

Pareto Efficient Taxation with Learning by Doing

Marek Kapička

University of California, Santa Barbara
and CERGE-EI, Prague

November 10, 2015

VERY PRELIMINARY AND INCOMPLETE!

Abstract

I study the Pareto efficient income taxation in a Mirrlees economy with human capital formation. I provide a general framework for analyzing the problem, show that human capital formation effectively makes preferences nonseparable over labor supply, and derive a tax formula that holds in any efficient allocation. I compare it with the optimal tax formula in a Ramsey economy, and show that they differ because the Ramsey planner does not take into account dynamic changes in the earnings distribution.

I show that a model with learning-by-doing, as well as a model with learning-or-doing are special cases of this framework, and compare their implications for the efficient tax structure. I show that in both models the optimal marginal tax rates decrease with age, despite the fact that both models respond differently to any given tax change.

J.E.L Codes: E6, H2

Keywords: optimal taxation, learning by doing, human capital

1 Introduction

In this paper I study Pareto efficient allocations and taxes in dynamic Mirrlees economies where agents' human capital is endogenous and unobservable. To that end, I specify a very general framework that is suitable to study such problems, and study its properties. The framework does not explicitly include human capital formation, but features a utility function that is nonseparable in labor effort over time. I show that the two of the best known models of human capital formation, learning-or-doing model and learning-by-doing model, emerge as special cases. That is, endogenous human capital formation in both models induces reduced-form preferences over sequences of labor effort that are nonseparable. The intuition for that is relatively straightforward: In the learning-by-doing model past labor effort shows up directly in the current period utility function through the accumulated human capital. In the learning-or-doing model the nonseparability is an implication of the fact that time spent accumulating human capital is endogenous and interacts with work effort.

I extend the framework of Kapička (2014) to allow for income effects and arbitrary Pareto weights and derive a novel simple condition that shows how the marginal income taxes change over time in any Pareto efficient allocation. The condition does not directly depend on either the Pareto weights (equivalently, on the social welfare function), or on the distribution of abilities. It only depends on a set of coefficients that determine how the required information rent in any period j responds to changes in labor effort in any period t . The social planner will tend to decrease the marginal income taxes and increase labor effort in period t if higher labor effort decreases the information rent in period j . In other words, if labor effort in period t is complementary with labor effort in period j , marginal taxes should be lower. Changes in the interactions among labor effort in different periods determine the intertemporal profile of the optimal marginal income taxes.

How exactly the labor effort interacts over time obviously depends on the technology for human capital formation. I first study the implications of a canonical learning-by-doing model (Imai and Keane, 2004) for the optimal taxation in a life-cycle economy. I calibrate the economy to reproduce the observed life-cycle profiles

of hours worked and wages, as in Wallenius (2011). I decompose the determinant of the optimal tax rate into the contemporaneous effect on the information rent, the overall effect on past labor effort, and the overall effect on the future labor effort. I call the effect on past labor effort an anticipation effect (since the individual changes its labor effort in response to an anticipated changes in future tax rates), and the effect on the future labor effort an accumulation effect (since the individual changes its labor effort due to changes in the accumulated stock of human capital).¹ I find that the anticipation effect is negative because an increase in the current labor effort increases the benefit from higher human capital, and is complementary with labor effort in the past. Moreover, the anticipation effect becomes stronger with age, contributing to a decreasing profile of the optimal tax rates. The accumulation effect is also negative, because an increase in the current labor effort increases the stock of human capital in the future, which in turn improves incentives to work in the future. Contrary to the anticipation effect, the accumulation effect becomes weaker with age, and contributes to an increase in the marginal tax rates over time. The accumulation and the anticipation effects thus move in the opposite direction. Both of them are, however, dominated by the contemporaneous effect. The contemporaneous effect is positive and declines with age. An intuitive explanation is that the own-Frisch elasticity decreases with age, because taxes in the initial periods have only a small effect on current hours worked. Later in the life-cycle the contemporaneous effect weakens as labor effort elasticity goes up. Since the contemporaneous effect dominates, the marginal income taxes decrease with age. While the findings hold for a range of plausible parameter values, they are not completely general. I show that by constructing a simple example where the optimal marginal tax rates are constant from the second period onwards, and an example where they cycle.

A unified framework also allows me to compare the implications of the learning-by-doing model with the implications of a learning-or-doing model (Ben-Porath, 1967), where individuals can invest in human capital by training on the job. The

¹The terminology is similar to the terminology used in Best and Kleven (2013), but the definition is different, because the effects are defined directly in terms of the information rent, and not in terms of the elasticities.

model with learning-or-doing technology has been studied in detail in Kapička (2014). The comparison is interesting, because both the technology of human capital formation in both models is very different. In the learning-by-doing model the current labor effort is complementary with labor effort in other periods, while in the learning-or-doing model the time spent working competes with the time spent accumulating human capital, and the current labor effort can in effect become a substitute for labor effort in other periods. As a result, both models have a very different response to any given change in the tax rates. For example, a temporary increase in the marginal income tax rate leads to a decrease in the future labor effort in the learning-by-doing model, but in the learning-or-doing model the future labor effort increases.² It is then intuitive to expect that those differences will manifest themselves in different intertemporal profiles of marginal income taxes. I thus calibrate the learning-or-doing model to reproduce the same set of facts as the learning-by-doing model and show that, despite differences in the underlying mechanisms, the learning-or-doing model also predicts a decreasing pattern in the optimal marginal tax rates over time. Inspecting the contribution of each of the three effects, however, reveal the underlying differences in both models. Most importantly, the contemporaneous effect contributes to a decrease of the marginal income taxes because labor effort elasticity now decreases with age. However, the contemporaneous effect is no longer dominant. The anticipation and accumulation effects are not monotone over time, and they can be both positive or negative. Taken together, however, they prescribe a decreasing pattern of the marginal tax rates, and their contribution dominates the contemporaneous effect.

The intertemporal profile of the optimal marginal income taxes that arises in the Mirrlees model is similar to the intertemporal profile of the optimal marginal income taxes in the representative agent Ramsey model. I show, however, that they are not identical unless the coefficient of relative risk aversion is equal to one. To understand the differences between both models, I use a dual approach, where the government chooses the tax functions directly, rather than inferring the optimal tax structure indirectly from the efficient allocations, which is the primal approach. The dual approach

²Similarly, Cossa, Heckman, and Lochner (1999) finds that both models respond differently to changes in the wage subsidies, for example in the Earned Income Tax Credit.

to Mirrlees optimal taxation has been pioneered by [Saez \(2001\)](#), and extended by [Best and Kleven \(2013\)](#).³ It has its advantages in that one expresses the optimal tax formulas directly in terms of labor supply elasticities, and the costs and benefits of choosing a tax rate are more transparent. On the other hand, it loses its tractability if there is more than two periods, and the role of the underlying model structure is less clear.

I therefore limit attention to a two period economy when using the dual approach. I show that the key difference between the Mirrlees approach and the Ramsey approach is that the Mirrlees planner takes into account that the distribution of earnings changes in response to a change in taxes, while the Ramsey planner does not. More precisely, it is the difference in response of the earnings distribution over time that differentiates Mirrlees from Ramsey: if the distribution of earnings responds identically over time, the Mirrlees optimal tax formula is identical to the Ramsey formula. This happens precisely when the coefficient of relative risk aversion equals one, confirming the findings from the primal approach. It is also worth noting that the insights into differences between Ramsey and Mirrlees taxes apply more generally, although they are especially useful in models with human capital.

The paper follows the Mirrlees approach to the optimal income taxation ([Mirrlees, 1971, 1976](#)). The Mirrlees approach has been recently extended to dynamic environments by [Golosov, Kocherlakota, and Tsyvinski \(2003\)](#), [Kocherlakota \(2005\)](#), [Farhi and Werning \(2012\)](#), [Golosov, Tsyvinski, and Troshkin \(2013\)](#), and many others. Most of the literature on dynamic Mirrlees taxation, however, assumes that abilities are exogenously given, and abstracts from endogenous human capital formation. Papers that are most related to this are [Best and Kleven \(2013\)](#) and [Kapička \(2014\)](#). [Best and Kleven \(2013\)](#) study a dual Mirrlees problem in a two period economy with learning-by-doing technology, and parameterize the technology to match empirically relevant interdependencies in labor effort (called "career elasticity"). Relatively to [Best and Kleven \(2013\)](#), I provide a more general framework for the optimal tax problems with unobservable human capital formation, which encompasses both the learning-by-doing and learning-or-doing models. I also focus on the primal approach that is

³The distinction between primal and dual approaches is well established in the Ramsey literature. It is not usually used in the Mirrlees literature, despite the fact that it applies equally well.

suitable for models with more than two periods. The general framework restricted to economies with no income effects has been studied by Kapička (2014), but only for the learning-or-doing technology. Finally, several papers have analyzed optimal taxation in models where human capital formation is observable, and gives rise to a joint problem of finding the optimal income tax and human capital subsidies. Bovenberg and Jacobs (2005), Stancheva (2014) and Findeisen and Sachs (2014) study optimal taxation and educational subsidies in a model where education costs physical resources rather than time, and Boháček and Kapička (2008) study optimal taxation and educational subsidies in a learning-or-doing model.

2 The Model

The general formulation is as follows. The agents live for $T + 1$ periods, where T is finite. They work $z_t \in \mathbb{R}_+$ hours and consume $c_t \in \mathbb{R}_+$ at age t . The utility function is given by

$$W(c^T, z^T) = \sum_{t=0}^T \beta^t U(c_t) - V(z^T), \quad (1)$$

where $c^T = (c_t)_{t=0}^T$ and $z^T = (z_t)_{t=0}^T$. The function $U : \mathbb{R}_+ \rightarrow \mathbb{R}$ is increasing and concave, and the function $V : \mathbb{R}_+ \rightarrow \mathbb{R}$ is increasing and convex in all arguments.

Each individual is associated with an ability level $\theta \in [\underline{\theta}, \bar{\theta}] = \Theta$, which does not change with age. The ability is drawn from a distribution function F , which is differentiable and has density f . Agent's ability together multiplied by hours worked determine the number of the efficiency units of labor supplied. The earnings in period t are then

$$y_t = w_t \theta z_t, \quad (2)$$

where w_t is the wage rate per efficiency unit of labor.

