Optimal Income Taxation and Human Capital Accumulation

Marek Kapicka*

Department of Economics

University of California

3052 North Hall, Santa Barbara, CA 93106.

e-mail: mkapicka@econ.ucsb.edu

phone: (805) 893 2530

March 7, 2005

Abstract

This paper characterizes optimal income taxes in a dynamic economy with human capital accumulation. I extend the tools and insights developed by Mirrlees (1971) into a dynamic framework. I analyze the relationship between optimal taxes in economies with and without endogenous human capital. Two qualitative reasons why the optimal tax codes will differ are identified. I perform numerical simulations to calculate the quantitative relevance of endogenous human capital formation for optimal tax policy. I find that endogenous human capital lowers marginal tax rates by as much as 16% on average, as compared to a static model without human capital.

J.E.L Codes: E6, H2

Keywords: optimal taxation, income taxation, human capital

---

[*] I would like to thank Nancy L. Stokey for her advice and encouragement and Fernando Alvarez, Robert E. Lucas Jr., Narayana Kocherlakota and numerous seminar participants for their comments and suggestions. I acknowledge financial support from The University of Chicago. All errors are mine.

1 Introduction

This paper analyzes optimal income taxes in a dynamic heterogeneous agent economy where people can accumulate human capital over time. Income taxes therefore affect two margins. As is well known from the static models, they distort the labor-leisure trade-off. In addition to that, they discourage investment in human capital. The government's ability to redistribute resources from more able people to less able ones is thus limited even more severely than in a model without human capital.

The effect of income taxes on human capital formation has been much studied.[1] The literature typically tries to assess the effects of a given tax code but stops short of asking the related question: what does the optimal income tax look like? The contribution of this paper is to answer this normative question. I solve for optimal tax policies in a steady state, relate them to optimal tax policies in an economy without human capital, and offer several new insights into the problem of optimal taxation.

Modern analysis of the optimal income tax was initiated by the seminal contributions of Mirrlees [22],[23] and further developed, for instance, by Diamond [9] or Saez [25]. This analysis abstracts entirely from human capital and any dynamic issues. Several papers adopted the approach of Mirrlees to incorporate education into the analysis.[2] However, their models typically considered only once-and-for-all investment in education and thus remained essentially static. They are therefore not able to capture the full effects of a given tax policy. Diamond and Mirrlees [10] analyze a model with human capital and show that

---

[1] See Heckman, Lochner and Taber [16], Erosa and Koreshkova [11] or Caucutt, Imrohoroglu and Kumar [7] and others.

[2] See for instance references in Tuomala [26] or more recently Brett and Weymark [5].

marginal tax rates will be lower if human capital is endogenous. However, their model is also static and human capital thus looks more like effort. Their contribution is that they derive the results analytically and show the connection with the Le Chatelier principle.

Recent research, most notably Golosov, Kocherlakota and Tsyvinski [13], Kocherlakota [21], Werning [28], Albanesi and Sleet [1] and Battaglini and Coate[2] extended the Mirrleesian framework into a dynamic economy with physical capital. Their work differs from mine, however, because they allow the tax levied in any period to depend either on the agent's entire income history or on current income and wealth. In my analysis I will restrict the income tax to depend only on current income.

Conesa and Krueger [8] and Benabou [4] present another attack on the problem of optimal taxation in a dynamic model. They parameterize tax functions and solve for the tax policy that maximizes utility in a steady state. There are two reasons why the approach adopted in this paper is superior to their approach. First, I do not restrict the set of tax functions considered. The second difference is conceptual. Maximizing steady state utility is quite different from maximizing utility and only then looking at the steady state: the former ignores the transition path. The second criterion is one a benevolent government would adopt. (The difference is essentially the same as a difference between golden rule and modified golden rule in a typical capital accumulation problem.)

Intuition suggests that, if incentives to invest in human capital are taken into account, labor supply will be more flexible in the long run and hence less redistribution will be optimal. On the other hand, human capital can increase inequality across agents and hence favor more redistribution. The first argument is shown to prevail in the long run, assuming that the government has the ability to commit to a sequence of future tax schedules. In this case, the optimal tax codes in a static and dynamic model differ for two reasons.

First reason is related to the fact that, at the optimum, human capital introduces increasing returns in the function that relates an agent's supply of effective labor to her time input. Abler people will invest more in human capital but will also work more. Thus, if people behave optimally, their output is strictly convex in their nonleisure time (although at any single period it is linear in labor supply, as usual).[3] This simple observation implies that the output is more sensitive to any distortion created by an income tax and the government will optimally set marginal tax rates lower than it would otherwise.

The second reason stems from the dynamic nature of investment in human capital. While the costs of human capital investment are borne by the agent today, the benefits do not appear until the next period. Since agents discount future consumption, it is harder to provide proper incentives for the government. However, as we shall show, this effect disappears when the discount factor approaches one.

I perform numerical simulations to quantitatively assess the importance of these two differences. The main finding is that the intuition is confirmed. When human capital is endogenous and compensated labor supply elasticity is 0.5, the average marginal tax rates are 61.1%. In the static economy without human capital we have average marginal tax rates are 77.2%. Thus, the optimal marginal tax rates are on average 16% lower. If we consider lower labor elasticity then the difference shrinks but is still significant.

On the technical side, I prove a version of the taxation principle for our economy. I

---

[3] This per se is a well known result from literature studying gains from specialization (see Becker and Murphy [3]).

show that the problem can be conveniently reformulated as a direct revelation game where social planner assigns consumption and output allocation to the agents directly. We can then recover the tax function and it is shown that this tax function indeed solves the tax problem. The direct revelation game is necessary for both analytical and numerical solutions to the problem, but it also provides us with deeper insights into the efficiency properties of an income tax. It reveals that the solution of the direct revelation game takes a very particular form: the allocations depend only on the current report of the agent and not on any previous information. This is obviously an inefficient way of redistributing resources, in the sense that it is inferior to the problem of looking for second best allocations. In other words, an income tax is an inefficient way of redistributing resources.

The paper is organized as follows. The next section sets up the model. Section 3 focuses on the tax problem. Section 4 considers direct revelation game as an alternative equivalent formulation of the tax problem. In Section 5 I simplify the problem by using the first order approach. Section 6 focuses on the steady state, derives conditions that the solution must satisfy and specializes the results further for the Cobb-Douglas human capital production function. Section 7 looks at the qualitative properties of the solution and relates it to the solution of a static problem where human capital considerations are absent. Numerical simulations and computed optimal tax codes are presented in Section 8. Section 9 concludes. The Appendix contains most of the proofs.

2 The Model

Time is discrete, $t \geq 0$. There is a measure 1 of agents in the economy. Each individual is associated with a skill level $\theta \in [0, \bar{\theta}] = \Theta$.

I will assume this skill level does not change over time. This is certainly a very restrictive condition, but is necessary to keep the model tractable.[4] Distribution of skills is given by a distribution F. We assume that F is twice differentiable and has density $f(\theta)$. The skills are private information of each agent: only she knows her own ability. The skills affect earnings of the agent in a way specified below.

Each agent is endowed with one unit of time. At each period, time can be divided between leisure, work and time spent by human capital accumulation. Denote working time as l_t and time spent by accumulating human capital by s_t.[5] Thus $1 - l_t - s_t$ is time spent by consuming leisure at period t.

Period utility of each person is given by a function $U : \mathbf{R}_+ \times [0,1] \to \mathbf{R}$ and depends on consumption and leisure. We assume that U is continuously differentiable on $\mathbf{R}_{++} \times (0,1]$, strictly increasing and strictly concave. We also assume that U is additively separable. The discount factor is $\beta \in (0,1)$ and an individual evaluates consumption and leisure sequences

---

[4] It is worth noting that economists can in principle solve dynamic incentive problem only under two extreme assumptions: constant shocks and iid. shocks. Anything in between is significantly harder to solve (See Fernandes and Phelan [12] or Kapicka [19]). From these two extremes, constant abilities seems to be better choice.

[5] It is necessary to interpret time spent by accumulating human capital quite extensively. This does not include only time spent in schools but also other activities that increase individual's human capital, i.e. on-the-job training.

according to

$$\sum_{t\geq 0}\beta^t U(c_t, 1-l_t-s_t).$$

Each individual starts with initial human capital h_0. To reduce the complexity of the model, I assume that h_0 is identical for all people and is observed by the government. Agent's human capital at the beginning of period $t+1$ is denoted h_{t+1}. It depends on time spent accumulating it in previous period s_t, previous level of human capital h_t and is given by a human capital accumulation function $G : \mathbf{R}_+ \times [0,1] \to \mathbf{R}_+$:

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

We assume that G is continuously differentiable on $\mathbf{R}_{++} \times (0,1]$, strictly increasing and strictly concave and that $\lim\limits_{s\to 0} G_s(s,h) = +\infty$ for $h > 0$. Moreover, we assume that $G(h,0) \in (0,h)$ if $h > 0$ and that there is $\bar{h}$ such that $G(\bar{h},1) \leq \bar{h}$.

Human capital affects production abilities of the agent. I assume an efficiency unit specification: a person with human capital h_t, skills θ and working l_t hours produces $y_t = \theta h_t l_t$ at time $t \geq 0$.

I assume that the tax structure must balance the government budget each period. There is no outside borrower or lender and no storage is possible. It is supposed that the government has expenditures E in each period and that the expenditures are constant over time. The government is supposed to maximize expected discounted utility of an agent that is yet to draw his ability level from the distribution F. It should be noted that no argument in the paper relies on this assumption and it can be easily generalized.

In the next section I will specify the problem of our interest. In this environment the government selects a sequence of tax functions and the agent then decides how much time to spend schooling, working and how much to consume. This problem is called the *Tax Problem.* After that I specify another problem, called the *Direct Revelation Game* where we replace government with a social planner and specify a different mechanism: agents report their type to social planner. Social planner, instead of imposing tax functions, assigns consumption and output directly, as a function of agents' reports. It is easy to show the equivalence between these two formulations: solving one is equivalent to solving the other. The reason why the problem is reformulated as a direct revelation game is that the direct revelation game will turn out to be easier to solve. However, it will also give us deeper insight into efficiency of income tax as a tool for redistribution.

3 Tax Problem

The government observes only the output and consumption of each agent. It designs a tax structure which is, due to the information asymmetry, anonymous: it depends only on agent's income and not on her type. A tax policy is defined as follows:

Definition 1 *A tax policy is a sequence of increasing and upper semicontinuous functions* $\Gamma_t : \mathbf{R}_+ \to \mathbf{R}_+$, *all* $t \geq 0$.

