

Education Policies and Taxation without Commitment *

Sebastian Findeisen

University of Mannheim, CEPR

Dominik Sachs

European University Institute, CEPR

This version: August 27, 2016

Abstract

We study the implications of limited commitment on education and tax policies chosen by benevolent governments. Individual wages are determined by both innate abilities and education levels. Consistent with real world practices, the government can decide to subsidize different levels of education at different rates. The lack of commitment influences the optimal structure of education subsidies. The direction of the effect depends on the design of labor taxes. With linear labor tax rates and a transfer for redistribution, subsidies become more progressive. By contrast, if the government is only constrained by informational asymmetries when designing taxes, subsidies become more regressive.

***JEL-classification:* H21, H23, I21**

Keywords: Education Policies, Time-Inconsistency, Taxation, Inequality

*Contact: findeisen@uni-mannheim.de, dominik.sachs@eui.eu. We thank Friedrich Breyer, Leo Kaas, Fabrizio Zilibotti, Josef Zweimueller and seminar participants in Konstanz and Zurich for helpful comments and discussions, and Bela Figge for research assistance. Sebastian Findeisen acknowledges the support of the University of Zurich (Forschungskredit of the University of Zurich, grant no. 53210603). Dominik Sachs' research was partly funded by a post-doc fellowship of the Fritz-Thyssen Stiftung and the Cologne Graduate School in Management, Economics and Social Sciences.

1 Introduction

Public finance economists have long recognized that the challenges involved in the design of optimal education policies and income tax systems are intimately related. Income taxation influences the incentives to invest in education.¹ Education subsidies and policies, in turn, influence the choice of an optimal income tax system as they have a direct effect on both the level and the distribution of wages. Many papers have studied the design of education and tax policies jointly from a normative perspective – see, for example, Bovenberg and Jacobs (2005) for a state-of-the-art treatment in a heterogeneous agent model.² This strand of literature assumes that individuals rationally make human capital investment decisions, reacting to incentives set by the tax code and education subsidies. Importantly, the government fully commits to the income tax schedule that it announces before education decisions are made.

Boadway et al. (1996) have drawn attention to the issue of time-consistency, in the spirit of Kydland and Prescott (1977), inherent in the design of optimal tax and education policies. If the government lacks a device to credibly commit to tax policies at the time individuals make education decisions, this can dramatically depress the incentives of young individuals to invest into human capital. In their framework, they show that this underinvestment arises and make a case for mandatory education as a second-best policy in the presence of commitment problems.

This paper looks at the implications of limited commitment and policy credibility on education and tax policies from a new perspective. Consistent with real world practices, the government can decide to subsidize different levels of education at different rates. The idea here is that governments typically intervene at primary, secondary and tertiary education levels. However, the rate at which these different education levels are subsidized is very different. We formalize this by allowing the government to set a nonlinear schedule of education subsidies. We derive our results in a transparent and simple heterogeneous

¹See Abramitzky and Lavy (2012) for recent quasi-experimental evidence on the negative effect of redistributive taxation on education investment. More structural and model based approaches as the classic work by Trostel (1993) also have found big effects of income taxation on human capital investment.

²See Richter (2009) for a recent treatment in a Ramsey setting with a representative agent. See Da Costa and Maestri (2007) and Anderberg (2009) for a Mirrlees treatment with ex-ante homogeneous agents and uncertainty.

agent model with two types (Stiglitz 1982). Consistent with empirical evidence, individual wages are determined by both innate abilities and education levels. We show that the effect of a lack of commitment depends on the structure of the labor tax. First, we analyze a linear labor income tax schedule with a lump-sum transfer as in Sheshinski (1972). Second, we study income taxation in the tradition of Mirrlees (1971), where the planner is only constrained by informational asymmetries – we often refer to the latter case as nonlinear taxation.

Linear Labor Tax Rates. We start with the benchmark of full commitment. The optimal linear income tax rate takes into account education incentives and is lower than in Sheshinski (1972). Education subsidies for the high types are set such that a first-best rule for education is fulfilled: the subsidy corrects for the fiscal externality (Bovenberg and Jacobs 2005). For the low type, in addition to this correction of the fiscal externality, education is downwards distorted at the margin to relax the incentive constraint of the high type.

With commitment problems, tax promises of the government lack credibility and individuals rationally anticipate that the government might re-optimize after education is sunk. More concretely, the government can deviate from its announcements but this induces some output costs capturing the idea of a reputational loss. In equilibrium, however, deviation does not occur.³

The optimal tax rate is larger than with full commitment. The intuition is that for the deviating government education is sunk and, hence, it taxes labor at a high rate (as if education decisions were exogenous). On the equilibrium path, the government anticipates that the full commitment tax rate is not credible and sets a higher tax rate to make deviation less attractive at the margin.

Education subsidies become more progressive compared to the full commitment benchmark. Key is that in the case of deviation, the government chooses a higher tax rate. A more unequal distribution of wages makes such a deviation and the higher tax rate more attractive for the planner. This is because the planner values redistribution from high earners

³Farhi et al. (2012) show how to microfound such an output loss in a dynamic repeated game, where a deviation today brings a reputational cost borne in the future, because of depressed investment of future generations.

to low earners and the incentive to re-optimize – and set a higher tax rate – increases when the wage differential between the skill groups is large. For education policies this means that a higher education subsidy for low types and a lower education subsidy for high types will help to limit the commitment problem by compressing the distribution of education and, ultimately, wages in the next period.

Nonlinear Labor Tax Rates. We first study the benchmark case with full commitment. In this case the low type faces a positive marginal tax rate on labor income whereas the high type faces a zero marginal tax rate. The latter is a standard result in the theory of nonlinear taxation and we address the intuition of it in more detail in the main body. As a consequence, only education of the low type is subsidized at the margin. For the high type there is no reason to subsidize education as there is no tax distortion.

We then introduce lack of commitment and show that the main result from the linear tax case is reversed when labor taxes are nonlinear. Limited commitment to tax policies makes education policies more *regressive*, when the government uses nonlinear labor income taxation.

The key to this result lies again in the different labor tax policies chosen by the deviating government (which treats education as sunk) and the government sticking to its promises. In the case of deviation, the government re-optimizes after having learned the type of each individual and redistributes with type-dependent lump-sum taxes and therefore sets zero labor wedges. By contrast, the non-deviating government that sets tax policies in the first period, sets a positive labor wedge for the low type. This implies that labor supply of the low type is always higher in the deviation case. More education for the low type leads hence to a higher output gain for a deviating government as compared to the non-deviating government. To make this deviation less attractive and resist its own temptation to re-optimize, the government discourages education of the low type at the margin, making education policies more regressive. The labor supply of the high type is not distorted in both cases and so there is no difference in labor supply in both cases. As a consequence her education distortion is zero also with partial commitment; distorting this decision cannot dampen the temptation to deviate.

Literature. This paper is related to Farhi et al. (2012) who consider capital taxation without commitment. They establish an important benchmark: lack of commitment makes savings taxes progressive. The important difference between *human capital* and savings is that a more compressed wealth distribution makes a deviation always less likely whereas a more compressed human capital (wage) distribution can make a deviation less but also more likely, depending on labor taxes. In the case with physical capital, wealth can always be taxed and redistributed directly by the government. Human capital in contrast is taxed indirectly only by the labor income tax, which creates a labor supply distortion. We show that the effect of limited commitment on education policies will depend on how labor taxes are set, and with nonlinear labor taxation one obtains in fact a reversal of the result that more inequality worsens credibility, because more inequality makes a deviation less tempting.

Related papers are also found in the literature analyzing the dynamics and stability of redistributive policies, especially the articles by Hassler et al. (2003, 2005). Here current and expected redistributive policies also influence the productivity distribution of future voters by influencing human capital investments. Our paper is less interested in the rich dynamics that those papers characterize, but adds education subsidies to the picture, which interact with redistributive labor tax policies in equilibrium.

This paper also relates to the work on time inconsistency and education policies by Konrad (2001) and Andersson and Konrad (2003).⁴ Konrad (2001) shows how the time inconsistency problem is alleviated by the presence of private information in an optimal taxation framework with idiosyncratic uncertainty. In particular, he shows that the strong no-education result obtained in Boadway et al. (1996) no longer applies, as with private information some rents of education are still captured by individuals, preserving some incentives to invest in education.⁵

Lastly, Andersson and Konrad (2003) investigate education policies chosen by extortory governments lacking commitment and how migration and tax competition affect

⁴In a related paper, Pereira (2009) studies linear education subsidies and shows that this subsidy offsets some of the excessive redistribution from income taxes, when the government lacks commitment.

⁵Poutvaara (2003) shows that redistribution without commitment may still involve more education than in the laissez-faire if the insurance effect of taxes is important.

policies.⁶ We depart from these papers by placing our focus on nonlinear education subsidies as used in the real world.

2 Environment

We consider a two-period model, where ex-ante heterogeneous agents make an educational investment in the first period. In the second period, they make a labor-leisure decision. More formally, there are two types of agents to which we refer to as low ability and high ability type. Their masses are f_l and f_h with $f_l + f_h = 1$ and the type is private information. In period 1, they make a monetary educational investment e . The wage w they earn in period 2 is a function of innate type and education, i.e. $w_i(e)$ for $i = l, h$.

We impose three intuitive assumptions on the wage function. First, education is productive and raises wages $\frac{\partial w_i(e)}{\partial e} > 0$ for $i = l, h$. Second education and innate ability are complements implying higher marginal returns to education for the higher innate type: $\frac{\partial w_h(e)}{\partial e} - \frac{\partial w_l(e)}{\partial e} > 0$. Finally, innate abilities positively influence wages for a given level of education: $w_h(e) - w_l(e) > 0$. None of these assumptions is needed for most of the results we derive in the sense that all formulas are valid if we deviate from those assumptions. These assumptions ease the understanding of the model, however, and have strong empirical support. E.g., Card (1999) provides a comprehensive review of the literature estimating the *causal* effect of education on earnings. Carneiro and Heckman (2005) and Lemieux (2006), among others, document complementarity between innate skills and formal education. Taber (2001) and Hendricks and Schoellman (2012) suggest that much of the rise in the college premium may be attributed to a rise in the demand for unobserved skills, which are predetermined and independent of education.

