

The Dynamics of Optimal Taxation when Human Capital is Endogenous

Marek Kapička*

Department of Economics

University of California, Santa Barbara

November 28, 2011

Abstract

This paper characterizes the dynamics of optimal income taxes in a multiperiod 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 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

*E-mail: mkapicka@econ.ucsb.edu.

1 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.

Since human capital is unobservable, the incentives to accumulate human capital can be provided only indirectly, through variations in the marginal income tax rates over time and across agents. The intertemporal variations in the marginal income tax rate are driven by the following considerations: Conditional on labor supply, investment in human capital is smaller than the socially optimal investment. The reason is that a positive marginal income tax rate creates a wedge between private returns and social returns. However, since human capital is unobservable, the social planner cannot separate the labor supply and schooling decisions. If human capital investment is a complement to labor supply, labor supply will be encouraged relative to an economy without human capital and the marginal income taxes will be lower. On the other hand, if human capital investment is a substitute to the labor supply, labor supply will be discouraged relative to an economy without human capital and marginal income taxes will be higher. The intertemporal profile of the marginal income taxes thus depends on how strong the substitutability and complementarity effects are at any point in time.

To study the intertemporal variations in the marginal income tax rates I first solve three simple examples. In the first two examples, labor supply and schooling are separable in the utility function and they are not substitutes to each other. In the first one, human capital fully depreciates each period, and human capital investment is complementary to the labor supply in the next period. The only exception is human capital in the initial period, which is already fixed. I show that the marginal income tax rates for all agents decrease from the second period onwards, and stay constant after that.

The second example assumes no depreciation in human capital, and assumes that indi-

viduals live for three periods. Contrary to the first example, the marginal income tax rates strictly decrease between the last two periods as well. The reason is that labor supply in the third period is complementary to investment in human capital in both the first and second period, while labor supply in the second period is complementary only to investment in human capital in the first period. Thus, labor supply in the third period should be encouraged more, and the marginal income tax rate should be lower.

The third example differs from the first two examples in that utility is non-separable in labor supply and schooling. The example thus features both complementarity and substitutability effects. In particular, schooling is a substitute to current period labor supply since higher labor supply increases the marginal disutility from schooling. It is also a complement to next period labor supply. Hence, only the substitutability effect remains in the initial period, and only the complementarity effect remains in the third period. As a result, I show that the marginal income tax rates again strictly decrease over time. But the third reform indicates that this pattern may not hold in general. Nonseparability of schooling and labor supply also implies that labor supply elasticity depends on the amount of schooling the agent takes. If schooling decreases, the elasticity of labor supply will decrease as well. This provides a reason to increase the marginal income tax rates over time. In this example, changes in the elasticity of labor supply are not strong enough to overweight the complementarity and substitutability effects, but in other situations that may as well be true.

All three examples indicate that, unless the changes in the elasticity of labor supply are large enough, the optimal tax reform in a more general economy will start with relatively high marginal income tax schedules, and will decrease them over time. In the short run, past human capital investments are already made, and only the substitution effect remains, as in the third example. Thus, marginal income taxes should be relatively high to encourage human capital investment and future labor supply. Over time, the complementarity effect becomes stronger, as in the second and third example, and the marginal income tax schedules should decrease.

In the numerical example I analyze the optimal allocations and tax policies in a more general economy. 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.

1.1 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 \(2005\)](#), [Golosov, Tsyvinski, and Troshkin \(2010\)](#), [Farhi and Werning \(2010\)](#)) 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. [Shourideh \(2010\)](#) studies optimal allocation with entrepreneurial risk in a multiperiod setting. 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 setting with observable human capital that the problem 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 when human capital is unobservable.

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.¹ 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

¹See 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.

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 \(1998\)](#) and [Saez \(2001\)](#) 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.² As well, it is now irrelevant whether individuals can borrow and save or not, because they are indifferent about the timing of consumption, and care only about its present value.

2 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 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. The following is assumed:

Assumption 1 *V 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 also that the period utility is linear in consumption: there are no income effects on leisure.

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 $G : \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}_+$:

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

We assume the following:

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

Assumption 2 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 $S(h_t, h_{t+1})$. Assumption (2) implies that S is strictly convex, strictly decreasing in its first argument, strictly increasing in its second argument, and differentiable.

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 θh_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 information 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.

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.

2.1 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 $h_{t+1} : \Theta^2 \rightarrow \mathbb{R}_+$ specifies recommended amount of human capital at the beginning of period $t + 1$. In particular, $h_{t+1}(\theta, \hat{\theta})$ specifies the next period human capital of an agent with ability θ who has reported $\hat{\theta}$.³

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. An agent with ability θ who has reported $\hat{\theta}$ then chooses a sequence of human capital allocations $\hat{h} = \{\hat{h}_{t+1}\}_{t=0}^T$ with $\hat{h}_{t+1} : \Theta^2 \rightarrow \mathbb{R}_+$

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

that maximizes her lifetime utility or, equivalently, minimizes the utility costs of delivering the output sequence $\{y_t(\hat{\theta})\}_{t=0}^T$:

$$\hat{h}(\theta, \hat{\theta}) \in \arg \min_{\tilde{h}} \sum_{t=0}^T \beta^t V \left(\frac{y_t(\hat{\theta})}{\theta \tilde{h}_t}, S(\tilde{h}_t, \tilde{h}_{t+1}) \right). \quad (1)$$

The utility of a θ - type agent who reports $\hat{\theta}$ is

$$W_\sigma(h_0, \theta; \hat{\theta}) = u(\hat{\theta}) - \sum_{t=0}^T \beta^t \left[V \left(\frac{y_t(\hat{\theta})}{\theta \hat{h}_t(\theta, \hat{\theta})}, \hat{s}_t(\theta, \hat{\theta}) \right) - V \left(\frac{y_t(\hat{\theta})}{\theta \hat{h}_t(\hat{\theta}, \hat{\theta})}, \hat{s}_t(\hat{\theta}, \hat{\theta}) \right) \right],$$

where, to simplify notation, $\hat{s}_t(\theta, \hat{\theta}) \equiv S(\hat{h}_t(\theta, \hat{\theta}), \hat{h}_{t+1}(\theta, \hat{\theta}))$. The incentive compatibility constraint requires the allocation to be such that θ - type agent prefers to report her own type to any other report and chooses the recommended human capital sequence. Thus, an allocation σ is *incentive compatible* if, for all $\theta \in \Theta$, \hat{h} solves (1) for all $\theta, \hat{\theta} \in \Theta$, and

$$W_\sigma(h_0, \theta; \theta) \geq W_\sigma(h_0, \theta; \hat{\theta}) \quad \forall \hat{\theta} \in \Theta. \quad (2)$$

Note that $W_\sigma(h_0, \theta; \theta) = u(\theta)$ by construction. Also, for any incentive compatible allocation, the human capital sequence chosen is $\hat{h}(\theta, \theta)$, and it will be denoted simply by $h(\theta)$. Let Σ^{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 $\sum_{t=0}^T \beta^t c_t = u + \sum_{t=0}^T \beta^t V(l_t, s_t)$, the revenue maximization problem can be written as

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

subject to the promise keeping constraint

$$u(\theta) \geq \underline{U}(\theta) \quad \forall \theta \in \Theta. \quad (3)$$

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.⁴

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 $\underline{\theta}$.

3 Characterizing incentive compatible allocations

The definition of incentive compatibility involves two conditions: one is that the agent must follow the recommended human capital sequence, and one is that the agent must find it optimal to report truthfully her productivity. I will now provide a sharper characterization of the first condition.

3.1 The Euler Equation

The first-order condition to the minimization problem (1) yields the following Euler equation:

$$V_{s,t} \frac{1}{G_{s,t}} \geq \beta V_{l,t+1} \frac{y_{t+1}(\hat{\theta})}{\theta \hat{h}_{t+1}(\theta, \hat{\theta})^2} + \beta V_{s,t+1} \frac{G_{h,t+1}}{G_{s,t+1}}, \quad = \text{ if } s_t(\theta, \hat{\theta}) > 0, \quad (4)$$

where, to reduce notation, $V_{s,t} = V_s(\frac{y_t(\hat{\theta})}{\theta \hat{h}_t(\theta, \hat{\theta})}, s_t(\theta, \hat{\theta}))$, $G_{s,t} = G_s(\hat{h}_t(\theta, \hat{\theta}), s_t(\theta, \hat{\theta}))$, and similarly for $V_{l,t}$, and $G_{h,t}$.

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

Equation (4) is the Euler equation in human capital investment. The left hand side are the marginal costs of increasing human capital next period by one unit: To increase human capital by one unit, one needs to invest $\frac{1}{G_{s,t}}$ units of schooling time, and the marginal costs are then $V_{s,t}\frac{1}{G_{s,t}}$. The right hand side shows that there are two types of private benefits from such an investment. First, higher investment in human capital reduces labor supply next period by $\frac{y_{t+1}(\hat{\theta})}{\theta\hat{h}_{t+1}(\theta,\hat{\theta})^2}$ and the marginal benefit is equal to $V_{l,t+1}\frac{y_{t+1}(\hat{\theta})}{\theta\hat{h}_{t+1}(\theta,\hat{\theta})^2}$. Second, higher investment in human capital today decreases the amount of next period schooling time required to achieve a given level of human capital two periods ahead by $\frac{G_{h,t+1}}{G_{s,t+1}}$. This marginal benefit is then given by $V_{s,t+1}\frac{G_{h,t+1}}{G_{s,t+1}}$. 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 are equal to the marginal benefits of doing so. Note that the possibility of a corner solution cannot be dispensed with, since it is possible that the social planner will choose $y_{t+1}(\hat{\theta})$ for all future periods for some $\hat{\theta}$, in which case the corner solution is likely to apply (depending on the properties of the human capital accumulation function).

The Euler equation (4) is necessary for optimality of \hat{h} . Together with a transversality condition (if $T = \infty$), it is sufficient:⁵

Lemma 3 *If \hat{h} solves (1) then (4) holds. Conversely, if (4) holds and, if $T = \infty$,*

$$\lim_{t \rightarrow \infty} \beta^t \left[V_{l,t} \frac{y_t(\hat{\theta})}{\theta \hat{h}_t(\theta, \hat{\theta})} + V_{s,t} \frac{\hat{h}_t(\theta, \hat{\theta}) G_{h,t}}{G_{s,t}} \right] = 0, \quad (5)$$

then $\hat{h}(\theta, \hat{\theta})$ solves (1).

The transversality condition applies only if T is infinite. Even then, it provides only mild restrictions on the nature of the problem. It will always be satisfied if all the partial derivatives in V_l , V_s , as well as $\frac{hG_h}{sG_s}$ are bounded.