If an agent produces y_t, her post-tax income is $\Gamma_t(y_t)$. The tax function therefore depends only on current output and not on the past history of outputs. The assumption that a tax policy is increasing is rather innocuous since no agent will choose to produce an income level at the decreasing part of income function. Let $\mathfrak{G}$ be the set of all policies that satisfy these assumptions.

3.1 Agent's Problem

The timing is as follows. At the beginning of period t, the agent starts with human capital h_t. She decides how to allocate her time among leisure, labor l_t and schooling s_t. After that she produces output y_t. She is taxed and consumes whatever is left after taxation. The agent takes the sequence of tax functions $\mathbf{\Gamma}$ as given. Her objective is to solve the following problem:

$$\max_{\{l_t, s_t, h_{t+1}\}} \sum_{t\geq 0} \beta^t U(c_t, 1 - l_t - s_t)$$

$$\begin{aligned} \text{s.t.} \qquad c_t &= \Gamma_t(\theta h_t l_t)) \\ h_{t+1} &= G(h_t, s_t) \\ 0 &\leq s_t, l_t \leq 1 \end{aligned}$$

and taking h_0 as given. The first constraint incorporates the assumption that there are no capital markets in the economy. One can show, however, that if we allow the government to use capital taxes then this assumption is innocuous and can be relaxed. The reason is that, in the optimum, the capital income taxes would be set in such a way to induce the agent to have zero savings. The optimal capital income taxes would be nonzero along the transition path but they would converge to 0 as the system converges to the steady state. [6]

Define the optimal policy correspondences to be $l_t^{\mathbf{\Gamma}}(\theta)$, $s_t^{\mathbf{\Gamma}}(\theta)$ and $h_{t+1}^{\mathbf{\Gamma}}(\theta)$. Notice that they depend on the whole tax sequence $\mathbf{\Gamma}$. One can show that the optimal policy correspondences

---

[6] This intuition behind nonzero capital taxes along the transition is different from the intuition in Golosov, Kocherlakota, Tsyvinski [13], Kocherlakota [21] or Albanesi and Sleet [1]. In our model, exogenous restrictions on tax instruments prevent the consumption being smoothed optimally. This creates incentives for private savings or borrowing and they need to be discouraged by capital tax.

are nonempty and upper hemicontinuous. The proof is omitted here but can be found in [18].

For any $\mathbf{\Gamma} \in \mathfrak{G}$ and any optimal policy correspondences $l_t^{\mathbf{\Gamma}}(\theta)$ and $s_t^{\mathbf{\Gamma}}(\theta)$ we can construct optimal output and consumption policy correspondences by $y_t^{\mathbf{\Gamma}}(\theta) = \theta h_t^{\mathbf{\Gamma}}(\theta) l_t^{\mathbf{\Gamma}}(\theta)$ and $c_t^{\mathbf{\Gamma}}(\theta) = \Gamma_t(y_t^{\mathbf{\Gamma}}(\theta))$ where the argument is agent's type. It follows that these correspondences are also upper hemicontinuous.

We cannot guarantee that the solution to the agent's problem is unique, however. That would require concavity of the tax function - an assumption which is hard to justify. Essential uniqueness is also hard to prove since it typically relies on monotonicity of agent's policy functions and these do not necessarily hold in the model. To see why monotonicity does not hold, consider for instance the function $h_{t+1}^{\mathbf{\Gamma}}(\theta)$. It seems natural that this function will be increasing in type, but this intuition may be easily false. Consider a case, studied in section (6.1), where we have a steady state result that $h(\theta) \sim l(\theta)$. Thus, human capital will be increasing only if labor supply is increasing with type. But this does not necessarily hold even in a static model and may not hold here either.[7] For all these reasons, we will restrict attention to an upper semicontinuous selection from the functions $y_t^{\mathbf{\Gamma}}$ and $c_t^{\mathbf{\Gamma}}$.

---

[7] To see an example with y_t^{Γ} non increasing, consider a two period model and the following tax system. There is no tax in the first period and 100% tax in the second period, except for income $\bar{y} > 0$. Then the agents with relatively low θ will decide not to work in the second period and will therefore not invest in human capital. Relatively high θ agents will decide to work $\bar{y}$ and invest in human capital. Consequently, an agent with θ slightly lower than the marginal agent will work in the first period more than an agent with θ slightly above marginal θ since she has discontinuously lower investment in human capital.

3.2 Government Problem

The government problem is to maximize his objective function taking policy functions of households as given. The government takes agents' incentives into account only implicitly, through their policy functions.

The instrument the government has is a sequence of tax functions $\mathbf{\Gamma} \in \mathfrak{G}$. In its choice, it faces only one type of constraint. Any solution must satisfy the period aggregate resource constraint. Thus, the government solves

$$\max_{\mathbf{\Gamma}\in\mathfrak{G}} \int \sum_{t\geq 0} \beta^t U[c_t^{\mathbf{\Gamma}}(\theta), 1 - l_t^{\mathbf{\Gamma}}(\theta) - s_t^{\mathbf{\Gamma}}(\theta)] f(\theta) d\theta$$

$$\text{s.t.} \qquad E + \int c_t^{\mathbf{\Gamma}}(\theta) f(\theta) d\theta \leq w \int y_t^{\mathbf{\Gamma}}(\theta) f(\theta) d\theta, \quad t \geq 0$$

where w is a constant. It can be interpreted as a real wage per unit of labor. Denote the solution by $\mathbf{\Gamma}^*$.

4 Direct Revelation Game

I now show that we can reformulate the problem as a direct revelation game: agents report their type to a social planner and the social planner assigns consumption and output based on the report.

Definition 2 *Social planner's allocation A is given by upper semicontinuous functions $c_t(.) : \Theta \to R$ and $y_t(.) : \Theta \to R$ for all $t \geq 0$.*

We allow the social planner to offer discontinuous allocations, but we require these functions to be upper semicontinuous. Let $\mathfrak{A}$ be the set of all allocations satisfying this assumption.

The agent reports her type repeatedly in the beginning of every period. Thus, the report is defined as follows:

Definition 3 *A report is a sequence* $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_t\}_{t=0}^{\infty} \in \Theta^{\infty}$.

Truthful report for θ- type agent is a constant sequence θ^{∞}. Several remarks regarding this definition are worth mentioning. First, although the domain is the same as in previously defined functions $c_t^{\mathbf{\Gamma}}$ and $y_t^{\mathbf{\Gamma}}$ the meaning of the argument is very different. In social planner's allocations c_t and y_t the argument is agent's report of his type, which might be truthful or not. In case of $c_t^{\mathbf{\Gamma}}$ and $y_t^{\mathbf{\Gamma}}$ the argument is agent's true type - since these are policy functions of the agent. Second, it is meaningful to define social planner's allocations only for those variables that are observable and can be therefore enforced. This is not the case of time allocations and therefore they are not included in the definition. Third, notice how the fact that income tax depends only on current output is translated into the definition of social planner's allocations: they depend only on current report and not on the whole history of reports.[8] The social planner is thus throwing away a lot of information he might in general use to redistribute resources.

4.1 Agent's Problem

The timing is as follows: At the beginning of period t the agent starts with human capital h_t. She reports her type $\hat{\theta}_t$ and is assigned output $y_t(\hat{\theta}_t)$ and consumption $c_t(\hat{\theta}_t)$. She works to produce assigned output and decides about schooling level s_t.

The agent can be thought of as solving a two stage problem. In the first step, we fix the

---

[8] This step will be justified later.

report $\hat{\boldsymbol{\theta}}$. For any given report $\hat{\boldsymbol{\theta}}$ the agent chooses time allocation that maximizes her utility and is consistent with her report (in a sense that it delivers assigned output allocations). The agent thus solves

$$\begin{aligned} V(\theta; \hat{\boldsymbol{\theta}}) &= \max_{\{s_t, h_{t+1}\}} \sum_{t \geq 0} \beta^t U(c_t(\hat{\theta}_t), 1 - \frac{y_t(\hat{\theta}_t)}{\theta h_t} - s_t) \\ \text{s.t.} \quad h_{t+1} &= G(h_t, s_t) \\ 0 &\leq s_t \leq 1 - \frac{y_t(\hat{\theta}_t)}{\theta h_t} \quad t \geq 0 \end{aligned}$$

and taking h_0 as given. If there is no sequence $\{s_t, h_{t+1}\}_{t \geq 0}$ that satisfies the constraints[9], set $V(\theta; \hat{\boldsymbol{\theta}}) = -\infty$. For all other cases define the optimal policy functions that solve the problem as $\tilde{s}_t(\theta; \hat{\boldsymbol{\theta}})$ and $\tilde{h}_{t+1}(\theta; \hat{\boldsymbol{\theta}})$. The functions depend on agent's true type and the reports. We can show that the optimal policy functions are correctly defined and that they are upper semicontinuous in all their arguments. The proof is fairly straightforward and can be found in [18].

In the second stage, the agent chooses her report $\hat{\boldsymbol{\theta}}$ to maximize utility:

$$\max_{\hat{\boldsymbol{\theta}} \in \Theta^\infty} V(\theta; \hat{\boldsymbol{\theta}})$$

In what follows we make use of the revelation principle. It is without loss of generality to restrict attention to allocations that are incentive compatible, i.e. to allocations where agents prefer to report their type truthfully:

Definition 4 (Incentive compatibility) *Social planner's allocation $A \in \mathfrak{A}$ is incentive compatible if for almost all $\theta \in \Theta$*

$$V(\theta; \theta^\infty) \geq V(\theta; \hat{\boldsymbol{\theta}}) \quad \forall \hat{\boldsymbol{\theta}} \in \Theta^\infty \tag{1}$$

---

[9] Suppose an agent with low θ reported very high $\hat{\theta}_t$ and is assigned very high output $y_t(\hat{\theta}_t)$. There could be no way he can produce such high output.

The set of all incentive compatible allocations is denoted by $\mathfrak{A}^{IC}$. It is straightforward to see that $\mathfrak{A}^{IC} \subseteq \mathfrak{A}$. It is also easy to see that an allocation is incentive compatible only when both $y_t(\theta)$ and $c_t(\theta)$ are increasing (or decreasing).

It remains to show the equivalence between the tax problem and the direct revelation game - a version of the taxation principle. This is an important step since it justifies the whole approach employed in subsequent sections. The result, which is proved in [18], is the following:

Proposition 1 (Taxation Principle) *Take any* $\mathbf{\Gamma} \in \mathfrak{G}$. *Consider social planner's allocation* A *defined by* $c_t(\theta) = c_t^{\Gamma}(\theta)$ *and* $y_t(\theta) = y_t^{\Gamma}(\theta)$ *for all* $t \geq 0$ *for almost all* θ. *Then* $A \in \mathfrak{A}^{IC}$.

