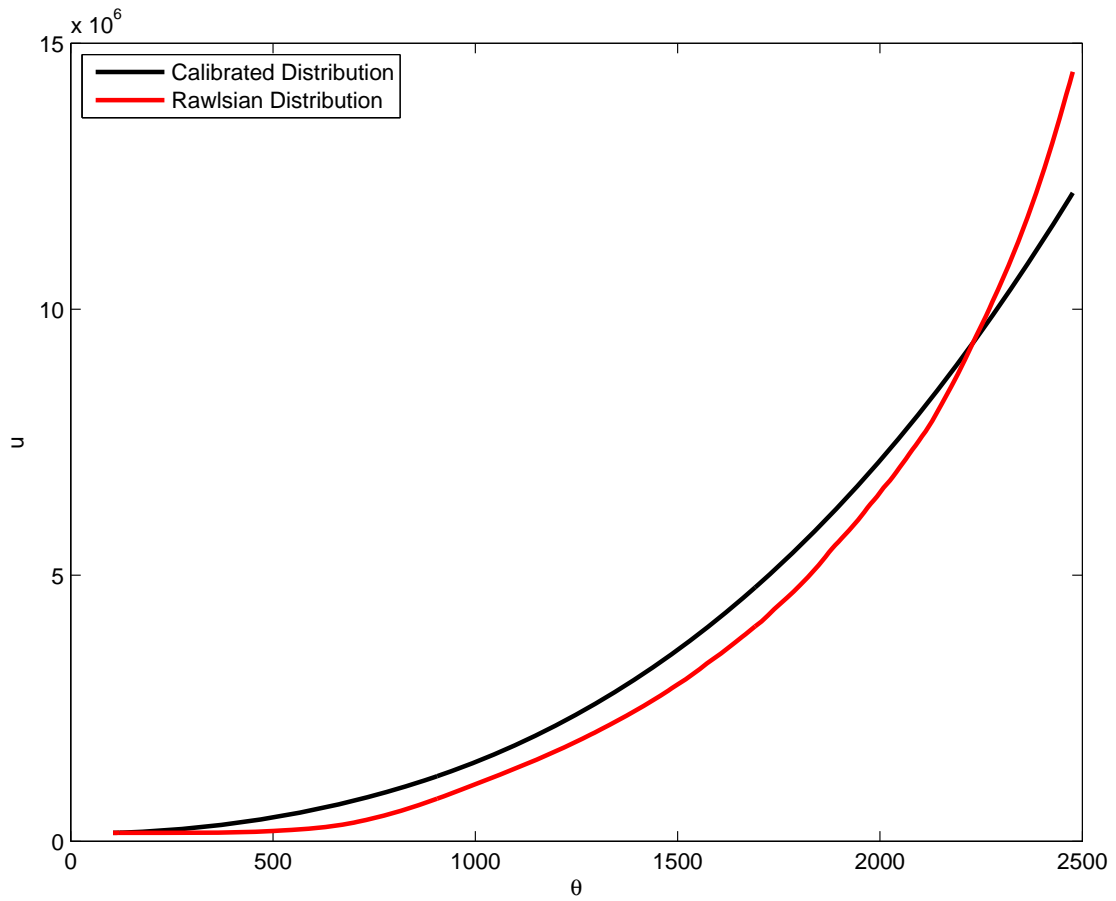


FIG. 3. Utility Distribution, Rawlsian Reform

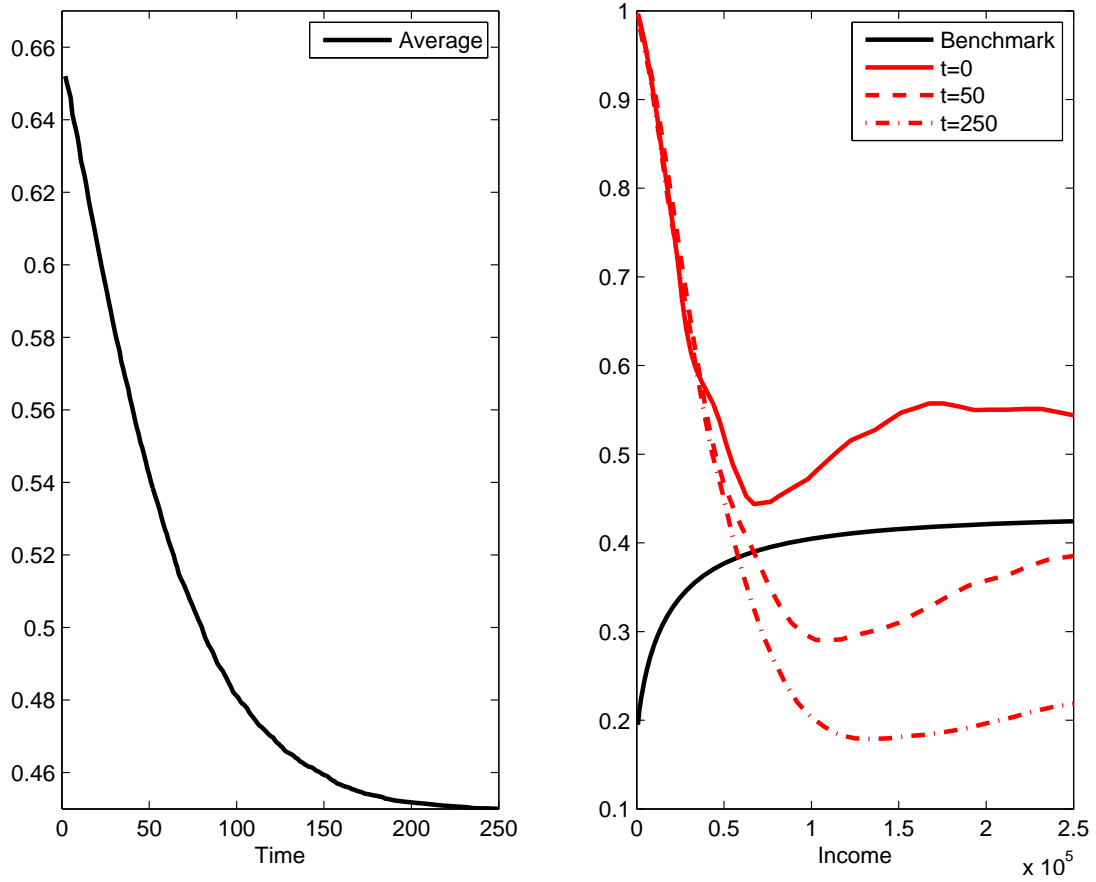