2.1 Pareto efficient allocations

I will adopt a standard assumption that consumption c^T and income y^T is observable by the planner. Hours worked z and idiosyncratic productivity θ are, on the other hand, a private information of the agent. The agents submit, at the beginning of period zero, a report $\theta \in \Theta$ to the social planner. The planner chooses consumption $c^T(\theta) = (c_t(\theta))_{t=0}^T$ and earnings $y^T(\theta) = (y_t(\theta))_{t=0}^T$ as functions of the reports. The allocation must be feasible, i.e. to satisfy the present value budget constraint:

$$\int_{\underline{\theta}}^{\bar{\theta}} \sum_{t=0}^T R^{-t} c_t(\theta) f(\theta) d\theta \leq \int_{\underline{\theta}}^{\bar{\theta}} \sum_{t=0}^T R^{-t} y_t(\theta) f(\theta) d\theta, \quad (3)$$

where $R \geq 1$ is the interest rate. Incentive compatibility requires that the allocation for a θ -type agent is preferred to the allocation for any other type:

$$W\left(c^T(\theta), \frac{y^T(\theta)}{\theta}\right) \geq W\left(c^T(\hat{\theta}), \frac{y^T(\hat{\theta})}{\theta}\right) \quad \forall \theta, \hat{\theta} \in \Theta. \quad (4)$$

A necessary condition for incentive compatibility is

$$W\left(c^T(\theta), \frac{y^T(\theta)}{\theta}\right) = \underline{W} + \int_{\underline{\theta}}^{\theta} \sum_{t=0}^T V_{z_t} \left(\frac{y^T(\varepsilon)}{\varepsilon} \right) \frac{y_t(\varepsilon)}{\varepsilon^2} d\varepsilon \quad \forall \theta \in \Theta. \quad (5)$$

where $\underline{W} = W\left(c^T(\underline{\theta}), \frac{y^T(\underline{\theta})}{\underline{\theta}}\right)$ is the lifetime utility of an agent with the lowest ability $\underline{\theta}$. The last term on the right-hand side of the envelope condition (5) is the overall information rent from having ability θ . The information rent is the total of a sequence of period information rents $V_{z_t} \frac{y_t}{\theta^2}$.

The planner is constrained by the incentive constraint. In what follows, I will only constrain the social planner by the envelope condition, thus solving for a relaxed problem. The envelope condition is, however, sufficient for the incentive compatibility under the following conditions:

Lemma 1. *If (5) holds and $y^T(\theta)$ is increasing in θ then (4) holds.*

Proof. Let $w(\theta, \hat{\theta}) = W\left(c^T(\hat{\theta}), \frac{y^T(\hat{\theta})}{\theta}\right)$. Evaluating (5) at $\theta, \hat{\theta}$ and subtracting, one gets

$$\begin{aligned} w(\theta, \theta) - w(\hat{\theta}, \hat{\theta}) &= \int_{\hat{\theta}}^{\theta} \sum_{t=0}^T V_{z_t} \left(\frac{y^T(\varepsilon)}{\varepsilon} \right) \frac{y_t(\varepsilon)}{\varepsilon} \frac{d\varepsilon}{\varepsilon} \geq \int_{\hat{\theta}}^{\theta} \sum_{t=0}^T V_{z_t} \left(\frac{y^T(\hat{\theta})}{\varepsilon} \right) \frac{y_t(\hat{\theta})}{\varepsilon} \frac{d\varepsilon}{\varepsilon} \\ &= w(\theta, \hat{\theta}) - w(\hat{\theta}, \hat{\theta}). \end{aligned}$$

The inequality follows from the assumption that $y_t(\theta)$ increases in θ for all $t = 0, 1, \dots, T$ and that V is convex. ■

The agents are assigned Pareto weights $p(\theta) \geq 0$. I define P to be the cumulative distribution function, and normalize the weights so that $P(\bar{\theta}) = 1$. A natural benchmark is to have $P = F$, in which case the objective function of the social planner is the expected utility of the agent. More generally, I will assume that $P \geq F$, which means that the planner puts higher weights on lower ability agents.

The social planner in the relaxed Mirrlees problem chooses an allocation (c^T, y^T) to maximize

$$\int_{\underline{\theta}}^{\bar{\theta}} W\left(c^T(\theta), \frac{y^T(\theta)}{\theta}\right) p(\theta) d\theta. \quad (6)$$

subject to (3) and (5). The *efficient allocation* is an allocation that attains the maximum of the Pareto problem.

Conditions for Efficiency

Define the *intratemporal wedge* $\tau^T = (\tau_t)_{t=0}^T$ as one minus the marginal rate of substitution between consumption and earnings:

$$\tau_t = 1 - \frac{V_{z_t}(z^T)}{\beta^t U'(c_t) \theta}.$$

The intratemporal wedge corresponds to the marginal income tax rate, and I will occasionally refer to it as such.⁴ Next theorem derives a simple characterization of the dynamics of the intratemporal wedge for a given agent.

⁴The implementation problem is standard, and is omitted from the paper.

Theorem 2. *The efficient intratemporal wedges in the Mirrlees economy satisfy*

$$\frac{\frac{\tau_j(\theta)}{1-\tau_j(\theta)}}{\frac{\tau_t(\theta)}{1-\tau_t(\theta)}} = \frac{1 + \rho_j(\theta)}{1 + \rho_t(\theta)}, \quad (7)$$

where $\rho_t = \sum_{k=0}^T \rho_{t,k}$ and $\rho_{t,k} = \frac{V_{z_t z_k} z_k}{V_{z_t}}$.

Proof. Fix θ , and denote the efficient allocation for the θ -type agent by c^{*T}, z^{*T} . Consider a perturbation that minimizes the cost of an allocation to the θ -type agent, while keeping both his utility and his marginal information rent unchanged:

$$\min_{c^T, z^T} \Delta = \sum_{t=0}^T R^{-t} (c_t - \theta z_t) - \sum_{t=0}^T R^{-t} (c_t^* - \theta z_t^*)$$

subject to

$$W(c^T, z^T) = W(c^{*T}, z^{*T}) \quad (8)$$

$$\sum_{t=0}^T V_{z_t} \left(z^T \right) \frac{z_t}{\theta} = \sum_{t=0}^T V_{z_t} \left(z^{*T} \right) \frac{z_t^*}{\theta}. \quad (9)$$

Because neither the lifetime utility nor the information rent change, any perturbation will continue to satisfy the envelope condition (5). Moreover, since the efficient allocation c^{*T} and z^{*T} is feasible and delivers $\Delta = 0$, any solution to the cost minimization program above can only save resources and will satisfy the resource constraint (3).

Since c^{*T} and z^{*T} is efficient, it must be the solution to the cost minimization problem. Taking the first-order conditions and rearranging, one obtains

$$\frac{\theta \beta^t U'(c_t^*)}{V_{z_t}(z^{*T})} - 1 = \mu(1 + \rho_t), \quad (10)$$

where μ is the Lagrange multiplier on the constraint (9). The left-hand side is equal to $\frac{\tau_j(\theta)}{1-\tau_j(\theta)} / \frac{\tau_t(\theta)}{1-\tau_t(\theta)}$ by the definition of the intratemporal wedge. Taking the ratio of (10) for two different periods, one obtains (7). ■

The most important property of equation (7) is that it holds for any Pareto weights p . Equation (7) thus provide necessary conditions for a Pareto efficient tax structure. Since it is expressed in terms of a ratio of the wedges, knowing one intertemporal

wedge, say τ_0 , together with the underlying structure, is enough to compute the whole sequence of wedges.

One of the factors in determining the optimal intratemporal wedge in a given period is how responsive the information rent is to the changes in the labor supply. If the information rent is relatively responsive, the social planner has to award the agent with relatively high consumption in order to induce him to increase labor supply, which is costly. Nonseparability in labor supply brings about interdependencies between labor supply and information rent across periods. Specifically, the coefficient $\rho_{t,j}$ measures the effect that a change in hours worked in period t has on the information rent in period j , and ρ_t measures the effect that a change in hours worked in period t has on the overall information rent. As we will see later, the coefficients $\rho_{t,j}$ may be both positive or negative: if they are negative, an increase in hours worked in period t increases incentives to work in period j , and thus reduces consumption needed to compensate the agent for higher work effort. The equation (7) shows that changes in taxes over time depend only on changes in the associated information rents.

To further simplify the equation (7), I decompose the overall effect ρ_t into three components: The effect of a change in the tax rate on the contemporaneous information rent $\rho_{t,t}$, the effect of a change in hours worked on the incentives to supply labor in all the previous periods $\rho_t^- = \sum_{j=0}^{t-1} \rho_{t,j}$ (the anticipation effect), and the effect of a change in hours worked on the incentives to supply labor in all the future periods $\rho_t^+ = \sum_{j=t+1}^T \rho_{t,j}$ (the accumulation effect). The anticipation effect is the overall effect that a given change in future hours worked have on labor supply in previous periods. As the name suggests, it is nonzero, because the agent anticipates future changes and adjusts labor supply accordingly. The accumulation effect will typically work through human capital accumulation: a current change in hours worked will change incentives to accumulate human capital, which will affect future labor supply.

$$\frac{\frac{\tau_j}{1-\tau_j}}{\frac{\tau_t}{1-\tau_t}} = \frac{1 + \rho_j^- + \rho_{j,j} + \rho_j^+}{1 + \rho_t^- + \rho_{t,t} + \rho_t^+}. \quad (11)$$

Changes in the optimal tax rates over time can thus be characterized in terms of changes in the own effects, anticipation effects, and accumulation effects.

The level of optimal intratemporal wedges is determined not only by the parameters γ , but also by other factors: Pareto weights, distribution of abilities, or curvature of the utility function. Those additional factors affect the intratemporal wedges symmetrically, and thus do not enter the relative intratemporal wedges in equation (7). If the utility is linear in consumption then the level of taxes is simply given by

$$\frac{\tau_t(\theta)}{1 - \tau_t(\theta)} = [1 + \rho_t(\theta)] \frac{P(\theta) - F(\theta)}{\theta f(\theta)}. \quad (12)$$

In models where V is additively separable, the term ρ_t has a simple expression: $\rho_{t,k}$ is zero if $k \neq t$, and $\rho_{t,t}$ is the inverse of Frisch elasticity of labor in period t , $\epsilon_{t,t} = \frac{d \log z_t}{d \log w_t}$. Thus, one obtains a simple formula $\frac{\tau_j}{1 - \tau_j} = \frac{1 + \epsilon_{j,j}^{-1}}{1 + \epsilon_{t,t}^{-1}}$. With nonseparable V one cannot express the optimal tax formula as easily in terms of the Frisch elasticities. However, one can show, that the matrix $\rho = (\rho_{t,k})$ is the inverse of the matrix of the Frisch elasticities $\epsilon = (\epsilon_{t,k})$, where $\epsilon_{t,k} = \frac{d \log z_t}{d \log w_k}$. Thus, it is in principle possible to obtain the optimal intratemporal wedges directly as a function of the elasticities, but the procedure is not very intuitive and practical beyond two period models, since one has to work with an inverse of the matrix of Frisch elasticities. Moreover, as we shall see in the example below, the matrix ρ sometimes takes a very simple and intuitive form, while the matrix of elasticities ϵ does not.