Conversely, take any $A \in \mathfrak{A}^{IC}$. *Then there exists a tax regime* $\mathbf{\Gamma}$ *such that for almost all* θ, $c_t^{\mathbf{\Gamma}}(\theta) = c_t(\theta)$, $y_t^{\mathbf{\Gamma}}(\theta) = y_t(\theta)$ *for all* $t \geq 0$ *and* $\mathbf{\Gamma} \in \mathfrak{G}$.

4.2 Social Planner's Problem

Social Planner is constrained not only by the resource constraint, but also by the incentive compatibility constraint for each agent. He chooses $A \in \mathfrak{A}$ to solve

$$\max_{A \in \mathfrak{A}} \int_{\Theta} \sum_{t \geq 0} \beta^t U[c_t(\theta), 1 - \frac{y_t(\theta)}{\theta h_t(\theta)} - s_t(\theta)] f(\theta) d\theta$$

$$\begin{aligned} s.t. \qquad E + \int_{\Theta} c_t(\theta) f(\theta) d\theta &\leq w \int_{\Theta} y_t(\theta) f(\theta) d\theta \\ V(\theta; \theta^{\infty}) &\geq V(\theta; \hat{\boldsymbol{\theta}}) \quad \forall \theta \in \Theta, \hat{\boldsymbol{\theta}} \in \Theta^{\infty} \end{aligned}$$

where we have simplified the notation by writing $s_t(\theta) = \tilde{s}_t(\theta; \theta^{\infty})$ and $h_t(\theta) = \tilde{h}_t(\theta; \theta^{\infty})$. Denote the solution as A^*. The Taxation Principle1 implies that $c_t^{\mathbf{\Gamma}^*}(\theta) = c_t^*(\theta)$ and $y_t^{\mathbf{\Gamma}^*}(\theta) =$

$y_t^*(\theta)$ for all $t \geq 0$. This means, the solution to social planner's problem coincides with the solution to the tax problem.

5 First Order Approach

The usual approach in the optimal taxation literature is to replace the incentive compatibility constraint with an envelope condition describing how individual utility varies across types, if the agents choose optimal reports. We will proceed along these lines. Before doing so, however, we will show that the incentive compatibility condition 1 can be equivalently written as a series of time dependent incentive compatibility constraints. Define a lifetime utility of a $\theta-$type agent when she reports some feasible sequence $\hat{\boldsymbol{\theta}}$ and starts with human capital h_t from period t on to be

$$\begin{aligned}
\tilde{V}_t(h_t, \theta; \hat{\boldsymbol{\theta}}) &= \max_{\{s_j, h_{j+1}\}_{j \geq t}} \sum_{j=t}^{\infty} \beta^{j-t} U(c_j(\hat{\theta}_j), 1 - s_j - \frac{y_j(\hat{\theta}_j)}{\theta h_j}) \\
\text{s.t.} \quad h_{j+1} &= G(h_j, s_j) && (2) \\
0 &\leq s_j \leq 1 - \frac{y_j(\hat{\theta}_j)}{\theta h_j} \qquad j \geq t && (3)
\end{aligned}$$

and taking h_t as given. The value function of a $\theta-$type agent who reports optimally from period t on and her human capital at period t happens to be the optimal one is defined as $V_t(\theta) = \tilde{V}_t(h_t(\theta), \theta; \theta^{\infty})$. It is easy to show that an allocation is incentive compatible if and only if $V_t(\theta) \geq \tilde{V}_t(h_t(\theta), \theta; \hat{\boldsymbol{\theta}})$ for all $\hat{\boldsymbol{\theta}} \in \Theta^{\infty}$ and for all $t \geq 0$.[10] Necessary conditions for incentive compatibility are spelled out in the next Proposition. Note that we impose an assumption that $h_t(\theta)$ is differentiable almost everywhere.

---

[10] Note that we keep the initial period human capital at the optimal level $h_t(\theta)$.

Proposition 2 *Suppose* $h_t(\theta)$ *is differentiable a.e. Then the value function* $V_t(\theta)$ *satisfies*

$$V_t(\theta) = V_t(0) + \int_0^\theta \sum_{j=t}^{\infty} \beta^{j-t} U_{l_j} l_j \frac{d\varepsilon}{\varepsilon} + \int_0^\theta U_{l_t}(\frac{l_t}{h_t} + \frac{G_{h_t}}{G_{s_t}})\frac{dh_t}{d\theta} d\varepsilon. \tag{4}$$

Proof. See the Appendix. ■

The first order condition with respect to time allocations s_t and h_{t+1} is derived from the stage 1 problem and so is taken for a fixed reporting strategy. It is therefore independent of the differentiability properties of the social planner's allocation. After some algebra we get

$$\frac{U_{l_t}}{G_{s_t}} \geq \beta[U_{l_{t+1}} \frac{G_{h_{t+1}}}{G_{s_{t+1}}} + U_{l_{t+1}} \frac{l_{t+1}}{h_{t+1}}] \tag{5}$$

and the equation holds with equality if $s_t > 0$. Left hand side represent current costs of increasing next period human capital by one unit. The right hand side shows that there are two types of benefits from such human capital investment. First term reflects the fact that an individual can reduce his schooling time by $\frac{G_{h_{t+1}}}{G_{s_{t+1}}}$ units of time next period. Second term is the reduction of time required to produce output y_{t+1}. It is easy to see that this condition is necessary and sufficient for the optimum in the first stage optimization problem.

Unfortunately, very little can be said about the dynamics of optimal tax policies along the transition path. In each period there will be a fraction of people who decide not to work. Due to our assumptions on human capital production function, schooling $s_t(\theta)$ will be strictly positive as long as there is some $\tau \geq t+1$ such that $y_\tau(\theta) > 0$. Human capital, schooling, output and consumption will be continuous functions for all $t \geq 0$. This, together with the first order conditions for the social planner's problem are presented as an intermediate step in the proof of Proposition 3 in the Appendix.

6 Steady State Analysis

Steady state assumption implies that allocations and the optimal policy functions are constant over time: $y_t(\theta) = y(\theta)$, $c_t(\theta) = c(\theta)$, $l_t(\theta) = l(\theta)$ and $h_{t+1}(\theta) = h(\theta)$. A constant fraction of people will not work in steady state and their human capital will depreciate to 0. Let $\theta_0 = \inf[\theta : h(\theta) > 0]$ represent the threshold value of θ.

An additional complication is presented by the fact that $h(\theta)$ may be discontinuous at θ_0, although h_t is continuous for all $t \geq 0$. Consequently, $u(\theta)$ will also be discontinuous. To see this note that, as long as $c(\theta)$ is strictly increasing,

$$\lim_{\theta \to \theta_{0-}} V(\theta) = \frac{U(c(0), 1)}{1 - \beta} = \max_{\hat{\boldsymbol{\theta}}} \tilde{V}(0, \theta_0; \hat{\boldsymbol{\theta}}) < \max_{\hat{\boldsymbol{\theta}}} \tilde{V}(h(\theta_0), \theta_0; \hat{\boldsymbol{\theta}}) = V(\theta_0)$$

because $h(\theta_0) > 0$. The intuition is that if θ_0-type agent decides to stop schooling, he can still work over transitional period and secure himself strictly higher utility than an agent with no human capital. We will reflect the discontinuity by writing the steady state incentive compatibility constraint as

$$u(\theta) = u(\theta_0) + \int_{\theta_0}^{\theta} U_l l \frac{d\varepsilon}{\varepsilon} + (1 - \beta) \int_{\theta_0}^{\theta} U_l \left(\frac{l}{H(s)} + \frac{G_h}{G_s}\right) \frac{dh}{d\varepsilon} d\varepsilon \tag{6}$$

where we have defined $u(\theta) = V_t(\theta)(1 - \beta)$ to be a period utility function. Steady state Euler equation becomes

$$l = \frac{1 - \beta G_h}{\beta G_s} h. \tag{7}$$

Next theorem states the main result of this section. It finds necessary conditions that the solution to the social planner's problem must satisfy in the steady state. The algebra that leads to this result is stated in the proof of this proposition.

Proposition 3 *Suppose that $y(\theta), c(\theta)$ are allocations that solve social planner's problem in a steady state. Then $y(\theta), c(\theta), s(\theta), l(\theta)$ and $u(\theta)$ satisfy, in addition to the envelope condition 6 and the Euler equation 7, the following set of equations:*

$$\begin{aligned}
\int\limits_{\theta}^{\infty} f(\frac{1}{U_c} - \lambda)d\varepsilon &= \theta^2 f \frac{wh - \frac{U_l}{\theta U_c}}{U_l - U_{ll}l} + \rho\lambda\theta \frac{h\frac{1-G_h}{G_s}U_{ll} + \frac{\gamma_h}{\gamma_l}(U_l - U_{ll}l)}{U_l - U_{ll}l} \\
\int\limits_{\theta}^{\infty} f(\frac{1}{U_c} - \lambda)d\varepsilon &= \theta^2 f \frac{wh - \frac{U_l}{\theta U_c}}{-U_{ll}l} - \rho\lambda\theta \frac{\Psi + \frac{\gamma_h}{\gamma_l}U_{ll}l\frac{1-\beta G_h}{G_s}}{-U_{ll}l\frac{1-\beta G_h}{G_s}} \\
\int\limits_{\Theta} (\frac{1}{U_c} - \lambda)f(\theta)d\theta &= 0
\end{aligned}$$

where λ is the inverse of Lagrange multiplier on resource constraint, ρ is (appropriately normalized) Lagrange multiplier on the Euler equation, $\Psi = h(1 - G_h)(\frac{\partial\eta}{\partial h'} + \beta\frac{\partial\eta}{\partial h}) + \eta[1 - \beta G_h - \beta h(G_{hh} + G_{hs}\frac{1-\beta G_h}{\beta G_s})]$ and $\eta = \frac{U_l}{G_s}$.

Proof. See the Appendix. ■

Generality of these equations prevents us from giving more precise statements about the nature of the solution. In the next subsection I specify the functional form of the production function to be Cobb-Douglas and show that our steady state analysis will be considerably simplified.

6.1 Cobb-Douglas Production Function

Assume that the law of motion for human capital has the following functional form:

$$G(h, s) = (1 - \delta)h + h^{\alpha_1}s^{\alpha_2} \quad 0 < \alpha_1, \alpha_2 < 1. \tag{8}$$

Depreciation of human capital is given by δ. Parameters α_1 and α_2 measure efficiency of each input in production. We can easily verify that this production function satisfies all the conditions required for production functions.