⁵The proof is straightforward and is available upon request. All other proofs are in Appendix 1.

3.2 First Order Approach

The incentive compatibility constraint (2) will now be further characterized 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 σ is given by $\sigma = (u, l, h')$.

Necessary conditions for incentive compatibility are given by the envelope condition in θ :

Proposition 4 *If an allocation is incentive compatible then*

$$u(\theta) = \int_{\underline{\theta}}^{\theta} \sum_{t=0}^T \beta^t V_{l,t}(\varepsilon) l_t(\varepsilon) \frac{d\varepsilon}{\varepsilon} + u(\underline{\theta}), \quad (6)$$

The proof of the envelope condition (6) is a fairly straightforward modification of the proof that can be found in (Kapička 2006) or in (Mirrlees 1986), and is omitted. Equation (6) 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 $u(\underline{\theta})$.

Constraints (4) and (6) 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 (6) 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 (3).

Let Σ^{FOC} be the set of allocations that satisfy the first order conditions (4) and (6). The constraints (4) and (6) are necessary for an allocation to be incentive compatible, but they are

not sufficient: an allocation might satisfy them, yet there would be a profitable deviation in the agent's report.⁶ This problem is well known from the static optimal taxation literature. 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. Next proposition therefore presents conditions that are sufficient for an allocation to be incentive compatible.

Proposition 5 *Suppose that an allocation satisfies (4) and (6). If, in addition,*

$$\Omega(\theta, \hat{\theta}) = \sum_{t=0}^T \beta^t V_l \left(\frac{y_t(\hat{\theta})}{\theta \hat{h}_t(\theta, \hat{\theta})}, S(\hat{h}_t(\theta, \hat{\theta}), \hat{h}_{t+1}(\theta, \hat{\theta})) \right) \frac{y_t(\hat{\theta})}{\theta \hat{h}_t(\theta, \hat{\theta})}$$

increases in $\hat{\theta}$ for all $\theta \in \Theta$ then the allocation is incentive compatible.

The monotonicity condition of Proposition (5) will hold for instance if both $\frac{y_t(\hat{\theta})}{\theta \hat{h}_t(\theta, \hat{\theta})}$ and $S(\hat{h}_t(\theta, \hat{\theta}), \hat{h}_{t+1}(\theta, \hat{\theta}))$ are increasing in $\hat{\theta}$ for all θ .⁷ However, the unfortunate feature of those requirements is that they involve an off-equilibrium object $\hat{h}(\theta, \hat{\theta})$. In general, one thus cannot verify the monotonicity of Ω by examining the equilibrium allocation only. However, as we shall see, in some special cases one can find more primitive conditions on only the equilibrium allocation that will imply the monotonicity of Ω .

3.3 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 Σ^{FOC} :

$$\hat{\Psi}(\underline{U}, h_0) = \min_{\sigma \in \Sigma^{FOC}} \int_{\Theta} \left\{ \sum_{t=0}^T \beta^t [y_t(\theta) - V(l_t(\theta), S(h_t(\theta), h_{t+1}(\theta)))] - u(\theta) \right\} f(\theta) d\theta, \quad (7)$$

subject to the promise keeping constraint (3).

⁶Given that $\hat{h}(\theta, \hat{\theta})$ is optimal conditional on θ and $\hat{\theta}$ by Lemma (3), joint deviations in both the report and human capital (from \hat{h}) are never optimal, unlike for instance [Kocherlakota \(2005\)](#).

⁷[Shourideh \(2010\)](#) derives similar conditions in a model with risky physical capital. See also [Pavan, Segal, and Toikka \(2009\)](#) for a more general formulation of the sufficiency conditions.

Let $\mu(\theta)$ be the Lagrange multiplier on the Envelope condition (6). Define a *cumulative distortion function* X_μ 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 social planner's preferences toward redistribution. To see the intuition behind X_μ , consider a small increase in the labor supply of an agent with ability θ . Such a change tightens the incentive compatibility constraint (6) for all the agents with higher abilities, and the social planner's costs of doing so are given by the average of the Lagrange multipliers on the promise keeping constraint (3) for all such agents. There is $1 - F(\theta)$ people with higher abilities, and the change in labor supply directly affects $f(\theta)$ people. Note that in the case of Rawlsian preferences $\mu(\theta) = 1$ for all $\theta > \underline{\theta}$, the second term is equal to one, and $X_\mu(\theta) = \frac{1 - F(\theta)}{\theta f(\theta)}$.

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:

Theorem 6 *An allocation σ^* , together with the Lagrange multipliers on the envelope condition μ^* , solves the relaxed social planner's problem if and only if satisfies the envelope condition (6), the promise keeping constraint (3), and for all $\theta \in \Theta$,*

$$[l^*(\theta, h_0), h^*(\theta, h_0)] \in \arg \max_{l, h} \sum_{t=0}^T \beta^t [\theta h_t l_t - V_t - V_{l,t} l_t X_{\mu^*}(\theta)], \quad (8)$$

subject to

$$\frac{V_{s,t}}{G_{s,t}} \geq \beta(V_{l,t+1} \frac{l_{t+1}}{h_{t+1}} + V_{s,t+1} \frac{G_{h,t+1}}{G_{s,t+1}}), \quad = \text{ if } S(h_t, h_{t+1}) > 0 \quad (9)$$

and

$$\mu^*(\theta) [u(\theta) - \underline{U}(\theta)] = 0. \quad (10)$$

The maximization problem (8) shows the social planner's costs and benefits of a higher labor supply. An increase in the labor supply of an agent with ability θ has an obvious benefit in that it increases individual's production by $\theta h_t l_t$ and an obvious cost in that it decreases the utility of the agent by V_t . In the absence of private information frictions, those would be the only effects, and one would set $\theta h_t = V_{l,t}$. With private information about skills and human capital, there are two additional considerations. First, increasing the labor supply tightens the incentive compatibility constraint for all the agents with higher skills, and requires the social planner to give them more utility. The amount of the extra utility needed is given by a change in $V_{l,t} l_t$, and the costs of doing so are determined by the cumulative distortion $X_\mu(\theta)$. With unobservable human capital, an increase in labor supply also affects the individual's incentives to accumulate human capital, and those effects are captured by the changes in the Euler Equation (9). In the optimum, all the marginal benefits are equated with all the marginal costs. A similar reasoning applies to the choice of human capital.

A computational and analytical advantage of theorem (6) is that the relaxed social planner's problem can be broken into a continuum of separate maximization problems (8), connected only through the cumulative distortion function. This is convenient since computing the cumulative distortion function may be complicated, but it is essentially a static problem. On the other hand, all the dynamics in the model is captured in the maximization problem (8), where only $X_\mu(\theta)$ needs to be known.

4 Some Examples Solved

Analytical results for the maximization problem (8) 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. The first two examples

focus on economies where human capital is never a substitute to labor supply and only the complementarity effect applies. The third example, on the other hand, studies an economy where both the substitutability and complementarity effects are in play.

All the three examples rely on the first-order approach. However, I will give relatively mild and easily verifiable sufficient conditions that guarantee its validity. In what follows, the marginal income tax rate refers to the marginal rate of substitution between consumption and output, $\tau_t(\theta) = 1 - \frac{V_{l,t}(\theta)}{\theta h_t(\theta)}$, and $X_\mu(\theta) > 0$ is an arbitrary cumulative distortion of an agent with ability θ .⁸

4.1 Example 1: Separable Utility, Full Depreciation

Start by assuming that the disutility of labor and schooling is additively separable and human capital fully depreciates each period:

Assumption 7 $V(l, s) = \frac{l^{1+\nu^{-1}}}{1+\nu^{-1}} + \frac{s^{1+\epsilon^{-1}}}{1+\epsilon^{-1}}$ and $G(h, s) = s$.

Here ν is the Frisch elasticity of labor supply, and ϵ determines how responsive schooling is to changes in taxes.⁹ The Euler equation (4) then reduces to

$$s_t(\theta, \hat{\theta})^{1+\epsilon^{-1}} = \beta \left(\frac{y_{t+1}(\hat{\theta})}{\theta s_t(\theta, \hat{\theta})} \right)^{1+\nu^{-1}}. \quad (11)$$

This equation can be easily solved for $s_t(\theta, \hat{\theta})$ as a function of $y_{t+1}(\hat{\theta})$ and of θ . Doing so and eliminating schooling from the problem yields the following equivalence result:

Lemma 8 *Suppose that Assumption (7) holds. Then the efficient consumption and output allocations are identical to the efficient consumption and output allocations in an economy*

⁸It can be shown that the efficient allocation can be implemented by an income tax that is possibly history dependent. The implementation is similar to the one used in [Werning \(2007\)](#) and [Boháček and Kapička \(2008\)](#), and is omitted.

⁹The economy has a convenient closed form solution for the first best allocations, where $l_0 = h_0^{\nu^{-1}} \theta^{\nu^{-1}}$, and for $t > 0$, $l_t = k_l \theta^{\nu^{-1} \frac{1+\epsilon^{-1}}{1-\nu^{-1}\epsilon^{-1}}}$ and $h_t = k_h \theta^{\epsilon^{-1} \frac{1+\nu^{-1}}{1-\nu^{-1}\epsilon^{-1}}}$ for some constants k_l and k_h . Those values are in general incompatible with both the promise keeping constraint (3) and the envelope condition (6).

with production functions $y_0 = \theta h_0 l_0$ and $y_t = \theta \kappa l_t$, $t \geq 1$, and with a utility function given by

$$c_0 - \frac{l_0^{1+\nu^{-1}}}{1 + \nu^{-1}} + \sum_{t=1}^T \beta^t \left(c_t - \frac{l_t^{1+\hat{\nu}^{-1}}}{1 + \hat{\nu}^{-1}} \right),$$

where $\hat{\nu} = \frac{2+\nu^{-1}+\epsilon^{-1}}{\nu^{-1}\epsilon^{-1}-1}$ and $\kappa = \beta^{\frac{1}{1+\epsilon^{-1}}}(1 + \epsilon^{-1})^{\frac{-1}{1+\hat{\nu}^{-1}}}$.