3 Mirrlees vs Ramsey

The apparent simplicity of the formula (7) bears resemblance to the optimal tax formulas in a Ramsey problem, where the government imposes linear taxes on the commodities in the economy. In this section I compare both formulas, and show that they are, in general, different.

In what follows, I will restrict attention to preferences where the period utility

from consumption exhibits a constant relative risk aversion:

$$U(c) = \frac{c^{1-\sigma} - 1}{1-\sigma},$$

with the usual limiting form $U(c) = \ln c$ when $\sigma = 1$. Consider a representative agent economy⁵ where the government chooses a sequence of linear taxes $(\tau^R)^T$ on labor. It is easy to show that in the current setting the optimal capital taxes will always be set to zero. Since my main objective is to study the intratemporal wedges, I will simplify the analysis by setting the capital income taxes to zero from the outset. A representative consumer maximizes its lifetime utility (1) subject to a budget constraint

$$\sum_{t=0}^T R^{-t} c_t \leq \sum_{t=0}^T R^{-t} (1 - \tau_t^R) z_t. \quad (13)$$

As in the Mirrlees economy, I will use a primal approach to optimal taxation. Substituting the first-order conditions from the agent's problem back to the budget constraint (13) to eliminate the taxes yields the following implementability constraint:

$$\sum_{t=0}^T \beta^t U'(c_t) c_t = \sum_{t=0}^T V_{z_t} z_t. \quad (14)$$

The Ramsey social planner chooses the allocation (c^T, z^T) to maximize the lifetime utility (1) subject to the implementability constraint (14) and the resource constraint:

$$\sum_{t=0}^T R^{-t} c_t \leq \sum_{t=0}^T R^{-t} z_t. \quad (15)$$

Next theorem characterizes the optimal taxes in the Ramsey economy:

⁵One might be understandably worried that a representative agent economy cannot be meaningfully compared to a heterogeneous agent economy in the previous subsection. It is possible to show, however, that the optimal tax ratio (7) can be also obtained as a limiting value in a two-type Mirrlees economy, when the two types approach each other and heterogeneity disappears.

Theorem 3. *The efficient intratemporal wedges in a Ramsey economy satisfy*

$$\frac{\frac{\tau_j^R}{1-\tau_j^R}}{\frac{\tau_t^R}{1-\tau_t^R}} = \frac{\sigma + \rho_j}{\sigma + \rho_t}. \quad (16)$$

Proof. Maximizing (1) subject to (15) and (14) yields the first-order conditions

$$\begin{aligned} \beta^t U'(c_t) &= \lambda R^{-t} - \mu \beta^t [U'(c_t) + U''(c_t)c_t] \\ V_{z_t} &= R^{-t} \lambda - \mu \left[\sum_{j=0}^T V_{z_t z_j} z_j + V_{z_t} \right], \end{aligned}$$

where λ and μ are the Lagrange multipliers on the resource constraint (15) and on the implementability constraint (14). Rearranging, one obtains

$$\frac{\beta^t U'(c_t)}{V_{z_t}} = \frac{1 + \mu(1 + \rho_t)}{1 + \mu(1 - \sigma)}.$$

Using the definition of τ_t and taking the ratio for two different periods, one obtains (16). ■

The optimal tax formula in the Ramsey economy (16) needs to be compared with the optimal tax formula in the Mirrlees economy (7). They are not, in general, identical, as the coefficient of a relative risk aversion σ enters the optimal tax formula in the Ramsey economy, but not in the Mirrlees economy. Only if $\sigma = 1$ then both formulas coincide. The optimal tax rates may differ in their levels, but will exhibit identical changes over time. If $\sigma > 1$ then the Ramsey planner will produce a less steep intertemporal profile of taxes than the Mirrlees planner (regardless of whether the taxes are increasing or decreasing). If $\sigma < 1$ then the Ramsey planner will produce a steeper profile of marginal income taxes.

3.1 A dual problem in two periods

The disadvantage of the primal approach is that formulas (7) and (16) are not very intuitive, and it is not very clear what makes the Ramsey formula different from the Mirrlees formula. To understand the difference between them, I will turn to the dual

problem, where the taxes are determined directly by the planner, rather than being inferred indirectly from the optimal allocations. The dual problem can be formulated in terms of the compensated and uncompensated elasticities of labor, and provides additional insights into the mechanisms that shape the optimal tax structure. The disadvantage is, that it does not have explicit solutions when there is more than two periods. For that reasons I will now restrict attention to a two period economy ($T = 1$). In what follows, $\epsilon_{i,j}^c = \frac{d \log l_i^c}{d \log w_j}$ denotes the compensated elasticity of labor supply, and $\epsilon_{i,j}^u = \frac{d \log l_i^u}{d \log w_j}$ denotes the uncompensated elasticity of labor supply.

Consider first a Ramsey economy. Modifying the standard textbook treatment with multiple consumption goods (Atkinson and Stiglitz, 1980, chapter 12) yields the following optimal tax formula:

Theorem 4. *The efficient intratemporal wedges in a 2-period Ramsey economy satisfy*

$$\frac{\frac{\tau_1^R}{1-\tau_1^R}}{\frac{\tau_0^R}{1-\tau_0^R}} = \frac{\epsilon_{0,0}^c - \epsilon_{1,0}^c}{\epsilon_{1,1}^c - \epsilon_{0,1}^c}. \quad (17)$$

Proof. Let $V^*(w_0, w_1)$ be the indirect utility function of the agent, a function of the after tax wages w_0 and w_1 . Similarly, let $z_t^u(w_0, w_1)$ be the optimal (uncompensated) labor demand. The Ramsey social planner is constrained by the resource constraint

$$\tau_0 w_0 z_0^u((1-\tau_0)w_0, (1-\tau_1)w_1) + R^{-1} \tau_1 w_1 z_1^u((1-\tau_0)w_0, (1-\tau_1)w_1) = G, \quad (18)$$

and maximizes the representative agent's utility

$$V^*((1-\tau_0)w_0, (1-\tau_1)w_1)$$

subject to the government budget constraint (18). Taking the first-order conditions and using Roy's lemma ($V_0^* = \alpha z_0$ and $V_1^* = \alpha R^{-1} z_1$) yields

$$\alpha z_t = \lambda \left[z_t - \tau_0 w_0 \frac{dz_0^u}{dw_t} - R^{-1} \tau_1 w_1 \frac{dz_1^u}{dw_t} \right], \quad t = 1, 2,$$

where λ is the Lagrange multiplier on the government budget constraint. Using the Slutsky equation, $\frac{dz_i^u}{dw_t} = \frac{dz_i^c}{dw_t} - z_t \frac{dz_i^u}{dM}$, where $\frac{dz_i^u}{dM}$ is the derivative of the uncompensated demand with

respect to income, one can rewrite the first-order conditions as

$$\zeta z_t = \tau_0 w_0 \frac{dz_0^c}{dw_t} + R^{-1} \tau_1 w_1 \frac{dz_1^c}{dw_t}, \quad t = 1, 2,$$

where $\zeta = 1 - \alpha/\lambda + \tau_0 w_0 \frac{dz_0^u}{dM} + R^{-1} \tau_1 w_1 \frac{dz_1^u}{dM}$ is a common constant. Solving for $\frac{dz_i^c}{dw_t}$, using the symmetry of the Slutsky matrix and the definition of the compensated elasticities yields the formula in the text. ■

According to the optimal tax formula (17), a higher own compensated elasticity decreases the optimal labor tax rate relative to the tax rate in the other period. Complementarity between labor supply in the two periods ($\epsilon_{1,0}^c > 0$ and $\epsilon_{0,1}^c > 0$) may either increase or decrease relative tax rates, depending on the relative size of the two cross-elasticities. The formula (17) is also identical for both separable and non-separable preferences over labor supply, because nonseparability manifests itself only indirectly, through the compensated elasticities. This has its costs and benefits. On one hand, one does not need to know the internal structure of the model to use (17). On the other hand, one cannot infer the contribution of nonseparable preferences for the optimal intertemporal profile of taxes. Finally, it can be easily shown that (17) is equivalent to (16).⁶

Consider now the Mirrlees economy. In what follows, let $\tilde{F}_t(y)$ be the equilibrium distribution of earnings in period t , and $\tilde{f}_t(y)$ be the corresponding density. Unlike the distribution of abilities, the distribution of earnings is neither constant over time, nor exogenously given. The dual problem yields the following formula for the Pareto efficient intratemporal wedges:

Theorem 5. *The efficient intratemporal wedges in a 2-period Mirrlees economy satisfy*

$$\frac{\frac{\tau_1}{1-\tau_1}}{\frac{\tau_0}{1-\tau_0}} = \frac{\tilde{f}_0(y_0)y_0\epsilon_{0,0}^c - \tilde{f}_1(y_1)y_1\epsilon_{1,0}^c}{\tilde{f}_1(y_1)y_1\epsilon_{1,1}^c - \tilde{f}_0(y_0)y_0\epsilon_{0,1}^c}. \quad (19)$$

Proof. The proof is, in its substance, similar to the proofs in Saez (2001) and Best and Kleven (2013), although I will provide a formal variational argument to make the proof more

⁶The result can be derived from the fact that the compensated elasticities are given by $\epsilon_{00}^c = [(1 + \kappa)\rho_1 + \sigma]D$, $\epsilon_{01}^c = -\sigma D$, $\epsilon_{10}^c = -\kappa\sigma D$ and $\epsilon_{11}^c = [(1 + \kappa)\rho_0 + \kappa\sigma]D$ for some constants κ and D .

symmetric to the proof of Lemma (4). Consider a government choosing tax functions $T_t(y)$ for $t = 0, 1$, and let $\tau_t(y) = T'_t(y)$ be the marginal tax rates. One can then write the tax function as $T_t(y) = \bar{T}_t + \int_0^y \tau_t(\zeta) d\zeta$. Write the indirect utility function of a θ type agent as $V^*(\bar{T}_0, \bar{T}_1, \tau_0, \tau_1, \theta)$. The government's budget constraint is

$$\begin{aligned} G &= \int_y T_0(y) d\tilde{F}_0(y) + R^{-1} \int_y T_1(y) d\tilde{F}_1(y) \\ &= \bar{T}_0 + R^{-1} \bar{T}_1 + \int_y \tau_0(y) [1 - \tilde{F}_0(y)] dy + R^{-1} \int_y \tau_1(y) [1 - \tilde{F}_1(y)] dy, \end{aligned} \quad (20)$$

where the second equality uses integration by parts. The planner's problem is to choose the constants $\bar{T}_t$ and the functions τ_t to maximize

$$\sum_{t=0,1} R^{-t} \int_y \left[U \left(y - \bar{T}_t - \int_0^y \tau_t(\zeta) d\zeta \right) - V \left(\frac{y}{\theta_t(y)} \right) \right] dP_t(y)$$

subject to the resource constraint (20). The functions $\theta_t(y)$ denotes the agent's type that produces earnings y in period t , and P_t denotes the distribution of Pareto weights over income levels.