Put $q = (\frac{1}{\delta})^{\frac{1}{1-\alpha_1}}$ and $\alpha = \frac{\alpha_2}{1-\alpha_1}$. Also, let $p = \frac{1-\beta(1-\delta(1-\alpha_1))}{\alpha_2\beta\delta}$. It is easy to show that $h(\theta) = qs(\theta)^\alpha$ and that the steady state Euler equation reduces to

$$l(\theta) = ps(\theta) \tag{9}$$

i.e. human capital and working time are related in constant proportions. What is important is that this proportion is constant for all agents, regardless of their abilities. In a steady state we can therefore work with only one variable summarizing all time spent by not consuming leisure. I will denote it by n. Thus, $n = (1+p)s$.[11]

7 Properties of Solution

I now explore general features of the solution. The idea of this section is to relate the optimal steady state allocations to the optimal allocations in a static economy without human capital. A static economy is an economy with output depending only on labor: $y = \varphi(l)$ for some function φ. A standard specification is $\varphi(l) = l$. This economy has been analyzed many times before and will be called a benchmark static economy.

What difference does the introduction of human capital into the environment make from the government's point of view? We can answer this in the following way. First, we try to find a static economy without human capital in which the social planner chooses the same optimal allocations of consumption, output and leisure as in the dynamic economy in a steady state? Suppose we find such static economy. Then we can look at its features and identify where exactly the human capital makes a difference, as compared to the benchmark static economy.

---

[11] This relationship obviously holds only in the steady state.

Before stating the main result of this paper (Proposition 4), I will discuss the intuition why we should expect that the optimal allocations and policies in the dynamic economy are different than in the benchmark static economy. There are two unique features of the dynamic model with human capital that cannot be found in the static model:

First, the presence of human capital introduces "increasing returns" into the output production function. The source of increasing returns is specialization. The agent can be viewed as allocating her non-labor time between two tasks: producing output directly through labor supply and indirectly through human capital. By equation (9) these two time allocations are complements. In fact, by virtue of the Cobb-Douglas human capital production function, they are related in fixed proportions. Thus, the output will be increasing more than linearly with respect to agent's non-leisure time:

$$y \sim hl \sim s^{\alpha} l \sim s^{1+\alpha} \sim n^{1+\alpha}$$

In words, marginal product of agent's output will appear to be increasing in her non-leisure time. This stands in contrast to the benchmark static model, where marginal product of output is constant for each agent: $y \sim n$.

Second, one of the properties of the model that distinguishes it from the static model is the presence of time dimension. Human capital investment bears utility costs today, but the benefits do not appear until next period and agents discount future consumption. On the other hand, in the static economy the returns are immediate. What matters the most is the discount factor β. I will show that as agents become more and more patient, this difference becomes arbitrarily small.

To show these results formally, recall that we have assumed additive separability of the utility function. Thus, we can write the utility function as

$$U(c, 1-n) = U^1(c) + U^2(1-n)$$

The following theorem finds a static economy that replicates the allocations in the dynamic one and makes the above intuitive conclusions more precise. In particular, it shows that they translate to the following properties of the static economy: increasing returns in production and stronger preferences for leisure.

Proposition 4 *The steady state consumption, output and leisure allocations and steady state optimal tax policy in the dynamic economy with human capital are identical to the consumption, output and leisure allocations and optimal tax policy in a static economy with*

i) $y = \theta l^{1+\alpha}$ *(increasing returns in labor supply)*

ii) $\hat{U}(c,n) = U^1(c) + \frac{(1+\alpha)p}{1+p} U^2(1-n)$ *(stronger preferences for leisure)*

iii)$\hat{w} = wq \frac{p}{(1+p)^{1+\alpha}}$

and the social planner assigns lower Pareto weights for all non-working people.

Proof. See the Appendix. ■

Note that the theorem does not imply that working times will be identical in both economies. This will certainly not hold, since in the dynamic economy a fraction of time is used for schooling. The static economy that possesses the properties stated in the Proposition (4) will be called the static economy corresponding to the steady state dynamic economy. When we want to look at steady state of some dynamic economy, it is enough to look at the corresponding static economy and this is what we will do in the numerical simulations.[12]

---

[12] Assuming that the correct Pareto weights are found. Note that if everybody works at the equilibrium then this is not a problem.

The third feature of the corresponding static economy is merely a normalization and so is quite uninteresting from the economic point of view. The fact that the social planner has different Pareto weights is closely connected with the fact that the marginal agent in the dynamic economy still has some human capital to enjoy along possible transition while this extra utility is nonexistent in a static economy. The first two features - increasing returns and stronger preferences for leisure - are the most interesting conclusions of the Proposition (4). They are reflections of the two unique features of the dynamic model discussed above. How should we expect that these features will be translated into the optimal tax schedule, as compared to our benchmark static economy?

i) Increasing returns in labor supply. The intuition suggests that marginal tax rates should be lower because it is more costly for the social planner to distort high skill agents. To see this, suppose that income tax distorts agent's decision and, as a result, she supplies $x\%$ less labor. With linear production function, output is also lower by $x\%$. However, for increasing returns production function the output lowers by $1-(1-x)^{(1+\alpha)} \sim (1+\alpha)x\%$ which is larger.

ii) Stronger preferences for leisure. Recall that $p = \frac{1-\beta(1-\delta(1-\alpha_1))}{\alpha_2\beta\delta} > 1$. Proposition (4) therefore shows that the dynamic economy looks in the steady state like the benchmark static economy but for a *different* utility function that puts more weight on utility of leisure. Agents thus look as if they had stronger preference for leisure than they have.

It is easy to show that the size of this effect depends crucially on the discount factor. If agents are patient enough (as β goes to 1), the magnitude of this effect is arbitrarily small ($\frac{(1+\alpha)p}{1+p}$ goes to 1).

8 Numerical Example

In this section I will numerically compute the optimal allocations in a steady state. The purpose of this experiment is to compare the steady state allocations with the allocations in the benchmark static economy, which are also computed. A difference between the optimal policy functions will tell us how far is the benchmark static model from the optimum and what is the quantitative significance of human capital accumulation. We will also compare the optimal tax code with the U.S. tax code and show how a potential tax reform should look like.

The methodology used in this section is similar to the one used by Saez [25]. The utility function used exhibits constant elasticity of labor supply,

$$U(c,n) = \log(c) - \log(1 + \frac{n^{1+k}}{1+k})$$

where the compensated elasticity of labor supply is given by $\frac{1}{k}$.

I will calibrate the distribution of skills in such a way that the resulting distribution of income, assuming we are in a steady state, will resemble the empirical distribution of income. Tax return data for 1992 are used for the empirical distribution of income. Since the data are not very smooth and this would present certain computational difficulties, I use a double Pareto-Lognormal distribution to approximate the empirical distribution of incomes. This distribution combines lognormal distribution with heavy paretian tails and replicates the empirical distribution reasonably well.[13] Figure 1 documents this by plotting both empirical

---

[13] See Jorgensen and Reed [17] for the definition of the distribution.

distribution of incomes and the approximation.

[INSERT FIGURE 1 HERE]

To calibrate the distribution of skills I first construct the U.S. income tax code. I use the NBER TAXSIM program to construct the effective federal marginal income tax schedule for calendar year 1992. The agent in the model is supposed to represent a household. Thus I restrict attention to married couples with two children. I adjust the income tax schedule by adding state income tax and sales tax as a linear tax.[14] What needs to be determined are government transfers, which are supposed to be independent of income and thus correspond to the negative of income tax at income equal to 0. Their value will be determined endogenously to balance the government budget. It is supposed that consumption to income ratio is 0.75. Thus, government expenditures are equal to 25% of total income. The calibration is based on the Proposition (4), so that we work with the corresponding static economy instead of the dynamic one.

For the utility function, the compensated elasticity is chosen to be 0.5 so that $k = 2$. I will later experiment with alternative value of $k = 4$. Parameterizing the human capital production function is more difficult. There is a very diverse evidence regarding both depreciation δ and parameters α_1 and α_2. The evidence for δ ranges from 0.0016 to 0.089, with most of the estimates concentrated around 0.04.[15] This is also the value I have chosen for parameter δ. Estimation in Browning, Hansen and Heckman [6] and Heckman, Lochner and Taber [15] indicates that human capital production function exhibits significant increasing

---

[14] Tax rates for the state income tax and sales tax were obtained by dividing government receipts from these taxes by labor income and consumption respectively. Their values were 2.78% and 7.06% .

[15] See the evidence in Browning, Hansen and Heckman [6] or Trostel [27].

returns, with α_1 being between 0.83 and 0.87 and α_2 around 0.94. However, this delivers α in the range of $5.6 - 7.3$ which seems to be an extremely high value. I have therefore chosen more conservative parameterization, with human capital production function that has constant returns to scale. Given that, α_1 and α_2 are both chosen to be $\frac{1}{2}$. This delivers $\alpha = 1$. I will later provide results for alternative values of α. The time period is one year and so the discount factor β was set equal to 0.96.

The results and comparisons with the U.S. economy are presented in figures 2 and 3. Figure 2 plots the optimal marginal tax rates against total output and compares them with the U.S. tax code. Note that, with the exception of the lowest income values, the shape and levels of the optimal tax code are relatively close to the U.S. values.

[INSERT FIGURE 2 HERE]

The optimal income tax code brings about significant changes in the optimal allocations. The average earnings stay almost constant as most of the changes in production cancel with each other. There is a significant increase in earnings inequality, Gini coefficient rises from 0.55 to 0.68. On the other hand, consumption Gini is almost unchanged at 0.44.

[INSERT FIGURE 3 HERE]

8.1 Comparison with the Benchmark Static Model

To compare the results with the benchmark static model, I have parameterized the static model along the same lines as the dynamic model. The distribution of skills was again chosen in such a way that empirical distribution of incomes, approximated by the double Pareto-Lognormal distribution, will be the same as the model distribution of incomes under the current tax code. Next table compares the optima in the human capital model and in the benchmark static model.

[INSERT TABLE 1 HERE]

Marginal tax rates are significantly lower in the dynamic economy with human capital. While the static economy prescribes that in the optimum average marginal tax rates are 77.2%, the dynamic economy with human capital has average marginal tax rates 61.1% in the optimum.[16] Thus, there is a difference of 16.1% between average marginal tax rates in these two economies. Figure 3 compares the optimal tax codes in the benchmark static economy, the dynamic economy with human capital and the U.S. tax code. An interesting conclusion from the graph is, that a model without human capital calls for a significant increase in the marginal tax rate, i.e. tax reform towards higher taxes. If human capital is taken into account, then the current tax code appears quite optimal. The optimal tax code prescribes only slightly lower marginal taxes.