We assume quasi-linear preferences. The utility functions are $U^1 = c^1$ in period 1 and $U^2 = c^2 - \Psi(h)$ in period 2, where h are hours worked. For simplicity, we assume that $\Psi(h) = \frac{h^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$, i.e. that Ψ exhibits a constant elasticity of labor supply ε . Before tax income is denoted by $y_i = w_i(e_i)h_i$ for $i = l, h$. Further we assume no discounting and a zero interest

⁶In a median voter framework, Poutvaara (2011) shows that generous subsidies for higher education may make the median voter of the future a college graduate, leading to lower taxes compared to a world with lower subsidies for high education. Relatedly, Poutvaara (2006) studies a median voter model with voting on social security benefits and higher public education. He shows that in the case with multiple equilibria, higher wage taxes are correlated with a higher provision of public education.

rate for notational convenience. Note that the allocation of intertemporal consumption is generally not pinned down with this utility function. We nevertheless distinguish between first and second period consumption as it will make a difference for the nonlinear tax problem without commitment.

We are considering redistributive taxation. That is, we are interested in the policies of a government that is interested in redistributing from the high type to the low type. To capture this redistributive concern, we set the Pareto weights \tilde{f}_l and \tilde{f}_h such that $\frac{\tilde{f}_l}{\tilde{f}_h} > \frac{f_l}{f_h}$. Thus, the government's objective is $\sum_{i=l,h} \tilde{f}_i (U_i^1 + U_i^2)$.

When deciding about the optimal degree of redistribution, the government has to take into account that taxes will (i) lower incentives to work and also (ii) lower incentives to invest in education. Concerning the sophistication of policy instruments, we consider two scenarios. In Section 3, we consider a planner that can use nonlinear education subsidies but only has access to linear labor income taxes. The revenue of this linear tax is used to finance the education subsidies and a lump-sum transfer in period 2. This captures a simple negative-income tax system with a linear marginal tax rate that has first been studied by Sheshinski (1972). In section 4, we assume that the government is only constrained by informational asymmetries in the tradition of the mechanism-design approach. This implies two changes: firstly, labor income tax rates can now be nonlinear. Secondly, in the case of deviation there is no informational asymmetry. A deviating government has all the information about types because types were revealed in the education period. A deviating government can therefore apply individualized lump-sum taxation.

3 Linear Tax Instruments

As a benchmark, we first look at the case with exogenous education in Section 3.1. We then study optimal policies with full commitment and endogenous education in Section 3.2, before we analyze the implications of limited commitment in Section 3.3.

3.1 Optimal Policies with Exogenous Education

Before looking at optimal policies in the different commitment scenarios, we look at the simple benchmark case of exogenous education where commitment issues do not arise.

For this purpose, consider a one period setting where education levels e_l and e_h are exogenous. In that case, the only relevant margin for the government, when choosing taxes, is the labor-leisure margin. Denote by t the linear tax rate and by T the lump-sum transfer. The problem of the government then simply is

$$\begin{aligned} \max_t \quad & \tilde{f}_l \left((1-t)w_l(e_l)h(t, w_l(e_l)) + T - \Psi[h(t, w_l(e_l))] \right) \\ & + \tilde{f}_h \left((1-t)w_h(e_h)h(t, w_h(e_h)) + T - \Psi[h(t, w_h(e_h))] \right) \end{aligned} \quad (1)$$

subject to a government budget constraint

$$T = t \left(f_l w_l(e_l) h(t, w_l(e_l)) + f_h w_h(e_h) h(t, w_h(e_h)) \right) \quad (2)$$

and optimal labor supply of the individuals

$$h(t, w_i) = \arg \max_h (1-t)h w_i(e_i) - \Psi(h).$$

The government only has to choose t optimally and thereby take into account how the transfer T is determined by the government budget constraint (2) and how individuals' hours worked h respond. It is then easy to show that the optimal linear tax rate t^{ex} , in this case with exogenous human capital, satisfies

$$\frac{t^{ex}}{1-t^{ex}} = \frac{(\tilde{f}_l - f_l) \left(\frac{y_h - y_l}{\bar{y}} \right)}{\varepsilon}, \quad (3)$$

where \bar{y} is average income $f_l y_l + f_h y_h$. The optimal tax rate is increasing in redistributive preferences $(\tilde{f}_l - f_l)$, increasing in inequality measured by $\frac{y_h - y_l}{\bar{y}}$ and decreasing in the elasticity of labor supply. The formula (3) is a variation for the optimal linear tax rate of

Sheshinski (1972).⁷ We refrain from providing a formal proof for this simple case as it is nested in the following formulas with endogenous educational attainment.

3.2 Optimal Policies with Full Commitment

3.2.1 The Government's Problem

We now consider the case where the educational decision is endogenous and the government can influence the decision of the agents by setting a nonlinear subsidy schedule. Thus, the government chooses a (nonlinear) subsidy function $S(e)$ and an income tax rate t subject to a government budget constraint and subject to behavioral responses of the individuals. Thus, formally we have:

$$\begin{aligned} \max_{t, S(\cdot)} \quad & \tilde{f}_l \left((1-t)w_l(e_l)h(t, w_l(e_l)) + T - \Psi[h(t, w_l(e_l))] - e_l + S(e_l) \right) \\ & + \tilde{f}_h \left((1-t)w_h(e_h)h(t, w_h(e_h)) + T - \Psi[h(t, w_h(e_h))] - e_h + S(e_h) \right) \end{aligned} \quad (4)$$

subject to a government budget constraint

$$T = t \left(f_l w_l(e_l)h(t, w_l(e_l)) + f_h w_h(e_h)h(t, w_h(e_h)) \right) - f_l S(e_l) - f_h S(e_h)$$

and optimal individual behaviour

$$\forall i = l, h : (e_i, h_i) = \arg \max_{e, h} (1-t)w_i(e)h + T - \Psi(h) - e + S(e). \quad (5)$$

This problem has some similarities to the problem in Stiglitz (1982), where a nonlinear tax schedule is chosen in an economy with two groups of individuals. By the revelation principle we can formulate the part of choosing $S(\cdot)$ as choosing e_l, e_h, c_l^1, c_h^1 directly, where c_l^1 and c_h^1 denote first period consumption. In that case, we can replace (5) by

$$h(t, w_i(e_i)) = \arg \max_h (1-t)hw_i(e_i) - \Psi(h) \quad (6)$$

⁷See Stantcheva (2013) for a similar formula in a discrete type setting.

and an incentive compatibility constraint⁸

$$\begin{aligned} & c_h^1 + (1-t)w_h(e_h)h(t, w_h(e_h)) - \Psi(h(t, w_h(e_h))) \\ & \geq c_l^1 + (1-t)w_h(e_l)h(t, w_h(e_l)) - \Psi(h(t, w_h(e_l))). \end{aligned} \quad (7)$$

Notice that in the incentive constraint (7) the deviation utility on the right-hand-side, the terms $w_h(e_l)$ and $h(t, w_h(e_l))$, show up. A deviating high-skilled agent receives the education level of the low skilled agent e_l . The wage she receives differs from the wage of the low skilled agent because of the effect of innate abilities on wages. To keep notation simple we denote the associated hour choice $h(t, w_h(e_l)) = h^c$ and associated income $y^c = h^c w_h(e_l)$, with a c for counterfactual as in equilibrium the high type will be truth-telling.

The government's problem now reads as:

$$\begin{aligned} \max_{c_l^1, c_h^1, t, e_l, e_h} & \tilde{f}_l \left(c_l^1 + (1-t)w_l(e_l)h(t, w_l(e_l)) + T - \Psi[h(t, w_l(e_l))] \right) \\ & + \tilde{f}_h \left(c_h^1 + (1-t)w_h(e_h)h(t, w_h(e_h)) + T - \Psi[h(t, w_h(e_h))] \right) \end{aligned} \quad (8)$$

subject to a government budget constraint

$$T = t \left(f_l w_l(e_l)h(t, w_l(e_l)) + f_h w_h(e_h)h(t, w_h(e_h)) \right) - f_l(c_l^1 + e_l) - f_h(c_h^1 + e_h), \quad (9)$$

and subject to (6) and (7), where we denote as η the Lagrangian multiplier of the incentive compatibility constraint. The Lagrangian and the first-order conditions are stated in Appendix A.1. Notice that in fact c_l^1 , c_h^2 and T are not pinned down uniquely. Due to the quasi-linearity of preferences, individuals are indifferent when to consume. Therefore, only the difference $c_h^1 - c_l^1$ is pinned down. However, without loss of generality, we focus on the solution with zero savings here.

⁸Since we assume $\tilde{f}_l > f_l$, we focus on downward redistributive taxation where only the incentive constraint of the high type is binding.

The solution (8) can then be implemented with a nonlinear subsidy function $S(\cdot)$ that has to yield the desired consumption levels, i.e. $S(e_l) = c_l^1 + e_l$ and $S(e_h) = c_h^1 + e_h$. In addition, we have to make sure that incentives for the level of education and labor supply are jointly optimal for the individual. This implies that – given the subsidy function – (5) has to hold. Naturally, infinitely many nonlinear subsidy schedules can implement the desired allocation, as in the nonlinear tax problem with two types of Stiglitz (1982). We will in the following be interested in those subsidy functions that are differentiable at e_l and e_h . In these cases, we know that the first-order condition for education of an individual can be rearranged as:

$$(1 - S'(e_i)) = (1 - t) \frac{\partial w_i(e_i)}{\partial e_i} h_i \quad \forall i = l, h.$$

In the following, we will therefore be interested in

$$s_i \equiv 1 - (1 - t) \frac{\partial w_i(e_i)}{\partial e_i} h_i \quad \forall i = l, h.$$

Having computed an optimal allocation, we can therefore infer the implicit marginal education subsidies s_l and s_h for this allocation. For simplicity, we will call s_l and s_h education subsidies in the remainder of this paper.⁹ Note also that throughout this paper, we only characterize *marginal* subsidies and not *average* subsidies.

3.2.2 Optimal Tax and Education Policies

We start by characterizing the optimal linear income tax rate. As the following proposition shows, the optimal linear tax rate is corrected by the endogeneity of education as compared to the optimal tax rate with exogenous education in equation (3).