Consider first a perturbation of the marginal tax rate $\tau_0(y)$ by δ_τ on a small interval $(y_0^*, y_0^* + \delta_{y_0})$. The marginal change in the objective function of the planner is

$$\Delta = \delta_\tau \delta_{y_0} \int_y U'(c_0(y)) dG_0 - \lambda \left[(1 - \tilde{F}_0(y_0^*)) \delta_\tau \delta_{y_0} - \int_y \tau_0(y) \delta_{\tilde{F}_0}(y) dy - R^{-1} \int_y \tau_1(y) \delta_{\tilde{F}_1}(y) dy \right],$$

where λ is the Lagrange multiplier on the government budget constraint and $\delta_{\tilde{F}_t}(y)$ is the implied change in the cdf of earnings in period t at earnings level y . We have $\delta_{\tilde{F}_0}(y) = 0$ if $y < y_0^*$ because individual with earnings below y_0^* are not affected by the perturbation, $\delta_{\tilde{F}_0}(y) = h(y_0^*) \theta \frac{dy_0^c}{dw_0} \delta_\tau$ if $y = y_0^*$ because individuals at y_0^* face a compensated change in wages, and $\delta_{\tilde{F}_0}(y) = h(y) \frac{dy_0}{dM} \delta_\tau$ if $y > y_0^*$ because individuals with earnings above y_0^* experience an income effect. Let $y_1^* = y_1(y_0^*)$ be the second period earnings of an agent that produces y_0^* in the initial period. We then have $\delta_{\tilde{F}_1}(y) = 0$ if $y < y_1^*$, $\delta_{\tilde{F}_1}(y) = h(y_0) \theta \frac{dy_1^c}{dw_0} \delta_\tau \delta_{y_1}$ if $y = y_1^*$, and $\delta_{\tilde{F}_1}(y) = h(y_0) \frac{dy_1}{dM} \delta_\tau \delta_{y_1}$ if $y > y_1^*$. Substituting the expressions in, using $\tilde{f}_0(y_0^*) \delta_{y_0} = \tilde{f}_1(y_1^*) \delta_{y_1}$ and setting $\Delta = 0$ yields the following optimality condition:

$$\zeta = \tilde{f}_0 \theta \tau_0 \frac{dy_0^c}{dw_0} + R^{-1} \tilde{f}_0 \theta \tau_1 \frac{dy_1^c}{dw_0},$$

where $\zeta = 1 - \tilde{F}_0 - \lambda^{-1} \int_y U'(c_0(y)) dG_0 + \int_{y_0^*} \tilde{f}_0 \tau_0 \frac{dy_0}{dM} dy + R^{-1} \int_{y_1^*} \tilde{f}_1 \tau_1 \frac{dy_1}{dM} dy$. A similar perturbation to the marginal tax function τ_1 yields a second condition

$$\zeta = \tilde{f}_1 \theta \tau_0 \frac{dy_0^c}{dw_1} + R^{-1} \tilde{f}_1 \theta \tau_1 \frac{dy_1^c}{dw_1}.$$

Combining and using the definition of compensated elasticities yields (19). ■

The formula (19) has an intuitive explanation: while the compensated elasticity enters the optimal tax formula because it measures the distortion of one individual, the overall distortion in the population is the compensated elasticity weighted by the density of earnings, and by the earnings level. One can think of the effective elasticity in the Mirrlees model to be the density and earnings weighted individual elasticities. Lower earnings density or lower income level will make the tax at that income level less distortive. It is the effective elasticity that determines the tax rates in the Mirrlees problem.⁷

Comparing (19) and (17), one obtains one of the key insights into why Mirrlees formulas and Ramsey formulas differ. In the Mirrlees problem, the effect of compensated elasticities is weighted by the equilibrium densities of earnings, and by the level of earnings. In contrast, in a Ramsey problem the elasticities are unweighted. The lack of distributional considerations in a representative agent Ramsey model is obvious. It is little less straightforward to see why the elasticities in a Ramsey mode are not weighted by the earnings level: since the tax is linear, both the distortions and the gains from a change in the tax rate are proportional to the earnings level, and it cancels out.

By dividing both the nominator and denominator of the formula (19) by $\tilde{f}_0 y_0$, one can see that the optimal tax formula in the Mirrlees economy is identical to the optimal tax formula in a Ramsey economy only if $\tilde{f}_0 y_0 = \tilde{f}_1 y_1$. What makes the Mirrlees problem different from the Ramsey problem is that taxes in different periods

⁷The formula (19) is again independent of the functional form specification. It can also be shown that it is identical to its appropriate counterpart in Proposition 1 of [Best and Kleven \(2013\)](#), although the equivalence is not obvious at first sight. The formula (19) is preferable for the purpose of this paper, as it clearly shows the dependence on earning densities, and its right-hand side is independent of the intratemporal wedges.

affect the distribution of earnings differently. It is worth stressing that the differences between Ramsey and Mirrlees formulae are driven by the intertemporal differences in the distribution of earnings, not of abilities, which are constant over time. Those two are obviously related. It is easy to show that $\tilde{f}_t y_t (1 + \epsilon_{t,0}^u + \epsilon_{t,1}^u) = f(\theta)\theta$. A change in θ represents a permanent change in wages, and so affects the density of earnings proportionately to the sum of $\epsilon_{t,0}^u$ and $\epsilon_{t,1}^u$. One can then rewrite (19) in terms of elasticities only as

$$\frac{\frac{\tau_1}{1-\tau_1}}{\frac{\tau_0}{1-\tau_0}} = \frac{(1 + \epsilon_{1,0}^u + \epsilon_{1,1}^u)\epsilon_{0,0}^c - (1 + \epsilon_{0,0}^u + \epsilon_{0,1}^u)\epsilon_{1,0}^c}{(1 + \epsilon_{0,0}^u + \epsilon_{0,1}^u)\epsilon_{1,1}^c - (1 + \epsilon_{1,0}^u + \epsilon_{1,1}^u)\epsilon_{0,1}^c}. \quad (21)$$

The Mirrlees formula (21) will then differ from the Ramsey formula (17) in the uncompensated elasticities of labor supply are changing over time. If, on the other hand, $\epsilon_{1,0}^u + \epsilon_{1,1}^u = \epsilon_{0,0}^u + \epsilon_{0,1}^u$, then the Mirrlees formula will be identical to the Ramsey formula. Since the formula (19) must be identical to the formula (7), one can infer that this will happen only if the coefficient of the relative risk aversion σ is equal to one. A direct computation verifies this: if $\sigma = 1$ then $\epsilon_{1,0}^u + \epsilon_{1,1}^u = \epsilon_{0,0}^u + \epsilon_{0,1}^u = 0$.

4 Learning by Doing

I will now introduce a model with a learning-by-doing type of human capital formation. I will then show that the model is a special case of the general framework of the previous section.

Let $n_t \in \mathbb{R}_+$ be the raw hours that the agent works, and h_t be the stock of human capital at the beginning of period t . Current earnings are given by

$$y_t = Q_t \theta_t h_t n_t, \quad (22)$$

where Q_t is the rental rate of human capital, and θ is again agent's ability. Human capital next period depends on current human capital and on current raw labor, and

is given by

$$h_{t+1} = G(h_t, n_t, t), \quad (23)$$

where the initial value h_0 is given. The human capital production function $G : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is increasing and concave in both arguments.

The utility function of the agent depends on the sequence of consumption and raw labor and is given by

$$\sum_{t=0}^T \beta^t [U(c_t) - X(n_t)], \quad 0 < \beta \leq 1. \quad (24)$$

The function $X : \mathbb{R}_+ \rightarrow \mathbb{R}$ is increasing and convex. The function U is as defined as in Section 2.

At first sight, the learning-by-doing model appears to have quite a different structure than the general model of Section 2. It can, however, be transformed into the general framework as follows. Let $z_t = h_t n_t$ be the *effective* labor that the household supplies. Then the production function has the same form as in equation (2), with $w = Q$. In addition, for a given sequence of the efficiency units of labor z^T the human capital is uniquely determined from the human capital production function as follows. Let $H_0(\emptyset) = h_0$, and define $H_t(z^{t-1})$ by

$$H_{t+1}(z^t) = G\left(H_t(z^{t-1}), \frac{z_t}{H_t(z^{t-1})}, t\right), \quad t = 0, 1, \dots, T$$

where $z^t = (z_j)_{j=0}^t$. The disutility function V is then simply given by

$$V(z^T) = \sum_{t=0}^T \beta^t X\left(\frac{z_t}{H_t(z^{t-1})}\right). \quad (25)$$

Differentiating V with respect to z^T yields

$$V_{z_t}(z^T) = \beta^t X' \left(\frac{z_t}{H_t(z^{t-1})} \right) \frac{1}{H_t(z^{t-1})} - \sum_{j=t+1}^T \beta^j X' \left(\frac{z_j}{H_j(z^{j-1})} \right) \frac{z_j}{H_j(z^{j-1})^2} \frac{\partial H_j(z^{j-1})}{\partial z_t} \quad (26)$$

where $\frac{\partial H_j(z^{j-1})}{\partial z_t} = G_z(h_t, z_t, t) \prod_{k=t+1}^{j-1} G_h(h_k, z_k, k)$ is the derivative of human capital with respect to the effective labor supply. The partial derivative (26) shows that an increase in the current effective labor supply has two effects on the overall disutility of working a sequence z^T . First, an increase in current effective labor increases the disutility of working in the current period. Second, it decreases the disutility of working in all the future periods.⁸

How about the second derivatives $V_{z_t z_k}$ and the corresponding coefficients ρ ? It is relatively easy to show that $\rho_{t,t}$ is strictly positive whenever G is strictly concave. Higher z_t not only increases the marginal disutility of working in the current period t , but an increase in z_t also weakens the beneficial effect on the future disutility of working. On the other hand, $\rho_{t,k}$ when $k \neq t$ may be positive or negative. An increase in z_k decreases the marginal disutility of working in period $t > k$ because the agent accumulates more human capital, but there are also indirect effects going through the second term in (26) which may go either way. If, however, the direct effect dominates, then $\rho_{t,k}$ will be negative whenever $k \neq t$.