FIG. 4. Marginal Income Tax Rates, Rawlsian Reform

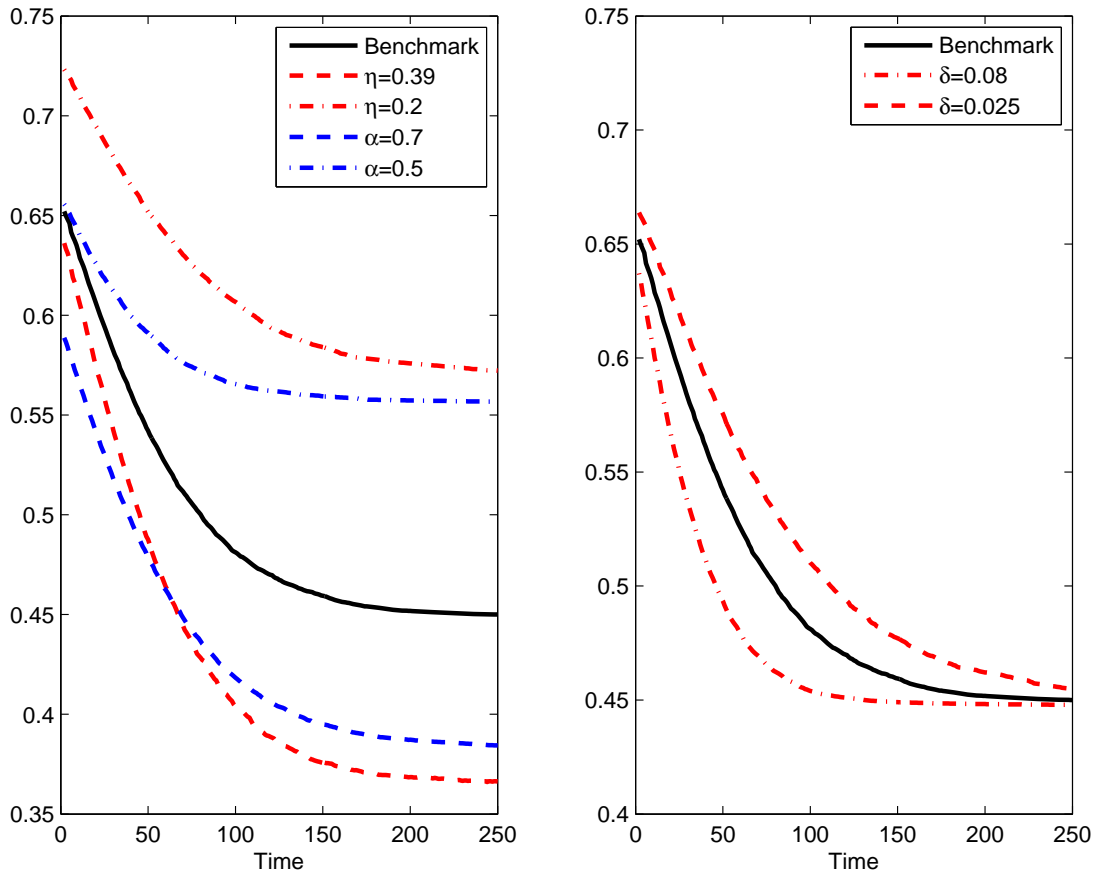


FIG. 5. Average Marginal Income Taxes, Rawlsian Reform