Proposition 1. *In a full-commitment economy, the optimal linear tax rate satisfies*

$$\frac{t^f}{1 - t^f} = \frac{(\tilde{f}_l - f_l) \left(\frac{y_h - y_l}{\bar{y}} \right) - \eta \left(\frac{y_h - y^c}{\bar{y}} \right)}{\epsilon},$$

⁹As is in the optimal taxation problem with discrete types, we can always pick a nonlinear subsidy schedule such that the first-order conditions of an individuals are also sufficient and her problem is concave. In order to ensure that locally linear subsidy schedules implement the desired allocation, further assumptions on $w_h(e)$ and $w_l(e)$ have to be made, see, e.g., Bovenberg and Jacobs (2005, p.2010) for a discussion of that in a similar framework.

where the multiplier satisfies $\eta = \tilde{f}_l - f_l$.

Proof. See Appendix A.1.1 □

The tax rate with endogenous education decisions is still increasing in income inequality and decreasing in the labor supply elasticity. As can be seen, there is an additional force given by $\eta \left(\frac{y_h - y^c}{\bar{y}} \right)$ in the numerator as compared to the case where education is taken as exogenous. It decreases the optimal tax rate, and the effect is stronger the bigger the difference $y_h - y^c$. y^c is the income level that the high type would attain when only taking the education level of the low type e_l . The difference, hence, captures the effect of a higher education level for the high type on her earnings. The more important the effect of education on earnings, the smaller the tax rate tends to be. Consider the one extreme case, where additional education does not change wages at all for the high-type, so $y_h = y^c$. In this case, there is no need for the optimal tax rate to take into account education incentives, and the formula collapses to the case with exogenous human capital. In the other extreme case, we would have $y_l = y^c$, so with the same education level both agents would receive the same wage. This would essentially eliminate agent heterogeneity and the optimal tax rate would be zero in a model without risk. The following corollary summarizes the above reasoning.

Corollary 2. *Let e_l^* and e_h^* be the solution to the problem (8). Then the respective optimal linear tax rate is smaller than the linear tax rate as defined by (3) for $e_l = e_l^*$ and $e_h = e_h^*$, i.e. $t^f(e_l^*, e_h^*) < t^{ex}(e_l^*, e_h^*)$.*

Proof. A change in t implies the same percentage change in y_h and y_l with a constant elasticity; thus the numerator and the denominator of $\frac{y_h - y_l}{\bar{y}} = \frac{y_h - y_l}{f_l y_l + f_h y_h}$ change by the same factor. Then, the corollary directly follows. □

Income taxes are not the only instrument of the government. Governments do rely on education subsidies to increase the incentives to invest into education. We now characterize optimal education subsidies.

Proposition 3. *In a full-commitment economy, education subsidies satisfy*

$$s_l^f = t^f \frac{\partial w_l(e_l)}{\partial e_l} h_l (1 + \epsilon) - \frac{\eta}{f_l} (1 - t^f) \left[h^c \frac{\partial w_h(e_l)}{\partial e_l} - h_l \frac{\partial w_l(e_l)}{\partial e_l} \right]$$

and

$$s_h^f = t^f \frac{\partial w_h(e_h)}{\partial e_h} h_h (1 + \epsilon).$$

Proof. See Appendix A.1.2 □

First, looking at the education subsidy for the low type one can see that there are two parts. The first term reflects the fiscal externality effect of private education decisions: the education decision of individuals imposes an externality on the government budget as individuals with higher education pay higher taxes. The government internalizes this fiscal externality by subsidizing education in a Pigouvian way. As the formula reveals, the larger the labor supply elasticity is, the larger the subsidy. Intuitively, the stronger individuals' working hours react to wage increases, the larger is the fiscal externality on the government budget. Relatedly, the subsidy increases in the marginal return of education $\frac{\partial w_l(e_l)}{\partial e_l}$ and in the income tax rate. Notice that even if labor supply is not distorted ($\epsilon = 0$), the education decision would be distorted by the tax rate because individuals only reap $(1 - t)$ of the financial gains from education.

The second term captures the fact that innate abilities and education are complements. The marginal return to education is increasing in innate ability. As the government is redistributive, there is a force towards lowering education subsidies, as they tend to profit more the initially high types. Maldonado (2008) first has shown that in case of a complementarity between educational investment and innate ability, education should be taxed. See also Jacobs and Bovenberg (2011) for a discussion of this issue.¹⁰ For the high type only the fiscal externality part is present because a standard "no-distortion-at-the-top" result applies for the second part.

3.3 Optimal Policies with Lack of Commitment

We now look at economies, where the degree of commitment power of the government is allowed to differ, nesting the case from the previous section.

¹⁰Maldonado (2008) and Jacobs and Bovenberg (2011) also consider the case where educational returns are decreasing in ability and show that in this case, education should rather be subsidized (relative to a first-best rule). In line with empirical evidence, we focus on the case of educational returns that are increasing in innate ability (Carneiro and Heckman 2005, Lemieux 2006). Our results concerning the relation between commitment power and the progressivity of education subsidies is not affected by this assumption, however.

3.3.1 Costs of Deviating and the Commitment Technology

Following Farhi, Sleet, Werning, and Yeltekin (2012), we introduce output costs of deviation. This implies that the government lacks commitment and can always deviate from its announced tax rate. However, deviation will incur some output loss κ , which can be considered as a reduced form for a reputational loss. Farhi, Sleet, Werning, and Yeltekin (2012) show how to microfound such an output loss in a dynamic repeated game, where a deviation today brings a reputational cost borne in the future because of depressed investment of future generations.

Formally, this implies an additional credibility constraint on the government problem. It takes the form:

$$\mathcal{W}_{pc}^2 \geq \mathcal{W}_{dev}^2(e_l, e_h) - \kappa, \quad (10)$$

where \mathcal{W}_{pc}^2 is second period welfare under the assumption that the government sticks to its promise. $\mathcal{W}_{dev}^2(e_l, e_h)$ on the other is the second period welfare obtained if the government reneges on its tax promise and effectively takes the education levels as exogenous as in Section 3.1. Keep in mind that the government is not allowed to use type-dependent lump sum taxation in the linear tax case, in contrast to the nonlinear case studied below.

This form of deviation costs allows to flexibly capture different levels of *limited commitment*. At the one extreme end, when κ is zero, there is no way for the government to credibly commit and we arrive at the case with no commitment. This case is also studied in an earlier version of this paper (Findeisen and Sachs (2014)). At the other extreme end, when κ is above some positive threshold $\bar{\kappa} > 0$, all tax promises are fully credible and we arrive at the full-commitment solution of Section 3.2, which naturally achieves the highest welfare. In this section we focus on the intermediate cases where κ lies between zero and $\bar{\kappa}$.

Before we can study optimal policies under such a credibility constraint, it is important to understand what policies a deviating government follows. It can be shown that a deviating government would set the tax rate according to the same rule as studies in Section 3.1. The intuition is that for a deviating government education incentives are considered sunk. This is summarized in the following lemma.

Lemma 4. *A deviating government takes education levels as exogenous and therefore sets the linear tax rate according to*

$$\frac{t^{dev}}{1 - t^{dev}} = \frac{(\tilde{f}_l - f_l) \left(\frac{y_h - y_l}{\bar{y}} \right)}{\epsilon}. \quad (11)$$

3.3.2 Optimal Policies and Discussion

In comparison to the full-commitment problem in Section 3.2, the government has to respect the credibility constraint (10) in addition to all other constraints. We denote the Lagrangian multiplier on this credibility constraint as ζ . The Lagrangian function and the first-order conditions are stated in Appendix A.2. The following proposition shows the optimal income tax rate for this case.

Proposition 5. *In a partial-commitment economy, the optimal linear tax rate satisfies:*

$$\frac{t^{pc}}{1 - t^{pc}} = \frac{(\tilde{f}_l - f_l) \left(\frac{y_h - y_l}{\bar{y}} \right) - \frac{\eta}{1 + \zeta} \left(\frac{y_h - y^c}{\bar{y}} \right)}{\epsilon}, \quad (12)$$

Proof. See Appendix A.2.1 □

One can see how this case nests the full-commitment case, i.e. the optimal income tax rate from Proposition 1. If the credibility constraint is not binding for sufficiently high κ (hence $\kappa > \bar{\kappa}$), ζ is equal to zero and the government is able to implement the full-commitment tax rate. As discussed above, the second term in the numerator reflects how labor taxes are adjusted to provide education incentives and complement education subsidies. This effect is now scaled down by $\frac{1}{1 + \zeta}$. The more severe the commitment problem, the bigger ζ tends to be. This will make any tax promises less credible and, anticipating this, the government will set a higher, more credible tax rate. Next, we characterize the resulting education subsidies.

Proposition 6. *In a partial-commitment economy, education subsidies satisfy:*

$$s_l^{pc} = t^{pc} \frac{\partial w_l(e_l)}{\partial e_l} h_l (1 + \epsilon) - \frac{\eta}{f_l} (1 - t^{pc}) \left[h^c \frac{\partial w_h(e_l)}{\partial e_l} - h_l \frac{\partial w_l(e_l)}{\partial e_l} \right] + \frac{\zeta}{f_l} \left(\frac{\partial \mathcal{W}_{pc}}{\partial e_l} - \frac{\partial \mathcal{W}_{dev}}{\partial e_l} \right)$$

where $\frac{\partial \mathcal{W}_{pc}}{\partial e_l} - \frac{\partial \mathcal{W}_{dev}}{\partial e_l} > 0$ and

$$s_h^{pc} = t^{pc} \frac{\partial w_h(e_h)}{\partial e_h} h_h (1 + \epsilon) + \frac{\zeta}{f_h} \left(\frac{\partial \mathcal{W}_{pc}}{\partial e_h} - \frac{\partial \mathcal{W}_{dev}}{\partial e_h} \right).$$

where $\frac{\partial \mathcal{W}_{pc}}{\partial e_h} - \frac{\partial \mathcal{W}_{dev}}{\partial e_h} < 0$.