4.1 An Example Solved

I will now provide a closed form solution to a special case when the function X exhibits a constant Frisch elasticity of raw labor,

$$X(n) = b \frac{n^{1+\varepsilon}}{1+\varepsilon}, \quad (27)$$

⁸The first effect has to dominate for the effective labor supply to stay finite.

and the human capital production function takes the form

$$G(h_t, n_t) = A_t(h_t n_t)^\alpha. \quad (28)$$

For simplicity, I will also assume that $h_0 = 1$, although nothing depends on this.

When the production function takes the form given by (28), the law of motion for the human capital, as a function of the effective labor z , takes a particularly simple form: $h_{t+1} = A_t z_t^\alpha$. That is, next period human capital is only a function of a current effective labor supply (although it is a function of the whole history of the "raw" labor supply n_t). As a result, one can write the disutility function V as

$$V(z^T) = \frac{z_0^{1+\varepsilon}}{1+\varepsilon} + \sum_{t=0}^{T-1} \beta^{t+1} \frac{\left(\frac{z_{t+1}}{A_t z_t^\alpha}\right)^{1+\varepsilon}}{1+\varepsilon} \quad (29)$$

The function V is clearly nonseparable in z^T , unless $\alpha = 0$ and there is no human capital formation. Because the production function (28) exhibits increasing returns, special attention needs to be paid to conditions under which V is convex. Next lemma shows that V is convex if the production function is sufficiently concave:

Lemma 6. *The function $V(z^T)$ is strictly convex in z^T if $\alpha < \frac{\varepsilon}{1+\varepsilon}$.*

Proof. See Appendix A. ■

The coefficients $(\rho_{t,k})$ can be calculated directly. They can be expressed as a function of the growth rate of the disutility of raw labor supply $a_t = (n_{t+1}/n_t)^{1+\varepsilon}$:

$$\begin{aligned} \rho_{t,t-1} &= -\frac{\alpha(1+\varepsilon)}{1-\alpha\beta a_t} \quad t = 1, \dots, T \\ \rho_{t,t} &= \frac{\varepsilon + \alpha\beta[1 + \alpha(1+\varepsilon)]a_t}{1-\alpha\beta a_t} \quad t = 0, \dots, T \\ \rho_{t,t+1} &= -\frac{\alpha\beta(1+\varepsilon)a_t}{1-\alpha\beta a_t} \quad t = 0, \dots, T-1. \end{aligned}$$

The coefficient $\rho_{t,t}$ is strictly positive, as expected. The anticipation effect $\rho_{t,t-1}$, which measures the impact of time t tax rate on labor supply in period $t-1$ is, on the other hand, strictly negative: an increase in the effective labor supply in period t makes

the agent more willing to work in period $t - 1$, because the gain from accumulating human capital is larger. Likewise, the accumulation effect $\rho_{t,t+1}$ measuring the impact on the effective labor supply in period $t + 1$ is strictly negative, because an increase in effective labor in period t makes the agent more productive in period $t + 1$, and makes him more willing to work. An increase in the effective labor supply in period t thus increases the information rent in period t , decreases the information rent in adjacent periods, and leaves the information rent unchanged in all other periods.

Summing the rows of the matrix ρ ,⁹ one obtains the overall effect on the information rents:

$$\begin{aligned}\rho_0 &= \varepsilon + \alpha(1 + \varepsilon)\kappa \\ \rho_t &= \varepsilon - \alpha(1 + \varepsilon) \quad t = 1, 2, \dots, T,\end{aligned}$$

where $\kappa = \frac{\alpha\beta a_0}{1 - \alpha\beta a_0} > 0$. That is, the overall effect on the information rent is constant from period 1 onwards, and it is thus optimal to set the marginal tax rates constant. This does not mean, however, that the effects on the information rent are constant over time. The own effect $\rho_{t,t}$ is increasing in a_t : higher growth rate in raw labor supply n_{t+1}/n_t means that future labor supply is relatively more important. As a result, people respond less to the current tax rates, and $\rho_{t,t}$ increases. On the other hand, both the anticipation effects $\rho_{t,t-1}$ and the accumulation effects $\rho_{t,t+1}$ are negative and decreasing in a_t . An increase in the current labor supply will have a stronger effect on both next period labor supply (because next period labor supply is relatively large), and on previous period labor supply (increases gain from working in the previous period). Whenever the agents experience a higher growth rate in the raw labor supply (for example, at the beginning of one's life-cycle), the direct effect on the current information rent will be higher, but the negative effects on the information rent in adjacent periods will also be higher, but the changes in magnitudes are exactly compensating each other to produce a constant tax rates. The only exception is period 0: the incentive effect on the previous effective labor supply is missing, and the tax is going to be higher than in future periods. Note also that if $a_t \leq 1$, then the

⁹In this case, $\rho_t = \rho_{t,t-1} + \rho_{t,t} + \rho_{t,t+1}$, with appropriate modifications for $t = 0$ and $t = T$.

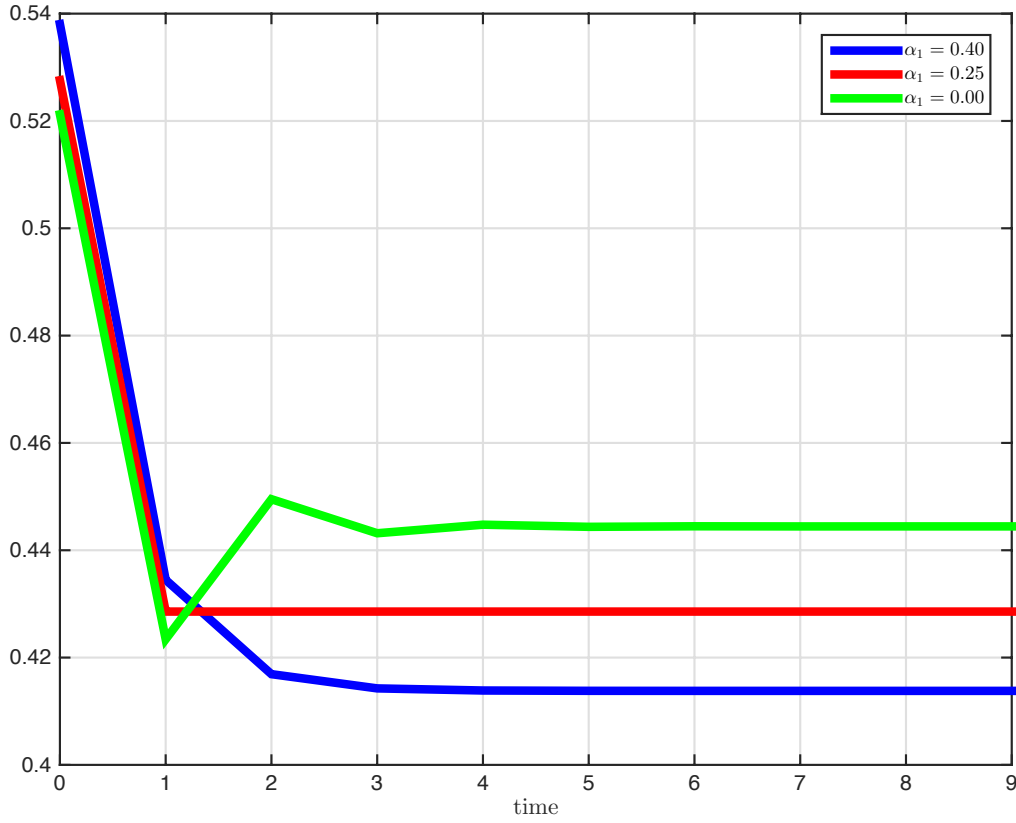


Figure 1: Optimal marginal tax rates for $G(h, n) = h^{\alpha_1} n^{\alpha_2}$, $\alpha_2 = 0.25$

accumulation effect is bigger than the anticipation effect. This will happen at the end of one's life-cycle, when $a_T = 0$. On the other hand, at the beginning of one's life-cycle, when the growth rate of labor is big, one would expect the accumulation effect to be dominated by the incentive effect.

Note also that higher α decreases τ_t for $t = 1, \dots, T$, but increases τ_0 . Higher α thus implies a larger drop in the marginal tax rates after the initial period. While one cannot say anything definitive about the level of taxes based on those considerations, it is reasonable to conjecture, that higher α will also imply lower tax rates in periods $t = 1, \dots, T$, and higher tax rates in the initial period. In particular, one can conjecture that, when compared to an economy with no human capital ($\alpha = 0$), the marginal taxes will be lower except for the initial period, when they will be higher. It follows from equation (??), that this will be true whenever the utility is linear in consumption.

The result of constant marginal tax rates is clearly an implication of the production function (28). What happens if one generalizes the production function to $G(h, n) =$

$h^{\alpha_1}n^{\alpha_2}$? Rewriting the law of motion for human capital in terms of the effective labor supply one obtains $h_{t+1} = h_t^{\alpha_1 - \alpha_2} z_t^{\alpha_2}$. Figure 4.1 shows the optimal marginal tax rates for $\alpha_2 = 0.25$, and for $\alpha_1 = 0.40$, $\alpha_1 = 0.25$, and for $\alpha_1 = 0$. If $\alpha_1 > \alpha_2$, which happens in the first case, then the current human capital has a persistent effect on the next period human capital.¹⁰ One would then expect that the missing incentive effect on the time zero human capital will persist too, with taxes declining gradually, rather than immediately after the initial period. This is indeed the case, as seen in Figure 4.1. If, on the other hand, $\alpha_1 < \alpha_2$ then, conditionally, on z , current human capital has a negative effect on the next period human capital. The optimal tax rates will then exhibit nonmonotonicities, with the optimal marginal tax rates increasing in some periods and decreasing in other periods.

5 Learning by Doing versus Learning or Doing

We will now compare the implications of the learning-by-doing model with the implications of a Ben-Porath model, where human capital and labor are two competing uses of time (Kapička (2014)). The Ben-Porath economy has the following structure. The utility function is the same as in the learning-by-doing model, i.e. it is given by equation (24). The hours worked n_t are now, however, divided between time spent producing output l_t and time spent accumulating human capital s_t :

$$n_t = l_t + s_t.$$

The output produced is then

$$y_t = \theta h_t l_t,$$

which is the same technology as in (22), but with only l_t being the time spent producing output.