It is easy to see that $\hat{\nu} > \nu$, and so the economy with human capital acts as if the labor supply was more elastic in all the periods except for period zero. Since the labor supply elasticity is one of the key determinants of the optimal marginal income tax rate, we have the following result:

Proposition 9 *If Assumption (7) holds then the optimal marginal income tax rates are given by*

$$\begin{aligned} \frac{\tau_0(\theta)}{1 - \tau_0(\theta)} &= (1 + \nu^{-1})X_\mu(\theta) \\ \frac{\tau_t(\theta)}{1 - \tau_t(\theta)} &= (1 + \hat{\nu}^{-1})X_\mu(\theta), \quad t = 1 \dots T, \end{aligned}$$

The proof uses Lemma (8) to determine the optimal marginal income tax rates in an isomorphic economy with no human capital and a higher elasticity of labor supply from period one onwards. The optimal tax problem in the isomorphic economy is essentially a sequence of static optimal tax problems. It is optimal to set $\theta h_t l_t^{-\nu_t^{-1}} - 1 = (1 + \nu_t^{-1})X_\mu$ ¹⁰. Here ν_t is the elasticity of labor supply in period t , equal to ν in period 0 and $\hat{\nu}$ in all the other periods. Since the transformation of the marginal income tax rate $\frac{\tau_t(\theta)}{1 - \tau_t(\theta)}$ is equal to $\theta h_t l_t^{-\nu_t^{-1}} - 1$, the result follows.

Rearranging the terms, one can also express a direct relationship between the tax rate in period zero and all the future periods as

$$\frac{\tau_t(\theta)}{1 - \tau_t(\theta)} = \frac{1 + \epsilon^{-1}}{2 + \nu^{-1} + \epsilon^{-1}} \frac{\tau_0(\theta)}{1 - \tau_0(\theta)}, \quad t = 1 \dots T.$$

¹⁰See [Diamond \(1998\)](#).

The size of the decrease in the marginal tax rates is independent of the cumulative distortion, and is therefore the same for all agents. The optimal marginal tax rate schedule thus simply shifts down from period 1 on. If the elasticity of schooling is larger or if the elasticity of labor supply is smaller, then the drop in the marginal tax rates is larger. Note also that the marginal income tax rate in period zero is, conditional on the cumulative distortion, identical to the marginal income tax rate one would obtain in a static economy with no human capital.

Additional insights into the social planner's problem can be obtained by analyzing it directly without using Lemma (8). In the absence of any connection between human capital and labor supply the social planner would set $\theta h_t l_t^{-\nu^{-1}} - 1 = (1 + \nu^{-1})X$ in all periods, not just in the initial period. However, in periods $t \geq 1$ there is a complementarity between labor supply and schooling. It is optimal for the social planner to increase labor supply above what he would have chosen in the absence of human capital considerations. The reason is that individuals only take into account private benefits from schooling, and private benefits are smaller than social benefits because of a positive marginal income tax. One can show that the marginal benefit from additional one percent of schooling is $\theta h_t l_t^{-\nu^{-1}} - 1 > 0$. It follows from the Euler equation (11) that a one percent increase in labor supply in period $t \geq 1$ increases schooling by $\frac{1+\nu^{-1}}{1+\epsilon^{-1}}$ percent in period $t - 1$. Taking this into account, the social planner will optimally set $(1 + \frac{1+\nu^{-1}}{1+\epsilon^{-1}})(\theta h_t l_t^{-\nu^{-1}} - 1) = (1 + \nu^{-1})X$. Rearranging, one gets the formula in proposition (9).

The discussion in the previous paragraph indicates that unobservability of human capital is essential in obtaining the decreasing time profile of the marginal income tax rates. If human capital were observable, the social planner could separate the choice of labor supply and schooling and would not have to change labor supply to provide incentives to accumulate human capital. He would set $\theta h_t l_t^\nu - 1 = (1 + \nu^{-1})X$ in all periods. At the same time, the social planner would equalize marginal cost of schooling to social marginal benefits. It is easy to show that the social planner would set $s_t^{1+\epsilon^{-1}} = \beta[1 + (1 + \nu^{-1})X]l_{t+1}^{1+\nu^{-1}}$. This clearly violates the Euler equation (11). Schooling subsidies would therefore be needed to increase incentives to invest in human capital.

Finally, note that full commitment to the optimal tax schedule is essential in generating the results. If the social planner lacks commitment, he will, at the beginning of each period, reoptimize, and choose the marginal income tax rate according to the first period formula $\theta h_t l_t^\nu - 1 = (1 + \nu^{-1})X$. Thus, the marginal income tax rates will be constant over time.

Due to Lemma (8), the following simple set of conditions guarantees sufficiency of the first-order approach:

Lemma 10 *Suppose that Assumption (7) holds. If $y_t(\hat{\theta})$ increases in $\hat{\theta}$ for all $t = 0, \dots, T$ then $\Omega(\theta, \hat{\theta})$ increases in $\hat{\theta}$ for all $\theta \in \Theta$.*

The proof is straightforward, and is omitted.

4.2 Example 2: Separable Utility, No Depreciation

A second example involves utility function that is additively separable as in Example 1, but assumes that human capital does not depreciate. To keep the problem tractable, there are only three periods.

Assumption 11 $V(l, s) = \frac{l^{1+\nu^{-1}}}{1+\nu^{-1}} + \frac{s^{1+\epsilon^{-1}}}{1+\epsilon^{-1}}$, $G(h, s) = hs^\alpha$, and $T = 2$.

The Euler equations are now given by

$$\begin{aligned} s_0(\theta, \hat{\theta})^{1+\epsilon^{-1}} &= \beta\alpha \left(\frac{y_1(\hat{\theta})}{\theta h_1(\theta, \hat{\theta})} \right)^{1+\nu^{-1}} + \beta^2\alpha \left(\frac{y_2(\hat{\theta})}{\theta h_2(\theta, \hat{\theta})} \right)^{1+\nu^{-1}} \\ s_1(\theta, \hat{\theta})^{1+\epsilon^{-1}} &= \beta\alpha \left(\frac{y_2(\hat{\theta})}{\theta h_2(\theta, \hat{\theta})} \right)^{1+\nu^{-1}}. \end{aligned}$$

Relative to the Euler equation in Example 1, the marginal benefit of the investment in human capital in period zero is now a sum of period one and two marginal benefits. The second Euler equation is, up to a constant α , identical to the Euler equation in Example 1, because the first period investment yields benefits only in one period.

There is no closed form solution that would reduce the problem into a simpler problem as in Lemma 8. We have the following result regarding the time profile of the marginal income taxes:

Proposition 12 *If Assumption (11) holds then the optimal marginal income tax rates satisfy*

$$\tau_0(\theta) > \tau_1(\theta) > \tau_2(\theta) > 0,$$

where

$$\frac{\tau_0(\theta)}{1 - \tau_0(\theta)} = (1 + \nu^{-1})X_\mu(\theta).$$

If human capital does not depreciate then the marginal income tax rates again decrease over time, but they decrease more in the second period. As in Example 1, it is optimal to increase incentives to invest in human capital by increasing labor supply beyond what the labor supply would be in the absence of human capital. Unlike in Example 1, labor supply in period 2 is complementary with schooling in both periods, while labor supply in period 1 is only complementary with schooling in period zero. In particular, the proof shows that the marginal income tax rates in periods 1 and 2 satisfy

$$\begin{aligned} \frac{\tau_1(\theta)}{1 - \tau_1(\theta)} &= (1 + \nu^{-1})(X_\mu(\theta) - \alpha\lambda_1), \\ \frac{\tau_2(\theta)}{1 - \tau_2(\theta)} &= (1 + \nu^{-1})(X_\mu(\theta) - \alpha\lambda_1 - \alpha\lambda_2), \end{aligned}$$

where λ_1 and λ_2 are the Lagrange multipliers associated with the Euler equations. The marginal income tax rates are decreasing in λ_1 in the second period, and in both λ_1 and λ_2 in the third period. This reflects the fact that taxes in the third period affect human capital investment in both previous periods, while taxes in the second period are unrelated to human capital investment in the second period. The rest of the proof shows that both λ_1 and λ_2 are strictly positive, implying that the social planner would like to increase schooling in excess of what the agents choose in both periods. Thus, labor supply in period two should be increased more than labor supply in period one and the marginal income tax rates strictly decrease over time.

There are two additional similarities with Example 1. First, the marginal income tax rate in period 0 is, conditional on the cumulative distortion, identical. This result follows from the fact that the utility function is additively separable in labor supply and schooling. Second, one can show that the decrease in the tax rate between the second and third period is again independent of the cumulative distortion, and is given by

$$\frac{\tau_2(\theta)}{1 - \tau_2(\theta)} = \frac{1 + \epsilon^{-1}}{1 + \epsilon^{-1} + \alpha(1 + \nu^{-1})} \frac{\tau_1(\theta)}{1 - \tau_1(\theta)}.$$

The similarity follows from the fact that the investment in human capital in the second period is used only in one period and thus behaves in a very similar way to the investment in human capital in Example 1. Thus, the incentives to decrease taxes are very similar as well.

Monotonicity of income is again sufficient for an allocation to be incentive compatible:

Lemma 13 *Suppose that Assumption (11) holds. If $y_0(\hat{\theta})$, $y_1(\hat{\theta})$ and $y_2(\hat{\theta})$ increase in $\hat{\theta}$ then $\Omega(\theta, \hat{\theta})$ increases in $\hat{\theta}$ for all $\theta \in \Theta$.*

4.3 Example 3: Non-separable 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 and there are only three periods.

Assumption 14 $V(l, s) = \frac{(l+s)^{1+\nu^{-1}}}{1+\nu^{-1}}$, $G(h, s) = s$, and $T = 2$.