Proof. See Appendix A.2.2 □

Whenever the credibility constraint is binding, the subsidies get adjusted by

$$\frac{\zeta}{f_l} \left(\frac{\partial \mathcal{W}_{pc}^2}{\partial e_l} - \frac{\partial \mathcal{W}_{dev}^2}{\partial e_l} \right) > 0 \quad (13)$$

and

$$\frac{\zeta}{f_h} \left(\frac{\partial \mathcal{W}_{pc}^2}{\partial e_h} - \frac{\partial \mathcal{W}_{dev}^2}{\partial e_h} \right) < 0 \quad (14)$$

respectively. This implies that whenever there is a commitment problem, the marginal value of education for the low type goes up as it strengthens the credibility of tax promises. The marginal benefit of high level education goes down instead. More education for the high type increases the temptation to renege on tax promises and increase the tax rate to redistribute.

The intuition for this result lies in the different tax rates on the equilibrium path and when deviating. On the equilibrium path, the tax rate is lower than when deviating. The deviating planner will take education as sunk and set a tax rate as in the problem with exogenous education (11). This makes education of the low type unambiguously more attractive on the equilibrium path than in the deviation case and therefore is a force for higher education subsidies for the low type. Why is education for the low type more valuable on the equilibrium path? First because the low type works more on the equilibrium path (since the tax rate is lower) which directly increases the benefits from more education. Second, the low type keeps a share $(1 - t)$ of her earnings whereas the share t is divided between the high and low type through the payment of the lump-sum transfer T . Consumption of the low type is valued more. Since $(1 - t)$ is higher on the equilibrium path, education of the low type is valued more also through this channel.

For the high type, these two forces are of opposite sign. The first channel is equivalent as for the low type: the high type works more on the equilibrium path which increases the returns to education. The second channel is of opposite sign, however, because the government values resources going to both types more (through the lump-sum transfer T) than resources going only to the high type. As we show in the appendix, the first effect always dominates for our functional form.

We now present a complementary intuition that we already discussed in the introduction. Therefore note that the difference in the tax rate *off* and *on* the equilibrium path captures the incentives to deviate for the planner in a single number. The former is given by (11) and the latter by (12). The incentive to deviate is stronger, the larger $\frac{\eta}{1+\zeta} \left(\frac{y_h - y^c}{y} \right)$. The difference $y_h - y^c$ in turn is increasing in the difference of wages. Thus, a more equal wage distribution makes this term smaller and therefore the tax rates on and off equilibrium more similar. In other words, a more equal wage distribution renders the deviation less attractive.

Taken together, the lack of commitment leads to more progressive education subsidies. For the low type, lack of commitment adds a force towards a higher education subsidy. For the high type, it adds a force towards lower subsidies. The larger the commitment problem, the larger ζ and the stronger this effect on the progressivity of education subsidies.

4 Nonlinear Labor Taxes

We now turn to policies that are only constrained by informational asymmetries. The difference to the previous section is twofold. On the one hand, the government can tax income at a nonlinear rate. The second issue concerns the deviation: if the government deviates from its announced policy path and reoptimizes, it does not face an information problem anymore because individuals revealed their type in the first period. It can therefore apply excessive redistribution through lump-sum taxes. Thus, in the deviation, marginal tax rates will be zero. As we lay out below, this drastically changes implications as compared to the case with linear taxes.

4.1 Optimal Policies with Exogenous Education

As a benchmark, we look at the case with only one period and exogenous levels of education also here. Individuals differ in their wages w_h and w_l . The government maximizes the usual social objective $\sum_{i=l,h} f_i (c_i - \Psi(h_i))$. Thereby it has to satisfy an incentive compatibility constraint:

$$c_h - \Psi(h_h) \geq c_l - \Psi\left(h_l \frac{w_l(e_l)}{w_h(e_h)}\right).$$

Further, it has to satisfy a resource constraint:

$$\sum_{i=l,h} f_i (w_i(e_i)h_i - c_i) \geq 0.$$

This is a standard problem in public finance that has first been extensively studied by Stiglitz (1982). As we show in Appendix B.1, optimal marginal tax rates in this case satisfy:

$$\tau_h = \frac{\eta}{w_l(e_l)f_l} \left(\Psi'(h_l) - \Psi' \left(h_l \frac{w_l(e_l)}{w_h(e_h)} \right) \frac{w_l(e_l)}{w_h(e_h)} \right) \quad (15)$$

$$\tau_h = 0 \quad (16)$$

with $\eta = \tilde{f}_l - f_l$. The low type faces a positive distortion that is increasing in redistributive preferences (i.e. $\tilde{f}_l - f_l$), inequality and the labor supply elasticity. The latter two are captured by $\left(\Psi'(h_l) - \Psi' \left(h_l \frac{w_l(e_l)}{w_h(e_h)} \right) \frac{w_l(e_l)}{w_h(e_h)} \right)$. This term is increasing in $\frac{w_h(e_h)}{w_l(e_l)}$ (hence inequality) and decreasing in the labor supply elasticity – the more convex Ψ , the higher the elasticity of labor supply.

The result in (15) is similar to the result about the optimal linear tax rate (3) in Section 3.1. The key difference is that marginal tax rate only applies to the low type. By contrast, the high productivity type faces a zero distortion. Thus, even in the Rawlsian case ($\tilde{f}_l = 1$) the high type would face a zero marginal tax rate. This ‘no-distortion at the top’ - result is a standard result in optimal taxation, see e.g. Mirrlees (1971), Stiglitz (1982) or Saez (2001). It seems counterintuitive at first sight. However, it does not say anything

about the average tax rate that the high productivity type pays. Inframarginal tax rates can be quite high. The reason why the marginal tax rate is zero is that it only applies to the very last (marginal) unit of income and would therefore not even raise revenue.¹¹ It would, however, distort the high types labor supply decision. Note, that this has already applied to education subsidies in Section 3.2 in Proposition 3.

4.2 Optimal Policies with Full Commitment

We now turn to the case with two periods, where educational investments are made in the first period, i.e. the counterpart to Section 3.2, but this time with nonlinear labor income taxes. In this case, the government's problem reads as:

$$\max_{c_l^1, c_l^2, c_h^1, c_h^2, h_l, h_h, e_l, e_h} \sum_{i=l, h} \tilde{f}_i(c_i^1 + c_i^2 - \Psi(h_i))$$

subject to the resource constraint

$$\sum_{i=l, h} f_i(w_i(e_i)h_i - c_i^1 - c_i^2 - e_i) \geq 0$$

and incentive compatibility

$$c_h^1 + c_h^2 - \Psi(h_h) \geq c_l^1 + c_l^2 - \Psi\left(h_l \frac{w_l(e_l)}{w_h(e_l)}\right).$$

The next proposition characterizes optimal policies.

Proposition 7. *In the full commitment case with nonlinear labor taxes, the optimal allocation has the following properties:*

- *First and second period consumption are not pinned down. Only $c_h^1 + c_h^2$ and $c_l^1 + c_l^2$ are pinned down.*
- *Labor wedges are still characterized by (15) and (16).*

¹¹An alternative interpretation, more in the mechanism design logic, is that distorting the labor supply of the high type does not lead to a relaxation of an incentive constraint because no other individual is indifferent between truth-telling and mimicking the high type.

- *Optimal education subsidies are given by.*

$$s_l^f = \tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l} + \frac{\eta}{\lambda f_l} \Psi' h_l \frac{\partial \left(\frac{w_l(e_l)}{w_h(e_l)} \right)}{\partial e_l} \quad \text{and} \quad s_h^f = 0$$

with $\eta = \tilde{f}_l - f_l$.

Proof. See Appendix B.2. □

First of all, consumption levels are not pinned down because individuals are indifferent between consuming in period 1 and 2. This is related to the discussion in Section 3.2. Importantly, this changes when we consider lack of commitment, see the discussion after Proposition 8.

Second, marginal labor income tax rates are governed by exactly the same forces as with exogenous education. Unlike in the case with linear taxes, incentives for educational investment are fully set through education subsidies.

Third, for education subsidies we find similar results for the low type as in the case with linear instruments. First, the fiscal externality term in the spirit of Bovenberg and Jacobs (2005) implies a subsidy for education. Second, there is a force towards a tax on education of the low type in order to relax the incentive constraint of the high type. For the high type, we obtain a zero subsidy. Given that the labor wedge for the high type is zero, there is no reason to subsidize education because of the fiscal externality logic. The incentive constraint effect is also not there because of the usual ‘no distortion at the top’ as just described in the previous section.

4.3 Optimal Policies with Lack of Commitment

We now study the limited commitment case. As in Section 3.3, if the planner deviates from his announced tax plan, κ units of outputs are lost. This is again captured by the constraint:

$$\mathcal{W}_{pc}^2 \geq \mathcal{W}_{dev}^2(e_l, e_h, c_l^1, c_h^1) - \kappa,$$

Formally defined in Appendix B.3, $\mathcal{W}_{dev}^2(e_l, e_h, c_l^1, c_h^1)$ is social welfare in the second period if the planner deviates from announced tax policies after education and consumption in the first period have taken place. \mathcal{W}_{pc}^2 is social welfare in the second period if the planner sticks to her announced tax policies.

As an intermediate result, it is helpful to consider the resulting tax policies in the case of deviation. Once education is sunk at the stage of re-optimizing, the planner can identify individuals and no longer faces an informational constraint. The planner can hence assign zero labor wedges for each type: $\tau_h^{dev} = \tau_l^{dev} = 0$. Let us denote by h_l^{dev} the labor supply level of the low type in this case. This level is equal to the first-best level without distortions.

Proposition 8. *In the partial commitment case with nonlinear labor taxes, the optimal allocation has the following properties:*

- Consumption for the low type across time is again indeterminate. For the high type, all consumption is front loaded to period 1.
- Labor wedges are still characterized by (15) and (16).
- Education wedges are

$$s_l^{dev} = \tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l} + \frac{\eta}{f_l} \Psi' h_l \underbrace{\frac{\partial \left(\frac{w_l(e_l)}{w_h(e_l)} \right)}{\partial e_l}}_{<0} + \zeta \underbrace{\frac{\tilde{f}_l}{f_l} \frac{\partial w_l(e_l)}{\partial e_l}}_{<0} (h_l - h_l^{dev}).$$

and

$$s_h^{dev} = 0.$$

with $\eta = \tilde{f}_l - f_l$.