8.2 Alternative Assumptions

I also provide results for alternative parameter assumptions. First, I consider a case with compensated elasticity of labor of 0.25. Thus, we set $k = 4$. Figure 4 reveals that optimal tax rates are significantly higher, as is to be expected. Average marginal tax rates in the dynamic model are now 65.2%. Moreover, the difference between the optimal tax schedule in the dynamic and benchmark static economy is now smaller, being only 9.8%. Thus, the

---

[16] The average is computed as a weighted average across the population in the steady state dynamic model.

quantitative conclusions depend to large extent on the compensated elasticity of labor.

[INSERT FIGURE 4 HERE]

Second, I consider alternative values for the degree of increasing returns. I vary the parameter α from 0 to 1. Figure 5 plots the difference between average marginal tax rates in both economies in dependence on α. One can see that this dependence is almost linear.

[INSERT FIGURE 5 HERE]

9 Conclusions

The paper explored a question how endogenous human capital formation influences optimal income tax policies. I have demonstrated that a model with individual heterogeneity and unobserved human capital investment can be conveniently analyzed by extending the standard Mirrleesian framework to a dynamic setting. Having restated the problem in terms of a direct revelation game, I was able to analyze it analytically and compare the results with results of a benchmark static model without human capital. I also parameterized the model to fit U.S. economy and performed numerical exercises.

Special attention was paid to the case when the human capital production function is Cobb-Douglas. Two reasons why the optimal tax code in the economy with human capital will differ from the optimal tax code in a static economy were qualitatively identified. One of the main results of the paper shows that, under certain conditions, the economy with human capital is identical to a static economy that possesses the following properties: First, it has increasing returns in production and second, agents have higher disutility of labor.

The economic intuition behind the first property is based on the equilibrium property that abler people will both work more and accumulate more human capital. Thus, their marginal product is increasing in their non-leisure time. Second reason is closely tied to the presence of time dimension in the model.

The model is parameterized for the U.S. economy. The results are compared both with the benchmark static economy and with the U.S. economy. The quantitative conclusion is that the optimal marginal tax rates are significantly lower in the dynamic economy with the human capital than in the static one. The difference is roughly between 10-16% on average, depending on the elasticity of labor supply.

References

[1] Albanesi, S. and Sleet, Ch. (2003). "Dynamic optimal Taxation with Private Information" (Discussion Paper 140, Federal Reserve Bank of Minneapolis).

[2] Battaglini, M. and Coate, S. (2003). "Pareto Efficient Income Taxation with Stochastic Abilities" (Princeton University, mimeo).

[3] Becker, G.S. and Murphy, K.M. (1993). "The Division of Labor, Coordination Costs and Knowledge" (In: Gary S. Becker, Human Capital,The University of Chicago Press).

[4] Benabou, R. (2002). "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?", *Econometrica* **70**, 481-517.

[5] Brett, C. and Weymark, J.A. (2001). "Financing education using optimal redistributive taxation" (Vanderbildt University, unpublished manuscript).

[6] Browning, M., Hansen, L.P. and Heckman, J.J. (2000). "Micro Data and General Equilibrium Models" (In J. B. Taylor and M. Woodford, eds, Handbook of Macroeconomics Vol 1a, Chapter 8).

[7] Caucutt, E.M., Imrohoroglu, S. and Kumar, K.B. (2001). "Does the Progressivity of Taxes Matter for Economic Growth?" (Discussion paper 138, Federal Reserve Bank of Minneapolis).

[8] Conesa, J.C. and Krueger, D. (2001). "On the Optimal Progressivity of the Income Tax Code" (Stanford University, unpublished manuscript).

[9] Diamond, P. (1998). "Optimal Income Taxation: an Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *The American Economic Review,* **88**, 83-95.

[10] Diamond, P. and Mirrlees, J.A. (2002). "Optimal Taxation and the Le Chatelier Principle".

[11] Erosa, A. and Koreshkova, T. (2001). "Progressive Taxation in a Dynastic Model of Human Capital" (Universitat Autonoma de Barcelona, unpublished manuscript).

[12] Fernandes, A. and Phelan, Ch. (2000). "A recursive formulation for repeated agency with history dependence", *Journal of Economic Theory,* **91,** 223-247.

[13] Golosov, M., Kocherlakota, N. and Tsyvinski, A. (2003). "Optimal Indirect and Capital Taxation", *Review of Economic Studies*, 70, 569-588.

[14] Heckman, J.J. (1975). "Estimates of a Human Capital Production Function Embedded in a Life-Cycle Model of Labor Supply" (In N. Terlekyj, ed., NBER Studies in Income and Wealth, **40**, New York, Columbia University Press).

[15] Heckman, J.J., Lochner, L. and Taber, Ch. (1998). "Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents", *Review of Economic Dynamics* **1**, 1-58.

[16] Heckman, J.J., Lochner, L. and Taber, Ch. (1998). "Tax Policy and Human-Capital Formation", *The American Economic Review*, **88**, Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association, 293-297.

[17] Jorgensen, M. and Reed, W.J. (2004). "The Double Pareto-Lognormal Distribution - A New Parametric Model for Size Distributions", *Com. Stats -Theory & Methods*, **33**.

[18] Kapicka, M. (2004). "Optimal Taxation and Human Capital Accumulation" (UCSB, unpublished manuscript).

[19] Kapicka, M. (2004). "Optimal Taxation with Persistent Shocks" (UCSB, unpublished manuscript).

[20] Kotlikoff, L., Smetters, K. and Walliser, J. (2001). "Finding a Way Out of America's Demographic Dilemma" (Boston University, unpublished manuscript).

[21] Kocherlakota, N. (2004). "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation" (Stanford University, unpublished manuscript).

[22] Mirrlees, J.A. (1971). "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies,* **38**, 175-208.

[23] Mirrlees, J.A. (1976). "Optimal Tax Theory: A Synthesis", *Journal of Public Economics*, **6,** 327-358.

[24] Mirrlees, J.A. (1986). "The Theory of Optimal Taxation" (In K.J. Arrow and M.D. Intriligator, eds., Handbook of Mathematical Economics, vol. III, chapter 24).

[25] Saez, E. (2001). "Using Elasticities to Derive Optimal Income Tax Rates", *Review of Economic Studies,* **68**, 205-229.

[26] Tuomala, M. (1990). Optimal Income Taxation and Redistribution (Clarendon Press: Oxford).

[27] Trostel, P A. (1993). "The Effect of Taxation on Human Capital", *Journal of Political Economy*, **101**, 327-350.

[28] Werning, I. (2001). "Dynamic Optimal Taxation" (MIT, unpublished manuscript).

10 Appendix

Proof of Proposition 2. Suppose $\theta_0 > 0$. Suppose in addition that $i > 1$. Take $\theta \geq \theta_0$ and define a sequence

$$\alpha_\eta(\theta) = \frac{1}{\eta}[\tilde{V}_t(h_t(\theta+\eta), \theta+\eta; \theta^\infty) - \tilde{V}_t(h_t(\theta), \theta; \theta^\infty)].$$

This sequence is differentiable in θ. We will show that there exists an integrable function α such that $|\alpha_\eta| \leq \alpha$ all η. By the mean value theorem, this sequence will be bounded by an integrable function if the partial derivative

$$\frac{\partial \tilde{V}_t}{\partial \theta} = \sum_{j=t}^{\infty} \beta^{j-t} U_l (1 - s_j - \frac{y_j}{\theta h_j}) \frac{l_j}{\theta} + U_{l_t} (\frac{l_t}{h_t} + \frac{G_{h_t}}{G_{s_t}}) h_t'$$

is bounded by some integrable function. To show this, note that the expression $\frac{l}{\theta}$ is bounded for $\theta \geq \theta_0$. Because U tends to $-\infty$ if l tends to $\bar{l}$ then there is some $l_{\max}$ such that $s_j^* + l_j \leq l_{\max}$ $\forall \theta \in \Theta$. Hence U_l is also bounded and so is $U_{l_j} \frac{l_j}{\theta}$ for $j \geq t$. By our assumptions, $G(h, 0) > 0$ if $h > 0$ and so h_t is bounded from below by some $h_{t0} > 0$. Consequently, the term $\frac{G_{h_t}}{G_{s_t}}$ is also bounded. Finally, $h_t'(\theta)$ is integrable since h is differentiable and so, by the mean value theorem, the sequence $|\alpha_\eta|$ is bounded by some integrable function α.Because

$$\lim_{\eta \to 0} \alpha_\eta(\theta) = \sum_{j=t}^{\infty} \beta^{j-t} U_l ((1 - s_j(\theta) - \frac{y_j(\theta)}{\theta h_j(\theta)}) \frac{l_j(\theta)}{\theta} + U_{l_t} (\frac{l_t(\theta)}{h_t(\theta)} + \frac{G_{h_t}(\theta)}{G_{s_t}(\theta)}) h_t'(\theta),$$

Lebesque Dominated Convergence theorem implies that for $\theta > \theta_0$,

$$\begin{aligned} \lim_{\eta \to 0} \int_{\theta_0}^{\theta} \alpha_\eta(\varepsilon) d\varepsilon &= \int_{\theta_0}^{\theta} \lim_{\eta \to 0} \alpha_\eta(\varepsilon) d\varepsilon \\ &= \sum_{j=t}^{\infty} \beta^{j-t} \int_{\theta_0}^{\theta} U_l (1 - s_j(\varepsilon) - \frac{y_j(\varepsilon)}{\varepsilon h_j(\varepsilon)}) \frac{l_j(\varepsilon)}{\varepsilon} d\varepsilon + \int_{\theta_0}^{\theta} U_{l_t} (\frac{l_t(\varepsilon)}{h_t(\varepsilon)} + \frac{G_{h_t}(\varepsilon)}{G_{s_t}(\varepsilon)}) h_t'(\varepsilon) d\varepsilon \end{aligned}$$