Note that ν is not equal to the elasticity of labor supply any more. The elasticity of labor supply is now endogenous, and is given by $\nu(1 + \frac{s_t}{l_t}) > \nu$ in period t . The Euler equations

are now given by

$$\begin{aligned} \left(\frac{y_0(\hat{\theta})}{\theta h_0} + s_0(\theta, \hat{\theta}) \right)^{\nu^{-1}} &= \beta \left(\frac{y_1(\hat{\theta})}{\theta s_0(\theta, \hat{\theta})} + s_1(\theta, \hat{\theta}) \right)^{\nu^{-1}} \frac{y_1(\hat{\theta})}{\theta s_0(\theta, \hat{\theta})^2} \\ \left(\frac{y_1(\hat{\theta})}{\theta s_0(\theta, \hat{\theta})} + s_1(\theta, \hat{\theta}) \right)^{\nu^{-1}} &= \beta \left(\frac{y_2(\hat{\theta})}{\theta s_1(\theta, \hat{\theta})} \right)^{\nu^{-1}} \frac{y_2(\hat{\theta})}{\theta s_1(\theta, \hat{\theta})^2}. \end{aligned}$$

Unlike in the previous two examples, y_0 now affects the choice of schooling in period zero. It is easy to see that l_0 is a substitute to s_0 in a sense that higher l_0 decreases the optimal s_0 . There is a similar substitutability between l_1 and s_1 . The marginal income tax rates are characterized as follows:

Proposition 15 *If Assumption (14) holds then the optimal marginal tax rates satisfy*

$$\tau_0(\theta) > \tau_1(\theta) > \tau_2(\theta) > 0,$$

where

$$\frac{\tau_0(\theta)}{1 - \tau_0(\theta)} < (1 + \nu^{-1})X_\mu(\theta).$$

The strictly decreasing pattern of the marginal income taxes is a result of two factors, changes in the substitutability and complementarity patterns between labor supply and schooling over time, and changes in the labor supply elasticity over time. In particular, period zero labor supply is a substitute for period zero schooling, period two labor supply is a complement to period one schooling, and period one labor supply is a substitute to period one schooling, and a complement to period zero schooling. As in the previous examples, individuals choose less than the socially optimal amount of schooling. The proof shows that the marginal income

tax rates satisfy

$$\begin{aligned}\frac{\tau_0(\theta)}{1 - \tau_0(\theta)} &= (1 + \nu_0^{-1})X_\mu + \lambda_1\nu_0^{-1}\frac{s_0}{l_0} \\ \frac{\tau_1(\theta)}{1 - \tau_1(\theta)} &= (1 + \nu_1^{-1})(X_\mu - \lambda_1) + \lambda_2\nu_1^{-1}\frac{s_1}{l_1} \\ \frac{\tau_2(\theta)}{1 - \tau_2(\theta)} &= (1 + \nu_2^{-1})(X_\mu - \lambda_2),\end{aligned}$$

where λ_1 and λ_2 are again Lagrange multipliers on the Euler equations, and ν_t is the elasticity of labor supply in period t . The Lagrange multipliers are shown to be strictly positive, implying again that individuals invest less than the socially optimal amount. Relative to what the social planner would choose in the absence of human capital considerations (conditional on the labor supply elasticity), taxes in period two are therefore lower, because of the complementarity of period three labor supply with period two schooling. Similarly, taxes in period 1 are higher, because of substitutability between period 1 labor supply and schooling. There are two opposing effects on the marginal income taxes in period 2, because both substitutability and complementarity effects are in play. The changes in labor supply elasticity work in the opposite direction, since labor supply elasticity decreases over time. The proof shows that this effect is not strong enough to revert the decreasing pattern of taxes over time. However, this example shows that there is a potential channel through which the decreasing pattern of the marginal income taxes can be reversed.

The second result shows that the first period marginal income tax is lower than what it would have been in the absence of human capital consideration (conditional on X_μ). This might be somewhat surprising, given that period zero labor supply is a substitute to period zero schooling. The result is due to the fact that the elasticity of labor supply in period zero is higher than ν . This effect decreases the marginal income tax rate. For CARA utility function, this effect turns out to be stronger than the substitutability effect.¹¹

To prove that the sufficiency condition (5) holds, one needs the monotonicity of income, but also restrictions on the parameter values:

¹¹However, the substitutability effect can be shown to dominate with CARA utility. Human capital considerations can therefore move the marginal income tax rates in the first period in both directions.

Lemma 16 *Suppose that Assumption (14) holds and that $2\nu + 4\nu^2 \geq 1$. If $y_0(\hat{\theta})$, $y_1(\hat{\theta})$ and $y_2(\hat{\theta})$ increase in $\hat{\theta}$ then $\Omega(\theta, \hat{\theta})$ increases in $\hat{\theta}$ for all $\theta \in \Theta$.*

Note that the parameter restriction is only sufficient, but not necessary.

Previous examples show several principles that most likely drive the intertemporal profile of marginal income taxes in a more general 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 third example, marginal income taxes will tend to decrease with time. Second, the intertemporal profile will depend on how fast human capital depreciate over time. Lower depreciation rate will tend to decrease the marginal income tax rates in more distant periods more, because it provides stronger incentives to accumulate human capital in all the previous periods. Third, the intertemporal profile will depend on how the elasticity of labor supply changes over time. As shown in the third example, the elasticity of labor supply will tend to fall, and so this effect will go in the opposite direction.

One can, however, determine the magnitude of those effects only through numerical simulations.¹²

Finally, note that in all three 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 three examples schooling is inefficiently low relative to labor supply, because people do not take into account all social benefits from schooling.

5 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

¹² 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.

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 β , the human capital production function parameters α and δ , 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 ν is closely linked to the Frisch elasticity of labor supply. Indeed, in a static model with no human capital, the elasticity of labor supply would be equal to ν . 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.¹³

To obtain a fraction of schooling attributed to reported hours worked I proceed as follows. According to [Francise 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 0.3.¹⁴ 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

¹³[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.

¹⁴ See [Pencavel \(1986\)](#), Table 1.22. For females, the labor supply elasticity is typically higher.

67% of schooling takes place on the job and is therefore included in the measure of labor supply. The resulting value of ν 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 α range from about 0.5 to over 0.9. The evidence regarding δ ranges from 0.0016 to 0.089.¹⁵ 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, in line with [Saez \(2001\)](#), 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 θ - 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.¹⁶ The consumption tax rate τ_c is set at 5.2%, following [Mendoza, Razin, and Tesar \(1994\)](#). The payroll tax rate τ_{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 κ_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 $\kappa_2 = 0.000226$.

¹⁵ See [Browning, Hansen, and Heckman \(1999\)](#), Table 2.3 and 2.4, [Huggett, Ventura, and Yaron \(2009\)](#) and [Trostel \(1993\)](#).

¹⁶See [Castañeda, Díaz-Giménez, and Ríos-Rull \(1999\)](#), [Conesa and Krueger \(2006\)](#), or [Conesa, Kitao, and Krueger \(2009\)](#).

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, 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.

5.1 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 (3), where the social planner needs to deliver a lifetime utility of at least \underline{u} , taking the initial distribution of human capital as given.¹⁷ 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 decline.¹⁸

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

¹⁷Calibrating 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.

¹⁸ 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.

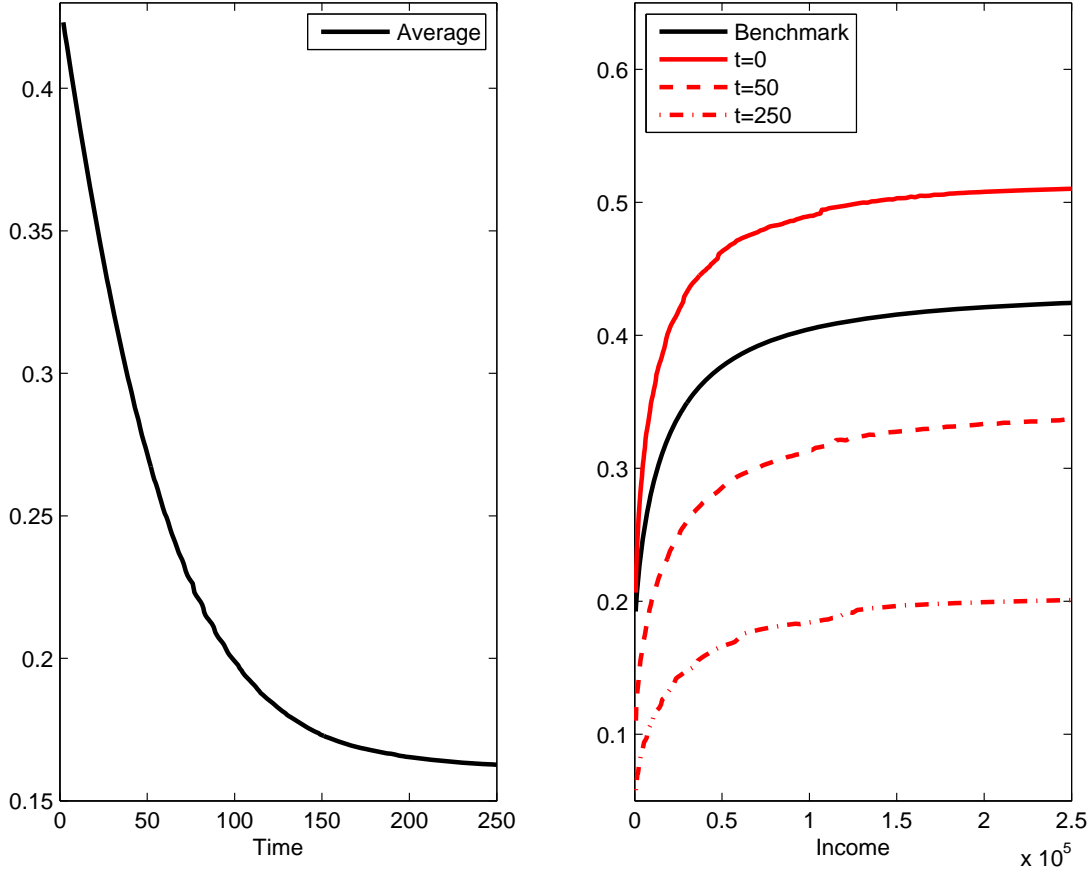


Figure 1: Marginal Income Tax Rates, Benchmark Reform

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.

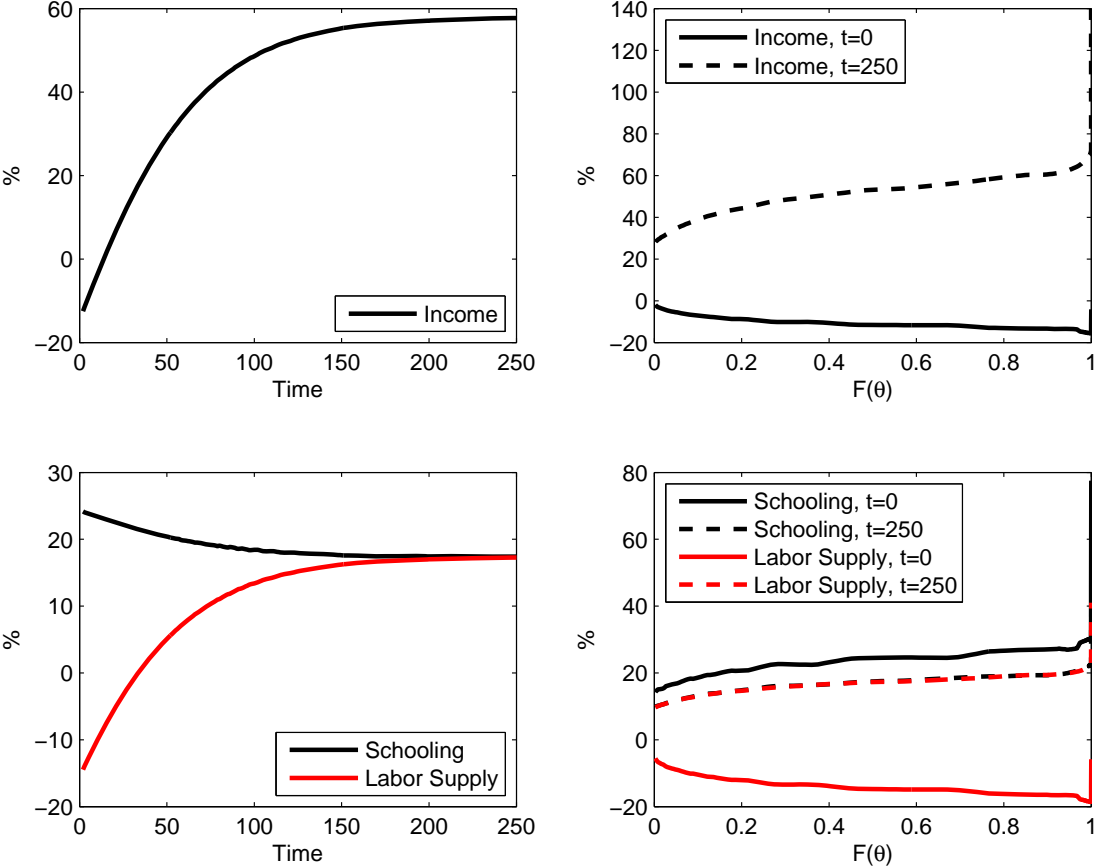


Figure 2: Income, Schooling and Labor Supply, Change from Pre-Reform

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 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.