The human capital production function depends now on time spent by investing

¹⁰The effect is not permanent as long as $\alpha_1 - \alpha_2 < 1$

in human capital. It is given by

$$h_{t+1} = G(h_t, s_t, t), \quad (30)$$

where h_0 is given, and $s_t \in \mathbb{R}_+$ is investment in human capital in period t . The function $G : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is increasing and concave in both arguments. 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 $S(h_t, h_{t+1})$. Properties of G imply that S is decreasing in its first argument, increasing in its second argument, and differentiable. Agent's human capital is publicly unobservable. The only exception is the initial human capital h_0 , which is identical for all agents and is publicly observable.¹¹

To map the Ben-Porath model to the theoretical framework of section 2, define the effective labor supply again by $z_t = h_t n_t$. The function $V(z^T)$ is then given by

$$V(z^T) = \max_{\{h_{t+1}\}_{t=0}^T} \sum_{t=0}^T \beta^t X\left(\frac{z_t}{h_t}, S(h_t, h_{t+1})\right), \quad (31)$$

where the dependence on the initial human capital h_0 is kept implicit. Properties of X imply that V is increasing in all arguments and twice differentiable. Convexity, however, is not guaranteed, and must be checked on a case-by-case basis.

Let $(H_t(z^T))_{t=0}^T$ be the human capital sequence that attains the minimum of (31). By the envelope theorem, we get

$$V_{z_t} = \beta^t X_{n,t} \left(\frac{z_t}{H_t(z^T)}, S\left(H_t(z^T), H_{t+1}(z^T)\right) \right) \frac{1}{H_t(z^T)}.$$

The maximization problem (31) yields a function V that is, in general, not separable in the effective labor supply, because human capital decision in any given period depends on all past and future values of labor supply. In particular, the optimal choice of human capital can be characterized by the following Euler equation that

¹¹Unlike the learning-by-doing model, it is possible to have a learning-or-doing model with observable human capital and unobservable hours worked. Such a model has been studied in Boháček and Kapička (2008).

equates marginal costs of investing in human capital to marginal benefits of doing so:

$$X_{s,t} \frac{1}{G_{s,t}} \geq \beta X_{n,t+1} \frac{z_{t+1}}{h_{t+1}^2} + \beta X_{s,t+1} \frac{G_{h,t+1}}{G_{s,t+1}}, \quad = \text{ if } s_t > 0, \quad (32)$$

where, to reduce notation, $X_{s,t} = X_s(\frac{z_t}{h_t}, s_t)$, $G_{s,t} = G_s(h_t, s_t)$, and similarly for $X_{n,t}$, and $G_{h,t}$. The Euler equation (32) is a second order difference equation that can be solved given the initial condition h_0 and a terminal condition $h_T = G(h_{T-1}, 0)$.

5.1 Parametrization of the Models

I will now calibrate the parameters of both models, and compare the optimal tax implications. I assume that the utility is linear in consumption, $U(c) = c$. I assume that the disutility from working takes the isoelastic form in 27. The human capital production function is

$$G(h, z, t) = (1 - \delta)h + Ae^{-dt}hz^\alpha$$

in the learning-by-doing model, and

$$G(h, s, t) = (1 - \delta)h + Ae^{-dt}hs^\alpha$$

in the learning-or-doing model. In both specifications the parameter δ is the rate of depreciation of human capital, A measures the initial efficiency of human capital investments, and d measures the rate at which the efficiency of human capital investments declines over time.

Since the efficiency condition (7) does not depend directly on the distribution of abilities and I will focus on the intertemporal profile of the optimal taxes, I will simplify the exercise as follows. I will choose the parameters of the utility function and of the human capital production functions in a representative agent economy with no heterogeneity. In choosing the parameters I closely follow Wallenius (2011). I will then solve for the optimal intertemporal profile of taxes for an arbitrary value

of $\frac{P-F}{\theta f}$ by using equation (12).

In both models I solve for a partial equilibrium where the agents maximize the utility function (24) subject to the production function for human capital, output production function, and the budget constraint

$$\sum_{t=0}^T R^{-t} c_t = \sum_{t=0}^T R^{-t} (1 - \tau_t) y_t + tr \quad (33)$$

taking R as given, where τ_t is the marginal tax rate and tr are government transfers that clear the government budget constraint.

Wallenius (2011) shows that, perhaps surprisingly, different values of ρ and δ have very little effect on the ability of the models to fit the data. I thus set $\alpha = 0.25$, which is consistent with the magnitude of what Best and Kleven (2013) call the behavioral career effect in a two period model. I also set $\delta = 0.03$, and assume that the marginal tax rate $\tau = 0.4$.

For both models I will then choose A, d, b, ϵ and β to minimize the sum of squares of the difference between hours and wages in the model and hours and wages in the data. I take the data on hours worked and wages directly from Wallenius (2011), who constructs the average life-cycle profiles of hours worked and wages using CPS data for the years 1976-2006 for ages 20-62.¹² In the learning-by-doing model hours and wages that correspond to measured hours and wages are simply n_t and h_t . In the learning-or-doing model, however, the notion of measured hours and wages is less straightforward. I assume that both l_t and s_t are included in the measured hours worked, and that measured wages are equal to the average compensation, i.e. total compensation $h_t l_t$ divided by total hours worked n_t .

Table 1 shows the resulting parameters for both models. Most of the parameter values are relatively similar, but there are slight differences. The parameter ϵ , which determines the intertemporal elasticity of substitution of labor is somewhat higher in the learning-or-doing model, and the efficiency of human capital investments declines faster in the learning-by-doing model. Figure 5.1 shows the model fit by plotting the resulting life-cycle profiles in hours and wages, and comparing them to the data.

¹²See Wallenius (2011) for details.

Table 1: Model Parameters

Parameter	Learning by Doing	Learning or Doing
α	0.2500	0.2500
δ	0.0300	0.0300
τ	0.4000	0.4000
β	0.9605	0.9647
ε	1.2022	1.4465
d	0.0389	0.0136
A	0.1161	0.1109
b	3.9977	2.9280

Both models are able to replicate well the hump shaped profiles observed in the data, although the fit is less satisfactory at the beginning of the life-cycle for hours worked, and at the end of life-cycle for wages.

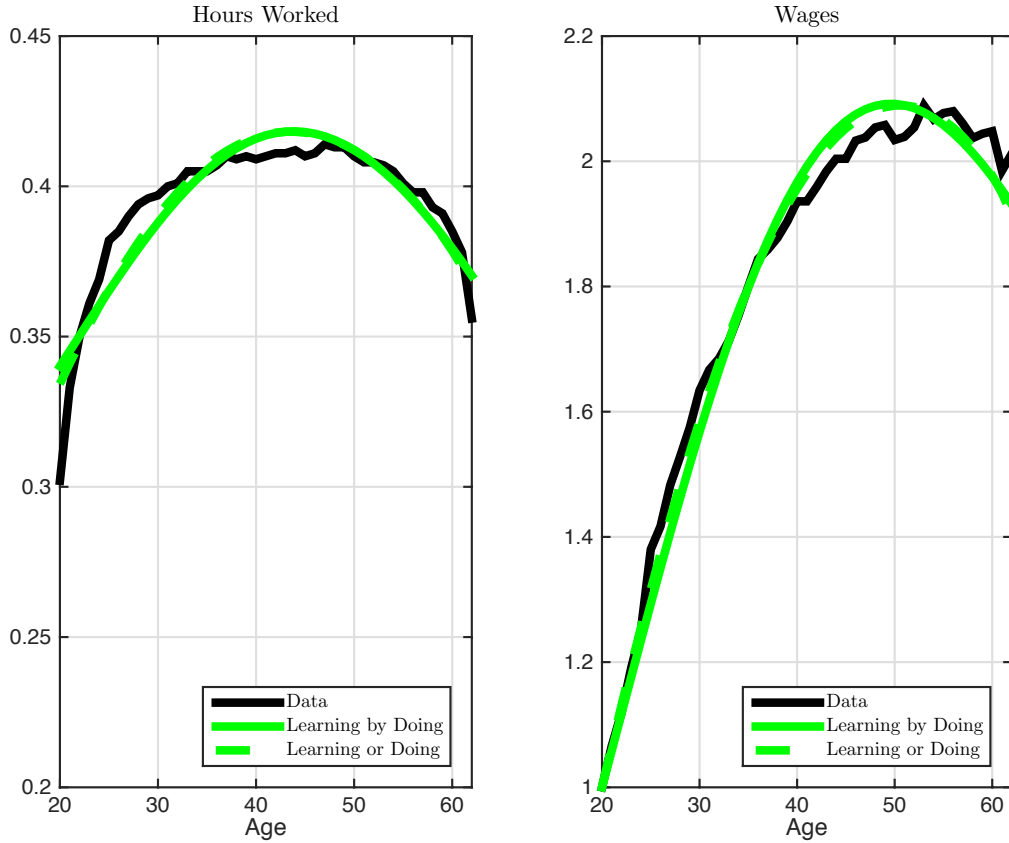


Figure 2: Hours worked and wages: data vs. models

Although both model are similar in terms of matching the data, they predict very different responses to changes in the tax rates. Consider a temporary increase in taxes. In both modes the current time spent producing output (n_t in the learning-

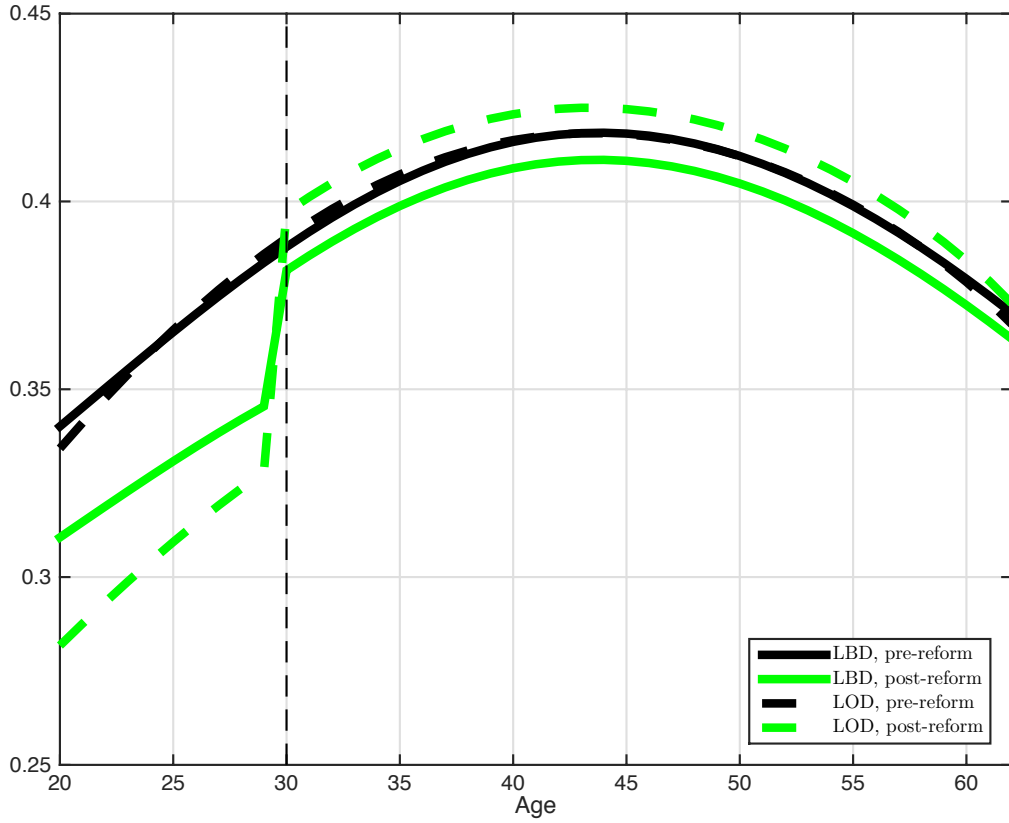


Figure 3: A temporary (age 20-29) marginal tax increase: response of hours worked