We also have

$$\begin{aligned}\eta\int_{\theta_0}^{\theta}\alpha_\eta(\varepsilon)d\varepsilon &= \int_{\theta_0}^{\theta}[\tilde{V}_t(h_t(\varepsilon+\eta),\varepsilon+\eta;\varepsilon^\infty)-\tilde{V}_t(h_t(\varepsilon),\varepsilon;\varepsilon^\infty)]d\varepsilon \\ &\le \int_{\theta_0}^{\theta}[V_t(\varepsilon+\eta)-V_t(\varepsilon)]d\varepsilon=\int_{0}^{\eta}[V_t(\theta+\tilde{\eta})-V_t(\theta_0+\tilde{\eta})]d\tilde{\eta} \qquad (10)\end{aligned}$$

where the inequality follows from the fact that ε^∞ may not maximize utility for $(\varepsilon+\eta)-$type agent and the last equality follows from the fact that integration over the interval $[\theta_0+\eta,\theta]$ cancels out. We now divide (10) by η and take the limits for η converging to 0 from left and for η converging to 0 from right. Note that the sign of the inequality in (10) changes if we divide by negative η and so we have

$$\lim_{\eta\to 0_+}\frac{1}{\eta}\int_{0}^{\eta}[V_t(\theta+\tilde{\eta})-V_t(\theta_0+\tilde{\eta})]d\tilde{\eta}\ge\lim_{\eta\to 0}\int_{\theta_0}^{\theta}\alpha_\eta(\varepsilon)d\varepsilon\ge\lim_{\eta\to 0_-}\frac{1}{\eta}\int_{0}^{\eta}[V_t(\theta+\tilde{\eta})-V_t(\theta_0+\tilde{\eta})]d\tilde{\eta}.$$

But the first and the last term both converge to $V_t(\theta)-V_t(\theta_i)$ and so

$$V_t(\theta)-V_t(\theta_0)=\sum_{j=t}^{\infty}\beta^{j-t}\int_{\theta_0}^{\theta}U_l((c_j(\varepsilon),1-s_j(\varepsilon)-\frac{y_j(\varepsilon)}{\varepsilon h_j(\varepsilon)})\frac{l_j(\varepsilon)}{\varepsilon}d\varepsilon+\int_{\theta_0}^{\theta}U_{l_t}(\frac{l_t(\varepsilon)}{h_t(\varepsilon)}+\frac{G_{h_t}(\varepsilon)}{G_{s_t}(\varepsilon)})h_t'(\varepsilon)d\varepsilon.$$

If the expression $U_l\frac{y}{\theta^2}$ is unbounded from below on $[0,\theta_0]$ then it must be true that $\lim_{\varepsilon\to 0}V_t(\varepsilon)=-\infty$. But this is not incentive compatible since $\Gamma_t(0)>0$ and so every agent can secure himself positive lifetime utility. Hence $U_l\frac{y}{\theta^2}$ is bounded on $[0,\theta_0]$ and we can set $\theta_0=0$. ■

Proof of Proposition 3. It will later be convenient to work with slightly more general version of the model where $y=\theta h^{\gamma_h}l^{\gamma_l}$. This specification will include both the dynamic model with human capital (with $\gamma_h=\gamma_l=1$) and a static model (with $\gamma_h=0$) and will be used to find a relationship between the two in subsequent proofs. We will also eliminate schooling time from the problem and work with human capital sequences directly. Define $S(h', h)$ as an "inverse" of the human capital production function: S satisfies $h' = G(h, S(h', h))$.[17] Let also $\psi(l, h, h') = \frac{1}{\gamma_l} U_l l$ and $\eta(l, h, h') = \frac{U_l}{G_s}$[18]. The Euler equation (5) can then be written as

$$h_{t+1}\eta_t \geq \beta[\gamma_h \psi_{t+1} + h_{t+1} G_{h_{t+1}} \eta_{t+1}] \tag{11}$$

and the Euler equation holds with equality if $h_{t+1}(\theta) > G(h_t(\theta), 0)$. We also need to treat formally the fact that incentive compatibility constraint involves first derivative of h_t. To do so, introduce an additional choice variable $q_t(\theta)$ which is related to human capital as follows:

$$h_{t+1}(\theta) = \int_0^\theta q_{t+1} d\varepsilon. \tag{12}$$

This incorporate our assumption that h_{t+1} is differentiable a.e., The incentive compatibility constraint (4) can then be written as

$$V_t(\theta) = V_t(0) + \int_0^\theta \sum_{j=t}^{\infty} \beta^{j-t} \psi_j \frac{d\varepsilon}{\varepsilon} + \int_0^\theta \phi_t q_t d\varepsilon.$$

where $\phi(l, h, h') = U_l(\frac{\gamma_h}{\gamma_l}\frac{l}{h} + \frac{G_h}{G_s})$. Finally, define consumption implicitly as a function of period utility, output and current and future human capital: $u = U(C(u, h, h', l), 1 - l - S(h, h'))$.[19] The relationship between period utility and lifetime utility is given by $u_t(\theta) = V_t(\theta) - \beta V_{t+1}(\theta)$ and so the incentive compatibility can be rewritten as

$$u_t(\theta) - u_t(0) = \int_0^\theta \psi_t \frac{d\varepsilon}{\varepsilon} + \int_0^\theta \phi_t q_t d\varepsilon - \beta \int_0^\theta \phi_{t+1} q_{t+1} d\varepsilon. \tag{13}$$

The objective of the social planner can be compactly written as

$$\max_{\{l_t, u_t, h_{t+1}, q_t\}_{t\geq 0}} \int_\theta \sum_{t=0}^{\infty} \beta^t u_t f d\theta$$

---

[17] It follows that $S_{h'} = \frac{1}{G_s}$ and $S_h = -\frac{G_h}{G_s}$.

[18] We use the shortcut $\psi_t = \psi(l_t, h_t, h_{t+1})$ and similarly for η.

[19] It follows that $C_u = \frac{1}{U_c}$, $C_l = \frac{U_l}{U_c}$, $C_{h'} = \frac{U_l}{U_c}\frac{1}{G_s}$ and $C_h = -\frac{U_l}{U_c}\frac{G_h}{G_s}$.

subject to resource constraint

$$w\int\limits_{\theta}\theta h_t^{\gamma_h} l_t^{\gamma_l} f d\theta \geq \int\limits_{\theta} C(u_t, h_t, h_{t+1}, l_t) f d\theta,$$

Euler equation (11), incentive compatibility constraint (13) and constraint (12). Denote the Lagrange multipliers on the resource constraint, Euler Equation, incentive compatibility constraint and on constraint (12) respectively as $\beta^t\hat{\lambda}_t$, $\beta^t\rho_t$, $\beta^t\mu_t$ and $\beta^t\pi_t$. Let also $M_t(\theta) = \int\limits_{\theta}^{\infty}\mu_t(\varepsilon)d\varepsilon$. First order condition in u_t and l_t are

$$\begin{aligned} f &= \hat{\lambda}_t \frac{1}{U_{c_t}} f - \mu_t & (14) \\ w\gamma_l\hat{\lambda}_t\theta h_t^{\gamma_h} l_t^{\gamma_l - 1} f &\leq \hat{\lambda}_t \frac{U_{l_t}}{U_{c_t}} f + [\frac{1}{\theta}\frac{\partial\psi_t}{\partial l_t} + \frac{\partial\phi_t}{\partial l_t} q_t] M_t(\theta) - \frac{\partial\phi_t}{\partial l_t} q_t M_{t-1} & (15) \\ & + \rho_t h_{t+1}\frac{\partial\eta_t}{\partial l_t} - \rho_{t-1}(\gamma_h\frac{\partial\psi_t}{\partial l_t} + h_t G_{h_t}\frac{\partial\eta_t}{\partial l_t}) \end{aligned}$$

and the last equation holds with equality if $l_t(\theta) > 0$. First order condition in h_{t+1} is

$$\begin{aligned} w\gamma_h\beta\hat{\lambda}_{t+1}\theta h_{t+1}^{\gamma_h - 1} l_{t+1}^{\gamma_l} f \leq \hat{\lambda}_t \frac{\eta_t}{U_{c_t}} f - \beta\hat{\lambda}_{t+1}\frac{\eta_{t+1}}{U_{c_{t+1}}} G_{h_{t+1}} f - \frac{\partial\phi_t}{\partial h_{t+1}} q_t M_t & & (16) \\ + [\frac{1}{\theta}\frac{\partial\psi_t}{\partial h_{t+1}} + \frac{\partial\phi_t}{\partial h_{t+1}} q_t - \beta\frac{\partial\phi_{t+1}}{\partial h_{t+1}} q_{t+1}] M_t + \beta(\frac{1}{\theta}\frac{\partial\psi_{t+1}}{\partial h_{t+1}} + \frac{\partial\phi_{t+1}}{\partial h_{t+1}} q_{t+1}) M_{t+1} \\ - \rho_{t-1}[\gamma_h\frac{\partial\psi_t}{\partial h_{t+1}} + h_t\frac{G_{hs_t}}{G_{s_t}}\eta_t + h_t G_{h_t}\frac{\partial\eta_t}{\partial h_{t+1}}] + \beta\rho_{t+1}h_{t+2}\frac{\partial\eta_{t+1}}{\partial h_{t+1}} \\ - \rho_t[\beta\gamma_h\frac{\partial\psi_{t+1}}{\partial h_{t+1}} + \beta G_{h_{t+1}}\eta_{t+1} + \beta h_{t+1}(G_{hh_{t+1}} - G_{hs_{t+1}}\frac{G_{h_{t+1}}}{G_{s_{t+1}}})\eta_{t+1} \\ + \beta h_{t+1} G_{h_{t+1}}\frac{\partial\eta_{t+1}}{\partial h_{t+1}} - \eta_t - h_{t+1}\frac{\partial\eta_t}{\partial h_{t+1}}] + \beta\pi_{t+1} \end{aligned}$$

and the equation holds with equality if $h_{t+1}(\theta) > G(h_t(\theta), 0)$. Finally, first order condition in q_{t+1} is

$$\int\limits_{\theta}^{\infty}\pi_{t+1}(\varepsilon)d\varepsilon = \phi_{t+1}(M_t - M_{t+1}). \tag{17}$$

Functions η, ψ, ϕ and C are all continuous in their arguments and f is continuously differentiable. h_0 is constant over all agents. Consequently, the functions $l_t, u_t, h_{t+1}, q_{t+1}$ and the Lagrange multipliers will also be continuous in θ for all $t \geq 0$. Constraints (13) in turn implies that u_t is differentiable in θ. Let Θ_t be a subset of Θ such that (15) and (16) bind. Then the first order conditions in turn imply that l_t, h_{t+1}, q_{t+1} are differentiable in θ on the interior of Θ_t. But since l_t, h_{t+1}, q_{t+1} are also differentiable on the interior of the complement of Θ_t, they are differentiable almost everywhere, except for points when agents stop working or investing in human capital.