5.2 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 (6) and (9). 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.

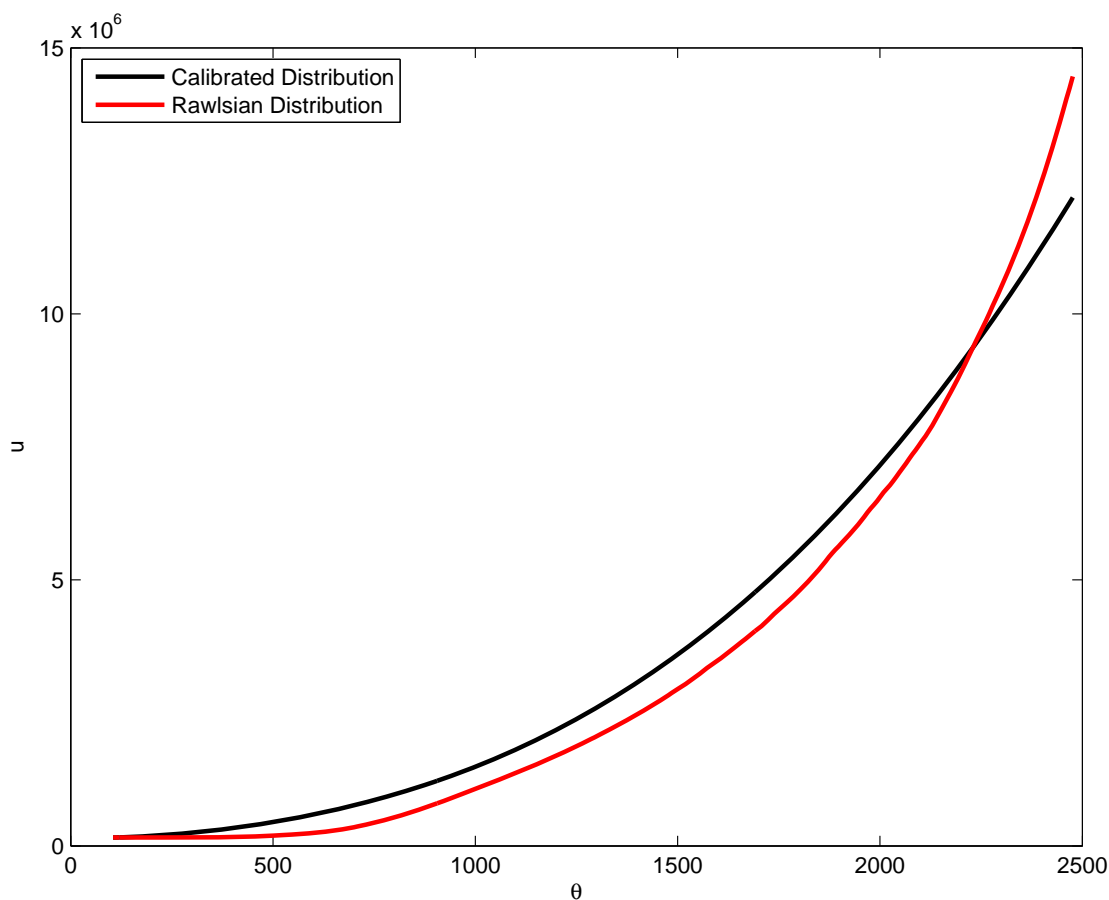


Figure 3: Utility Distribution, Rawlsian Reform

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 $X = \frac{1-F}{\theta f}$.¹⁹ 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.

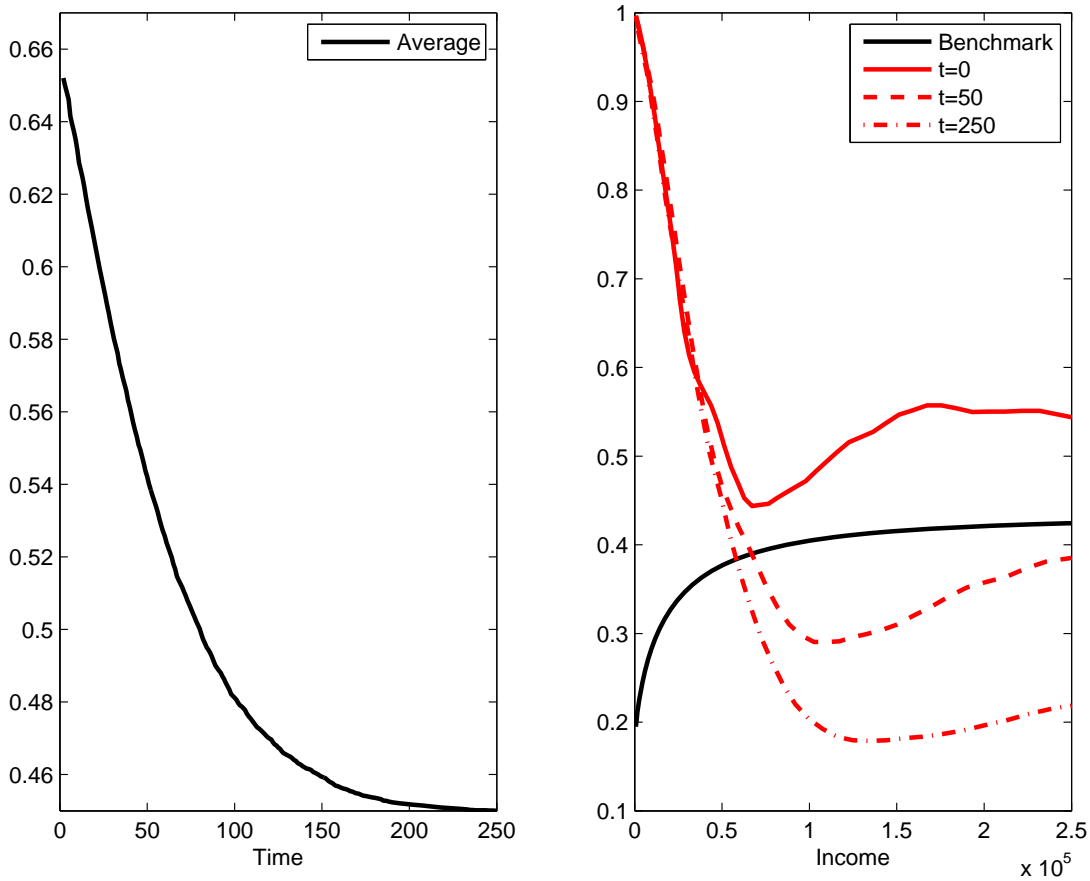


Figure 4: Marginal Income Tax Rates, Rawlsian Reform

The Rawlsian reform collects 2.155 times more resources than the benchmark tax system.

¹⁹ See [Saez \(2001\)](#) for the optimal marginal income tax schedule in a static setting, which has a similar shape.

That is, it provides savings equal to 115.5% of the present value of resources, or to \$10471 per person and year. The savings are thus 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. Since the social planner does not care about anyone beyond the least able agent, this is simply the maximum amount of resources he is able to extract from the agents. 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.

5.3 Alternative Assumptions

This section investigates the intertemporal profile of the marginal income taxes under alternative assumptions about the utility function parameter η and the human capital production function parameters δ and α . 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 η is set to a low and high value of 0.20 and 0.39, and when α 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 η corresponds to higher labor supply elasticity, and higher α 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.

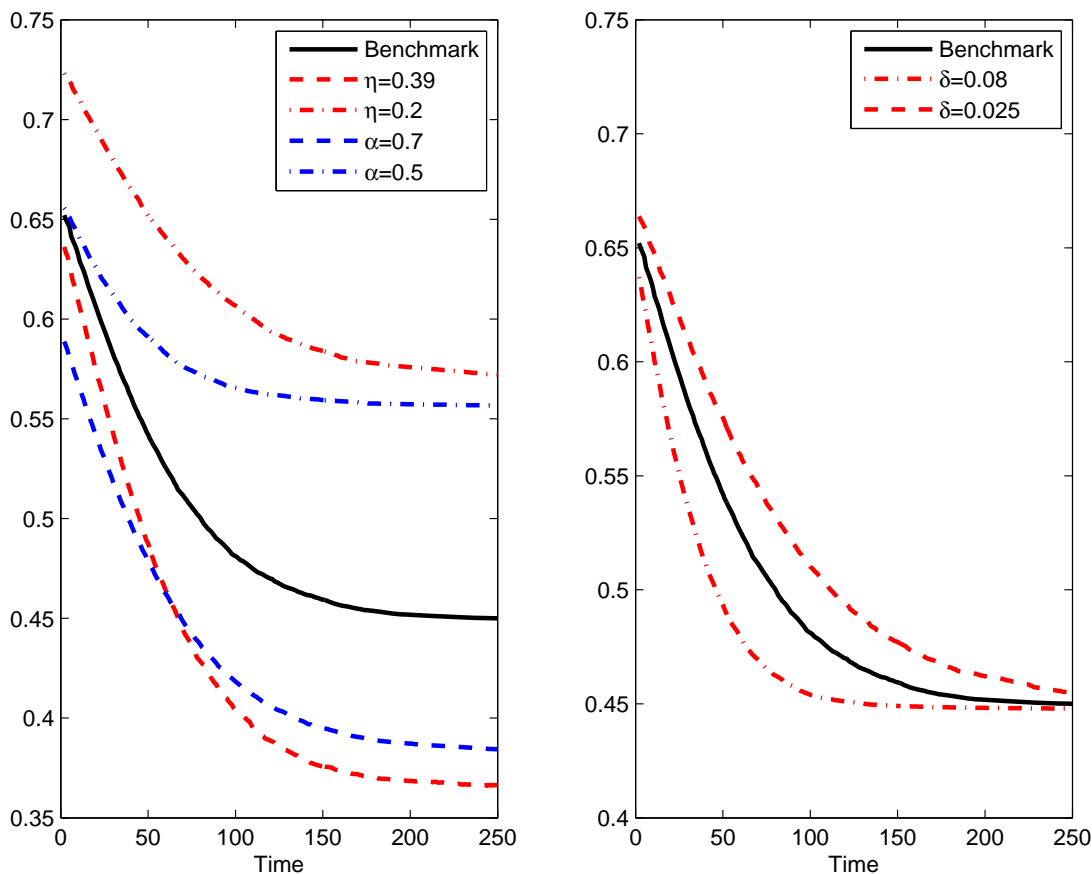


Figure 5: Average Marginal Income Taxes, Rawlsian Reform

6 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, it is no longer enough to determine