Proof. See Appendix B.3. □

A first interesting result is that consumption of the high type is completely front loaded. Intuitively, the individual is indifferent when to consume. For the government's temptation, however, it is better to give the high type all consumption in the first period such that these resources cannot be used for excessive redistribution in a deviation, which makes the

deviation less attractive.¹² For the low type in contrast, consumption is still indeterminate. In a deviation case, all of the resources are given to the low type, so timing of the low type consumption does not affect the credibility constraint. Also note that this implies that a non-negativity result on second period consumption of the high-type is binding in the optimal solution.

The main result concerns the impact of limited commitment on education policies. The results for the progressivity of the education subsidy schedule are in stark contrast to that of Section 3 with linear labor taxes. Here with nonlinear labor taxes, the lack of commitment leads to a lower education subsidy for the low type and a constant zero subsidy of the high type – education policies become more regressive as a consequence. Moreover, the larger ζ , the more binding the commitment constraint and the stronger the downward distortion of the education subsidy for the low type.

What is driving this? Key is the pattern of labor taxes in the deviation case and on the equilibrium path. In the deviation case the planner redistributes more resources; there is “excessive redistribution” as in the case with linear labor taxes. However and different from the case with linear labor taxes, this additional redistribution is not carried with more progressive taxes but with type dependent lump sum taxes. Labor distortions are zero. Therefore, labor supply is at its first best efficient level also for the low type in the deviation case. Thus, more education for the low type is valued higher in the deviation and therefore makes deviation more tempting at the margin. For the high type, this force is not there because the high type faces a zero labor supply distortion in both cases, on the equilibrium path and in the deviation case.¹³

Taken together, the lack of commitment leads to more regressive education subsidies. For the low type, lack of commitment adds a force towards a lower education subsidy. For the high type, the education subsidy stays at zero. This result of more regressive subsidies is increasing in ζ , i.e. the severeness of the commitment problem.

¹²Note that this was not the case for linear taxation. The reason is that in the linear case, if the planner front loads consumption for the high type, she has to decrease T in period 2 accordingly and also has to increase period 1 consumption for the low type. In case of deviation, the planner could then pay out a lower lump-sum transfer T as well. But this decrease of T due to front loading is the same on and off the equilibrium path and therefore does not relax or tighten the credibility constraint in the linear tax case.

¹³Note that in the linear tax case, we also discussed that a share $1 - t$ of the educational returns goes to the individual whereas the share t is reaped by the government. For the nonlinear tax case, these effects are not present because the planner is not constrained on how to use the additional resources through education.

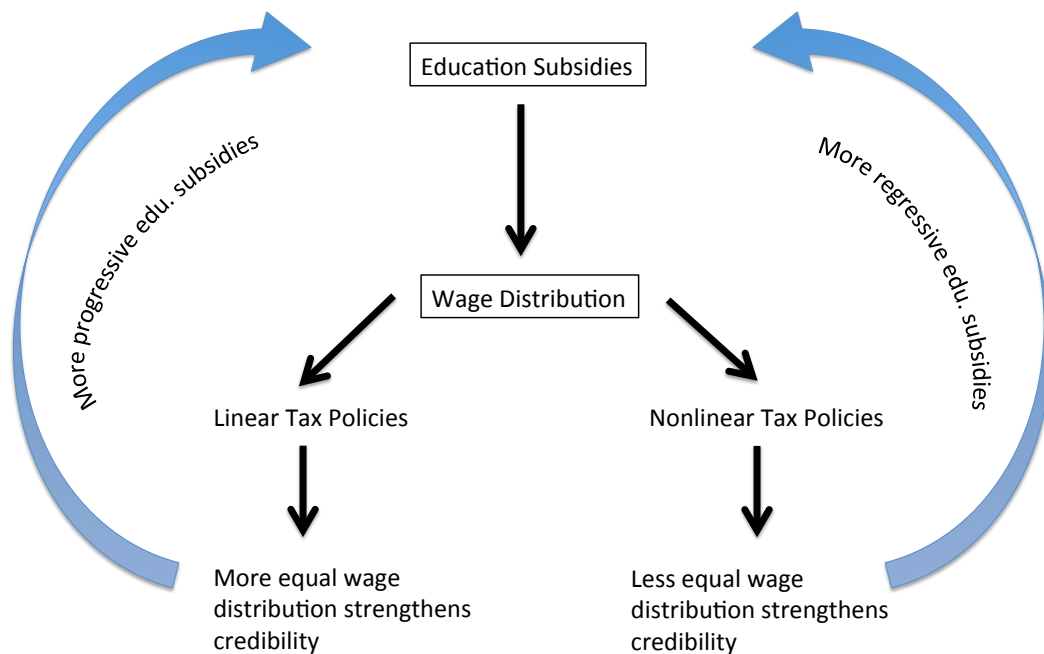


Figure 1: Link Between Education Subsidies and Policy Credibility

5 Discussion

Brief Summary. Figure 1 summarizes the main mechanisms behind the results. At the center of the commitment problem is the wage distribution. This is intuitive – at this stage when education decisions are sunk, deciding to re-optimize the tax code should only depend on the distribution of wages. Next, it has to be determined how the wage distribution affects the incentive to deviate for the government. Our contribution shows how this is connected to the structure of labor taxes. First, we have studied the case of a progressive tax system with a constant marginal rate and lump-sum redistribution. A more equal wage distribution strengthens credibility as the payoff of a deviation is increasing in wage inequality. The government tends to make education subsidies more progressive, relative to the full commitment case, to achieve a more equal distribution of wages.

In contrast, with nonlinear labor taxation where private information about types leads to deviations from the first-best, a more *unequal* wage distribution strengthens credibility. The payoff to a deviation is decreasing in wage inequality now. This is because in the

deviation case informational frictions disappear and the government can set zero labor wedges and implement the efficient level of labor supply. This implies labor supply for the low type will be higher in the deviation case. Labor supply for the high type is at the efficient level in both the deviation case and on the equilibrium path. Together this implies that a lower level of education and therefore a lower wage for the low type – i.e. a more unequal wage distribution – improves the commitment problem and makes tax promises set ex-ante more credible. The government hence implements more regressive education subsidies to achieve a more unequal wage distribution.

Differences To Capital Taxation and Limited Commitment. The results from the model with linear labor taxes are reminiscent of the results from Farhi, Sleet, Werning, and Yeltekin (2012). They show that a more equal distribution physical capital strengthens credibility leading to progressive savings taxation. The main difference between physical and human capital is that wealth or savings can be taxed directly by the government. Policies leading to less wealth inequality are more credible as they lower the chance of excessive wealth distribution later. Human capital in contrast is only taxed indirectly through the labor income tax. What we show is that policies leading to more wage inequality can be more or less credible, depending on the pattern of labor distortions.

Other Public Spending. In our analysis we have abstracted from other reasons to levy taxes than redistribution from rich to poor. One candidate would be public good provision. Let g be the amount of the public good and denote preferences in the second period by

$$U = c - v\left(\frac{y_i}{w_i(e_i)}\right) + u_i(g) \quad \forall i = l, h$$

in this case, where $u_i(g)$ is the type-dependent utility for the public good. Since public goods can be consumed by anybody, one can see that the level of public good spending g does not influence incentive constraints – the additional term just cancels out. Thus, in the nonlinear tax case, the optimal level of the public good is the same in the second and the first best. Consequently it would be the same on and off the equilibrium path and would not alter the commitment problem in any way. How would things be for linear taxes? Unless preferences for the public good are extremely strong and one ends up in

a corner solution where all tax revenue is used for g and $T = 0$, this will not influence deviation incentives. The reason is that if we are not in such a corner solution, a deviating government would use all the additional tax revenue (from setting $t = t^{dev}$ and not $t = t^{pc}$) to increase T – such as in the case without the public good.¹⁴

Empirical Content. We conclude this section with a brief discussion on the empirical content of the main mechanism of the model. First, the model assumes that individuals react to changes in taxes when undertaking human capital investments. This is behind the results with a constant tax rate, where the government sets lower taxes when taking education incentives into account.¹⁵ Abramitzky and Lavy (2012) provide quasi-experimental evidence on the negative effect of redistributive taxation on education investment. More structural and model based approaches as the classic work by Trostel (1993) also have found big effects of income taxation on human capital investment. A second operative margin in the theory is that education on the private level react to education subsidies. There is large empirical literature estimating this for college enrollment surveyed by Kane (2006) and Deming and Dynarski (2009). The consensus is that the magnitude of behavioral responses is sizable. Finally, the government needs to be aware of its future temptation to tax, when deciding on human capital subsidies in the present. Naturally, it is very challenging to come up with a credible research design to test this assumption. An additional difficulty comes from the fact that the model makes different predictions how lack of commitment influences education policies, depending on how labor taxes are designed. This why we leave a detailed investigation of those issues for further research.¹⁶

¹⁴Certainly one could think of other ways to model the public good such that it interacts with labor supply for example. But in this kind of standard way of modeling public goods in political economics or macroeconomics as additively separable (for example Song, Storesletten, and Zilibotti (2012)), our results would be unaffected by the presence of public goods.

¹⁵In the nonlinear case, optimal tax rates are described by the same formula with and without taking into account endogenous human capital investment – see Section 4.

¹⁶In an earlier version, we provided suggestive cross-country evidence for a more regressive incidence of education subsidies when the ability of a government to commit is high.

6 Conclusion

We build a simple model of education and tax policies when the government lacks full commitment. Individuals are born with heterogeneous abilities and undertake human capital investments early in their life and make labor supply decisions later. While we also characterize in detail labor tax policies, our main results concern the design of education policies. The impact of commitment frictions on education subsidies depends on the labor tax system: they become more progressive if labor tax rates are linear but more regressive if the labor tax system is nonlinear in the spirit of Mirrlees. Our paper complements earlier important work in the literature on the interaction of capital taxation and lack of commitment by Farhi et al. (2012). Their benchmark result that savings are taxed more progressively does not always generalize to human capital policies. Intuitively, in the case with physical capital, wealth can always be taxed and redistributed directly by the government. Human capital in contrast is taxed indirectly only by the labor income tax, which also creates a labor supply distortion. This interaction with the labor tax makes the results depend on what labor tax systems are available to the government. Future work might integrate wealth and human capital accumulation together with capital taxation and education policies into one model. We leave this for further research.