Steady state. We now impose a condition of steady state, where $u_t, l_t, h_{t+1}, \hat{\lambda}_t, \mu_t, \pi_t$ and ρ_t are all constant. We can drop the time subscript and simplify the equations by eliminating μ. Let $\lambda = \frac{1}{\hat{\lambda}}$ be the inverse of the Lagrange Multiplier on the resource constraint. Since equation (17) implies that $\pi = 0$, the steady state first order conditions become

$$\begin{aligned}
\frac{\partial \psi}{\partial l} \int_{\theta}^{\infty} f\left(\frac{1}{U_c} - \lambda\right) d\varepsilon &= \theta^2 f\left(w \gamma_l h^{\gamma_h} l^{\gamma_l - 1} - \frac{U_l}{\theta U_c}\right) - \rho\lambda\theta\left[h(1-G_h)\frac{\partial \eta}{\partial l} - \gamma_h \frac{\partial \psi}{\partial l}\right] \\
\left(\frac{\partial \psi}{\partial h'} + \beta \frac{\partial \psi}{\partial h}\right) \int_{\theta}^{\infty} f\left(\frac{1}{U_c} - \lambda\right) d\varepsilon &= \beta\theta^2 f\left(w \gamma_h h^{\gamma_h - 1} l^{\gamma_l} + \frac{U_l}{\theta U_c}\frac{1-\beta G_h}{\beta G_s}\right) + \rho\lambda\theta\Psi
\end{aligned}$$

where Ψ is defined by

$$\Psi = h(1-G_h)\left(\frac{\partial \eta}{\partial h'} + \beta \frac{\partial \eta}{\partial h}\right) + \eta\left[1 - \beta G_h - \beta h\left(G_{hh} + G_{hs}\frac{1-\beta G_h}{\beta G_s}\right)\right].$$

Using $\frac{\partial \psi}{\partial l} = U_l - U_{ll} l$, $\frac{\partial \phi}{\partial l} = \frac{\gamma_h}{\gamma_l}\frac{U_l}{h} - \frac{U_{ll}}{\beta G_s}$ and $\frac{\partial \psi}{\partial h'} + \beta \frac{\partial \psi}{\partial h} = -\frac{1}{\gamma_l} U_{ll} \frac{1-\beta G_h}{G_s}$ we can rewrite the equations as

$$\begin{aligned}
\frac{1}{\gamma_l}\int_{\theta}^{\infty} f\left(\frac{1}{U_c} - \lambda\right) d\varepsilon &= \theta^2 f \frac{w\gamma_l h^{\gamma_h} l^{\gamma_l - 1} - \frac{U_l}{\theta U_c}}{U_l - U_{ll} l} + \rho\lambda\theta \frac{h\frac{1-G_h}{G_s} U_{ll} + \frac{\gamma_h}{\gamma_l}(U_l - U_{ll} l)}{U_l - U_{ll} l} \qquad (18) \\
\frac{1}{\gamma_l}\int_{\theta}^{\infty} f\left(\frac{1}{U_c} - \lambda\right) d\varepsilon &= \theta^2 f \frac{w\gamma_l h^{\gamma_h} l^{\gamma_l - 1} - \frac{U_l}{\theta U_c}}{-U_{ll} l} - \rho\lambda\theta \frac{\Psi + \frac{\gamma_h}{\gamma_l} U_{ll} \frac{1-\beta G_h}{G_s}}{-U_{ll} l \frac{1-\beta G_h}{G_s}}
\end{aligned}$$

The proposition is proved by setting $\gamma_h = \gamma_l = 1$. ■

Proof of Proposition 4. It is straightforward to show that with Cobb-Douglas production function, steady state relationship between h and s is given by $h(s) = qs^{\alpha}$ where $q = (\frac{1}{\delta})^{\frac{1}{1-\alpha_1}}$ and $\alpha = \frac{\alpha_2}{1-\alpha_1}$. The relationship between labor supply and schooling time is deduced form the Euler Equation: $l = \frac{\gamma_l}{\gamma_h} ps$ where $p = \frac{1-\beta(1-\delta(1-\alpha_1))}{\beta\alpha_2\delta}$. We also have $\Psi = \frac{\beta p}{\alpha} s^{1-\alpha}(U_l - U_{ll}s)$ and $\frac{1-\beta G_h}{G_s} = \beta p s^{1-\alpha}$ and $h\frac{1-G_h}{G_s} = \frac{s}{\alpha}$. Using these expressions, we get that ρ solves

$$\rho = \frac{\alpha}{\lambda} f\theta \frac{\tilde{w} s^{\gamma_h \alpha + \gamma_l - 1} - \frac{U_l}{\theta U_c}}{U_l - U_{ll}(s+l)}$$

where $\tilde{w} = w\gamma_l^{\gamma_l}\gamma_h^{1-\gamma l} p^{\gamma_l - 1} q^{\gamma_h}$. Substituting back into 18 and replacing $s + l = (1 + \frac{\gamma_l}{\gamma_h} p)s$ by n we have

$$\int_{\theta}^{\bar{\theta}} f(\frac{1}{U_c} - \lambda) d\varepsilon = (\gamma_l + \gamma_h \alpha)\theta^2 f \frac{\bar{w} n^{\gamma_h \alpha + \gamma_l - 1} - \frac{U_n}{\theta U_c}}{U_n - U_{nn} n} \tag{19a}$$

with $\bar{w} = w\gamma_l^{\gamma_l}\gamma_h^{\gamma_h \alpha} p^{\gamma_l - 1} q^{\gamma_h}(\gamma_h + \gamma_l p)^{1-\gamma_h\alpha-\gamma_l}$.

Let $\theta_0 = \inf[\theta : h(\theta) > 0]$. The incentive compatibility constraint (13) can be written for $\theta \geq \theta_0$ in a differential equation form:

$$\frac{du}{d\theta} = \frac{p}{\gamma_h + \gamma_l p} \frac{U_n n}{\theta} - \frac{\gamma_h(1-p)}{\gamma_h + \gamma_l p} U_n \frac{dn}{d\theta}.$$

Write now $U(c, 1-n) = U^1(c) + \nu U^2(1-n)$ and $t = \log(\theta)$. We have $\frac{du}{dt} = U_c^1 c \frac{d\log c}{dt} - \nu U_n^2 n \frac{d\log n}{dt}$. and use it to eliminate u :

$$U_c^1 c \frac{d\log c}{dt} = \frac{p\nu}{\gamma_h + \gamma_l p} U_n^2 n [1 + (\gamma_l + \gamma_h \alpha)\frac{d\log n}{dt}]. \tag{21}$$

We now have two equations in two unknown functions $c(\theta)$ and $n(\theta)$. The parameters $(\gamma_l, \gamma_h, \alpha, p, w, \nu)$ in equations (19a) and (21) are bundled in such a way that the solution will be identical for all the parameter values that keep $\gamma_l + \gamma_h \alpha$, $\frac{p\nu}{\gamma_h + \gamma_l p}$ and $\frac{\bar{w}}{\nu}$ constant.

The dynamic economy with human capital has $\gamma_l = \gamma_h = 1$ while a static economy without human capital has $\gamma_h = 0$.

Suppose that $\hat{\gamma}_l, \hat{\nu}$ and $\hat{w}$ are the parameters of the static model. If the static economy is to replicate the allocation of the steady state dynamic economy, its parameters must satisfy

$$\hat{\gamma}_l = 1 + \alpha,\ \frac{\hat{\nu}}{\hat{\gamma}_l} = \frac{p\nu}{1+p},\ \frac{\hat{w}\hat{\gamma}_l}{\hat{\nu}} = \frac{wq}{\nu}(1+p)^{-\alpha}$$

and this implies that $\hat{\nu} = (1+\alpha)\frac{p\nu}{1+p}$ and $\hat{w} = wq\frac{p}{(1+p)^{1+\alpha}}$.

The final complication is associated with the fact that $h(\theta)$ will be discontinuous at θ_0 and so will be $u(\theta)$. To replicate this fact in a static economy note that, as shown in [24], we can write the incentive compatibility constraint (13) as a weak ($\geq$) inequality. The constraint will naturally be satisfied as an equality for continuous objective function. Suppose, however, that we introduce discontinuity by writing the objective function of the social planner in the static problem as

$$\int_{\Theta} \chi(\theta) U[c(\theta), 1 - n(\theta)] f(\theta) d\theta$$

where $\chi(\theta) = 1$ if $n(\theta) > 0$ and $\chi(\theta) = \bar{\chi} < 1$ if $n(\theta) = 0$. Then the incentive compatibility will hold as an inequality for appropriately chosen $\bar{\chi}$, satisfying $\bar{\chi} U(c(0), 1) = u(\theta_0)$. ■

Table 1

	benchmark static	human capital
avg. marginal tax rates	77.2%	61.1%
Gini of earnings	0.58	0.67
Gini of consumption	0.29	0.44
cons to earn ratio	0.728	0.756
transfer to avg. earnings	26.4%	28.2%