the present value of consumption (or, equivalently, a lifetime utility). The maximization problem (8) will also involve a choice of individual's consumption. Thus, the dimension of the problem increases and the problem becomes significantly harder to compute. Lemma (8) clearly applies for a more general utility and, consequently, the results of Example 1 can be extended to an economy with income effects.

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.

Finally, the assumption that the initial human capital is identical for everyone might be relaxed, but at a cost of adding additional complexity or undesirable features to the problem. In particular, if the initial human capital is observable but differs across people, the efficient allocations are likely to be such that people with higher initial human capital would pay higher taxes and would be worse off than people with lower initial human capital. It would be hard to give a reasonable interpretation to such policies. If, on the other hand, the initial human capital was a private information imperfectly correlated with abilities (which is the empirically relevant case, see [Huggett, Ventura, and Yaron \(2006\)](#)), an additional set of incentive constraints would be needed, making the problem much more complex. Having two simultaneous sources of heterogeneity, perhaps in a life-cycle framework, is certainly an interesting modification, and is left for future research.

References

- Albanesi, S. (2007). Optimal taxation of entrepreneurial capital with private information. Working paper, Columbia University. [5](#)
- Albanesi, S. and C. Sleet (2006). Dynamic optimal taxation with private information. *The 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. [26](#)
- Boháček, R. and M. Kapička (2008). Optimal human capital policies. *Journal of Monetary Economics* 55, 1–16. [5](#), [11](#), [15](#), [26](#)
- 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. [25](#)
- 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. [25](#)
- Conesa, J. C., S. Kitao, and D. Krueger (2009). Taxing capital? not a bad idea after all. *American Economic Review* 99(1), 25–48. [25](#)
- Conesa, J. C. and D. Krueger (2006). On the optimal progressivity of the tax code. *Journal of Monetary Economics* 53(7), 1425–1450. [25](#)
- Diamond, P. and J. Mirrlees (2002). Optimal taxation and the le chatelier principle. Working paper, MIT. [5](#)
- Diamond, P. A. (1998). Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review* 88, 83–95. [6](#), [16](#)
- Farhi, E. and I. Werning (2005). Inequality and social discounting. *Journal of Political Economy* 115(1), 365–402. [4](#)
- Farhi, E. and I. Werning (2010). Insurance and taxation over the life-cycle. Working paper, MIT. [4](#)
- Francise, N. and V. Ramey (2009). A Century of Work and Leisure. *American Economic Review: Macroeconomics* 1, 189–224. [24](#)
- Golosov, M., N. R. Kocherlakota, and A. Tsyvinski (2003). Optimal indirect and capital taxation. *The Review of Economic Studies* 70, 569–587. [4](#)

- Golosov, M., A. Tsyvinski, and M. Troshkin (2010). Optimal dynamic taxes. Working paper, Yale University. [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. [25](#)
- Grochulski, B. and T. Piskorski (2005). Optimal wealth taxes with risky human capital. Working paper, Federal Reserve Bank of Richmond. [5](#)
- Huggett, M., G. Ventura, and A. Yaron (2006). Human capital and earnings distribution dynamics. *Journal of Monetary Economics* 53, 265–290. [33](#)
- Huggett, M., G. Ventura, and A. Yaron (2009). Sources of lifetime inequality. Working paper, Georgetown University. [25](#)
- Imai, S. and M. P. Keane (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review* 45(2), 601–641. [24](#)
- Kapička, M. (2006). Optimal taxation and human capital accumulation. *Review of Economic Dynamics* 9, 612–639. [5](#), [11](#), [23](#)
- Keane, M. and K. Wolpin (1997). The career decisions of young men. *Journal of Political Economy* 105, 473–522. [5](#)
- Kocherlakota, N. R. (2005). Zero expected wealth taxes: A Mirrless approach to dynamic optimal taxation. *Econometrica* 73, 1587–1621. [4](#), [12](#)
- 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. [25](#)
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies* 38, 175–208. [4](#)
- Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics* 6, 327–358. [4](#)
- 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*. Elsevier. [4](#),

- Pavan, A., I. Segal, and Y. Toikka (2009). Dynamic mechanism design: Revenue equivalence, profit maximization, and information disclosure. Working paper, Stanford University. 12
- 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. 24
- Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies* 68, 205–229. 6, 25, 30
- Shourideh, A. (2010). Optimal taxation of entrepreneurial income: A mirrleesian approach to capital accumulation. Working paper, University of Minnesota. 5, 12
- Storesletten, K., C. Telmer, and A. Yaron (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics* 51, 609–633. 5
- Trostel, P. A. (1993). The effect of taxation on human capital. *Journal of Political Economy* 101, 327–350. 25
- Werning, I. (2007). Optimal fiscal policy with redistribution. *Quarterly Journal of Economics* 22(3), 925–967. 4, 5, 15

7 Appendix 1: Proofs

Proof of Proposition (5). First note that, by the fundamental theorem of calculus,

$$\frac{\partial}{\partial \theta} W_\sigma(h_0, \theta; \hat{\theta}) = \Omega(\theta, \hat{\theta}) - \Omega(\hat{\theta}, \hat{\theta}). \quad (12)$$

We then have

$$\begin{aligned} W_\sigma(h_0, \theta; \theta) - W_\sigma(h_0, \hat{\theta}; \hat{\theta}) &= u(\theta) - u(\hat{\theta}) = \int_{\hat{\theta}}^{\theta} \Psi(\varepsilon, \varepsilon) \frac{d\varepsilon}{\varepsilon} + u(\hat{\theta}) \\ &\geq \int_{\hat{\theta}}^{\theta} \Psi(\varepsilon, \hat{\theta}) \frac{d\varepsilon}{\varepsilon} + u(\hat{\theta}) = W_\sigma(h_0, \theta; \hat{\theta}) - W_\sigma(h_0, \hat{\theta}; \hat{\theta}). \end{aligned}$$

The first equality follows from the definition of u , the second one from the envelope condition (6), the third one from Proposition (5), and the fourth one from the (12). Thus, (2) holds. By Lemma (3), \hat{h} maximizes $W_\sigma(h_0, \theta; \hat{\theta})$, and by definition, $\hat{h}(\theta, \theta) = h(\theta)$. Hence the allocation is incentive compatible. ■

Proof of Theorem 6. Define Δ to be the set of all labor supply and human capital sequences (l, h') that satisfy the Euler equation (9) for all $t \geq 0$.²⁰

Necessity of the conditions is obvious. For sufficiency, let $\sigma^* = (u^*, l^*, h'^*)$, together with μ^* 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 (6) and the promise keeping constraint (3). 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 the fact that the envelope condition holds with equality for (u^*, l^*, h'^*) . The third equality follows from integration by parts.

It follows from (10) 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 (6) and (9). 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 $u(\theta) \geq \underline{U}(\theta)$.

²⁰Writing $(l, h') \in \Delta$ means that this property holds for all $\theta \in \Theta$.

Since (l^*, h^*) solves the problem by (8) and u^* maximizes $\xi_{\mu^*, u}$,

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

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 Lemma (8). Solving for $s_t(\theta, \hat{\theta})$ from the Euler equation that

$$s_t(\theta, \hat{\theta}) = \beta^{\frac{1}{2+\nu^{-1}+\epsilon^{-1}}} \left(\frac{y_{t+1}(\hat{\theta})}{\theta} \right)^{\frac{1+\nu^{-1}}{2+\nu^{-1}+\epsilon^{-1}}}.$$

Substituting back into the utility function, one gets that for any $t \geq 0$,

$$\begin{aligned}\frac{s_t(\theta, \hat{\theta})^{1+\epsilon^{-1}}}{1+\epsilon^{-1}} + \beta \frac{\left(\frac{y_{t+1}(\hat{\theta})}{\theta s_t(\theta, \hat{\theta})} \right)^{1+\nu^{-1}}}{1+\nu^{-1}} &= \left(1 + \frac{1+\epsilon^{-1}}{1+\nu^{-1}} \right) \frac{s_t(\theta, \hat{\theta})^{1+\epsilon^{-1}}}{1+\epsilon^{-1}} \\ &= \left(1 + \frac{1+\epsilon^{-1}}{1+\nu^{-1}} \right) \beta^{\frac{1+\epsilon^{-1}}{2+\nu^{-1}+\epsilon^{-1}}} \left(\frac{y_{t+1}(\hat{\theta})}{\theta} \right)^{\frac{(1+\nu^{-1})(1+\epsilon^{-1})}{2+\nu^{-1}+\epsilon^{-1}}} \\ &= \beta \frac{1}{1+\hat{\nu}^{-1}} \left(\frac{y_{t+1}(\hat{\theta})}{\theta \kappa} \right)^{1+\hat{\nu}^{-1}}.\end{aligned}$$

The preferences over consumption, schooling and labor supply in the economy with human capital can thus be expressed as indirect preferences over only consumption and labor supply l_t with technology given by $y_t = \kappa \theta l_t$ for $t = 1, \dots, T$. The preferences over c_0 , l_0 and technology in period zero are unchanged. It follows that the efficient allocations will also be identical in both economies. ■

Proof of Proposition 9. Lemma (8) implies that the social planner's problem can be

written as

$$\max_l \theta h_0 l_0 - \frac{l_0^{1+\nu^{-1}}}{1+\nu^{-1}} - l_0^{1+\nu^{-1}} X + \sum_{t=1}^T \beta^t \left\{ \theta \kappa l_t - \frac{l_t^{1+\hat{\nu}^{-1}}}{1+\hat{\nu}^{-1}} - l_t^{1+\hat{\nu}^{-1}} X \right\}.$$