by-doing model and l_t in the learning-or-doing model) goes down. This leads to a decrease in the stock of human capital in the learning-by-doing model, and a decrease in future hours worked. In the learning-or-doing model, however, lower l_t leads to an increase in the time spent accumulating human capital s_t , and increases future stock of human capital. As a result, future hours worked go up, in contrast to the learning-by-doing model. This is illustrated in Figure 5.1, where the marginal tax rate temporarily increases from 40% to 50% in the first 10 periods (age 20-29). For the remaining ages 30-62, the tax rates stays at 40%, as in the calibration. The figure shows that a temporary tax increase leads to a permanent decrease in hours worked in the learning-by-doing model, but a permanent increase in hours worked in the learning-or-doing model.¹³

¹³Similarly, a permanent future increase in the marginal tax rate increases current hours worked in the learning-or-doing model, but decreases them in the learning-by-doing model.

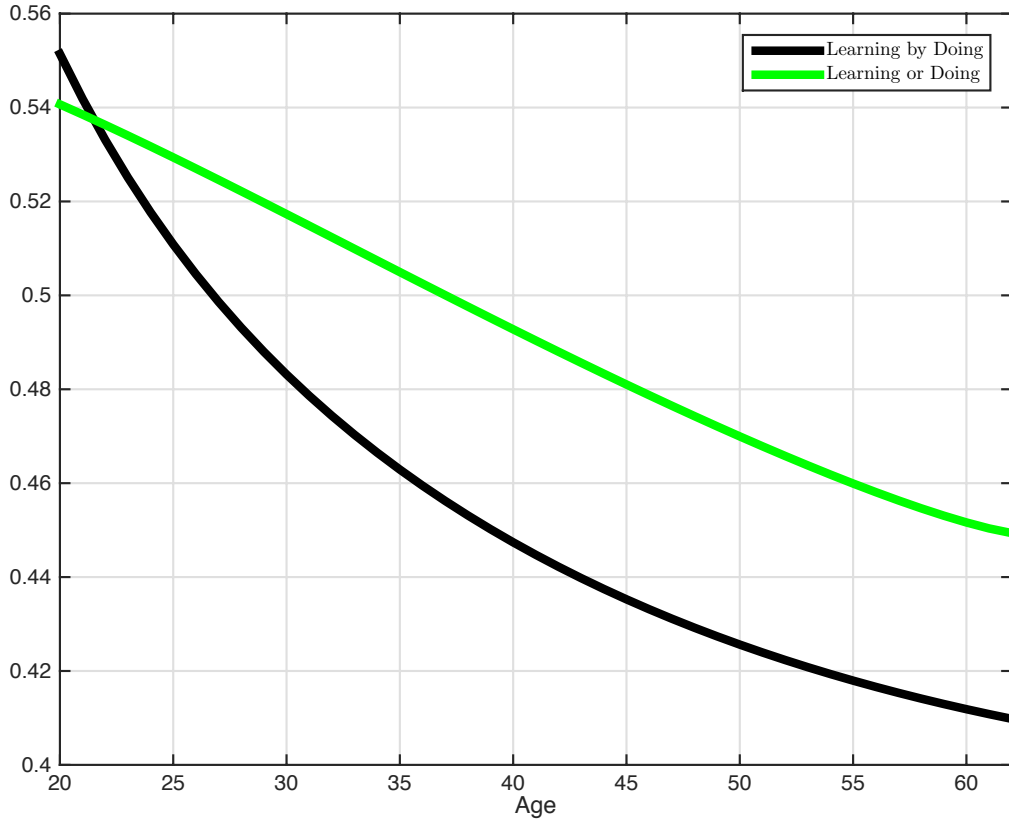


Figure 4: Optimal marginal tax rates

5.2 Pareto Efficient Tax Reform

Despite the fact that both models show a very different response to any given change in the tax rates, I will now show, that the optimal tax reform looks quite similar in both environments. In Figure 5.2 I show the optimal marginal tax rates for $\frac{G-F}{\theta_f} = 0.5$. In both models, the optimal marginal tax rates are decreasing over the life-cycle. The decrease is somewhat stronger in the learning-or-doing model, but in both cases the decrease is significant: In the learning by doing model taxes decline by about 12%, while in the learning-or-doing model they decline by about 8%.

Given that both models react very differently to a given change in taxes, it seems puzzling that they prescribe a similar intertemporal profile of the marginal tax rates. To understand this result, Figure 5.2 inspects the parameters ρ that determine the optimal marginal tax rates, and decomposes them into the own effects, the anticipation effects, and the accumulation effects. The figure shows that, while the overall effect ρ_t is similar in both models, its components behave very differently. Consider first

the own effect $\rho_{t,t}$. In the learning-by-doing the own effect is decreasing with age, but in the learning-or-doing model the own effect is increasing. It is helpful to think of the own effect as an inverse of the own-Frisch elasticity of labor supply. In the learning-by-doing model the own-Frisch elasticity is relatively low at the beginning of the life-cycle, because labor supply has benefits that extend beyond the current period. At the end of the life-cycle, the own-Frisch elasticity increases. Hence $\rho_{t,t}$ is decreasing with age. On the other hand, in the learning-or-doing model the own-Frisch elasticity is relatively large at the beginning of the life-cycle when time spent producing output is only a small fraction of the overall time spent working. At the end of the life-cycle, time spent by accumulating human capital goes to zero, and the own-Frisch elasticity decreases. As a result, $\rho_{t,t}$ is increasing with age.

The anticipation effect and the accumulation effect also behave very differently in both models, both in terms of their signs, and in terms of their changes over the life-cycle. In the learning-by-doing model, the accumulation effect is strongly negative: An increase in the current labor supply increases the incentive to supply labor in the future, as the discussion of Figure 5.1 has shown. This implies that $\rho_t^+ < 0$. Similarly, an increase in future labor increases incentives to supply labor today, and so $\rho_t^- < 0$. The magnitude of those effects depends on how many periods there is in the future and in the past, and so the accumulation effect ρ_t^+ increases with age, and the anticipation effect ρ_t^- decreases with age. Consider now the learning-or-doing model. An increase in the effective labor supply at age t has an ambiguous effect on the incentives to supply labor in the future. On one hand, it decreases schooling at age t , which decrease future incentives. On the other hand, it increases human capital investment in previous periods, which increases incentives to supply labor in the future. At early ages the first effect dominates, and so ρ_t^+ is positive, while at later ages the second effect dominates, making ρ_t^+ negative. Likewise, an increase in the effective labor supply at age t has an ambiguous effect on the incentives to supply labor in the past, because it increases human capital investments (decreasing incentives to supply labor), but the stock of human capital raises too, which in turn increases the incentives to supply labor. At early ages the first effect dominates, and

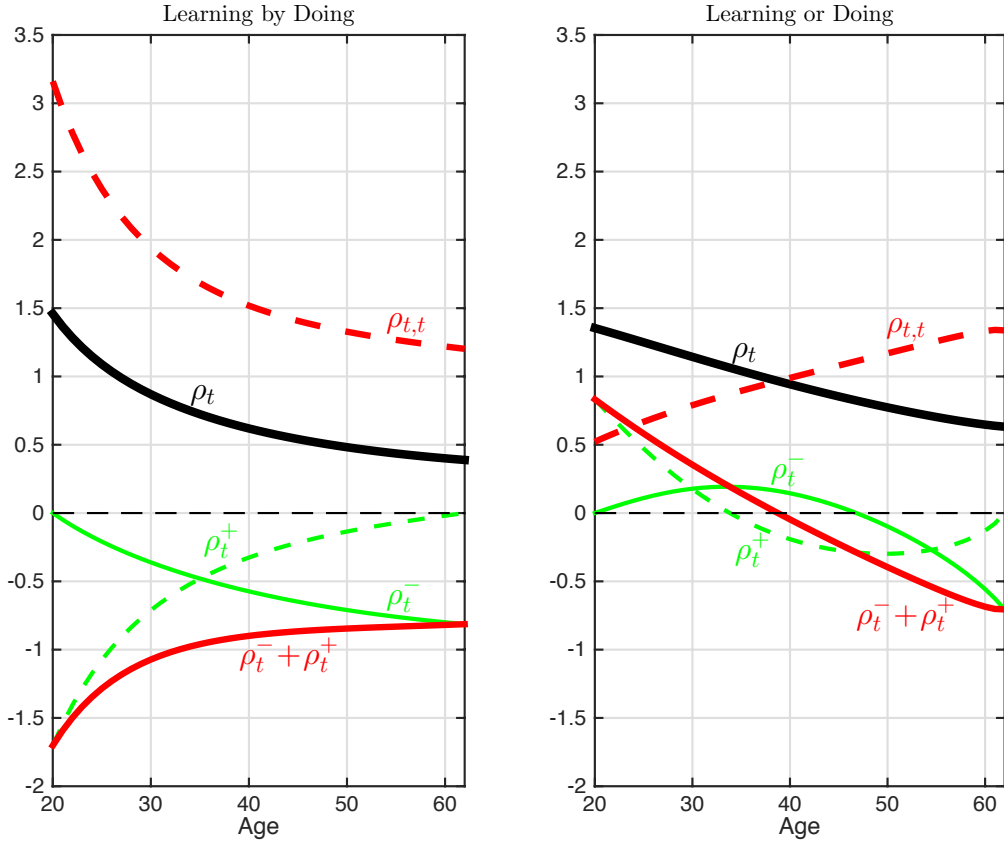


Figure 5: Own effects, anticipation effects, and accumulation effects

so ρ_t^- is positive, while at later ages the second effect dominates and so ρ_t^- is negative. Overall, the sum of ρ_t^- and ρ_t^+ is first positive, decreases with age, and ends up being negative at the end of the life-cycle.