A Linear Taxes

A.1 The Full-Commitment Planner

We first substitute the government budget constraint into the problem. The Lagrangian function then reads as

$$\begin{aligned} \mathcal{L} = & \tilde{f}_l \left(c_l^1 + (1-t)w_l(e_l)h(t, w_l(e_l)) - \Psi[h(t, w_l(e_l))] \right) \\ & + \tilde{f}_h \left(c_h^1 + (1-t)w_h(e_h)h(t, w_h(e_h)) - \Psi[h(t, w_h(e_h))] \right) \\ & + t \left(f_l w_l(e_l)h(t, w_l(e_l)) + f_h w_h(e_h)h(t, w_h(e_h)) \right) - f_l(c_l^1 + e_l) - f_h(c_h^1 + e_h) \\ & + \eta \left(c_h^1 + (1-t)w_h(e_h)h(t, w_h(e_h)) - \Psi(h(t, w_h(e_h))) - c_l^1 + (1-t)y^c - \Psi(h^c) \right) \end{aligned}$$

The first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial c_l^1} = \tilde{f}_l - f_l - \eta = 0$$

$$\frac{\partial \mathcal{L}}{\partial c_h^1} = \tilde{f}_h - f_h + \eta = 0$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} = & -\tilde{f}_l y_l - (1-\tilde{f}_l)y_h + f_l y_l + f_h y_h + t f_l w_l(e_l) \frac{\partial h_l}{\partial t} \\ & + t f_h w_h(e_h) \frac{\partial h_h}{\partial t} - \eta(w_h(e_h)h_h - w_h(e_l)h^c) = 0 \end{aligned} \quad (17)$$

$$\frac{\partial \mathcal{L}}{\partial e_l} = \tilde{f}_l(1-t) \frac{\partial w_l(e_l)}{\partial e_l} h_l + t f_l \frac{\partial w_l(e_l)}{\partial e_l} h_l(1+\epsilon) - f_l + \eta \left[-(1-t)h^c \frac{\partial w_h(e_l)}{\partial e_l} \right] = 0,$$

$$\frac{\partial \mathcal{L}}{\partial e_h} = \tilde{f}_h(1-t) \frac{\partial w_h(e_h)}{\partial e_h} h_h + t f_h \frac{\partial w_h(e_h)}{\partial e_h} h_h(1+\epsilon) - f_h + \eta(1-t) \frac{\partial w_h(e_h)}{\partial e_h} h_h = 0.$$

From the FOC for c_l^1 , one directly obtains $\eta = \tilde{f}_l - f_l$.

A.1.1 Proof of Proposition 1

Two manipulations of (17) yield

$$\left(\tilde{f}_l - f_l \right) (y_h - y_l) - \frac{t}{1-t} \left(f_l w_l(e_l) h_l \frac{\partial h_l}{\partial 1-t} \frac{1-t}{h_l} + f_h w_h(e_h) h_h \frac{\partial h_h}{\partial 1-t} \frac{1-t}{h_h} \right) - \eta (y^c - y_l).$$

Now use $\varepsilon = \frac{\partial h_h}{\partial 1-t} \frac{1-t}{h_h} = \frac{\partial h_l}{\partial 1-t} \frac{1-t}{h_l}$ and $\bar{y} = f_l y_l + f_h y_h$ and solve for $\frac{t}{1-t}$ to obtain the result.

A.1.2 Proposition 3

We start with the high type. Rewriting the FOC for e_h yields

$$\tilde{f}_h(1-t) \frac{\partial w_h(e_h)}{\partial e_h} h_h + t f_h \frac{\partial w_h(e_h)}{\partial e_h} h_h(1+\epsilon) - f_h + \left(f_h - \tilde{f}_h \right) (1-t) \frac{\partial w_h(e_h)}{\partial e_h} h_h = 0,$$

which yields

$$t f_h \frac{\partial w_h(e_h)}{\partial e_h} h_h(1+\epsilon) - f_h + f_h(1-t) \frac{\partial w_h(e_h)}{\partial e_h} h_h = 0.$$

This can be rewritten as

$$f_h(1-t) \frac{\partial w_h(e_h)}{\partial e_h} h_h = 1 - t \frac{\partial w_h(e_h)}{\partial e_h} h_h(1+\epsilon),$$

where the RHS is the definition of the implicit education subsidy for the high type.

Now we look at the low type. Rewriting the FOC for e_l yields:

$$\begin{aligned} & \tilde{f}_l(1-t) \frac{\partial w_l(e_l)}{\partial e_l} h_l + t f_l \frac{\partial w_l(e_l)}{\partial e_l} h_l(1+\epsilon) - f_l + \eta \left[-(1-t) h^c \frac{\partial w_h(e_l)}{\partial e_l} \right] \\ & + \eta(1-t) h_l \frac{\partial w_l(e_l)}{\partial e_l} - \eta(1-t) h_l \frac{\partial w_l(e_l)}{\partial e_l} = 0. \end{aligned}$$

Now use $\eta = \tilde{f}_l - f_l$ and obtain

$$t f_l \frac{\partial w_l(e_l)}{\partial e_l} h_l (1 + \epsilon) - f_l + \eta (1 - t) \left[h_l \frac{\partial w_l(e_l)}{\partial e_l} - h^c \frac{\partial w_h(e_l)}{\partial e_l} \right] - f_l (1 - t) h_l \frac{\partial w_l(e_l)}{\partial e_l} = 0.$$

Rearranging and again using the definition of the implicit subsidy yields the result.

A.2 The Partial-Commitment Planner

We first substitute the government budget constraint into the problem. The Lagrangian function then reads as

$$\begin{aligned} \mathcal{L} = & \tilde{f}_l \left((1 - t) w_l(e_l) h(t, w_l(e_l)) - \Psi[h(t, w_l(e_l))] \right) \\ & + \tilde{f}_h \left((1 - t) w_h(e_h) h(t, w_h(e_h)) - \Psi[h(t, w_h(e_h))] \right) \\ & + t \left(f_l w_l(e_l) h(t, w_l(e_l)) + f_h w_h(e_h) h(t, w_h(e_h)) \right) - f_l (c_l^1 + e_l) - f_h (c_h^1 + e_h) \\ & + \eta (c_h^1 + (1 - t) w_h(e_h) h(t, w_h) - \Psi(h(t, w_h)) - c_l^1 + (1 - t) y^c - \Psi(h^c)) \\ & + \zeta (\mathcal{W}_{pc}^2(e_l, e_h, t) - \mathcal{W}_{dev}^2(e_l, e_h) + \kappa,) \end{aligned}$$

where

$$\begin{aligned} \mathcal{W}_{dev}^2(e_l, e_h) = & \max_{t^{dev}} \tilde{f}_l (1 - t^{dev}) w_l(e_l) h_l - \Psi(h_l) + \tilde{f}_h (1 - t^{dev}) w_h(e_h) h_h - \\ & \left(\tilde{f}_l (1 - t^{dev}) w_l(e_l) h_l(w_l, t^{dev}) - \Psi(h_l(w_l, t^{dev})) + \tilde{f}_h (1 - t) w_h(e_h) h_h(w_h, t^{dev}) \right) \\ & + t (w_l(e_l) h_l f_l + w_h(e_h) h_h f_h) - t^d (w_l(e_l) h_l(w_l, t^{dev}) f_l + w_h(e_h) h_h(w_h, t^{dev}) f_h). \end{aligned}$$

For c_l and c_h we get the same FOC as in the full-commitment case. For t we get:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} = & (1 + \zeta) \left(-\tilde{f}_l y_l - (1 - \tilde{f}_l) y_h + f_l y_l + f_h y_h + t f_l w_l(e_l) \frac{\partial h_l}{\partial t} \right. \\ & \left. + t f_h w_h(e_h) \frac{\partial h_h}{\partial t} \right) - \eta (w_h(e_h) h_h - w_h(e_l) h^c) = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial e_l} = & \tilde{f}_l (1 - t) \frac{\partial w_l(e_l)}{\partial e} h_l + t f_l \frac{\partial w_l(e_l)}{\partial e_l} h_l (1 + \epsilon) - f_l + \eta \left[-(1 - t^F) h^c \frac{\partial w_h(e_l)}{\partial e_l} \right] \\ & + \zeta \frac{\partial w_l(e_l)}{\partial e_l} \left[\tilde{f}_l \left((1 - t^{pc}) h_l(e_l, t^{pc}) - (1 - t^{dev}) h_l(e_l, t^{dev}) \right) \right. \\ & \left. + f_l \left(t^{pc} h_l(e_l, t^{pc}) - t^{dev} h_l(e_l, t^{dev}) \right) (1 + \varepsilon_{h,w}) \right] = 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial e_h} = & \tilde{f}_h (1 - t) \frac{\partial w_h(e_h)}{\partial e} h_h + t f_h \frac{\partial w_h(e_h)}{\partial e_h} h_h (1 + \epsilon) - f_h + \eta (1 - t) \frac{\partial w_h(e_h)}{\partial e_h} h_h \\ & + \zeta \frac{\partial w_h(e_h)}{\partial e_h} \left[\tilde{f}_h \left((1 - t^{pc}) h_h(e_h, t^{pc}) - (1 - t^{dev}) h_h(e_h, t^{dev}) \right) \right. \\ & \left. + f_h \left(t^{pc} h_h(e_h, t^{pc}) - t^{dev} h_h(e_h, t^{dev}) \right) (1 + \varepsilon_{h,w}) \right] = 0. \end{aligned}$$

A.2.1 Proof of Proposition 5

Dividing the FOC for T by $1 + \zeta$ directly shows that the FOC is equivalent to the one in Appendix A.1; the only difference is that η is now replaced by $\frac{\eta}{1+\zeta}$. The proof is then equivalent to the proof in Appendix A.1.1

A.2.2 Proof of Proposition 6

The FOC for e_l and e_h are equivalent to those in Appendix A.1.2 apart from the additional terms multiplied by ζ . The steps are, however, the same as in A.1.2 and the additional terms multiplied with ζ then appear in the education subsidy formula as well.