The first order conditions in l_t are

$$\begin{aligned} \theta h_0 - l_0^{\nu^{-1}} - (1+\nu^{-1})l_0^{\nu^{-1}} X &= 0 \\ \theta h_t - l_t^{\nu^{-1}} - (1+\hat{\nu}^{-1})l_t^{\nu^{-1}} X &= 0, \quad t = 1, \dots, T. \end{aligned}$$

Rearranging and using the definition of τ_t , the result follows. ■

Proof of Proposition 12. Let λ_1 and λ_2 be the Lagrange multiplier on the Euler equations. The Lagrangean is

$$\begin{aligned} \max_{l_0, l_1, l_2, s_0, s_1} \theta h_0 l_0 - \frac{l_0^{1+\nu^{-1}}}{1+\nu^{-1}} - \frac{s_0^{1+\epsilon^{-1}}}{1+\epsilon^{-1}} - l_0^{1+\nu^{-1}} X_\mu + \beta \left[\theta h_0 s_0^\alpha l_1 - \frac{l_1^{1+\nu^{-1}}}{1+\nu^{-1}} - \frac{s_1^{1+\epsilon^{-1}}}{1+\epsilon^{-1}} - l_1^{1+\nu^{-1}} X_\mu \right] \\ + \beta^2 \left[\theta h_0 (s_0 s_1)^\alpha l_2 - \frac{l_2^{1+\nu^{-1}}}{1+\nu^{-1}} - l_2^{1+\nu^{-1}} X_\mu \right] \\ + \lambda_1 \left[\beta \alpha l_1^{1+\nu^{-1}} + \beta^2 \alpha l_2^{1+\nu^{-1}} - s_0^{1+\epsilon^{-1}} \right] + \beta \lambda_2 \left[\beta \alpha l_2^{1+\nu^{-1}} - s_1^{1+\epsilon^{-1}} \right]. \end{aligned}$$

The first order conditions in l_0, l_1 and l_2 are

$$\theta h_0 l_0^{-\nu^{-1}} - 1 = (1+\nu^{-1})X_\mu \tag{13}$$

$$\theta h_1 l_1^{-\nu^{-1}} - 1 = (1+\nu^{-1})(X_\mu - \alpha \lambda_1) \tag{14}$$

$$\theta h_2 l_2^{-\nu^{-1}} - 1 = (1+\nu^{-1})(X_\mu - \alpha \lambda_1 - \alpha \lambda_2) \tag{15}$$

The first order conditions in s_0, s_1 are:

$$[\lambda_2(1+\epsilon^{-1}) + 1]s_1^{\epsilon^{-1}} = \beta \alpha \theta h_0 s_0^\alpha s_1^{\alpha-1} l_2 \tag{16}$$

$$[\lambda_1(1+\epsilon^{-1}) + 1]s_0^{\epsilon^{-1}} = \beta \alpha \theta h_0 s_0^{\alpha-1} l_1 + \beta^2 \alpha \theta h_0 s_0^{\alpha-1} s_1^\alpha l_2. \tag{17}$$

The second part of the proposition follows directly from (13). To show the first part of the proposition, one needs to prove that $\lambda_1 > 0$, $\lambda_2 > 0$, and that $X_\mu - \alpha\lambda_1 - \alpha\lambda_2 > 0$. To show this, use the Euler equation in s_1 and the first-order condition (15) to simplify (16). One gets

$$\lambda_2 = \frac{1 + \nu^{-1}}{1 + \epsilon^{-1} + \alpha(1 + \nu^{-1})}(X_\mu - \alpha\lambda_1). \quad (18)$$

Eliminating λ_2 from equation (16), using the Euler equation in s_0 and the first-order conditions (14) and (17) one obtains

$$\lambda_1 = \frac{(1 + \nu^{-1})l_1^{1+\nu^{-1}} + \beta(1 + \hat{\nu}^{-1})l_2^{1+\nu^{-1}}}{[1 + \epsilon^{-1} + \alpha(1 + \nu^{-1})]l_1^{1+\nu^{-1}} + \beta[1 + \epsilon^{-1} + \alpha(1 + \hat{\nu}^{-1})]l_2^{1+\nu^{-1}}}X_\mu > 0,$$

where $1 + \hat{\nu}^{-1} = \frac{(1+\nu^{-1})(1+\epsilon^{-1})}{(1+\epsilon^{-1})+\alpha(1+\nu^{-1})}$. Rearranging, one also gets that $X_\mu - \alpha\lambda_1 > 0$. In turn, equation (18) implies that $\lambda_2 > 0$. Finally,

$$\begin{aligned} X_\mu - \alpha\lambda_1 - \alpha\lambda_2 &= X_\mu - \alpha\lambda_1 - \alpha \frac{1 + \nu^{-1}}{1 + \epsilon^{-1} + \alpha(1 + \nu^{-1})}(X_\mu - \alpha\lambda_1) \\ &= \frac{1 + \epsilon^{-1}}{1 + \epsilon^{-1} + \alpha(1 + \nu^{-1})}(X - \alpha\lambda_1), \end{aligned}$$

and so $X_\mu - \alpha\lambda_1 - \alpha\lambda_2 > 0$. ■

Proof of Lemma 13. By using the Euler equation in s_0 , one can write

$$\begin{aligned} \Omega(\theta, \hat{\theta}) &= \left(\frac{y_0(\hat{\theta})}{\theta h_0} \right)^{1+\nu^{-1}} + \beta \left(\frac{y_1(\hat{\theta})}{\theta h_1(\theta, \hat{\theta})} \right)^{1+\nu^{-1}} + \beta^2 \left(\frac{y_2(\hat{\theta})}{\theta h_2(\theta, \hat{\theta})} \right)^{1+\nu^{-1}} \\ &= \left(\frac{y_0(\hat{\theta})}{\theta h_0} \right)^{1+\nu^{-1}} + \frac{1}{\alpha} s_0(\theta, \hat{\theta})^{1+\epsilon^{-1}}. \end{aligned}$$

Thus, $\Omega(\theta, \hat{\theta})$ is increasing in $\hat{\theta}$ if both $y_0(\hat{\theta})$ and $s_0(\theta, \hat{\theta})$ are increasing in $\hat{\theta}$. This is in turn true if $y_0(\hat{\theta})$, $y_1(\hat{\theta})$ and $y_2(\hat{\theta})$ are all increasing in $\hat{\theta}$ and if $\tilde{s}_0(\theta, y_1, y_2)$ is increasing in y_1 and

y_2 , where $\tilde{s}_0(\theta, y_1, y_2)$ is defined by

$$\tilde{s}_0 \left(\theta, y_1(\hat{\theta}), y_2(\hat{\theta}) \right) = s_0(\theta, \hat{\theta}).$$

The rest of the proof shows that $\tilde{s}_0(\theta, y_1, y_2)$ is indeed increasing in y_1 and y_2 . It follows from the combination of both Euler equations that \tilde{s}_0 satisfies

$$\tilde{s}_0^{1+\epsilon^{-1}+\alpha(1+\nu^{-1})} = \beta\alpha \left(\frac{y_1}{\theta h_0} \right)^{1+\nu^{-1}} + \beta^2\alpha\kappa \left(\frac{y_2}{\theta h_0} \right)^{\frac{(1+\nu^{-1})(1+\epsilon^{-1})}{1+\epsilon^{-1}+\alpha(1+\nu^{-1})}} \tilde{s}_0^{\frac{\alpha^2(1+\nu^{-1})^2}{1+\epsilon^{-1}+\alpha(1+\nu^{-1})}},$$

where $\kappa = (\alpha\beta)^{-\frac{\alpha(1+\nu^{-1})}{1+\epsilon^{-1}+\alpha(1+\nu^{-1})}}$. By differentiating in y_1 one gets

$$\frac{\partial \tilde{s}_0}{\partial y_1} \frac{y_1}{\tilde{s}_0} = \frac{[1 + \epsilon^{-1} + \alpha(1 + \nu^{-1})] (1 + \nu^{-1}) \beta \alpha \left(\frac{y_1}{\theta h_0} \right)^{1+\nu^{-1}}}{\zeta \tilde{s}_0^{1+\epsilon^{-1}+\alpha(1+\nu^{-1})} + \alpha^2 (1 + \nu^{-1})^2 \beta \alpha \left(\frac{y_1}{\theta h_0} \right)^{1+\nu^{-1}}} > 0.$$

where $\zeta = (1 + \epsilon^{-1})^2 + 2\alpha(1 + \epsilon^{-1})(1 + \nu^{-1})$. Similarly, one gets

$$\frac{\partial \tilde{s}_0}{\partial y_2} \frac{y_2}{\tilde{s}_0} = \frac{(1 + \epsilon^{-1})(1 + \nu^{-1}) \beta^2 \alpha \kappa \left(\frac{y_2}{\theta h_0} \right)^{\frac{(1+\nu^{-1})(1+\epsilon^{-1})}{1+\epsilon^{-1}+\alpha(1+\nu^{-1})}}}{\zeta \tilde{s}_0^{1+\epsilon^{-1}+\alpha(1+\nu^{-1})} + \alpha^2 (1 + \nu^{-1})^2 \beta \alpha \left(\frac{y_1}{\theta h_0} \right)^{1+\nu^{-1}}} > 0.$$

■

Proof of Proposition 15. Let λ_1 and λ_2 be the Lagrange multipliers on the Euler

equations. The Lagrangean is

$$\begin{aligned} \max_{l,h} & \theta h_0 l_0 - \frac{(l_0 + s_0)^{1+\nu^{-1}}}{1 + \nu^{-1}} - (l_0 + s_0)^{\nu^{-1}} l_0 X_\mu \\ & + \beta \left[\theta s_0 l_1 - \frac{(l_1 + s_1)^{1+\nu^{-1}}}{1 + \nu^{-1}} - (l_1 + s_1)^{\nu^{-1}} l_1 X_\mu \right] \\ & + \beta^2 \left[\theta s_1 l_2 - \frac{l_2^{1+\nu^{-1}}}{1 + \nu^{-1}} - l_2^{1+\nu^{-1}} X_\mu \right] \\ & + \lambda_1 \left[\beta (l_1 + s_1)^{\nu^{-1}} l_1 - (l_0 + s_0)^{\nu^{-1}} s_0 \right] + \beta \lambda_2 \left[\beta l_2^{1+\nu^{-1}} - (l_1 + s_1)^{\nu^{-1}} s_1 \right]. \end{aligned}$$