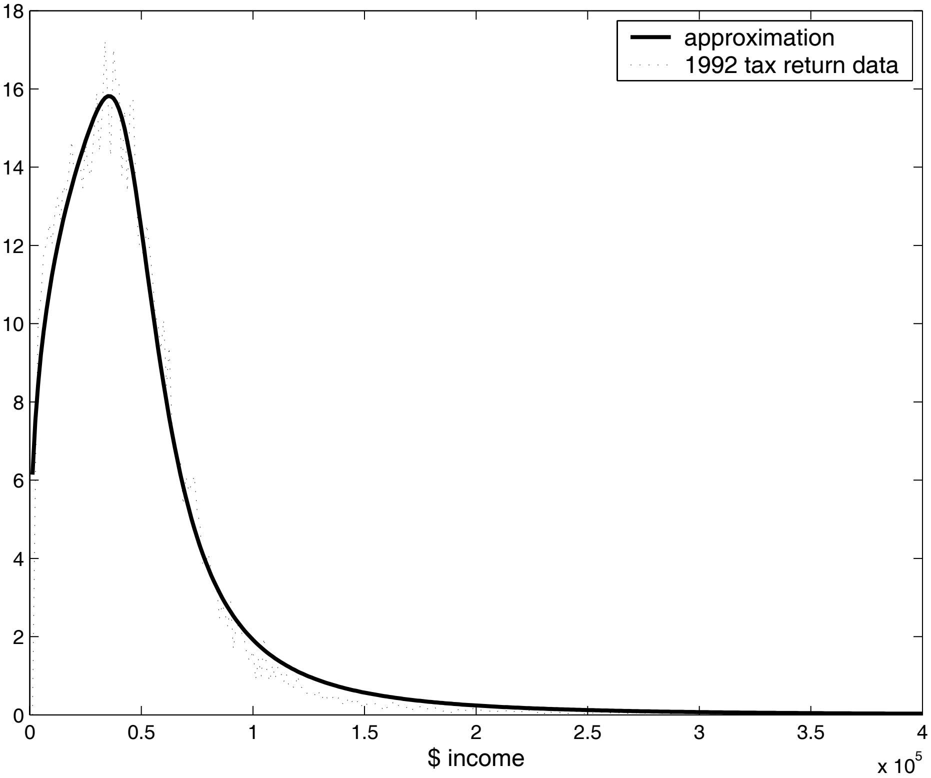

Figure 1: Earnings distribution and its approximation

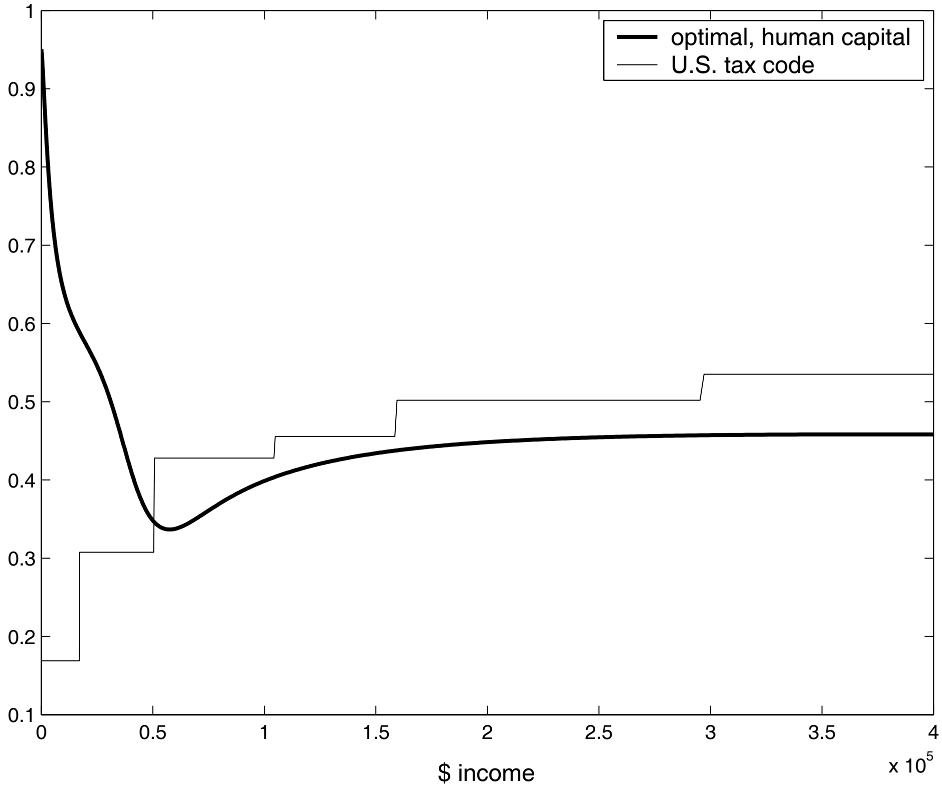

Figure 2: Marginal tax rates: optimum v.s. U.S.

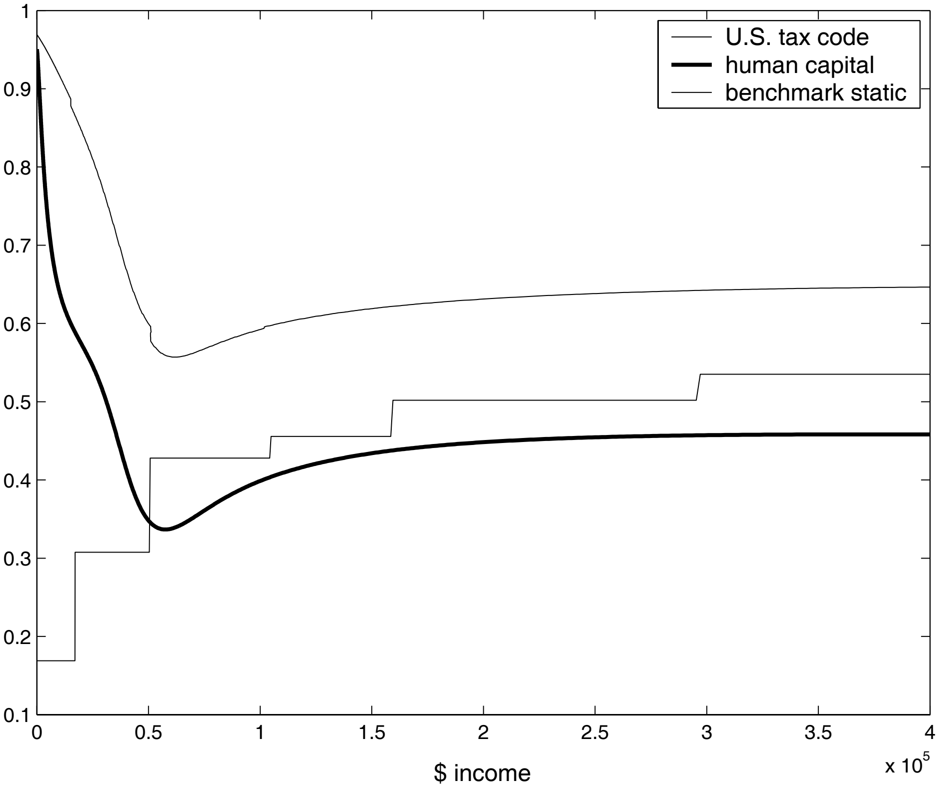

Figure 3: Marginal tax rates: human capital vs. static model

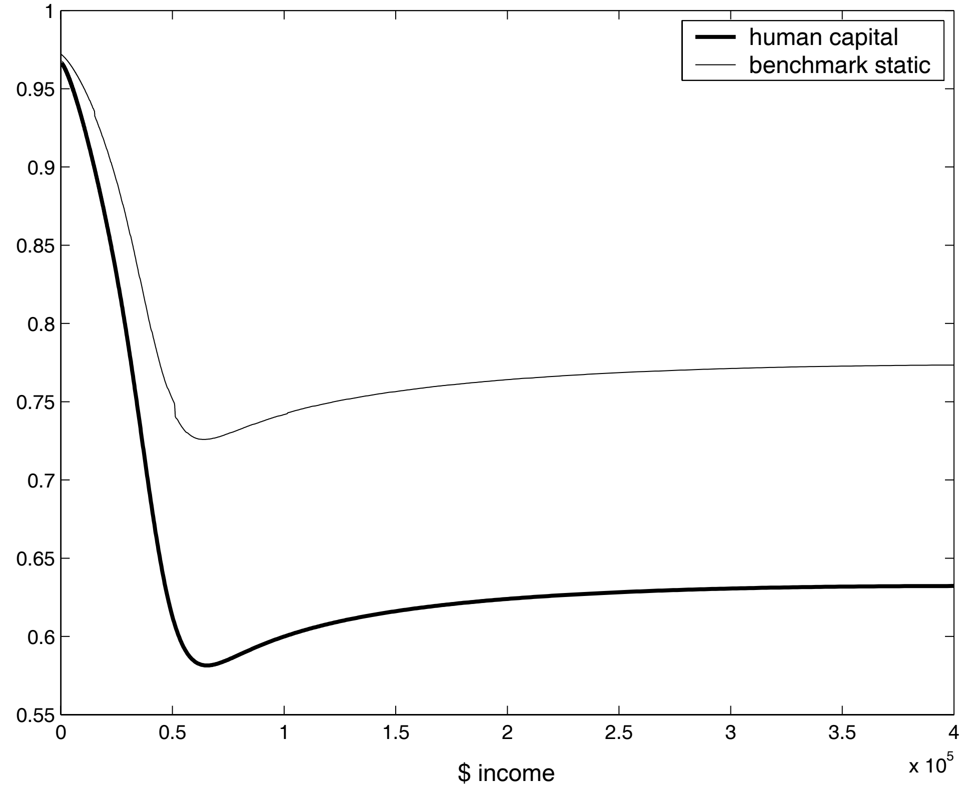

Figure 4: Optimal marginal tax rates, human capital vs. static model, $k = 4$.

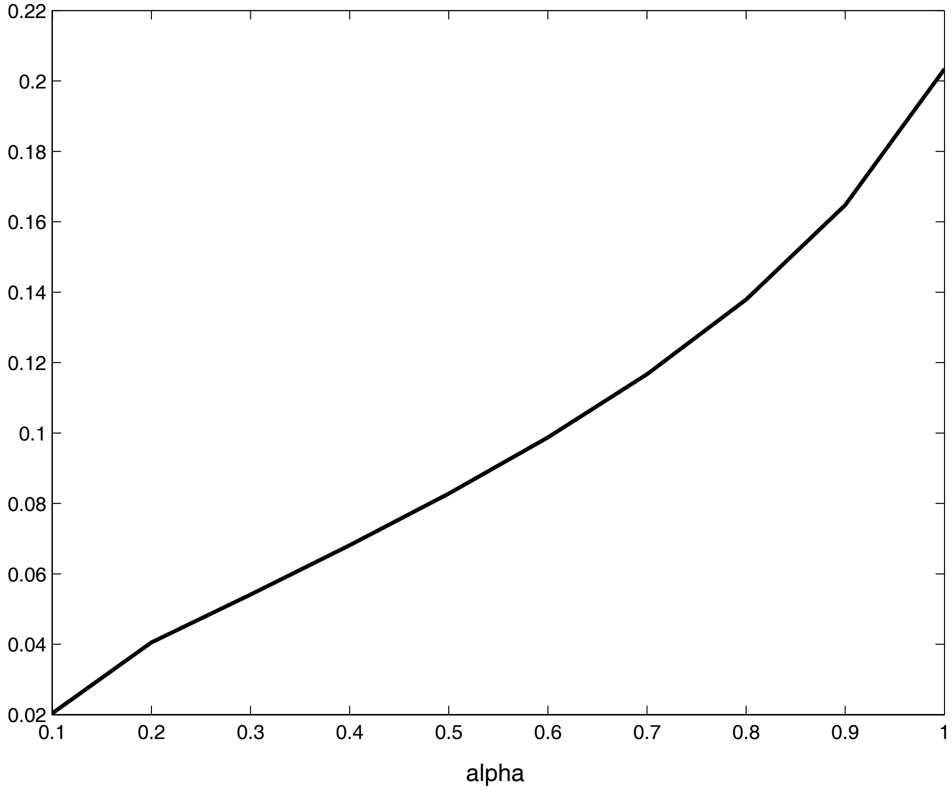

Figure 5: Average marginal tax rate difference

Figure Legends

Earnings distribution and its approximation

Marginal tax rates: optimum v.s. U.S.

Marginal tax rates: human capital vs. static model

Optimal marginal tax rates, human capital vs. static model, $k = 4$.

Average marginal tax rate difference