These additional terms in the formula for the education subsidy read as:

$$\zeta \frac{\partial w_i(e_i)}{\partial e_i} \left[\frac{\tilde{f}_i}{f_i} \left((1 - t^{pc})h_i(e_i, t^{pc}) - (1 - t^{dev})h_i(e_i, t^{dev}) \right) + \left(t^{pc}h_i(e_i, t^{pc}) - t^{dev}h_i(e_i, t^{dev}) \right) (1 + \varepsilon_{h,w}) \right]. \quad (18)$$

By assumption we have $\frac{\tilde{f}_i}{f_i} > 1$ and $\frac{\tilde{f}_h}{f_h} < 1$. In what follows we will write RF_i for $\frac{\tilde{f}_i}{f_i}$ to denote the relative Pareto weight and save on notation. We also simplify the notation for h and write $h_i(e_i, t^{dev}) = h_i^{dev}$ and similarly for the other expressions. Then (18) can be rearranged as:

$$\zeta \frac{\partial w_i(e_i)}{\partial e_i} \left((h_i^{pc} - h_i^{dev}) - [t^{dev}h_i^{dev} - t^{pc}h_i^{pc}] \left[\frac{1 + \varepsilon_{h,w}}{RF_i} - 1 \right] \right). \quad (19)$$

The sign of this term is equivalent to the sign of:

$$\frac{h_i^{pc} - h_i^{dev}}{t^{dev}h_i^{dev} - t^{pc}h_i^{pc}} - \left[\frac{1 + \varepsilon_{h,w}}{RF_i} - 1 \right] \quad (20)$$

if $h(t, w)t$ is increasing in t (which implies $t^{dev}h_i^{dev} - t^{pc}h_i^{pc} > 0$). The latter is the case if $\varepsilon_{h,t} > -1$. Note that $\varepsilon_{h,t} = -\frac{t}{1-t}\varepsilon$. One can show that for the Laffer tax rate, we have $\frac{t}{1-t} = \frac{1}{\varepsilon}$. As we are below the Laffer rate, we get $\varepsilon_{h,t} > -\frac{\varepsilon}{\varepsilon} = -1$. Thus, $h(t, w)t$ is increasing in t in the cases we consider.

Since $h_i = (w_i(1 - t))^\varepsilon$, (20) is $> (<)0$ if:

$$(1 - t)^\varepsilon - (1 - t^{dev})^\varepsilon > (<) \left(\frac{1 + \varepsilon}{RF_i} - 1 \right) (t^{dev} (1 - t^{dev})^\varepsilon - t (1 - t)^\varepsilon)$$

which is $> (<)0$ whenever

$$H(t) \equiv (1 - t)^\varepsilon \left(1 + t \frac{1 + \varepsilon}{RF_i} - t \right) > (<) (1 - t^{dev})^\varepsilon \left(1 + t^{dev} \frac{1 + \varepsilon}{RF_i} - t^{dev} \right) = H(t^{dev}).$$

We now have to show that $H(t) > H(t^{dev})$ for the low type and $H(t) < H(t^{dev})$ for the high type. We therefore take the derivative:

$$H'(t) = -\varepsilon(1-t)^{\varepsilon-1} \left(1 + t \frac{1+\varepsilon}{RF_i} - t\right) + (1-t)^\varepsilon \left(\frac{1+\varepsilon}{RF_i} - 1\right) = (1-t)^\varepsilon \left(\frac{1+\varepsilon}{RF_i} - 1 - \varepsilon \frac{1 + t \frac{1+\varepsilon}{RF_i} - t}{1-t}\right).$$

We need to show that it is < 0 for the low type and > 0 for the high type, which is equivalent to

$$(1-t) \frac{1+\varepsilon}{RF_i} - (1-t) - \varepsilon \left(1 + t \frac{1+\varepsilon}{RF_i} - t\right) < (>) 0$$

respectively, which is equivalent to

$$(1-t)(1+\varepsilon) - (1-t)RF_i - RF_i\varepsilon - t(1+\varepsilon)\varepsilon + tRF_i\varepsilon < (>) 0$$

and therefore

$$(1-t)(1+\varepsilon) - t(1+\varepsilon)\varepsilon < (>) RF_i ((1-t) + \varepsilon - t\varepsilon)$$

which yields

$$RF_i > (<) \frac{(1-t)(1+\varepsilon) - t(1+\varepsilon)\varepsilon}{(1-t)(\varepsilon+1)} = 1 - \frac{t}{1-t}\varepsilon.$$

As $RF_L > 1$, we directly see that this condition is always fulfilled for the low type. Importantly, it is fulfilled for any $t > 0$ and therefore we know that $H(t^{dev}) < H(t)$ for the low type.

How about our result for the high type? We need

$$RF_H < \frac{(1-t)(1+\varepsilon) - t(1+\varepsilon)\varepsilon}{(1-t)(\varepsilon+1)} = 1 - \frac{t}{1-t}\varepsilon$$

and

$$RF_H < \frac{(1-t)(1+\varepsilon) - t(1+\varepsilon)\varepsilon}{(1-t)(\varepsilon+1)} = 1 - \frac{t^{dev}}{1-t^{dev}}\varepsilon.$$

If both of these inequalities are fulfilled we can be sure that $H(t^{dev}) > H(t)$ for the high type. Since $t^{dev} > t$, the second is the stricter requirement. Inserting the formula for $\frac{t^{dev}}{1-t^{dev}}$ yields:

$$\frac{\tilde{f}_h}{f_h} < 1 - (f_h - \tilde{f}_h) \underbrace{\frac{y_h - y_l}{\bar{y}}}_{\equiv A}.$$

This is equivalent to

$$\tilde{f}_h(1 - f_h A) < f_h(1 - f_h A).$$

Thus, whenever $1 - f_h A > 0$, we have our result. Term A can be written as

$$A = \frac{w_h^{1+\varepsilon}(1-t)^\varepsilon - w_l^{1+\varepsilon}(1-t)^\varepsilon}{f_l w_l^{1+\varepsilon}(1-t)^\varepsilon + f_h w_h^{1+\varepsilon}(1-t)^\varepsilon} = \frac{w_h^{1+\varepsilon} - w_l^{1+\varepsilon}}{f_l w_l^{1+\varepsilon} + f_h w_h^{1+\varepsilon}}.$$

$f_h A < 1$ therefore implies

$$f_h w_h^{1+\varepsilon} - f_h w_l^{1+\varepsilon} < f_l w_l^{1+\varepsilon} + f_h w_h^{1+\varepsilon}$$

and hence

$$-f_h w_l^{1+\varepsilon} < f_l w_l^{1+\varepsilon}$$

which is always fulfilled.

B Nonlinear Taxes

B.1 Exogenous Education

The Lagrangian reads as

$$\begin{aligned} \mathcal{L} = & \sum_{i=l,h} \tilde{f}_i(c_i - \Psi(h_i)) + \lambda \sum_{i=l,h} f_i(w_i(e_i)h_i - c_i - e_i) \\ & + \eta \left(c_h - \Psi(h_h) - c_l + \Psi \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \right) \end{aligned}$$

Consumption levels are equivalent, therefore, let's just solve for $c_i = c_i^1 + c_i^2$.

$$\frac{\partial \mathcal{L}}{\partial c_l} = \tilde{f}_l - \lambda f_l - \eta = 0 \tag{21}$$

$$\frac{\partial \mathcal{L}}{\partial c_h} = \tilde{f}_h - \lambda f_h + \eta = 0 \tag{22}$$

which together yields $\lambda = 1$ and $\eta = \tilde{f}_l - f_l$.

$$\frac{\partial \mathcal{L}}{\partial h_l} = -\tilde{f}_l \Psi'_l + \lambda f_l w_l(e_l) + \eta \Psi' \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \frac{w_l(e_l)}{w_h(e_l)} = 0 \quad (23)$$

Solving (21) for λ inserting and then adding and subtracting $\eta \Psi'_l(h_l)$ yields:

$$-\Psi'_l(h_l) (\tilde{f}_l - \eta) + (\tilde{f}_l - \eta) w_l(e_l) + \eta \Psi' \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \frac{w_l(e_l)}{w_h(e_l)} - \eta \Psi'_l(h_l) = 0.$$

Dividing by w_l and using again $\lambda f_l = \tilde{f}_l - \eta$ and $\lambda = 1$ gives us

$$\tau_h = \frac{\eta}{w_l(e_l) f_l} \left(\Psi' \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \frac{w_l(e_l)}{w_h(e_l)} - \Psi'_l(h_l) \right)$$

i.e. a positive distortion for the low type.

$$\frac{\partial \mathcal{L}}{\partial h_h} = -\tilde{f}_h \Psi'_h + \lambda f_h w_h(e_h) - \eta \Psi'_h = 0 \quad (24)$$

Solving (22) for λ and inserting into (24) yields the usual no distortion at the top result.

B.2 Full Commitment

$$\begin{aligned} \mathcal{L} = & \sum_{i=l,h} \tilde{f}_i (c_i^1 + c_i^2 - \Psi(h_i)) + \lambda \sum_{i=l,h} f_i (w_i(e_i) h_i - c_i^1 - c_i^2 - e_i) \\ & + \eta \left(c_h^1 + c_h^2 - \Psi(h_h) - c_l^1 - c_l^2 + \Psi \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \right) \end{aligned}$$

From the Lagrangian one can directly see that c_i^1 and c_i^2 play exactly the same role and the allocation over time is therefore indeterminate. Thus, without loss of generality we can just solve for $c_i = c_i^1 + c_i^2$.

The first order conditions for c_h, c_l, h_h and h_l are completely identical to the case with exogenous education that is solved in Appendix B.1. The first-order conditions for education read as:

$$\frac{\partial \mathcal{L}}{\partial e_l} = \lambda \left(f_l h_l \frac{\partial w_l(e_l)}{\partial e_l} - f_l \right) + \eta \Psi' h_l \underbrace{\frac{\partial \frac{w_l(e_l)}{w_h(e_l)}}{\partial e_l}}_{<0} = 0 \quad (25)$$

and

$$\frac{\partial \mathcal{L}}{\partial e_h} = \lambda \left(f_h h_h \frac{\partial w_h(e_h)}{\partial e_h} - f_h \right) = 0. \quad (26)$$

Education wedges are defined by

$$s_i = 1 - (1 - \tau_i) h_i \frac{\partial w_i(e_i)}{\partial e_i}.$$

which yields $s_h = 0$ directly from (26). Given that there is no tax distortion on labor supply for the high type, there is no reason to counteract a tax distortion. Further, the education decision of the high type does not influence incentive constraints.