Let $\vartheta_t = \theta h_t (l_t + s_t)^{-\nu^{-1}} - 1$. The first order conditions in l_0, l_1, l_2 are

$$\vartheta_0 = \left(1 + \nu^{-1} \frac{l_0}{l_0 + s_0}\right) X_\mu + \lambda_1 \nu^{-1} \frac{s_0}{l_0 + s_0} \quad (19)$$

$$\vartheta_1 = \left(1 + \nu^{-1} \frac{l_1}{l_1 + s_1}\right) (X_\mu - \lambda_1) + \lambda_2 \nu^{-1} \frac{s_1}{l_1 + s_1} \quad (20)$$

$$\vartheta_2 = (1 + \nu^{-1}) (X_\mu - \lambda_2) \quad (21)$$

The first order conditions in s_0 and s_1 combined with the Euler equations imply

$$\vartheta_1 = \nu^{-1} \frac{l_0}{l_0 + s_0} X_\mu + \lambda_1 \left(1 + \nu^{-1} \frac{s_0}{l_0 + s_0}\right) \quad (22)$$

$$\vartheta_2 = \nu^{-1} \frac{l_1}{l_1 + s_1} (X_\mu - \lambda_1) + \lambda_2 \left(1 + \nu^{-1} \frac{l_1}{l_1 + s_1}\right). \quad (23)$$

As a preliminary result, I first show that λ_2 and $X_\mu - \lambda_1$ are both positive. To show this, eliminate ϑ_1 by using (20) and (22) and ϑ_2 by using (21) and (23) and sum both equations to get: Rearranging, one obtains

$$X_\mu - \lambda_1 = \frac{2 + \nu^{-1}}{2 + \nu^{-1} \frac{s_0}{l_0 + s_0}} \lambda_2. \quad (24)$$

Substituting into (20) and (22) and rearranging, one gets that $\lambda_2 > 0$. Equation (24) then implies that $X_\mu - \lambda_1 > 0$.

To show the first part of the proposition, note that by combining the first order conditions

one obtains $\vartheta_1 = \vartheta_0 - (X_\mu - \lambda_1)$ and $\vartheta_2 = \vartheta_1 - (X_\mu - \lambda_1 - \lambda_2)$. Since $X_\mu - \lambda_1 > X_\mu - \lambda_1 - \lambda_2$, the first part of the proposition will be proven if $X_\mu - \lambda_1 - \lambda_2 > 0$. Rearranging (24) one gets

$$X_\mu - \lambda_1 - \lambda_2 = \frac{\nu^{-1} \frac{l_0}{l_0 + s_0}}{2 + \nu^{-1} \frac{s_0}{l_0 + s_0}} \lambda_2 > 0,$$

which completes the proof of the first part of the Proposition. The second part of the Proposition follows from rewriting (19) as

$$\vartheta_0 = (1 + \nu^{-1})X_\mu - \nu^{-1} \frac{s_0}{l_0 + s_0} (X_\mu - \lambda_1) < (1 + \nu^{-1})X_\mu.$$

■

Proof. By using the Euler equation, one can write

$$\begin{aligned} \Omega(\theta, \hat{\theta}) &= \left(\frac{y_0(\hat{\theta})}{\theta h_0} + s_0(\theta, \hat{\theta}) \right)^{\nu^{-1}} \frac{y_0(\hat{\theta})}{\theta h_0} + \beta \left(\frac{y_1(\hat{\theta})}{\theta s_0(\theta, \hat{\theta})} + s_1(\theta, \hat{\theta}) \right)^{\nu^{-1}} \frac{y_1(\hat{\theta})}{\theta s_0(\theta, \hat{\theta})} \\ &\quad + \beta^2 \left(\frac{y_2(\hat{\theta})}{\theta s_1(\theta, \hat{\theta})} \right)^{1+\nu^{-1}} \\ &= \left(\frac{y_0(\hat{\theta})}{\theta h_0} + s_0(\theta, \hat{\theta}) \right)^{1+\nu^{-1}} + \beta^2 \left(\frac{y_2(\hat{\theta})}{\theta s_1(\theta, \hat{\theta})} \right)^{1+\nu^{-1}} \end{aligned}$$

Define $\tilde{s}_0(\theta, y_0, y_1, y_2)$ and $\tilde{s}_1(\theta, y_0, y_1, y_2)$ by $\tilde{s}_0(\theta, y_0(\hat{\theta}), y_1(\hat{\theta}), y_2(\hat{\theta})) = s_0(\theta, \hat{\theta})$ and $\tilde{s}_1(\theta, y_0(\hat{\theta}), y_1(\hat{\theta}), y_2(\hat{\theta})) = s_1(\theta, \hat{\theta})$, and let

$$\tilde{\Omega}(\theta, y_0, y_1, y_2) = \left(\frac{y_0}{\theta h_0} + \tilde{s}_0(\theta, y_0, y_1, y_2) \right)^{1+\nu^{-1}} + \beta^2 \left(\frac{y_2}{\theta \tilde{s}_1(\theta, y_0, y_1, y_2)} \right)^{1+\nu^{-1}}.$$

Clearly, $\Omega(\theta, \hat{\theta})$ is increasing in $\hat{\theta}$ if $y_0(\hat{\theta})$, $y_1(\hat{\theta})$ and $y_2(\hat{\theta})$ are all increasing in $\hat{\theta}$ and if $\tilde{\Omega}(\theta, y_0, y_1, y_2)$ increases in y_1 , y_2 , and y_3 . The rest of the proof shows that $\tilde{\Omega}$ satisfies those

properties. Differentiating $\tilde{\Omega}$ with respect to y_0 one obtains that $\tilde{\Omega}$ increases in y_0 if

$$\frac{\partial \tilde{s}_0}{\partial y_0} \frac{y_0}{\tilde{s}_0} > -\frac{y_0}{\theta h_0} \frac{1}{\tilde{s}_0} \frac{1}{(1+2\nu)\zeta_0}, \quad (25)$$

where $\zeta_0 = \frac{\frac{y_1}{\theta} + \tilde{s}_0 \tilde{s}_1}{(1+2\nu)\frac{y_1}{\theta} + (2+2\nu)\tilde{s}_0 \tilde{s}_1}$. Differentiating \tilde{s}_0 with respect to y_0 , we obtain

$$\frac{\partial \tilde{s}_0}{\partial y_0} \frac{y_0}{\tilde{s}_0} = -\frac{\frac{y_0}{\theta h_0} \tilde{s}_0^{1+2\nu}}{\zeta_1 \tilde{s}_0^{1+2\nu} + \zeta_2 \tilde{s}_0^{2+2\nu} + (2+2\nu)\zeta_0 \beta^\nu \left(\frac{y_1}{\theta}\right)^\nu},$$

where $\zeta_1 = \frac{(2\nu+4\nu^2-1)\frac{y_1}{\theta} + 2\nu(2+2\nu)\tilde{s}_0 \tilde{s}_1}{(1+2\nu)\frac{y_1}{\theta} + (2+2\nu)\tilde{s}_0 \tilde{s}_1}$ and $\zeta_2 = (2+2\nu)\frac{2\nu\frac{y_1}{\theta} + (1+2\nu)\tilde{s}_0 \tilde{s}_1}{(1+2\nu)\frac{y_1}{\theta} + (2+2\nu)\tilde{s}_0 \tilde{s}_1}$. Inequality (25) will thus hold if

$$\frac{(1+2\nu)\zeta_0 \tilde{s}_0^{2+2\nu}}{\zeta_1 \tilde{s}_0^{1+2\nu} + \zeta_2 \tilde{s}_0^{2+2\nu} + \zeta_3 \beta^\nu \left(\frac{y_1}{\theta}\right)^\nu} < 1.$$

Since $\zeta_1 > 0$ and $\zeta_3 > 0$, the inequality will be satisfied if $\zeta_2 > (1+2\nu)\zeta_0$. This in turn holds if $2\nu + 4\nu^2 \geq 1$.

Differentiating \tilde{s}_0 and \tilde{s}_1 in y_1 , one obtains that $\frac{\partial \tilde{s}_0}{\partial y_1} \frac{y_1}{\tilde{s}_0} > 0$ if $2\nu + 4\nu^2 \geq 1$, and that $\frac{\partial \tilde{s}_0}{\partial y_1} \frac{y_1}{\tilde{s}_0} < 0$. $\tilde{\Omega}$ thus increases in y_1 .

Finally, differentiating $\tilde{\Omega}$ in y_2 , one obtains that $\tilde{\Omega}$ increases in y_2 if

$$\frac{\partial \tilde{s}_0}{\partial y_2} \frac{y_2}{\tilde{s}_0} > -\frac{\tilde{s}_0 \tilde{s}_1}{\frac{y_1}{\theta}} \frac{\nu \frac{y_1}{\theta} + (1+\nu)\tilde{s}_0 \tilde{s}_1}{(1+2\nu)\frac{y_1}{\theta} + (2+2\nu)\tilde{s}_0 \tilde{s}_1}.$$

However, if $2\nu + 4\nu^2 \geq 1$ then one can show that $\frac{\partial \tilde{s}_0}{\partial y_2} \frac{y_2}{\tilde{s}_0} > 0$. Hence $\tilde{\Omega}$ increases in y_2 . ■

8 Appendix 2: Numerical Solution

To verify that an allocation is incentive compatible, the inequality (2) needs to hold. To do so, I compute the following Bellman equation for $t = 0 \dots \bar{t} - 1$:

$$\Omega_t(h, \theta_i, \theta_j) = \max_{h'} \{c_t(\theta_j) - V\left(\frac{y_t(\theta_j)}{\theta_i h}, S(h, h')\right) + \beta \Omega_{t+1}(h', \theta_i, \theta_j)\}, \quad (26)$$

where $\Omega_{\bar{t}}$ is a fixed point of the Bellman operator defined appropriately by the right hand side of (26), $i = 1 \dots N$ denotes agent's productivity index, and $j = 1 \dots N$ denotes agent's report index. Since $\Omega_0 = W_\sigma$,²¹ the incentive compatibility constraint is verified if

$$\Omega_0(h_0, \theta_i, \theta_i) \geq \Omega_0(h_0, \theta_i, \theta_j) \quad \forall i, j = 1 \dots N. \quad (27)$$

Computing (26) 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 (27) holds for all $i = 1 \dots N$ and $j = i - 5 \dots i + 5$. In addition, I verify that (27) 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 (27) is satisfied in all the computed cases.

²¹Strictly speaking, $\Omega_0 = W_\sigma$ only if \bar{t} is infinite. For \bar{t} high enough, Ω_0 is an appropriate numerical approximation of W_σ