The anticipation and accumulation effects taken together are thus increasing in the learning-by-doing model, and decreasing in the learning-or-doing model. In both cases their intertemporal profile is the *opposite* of the intertemporal profile of the own effect. The own effect, however, dominates in the learning-by-doing model, while the opposite is true in the learning-or-doing model. As a result, the total effect decreases with age in both models, and they deliver similar predictions about the optimal policies.

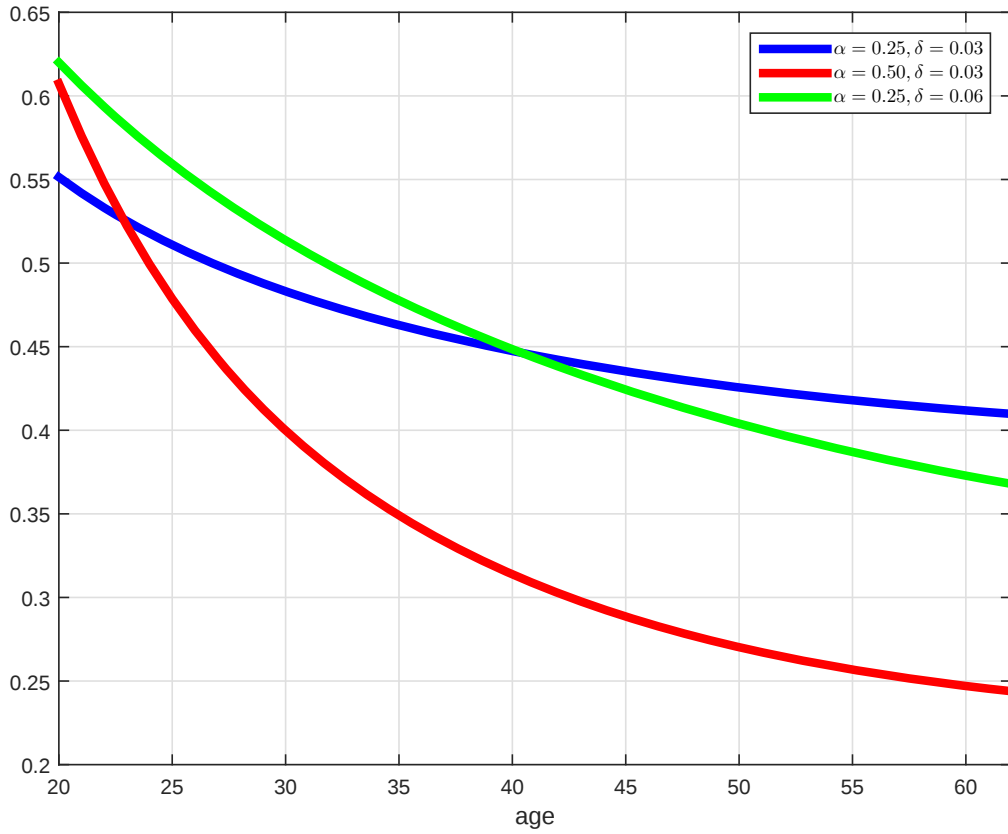


Figure 6: Learning-by-doing, alternative assumptions

5.2.1 Alternative assumptions

Figure 5.2.1 investigates the implications of alternative parameter values for the optimal marginal tax rates in the learning-by-doing model. Specifically, it shows the optimal marginal tax rates when α is 0.5 rather than 0.25, and when $\delta = 0.06$ rather than 0.03. In both cases, the marginal income tax rates are decreasing at a higher rate than in the benchmark. In the first case they decrease from 61% to 25%, while in the second case they decrease from 62% to 37%. The results reaffirm previous findings, and indicate that the conservative benchmark parametrization is a lower bound on the possible decline in the marginal tax rates.

6 Conclusions

This main goal of this paper is to provide a unifying framework for studying Pareto efficient income taxation with endogenous, unobservable, human capital formation. I use the framework to show that Mirrlees optimal taxation delivers results that are qualitatively different from what Ramsey taxation predicts.

The advantage of having a unifying framework is that it is easy to compare different models, and their normative implications. I do so for the learning-by-doing model, and the learning-or-doing model. The surprising finding is that both models have roughly similar predictions about the intertemporal profile of the optimal marginal tax rates: the tax rates should be decreasing over the life-cycle. The result is surprising, because both models use very different mechanisms of human capital formation, and exhibit very different reactions to changes in tax policies. I show that, although the result is similar in the end, there are different forces in play, and only when they are taken together a similar pattern emerges. The results have been obtained by using numerical simulations, and the results appear to be robust to alternative specifications.

In future work I plan to extend the results to environments with stochastic shocks to investigate how the optimal tax formulas will be modified when the disutility of work is nonseparable.

Appendix: Proofs

Before Proving Lemma 6, the following preliminary result is needed:

Lemma 7. *Suppose that A is a real symmetric tridiagonal matrix of the form*

$$A = \begin{pmatrix} a_0 & b_1 & 0 & \dots & 0 \\ b_1 & a_2 & b_2 & \dots & 0 \\ 0 & b_2 & a_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_T \end{pmatrix}$$

Then the matrix A is positive definite if $a_0 > 0$ and $a_t a_{t+1} > b_t^2$ for all $t = 0, 1, \dots, T-1$.

Proof. The matrix A is positive definite if all its leading principal minors are positive. The leading principal minors satisfy a well known recurrence relation

$$m_t = a_t m_{t-1} - b_t^2 m_{t-2}, \quad t = 0, \dots, T,$$

with $m_{-2} = 0$ and $m_{-1} = 1$. Both m_0 and m_1 are strictly positive given the assumptions. Now write $f_t = \frac{m_t}{m_{t-1}}$ and rewrite the recurrence relation as

$$f_t = a_t - \frac{b_t^2}{f_{t-1}}, \quad t = 0, \dots, T,$$

with $f_{-1} = \infty$. Both f_0 and f_1 are strictly positive given the assumptions. Suppose now that f_{t-2} and f_{t-1} are both positive. Then $f_{t-1} = a_{t-1} - \frac{b_{t-1}^2}{f_{t-2}} < a_{t-1}$, implying that f_t is positive if

$$a_t > \frac{b_t^2}{f_{t-1}} > \frac{b_t^2}{a_{t-1}}.$$

Thu, f_t is positive. Since a_{t-1} is positive, m_t is positive as well. ■

Proof of Lemma 6. The function $V(z^T)$ is strictly convex in z^T if all the principal minors M_t of its Hessian (V_{z_t, z_k}) are positive. The Hessian is a symmetric tridiagonal matrix with its nonzero elements given by

$$\begin{aligned} V_{z_t, z_t} &= \frac{\beta^t}{z_t^2} [\varepsilon x_t + \beta \alpha (1 + \alpha(1 + \varepsilon)) x_{t+1}] \\ V_{z_t, z_{t-1}} &= -\frac{\beta^t}{z_t z_{t-1}} \alpha (1 + \varepsilon) x_t, \end{aligned}$$

where $x_t = \left(\frac{z_t}{A_t z_{t-1}^\alpha} \right)^{1+\varepsilon}$ for $t = 0, 1, \dots, T+1$, where it is assumed that $z_{-1} = 1$ and $z_{T+1} = 0$ (implying $x_{T+1} = 0$). It is easy to show that the principal minors $(M_t)_{t=0}^T$ can be written as $M_t = \left(\prod_{k=0}^t z_k^{-2} \right) m_t$, with $(m_t)_{t=0}^T$ are the principal minors of the following matrix h that is given by

$$\begin{aligned} h_{t,t} &= \beta^t \varepsilon x_t + \beta^{t+1} \alpha (1 + \alpha(1 + \varepsilon)) x_{t+1} \\ h_{t,t-1} &= -\beta^t \alpha (1 + \varepsilon) x_t, \end{aligned}$$

and zero everywhere else. The matrix h is therefore real, symmetric, tridiagonal, and has $h_{0,0} > 0$. It follows from Lemma 7 that h is positive definite if

$$[\varepsilon x_t + \alpha\beta(1 + \alpha(1 + \varepsilon))x_{t+1}] [\varepsilon x_{t+1} + \alpha\beta(1 + \alpha(1 + \varepsilon))x_{t+2}] > \beta\alpha^2(1 + \varepsilon)^2 x_{t+1}^2,$$

which holds if $\varepsilon - \alpha(1 + \varepsilon) > 0$. ■

References

- Atkinson, A. B. and J. E. Stiglitz (1980). *Lectures on public economics*. New York, NY: McGraw-Hill. 14
- Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy* 75(4), 352–365. 3
- Best, M. C. and H. J. Kleven (2013). Optimal income taxation with career effects of work effort. Technical report, London School of Economics. 3, 5, 15, 17, 27
- Boháček, R. and M. Kapička (2008). Optimal human capital policies. *Journal of Monetary Economics* 55(1), 1–16. 6, 25
- Bovenberg, A. L. and B. Jacobs (2005). Redistribution and education are Siamese twins. *Journal of Public Economics* 89, 2005–2035. 6
- Cossa, R., J. J. Heckman, and L. Lochner (1999). Wage subsidies and skill formation: A study of the earned income tax credit. Technical report, University of Chicago. 4
- Farhi, E. and I. Werning (2012). Insurance and taxation over the life-cycle. *Review of Economic Studies* 80(2), 596–635. 5
- Findeisen, S. and D. Sachs (2014). Education and optimal dynamic taxation: The role of income-contingent student loans. Technical report, University of Zurich. 6
- Golosov, M., N. R. Kocherlakota, and A. Tsyvinski (2003). Optimal indirect and capital taxation. *The Review of Economic Studies* 70(3), 569–587. 5

- Golosov, M., A. Tsyvinski, and M. Troshkin (2013). Redistribution and social insurance. Working paper, Yale University. 5
- Imai, S. and M. P. Keane (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review* 45(2), 601–641. 2
- Kapička, M. (2014). Optimal Mirrleesian taxation with unobservable human capital formation. *AEJ: Macroeconomics*, forthcoming. 2, 4, 5, 6, 24
- Kocherlakota, N. R. (2005). Zero expected wealth taxes: A Mirrless approach to dynamic optimal taxation. *Econometrica* 73(5), 1587–1621. 5
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies* 38(2), 175–208. 5
- Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics* 6(4), 327–358. 5
- Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies* 68(1), 205–229. 5, 15
- Stancheva, S. (2014). Optimal taxation and human capital policies over the life cycle. Working paper, Harvard University. 6
- Wallenius, J. (2011). Human capital accumulation and the intertemporal elasticity of substitution of labor: How large is the bias? *Review of Economic Dynamics* 14, 577–591. 3, 26, 27