How about the low type. Divide (33) by λf_l and then add and subtract $\tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l}$ which yields:

$$h_l \frac{\partial w_l(e_l)}{\partial e_l} - 1 + \frac{\eta}{\lambda f_l} \Psi' h_l \frac{\partial \frac{w_l(e_l)}{w_h(e_l)}}{\partial e_l} + \tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l} - \tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l} = 0$$

and therefore we obtain

$$s_l = \tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l} + \frac{\tilde{f}_l - f_l}{\lambda f_l} \Psi' h_l \frac{\partial \frac{w_l(e_l)}{w_h(e_l)}}{\partial e_l}.$$

B.3 Partial Commitment

The problem of the government reads as

$$\begin{aligned} \max_{c_l^1, c_l^2, c_h^1, c_h^2, h_l, h_h, e_l, e_h} & \sum_{i=l,h} \tilde{f}_i (c_i^1 + c_i^2 - \Psi(h_i)) + \lambda \sum_{i=l,h} f_i (w_i(e_i) h_i - c_i^1 - c_i^2 - e_i) \\ & + \eta \left(c_h^1 + c_h^2 - \Psi(h_h) - c_l^1 - c_l^2 + \Psi \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \right) \\ & + \zeta \left(\sum_{i=1,2} \tilde{f}_i (c_i^2 - \Psi(h_i)) - W_D(e_l, e_h, c_l^1, c_h^1) \right) \end{aligned}$$

where

$$W_D(e_l, e_h, c_l^1, c_h^1) = \max_{c_l^2, c_h^2, h_l, h_h} \sum_{i=l,h} \tilde{f}_i (c_i^2 - \Psi(h_i)) + \lambda^D \sum_{i=l,h} f_i (w_i(e_i) h_i - c_i^1 - c_i^2 - e_i).$$

It can easily be shown that for the marginal value of public funds λ^D in the deviation case, we have:

$$\lambda^D = \frac{\tilde{f}_l}{f_l}.$$

The first-order conditions for consumption of the low type are

$$\frac{\partial \mathcal{L}}{\partial c_l^1} = \tilde{f}_l - \lambda f_l - \eta - \zeta \frac{\partial W_D}{\partial c_l^1} = 0 \quad (27)$$

$$\frac{\partial \mathcal{L}}{\partial c_l^2} = \tilde{f}_l (1 + \zeta) - \lambda f_l - \eta = 0 \quad (28)$$

Since $\frac{\partial W_D}{\partial c_l^1} = -\lambda f_l = -\tilde{f}_l$, consumption allocation across time is indeterminate for the low type.

The first-order conditions for the high type's consumption are given by

$$\frac{\partial \mathcal{L}}{\partial c_h^1} = \tilde{f}_h - \lambda f_h + \eta - \zeta \frac{\partial W_D}{\partial c_h^1} = 0 \quad (29)$$

$$\frac{\partial \mathcal{L}}{\partial c_h^2} = \tilde{f}_h (1 + \zeta) - \lambda f_h + \eta < 0 \quad (30)$$

Since $\frac{\partial W_D}{\partial c_h^1} = -\lambda f_h = -\tilde{f}_l \frac{f_h}{f_l}$, there is a corner solution with all consumption in first period for the high type, which is why (30) should be smaller than 0, i.e. a zero-consumption constraint must be binding.

Next, we turn to the first-order conditions for labor supply

$$\frac{\partial \mathcal{L}}{\partial h_l} = -\tilde{f}_l \Psi'_l (1 + \zeta) + \lambda f_l w_l(e_l) + \eta \Psi' \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \frac{w_l(e_l)}{w_h(e_l)} = 0 \quad (31)$$

Solving (28) for λ inserting and then adding and subtracting $\eta \Psi'_l$ yields:

$$-\Psi'_l \left(\tilde{f}_l (1 + \zeta) - \eta \right) + \left(\tilde{f}_l (1 + \zeta) - \eta \right) w_l(e_l) + \eta \Psi' \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \frac{w_l(e_l)}{w_h(e_l)} - \eta \Psi'_l = 0$$

Dividing by w_l and using again $\lambda f_l = \tilde{f}_l (1 + \zeta) - \eta$ gives us

$$\tau_h = \frac{\eta}{\lambda w_l(e_l) f_l} \left(\Psi' \left(h_l \frac{w_l(e_l)}{w_h(e_l)} \right) \frac{w_l(e_l)}{w_h(e_l)} - \Psi'_l \right)$$

i.e. a positive distortion for the low type.

$$\frac{\partial \mathcal{L}}{\partial h_h} = -\tilde{f}_h \Psi'_h (1 + \zeta) + \lambda f_h w_h(e_h) - \eta \Psi'_h = 0 \quad (32)$$

Solving (30) for λ and inserting into (32) yields the usual no distortion at the top result.

Next, we turn to the first-order conditions for education:

$$\frac{\partial \mathcal{L}}{\partial e_l} = \lambda \left(f_l h_l \frac{\partial w_l(e_l)}{\partial e_l} - 1 \right) + \underbrace{\eta \Psi' h_l \frac{\partial \frac{w_l(e_l)}{w_h(e_l)}}{\partial e_l}}_{<0} - \zeta \frac{\partial W_D}{\partial e_l} = 0 \quad (33)$$

$$\frac{\partial \mathcal{L}}{\partial e_h} = \lambda \left(f_h h_h \frac{\partial w_h(e_h)}{\partial e_h} - 1 \right) - \zeta \frac{\partial W_D}{\partial e_h} = 0. \quad (34)$$

Combining (27) and (29) yields

$$\lambda = 1 + \zeta \left(\tilde{f}_l + \frac{\tilde{f}_l}{f_l} f_h \right) = 1 + \zeta \tilde{f}_l \left(1 + \frac{f_h}{f_l} \right)$$

Inserting this into (27) yields $\eta = \tilde{f}_l - f_l$.

Dividing (33) by f_l and then adding and subtracting $\tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l}$ which yields:

$$s_l = \tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l} + \frac{\eta}{f_l} \Psi' h_l \underbrace{\frac{\partial \frac{w_l(e_l)}{w_h(e_l)}}{\partial e_l}}_{<0} + \zeta \frac{\tilde{f}_l}{f_l} \frac{\partial w_l(e_l)}{\partial e_l} (h_l - h_l^*).$$

References

- ABRAMITZKY, R., AND V. LAVY (2012): "How Responsive is Investment in Schooling to Changes in Redistributive Policies and in Returns?," *Working Paper*.
- ANDERBERG, D. (2009): "Optimal Policy and the Risk Properties of Human Capital Reconsidered," *Journal of Public Economics*, 93(9-10), 1017–1026.
- ANDERSSON, F., AND K. A. KONRAD (2003): "Human capital investment and globalization in extortionary states," *Journal of Public Economics*, 87(7), 1539–1555.

- BOADWAY, R., N. MARCEAU, AND M. MARCHAND (1996): "Investment in education and the time inconsistency of redistributive tax policy," *Economica*, pp. 171–189.
- BOVENBERG, L., AND B. JACOBS (2005): "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 89(11-12), 2005–2035.
- CARD, D. (1999): "The causal effect of education on earnings," *Handbook of labor economics*, 3, 1801–1863.
- CARNEIRO, P., AND J. HECKMAN (2005): "Human Capital Policy," *Inequality in America: What Role for Human Capital Policies*.
- DA COSTA, C. E., AND L. J. MAESTRI (2007): "The risk Properties of Human Capital and the Design of Government Policies," *European Economic Review*, 51(3), 695–713.
- DEMING, D., AND S. DYNARSKI (2009): "Into college, out of poverty? Policies to increase the postsecondary attainment of the poor," *NBER Working Paper 15387*.
- FARHI, E., C. SLEET, I. WERNING, AND S. YELTEKIN (2012): "Non-linear Capital Taxation Without Commitment," *The Review of Economic Studies*, 79(4), 1469–1493.
- FINDEISEN, S., AND D. SACHS (2014): "Education policies and taxation without commitment," *Mannheim Working Paper Series. Working Paper 14-16*.
- HASSLER, J., P. KRUSELL, K. STORESLETTEN, AND F. ZILIBOTTI (2005): "The dynamics of government," *Journal of Monetary Economics*, 52(7), 1331–1358.
- HASSLER, J., J. V. RODRÍGUEZ MORA, K. STORESLETTEN, AND A. ZILIBOTTI (2003): "The survival of the welfare state," *The American Economic Review*, 93(1), 87–112.
- HENDRICKS, L., AND T. SCHOELLMAN (2012): "Student Abilities During the Expansion of US Education, 1950-2000," *Working Paper*.
- JACOBS, B., AND A. BOVENBERG (2011): "Optimal taxation of human capital and the earnings function," *Journal of Public Economic Theory*, 13(6), 957–971.
- KANE, T. J. (2006): "Public intervention in post-secondary education," *Handbook of the Economics of Education*, 2, 1369–1401.

- KONRAD, K. A. (2001): "Privacy and time-consistent optimal labor income taxation," *Journal of Public Economics*, 79(3), 503–519.
- KYDLAND, F. E., AND E. C. PRESCOTT (1977): "Rules rather than discretion: The inconsistency of optimal plans," *The Journal of Political Economy*, pp. 473–491.
- LEMIEUX, T. (2006): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?," *The American Economic Review*, pp. 461–498.
- MALDONADO, D. (2008): "Education Policies and Optimal Taxation," *International Tax and Public Finance*, 15(2), 131–143.
- MIRRELES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38(2), 175–208.
- PEREIRA, J. (2009): "Essays on Time-Consistent Fiscal Policy," Ph.D. thesis, European University Institute.
- POUTVAARA, P. (2003): "Investment in education and redistributive taxation without precommitment," *FinanzArchiv/Public Finance Analysis*, pp. 177–188.
- (2006): "On the political economy of social security and public education," *Journal of Population Economics*, 19(2), 345–365.
- (2011): "The expansion of higher education and time-consistent taxation," *European Journal of Political Economy*, 27(2), 257–267.
- RICHTER, W. F. (2009): "Taxing education in Ramsey's tradition," *Journal of Public Economics*, 93(11), 1254–1260.
- SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.
- SHESHINSKI, E. (1972): "The optimal linear income-tax," *The Review of Economic Studies*, 39(3), 297–302.

- SONG, Z., K. STORESLETTEN, AND F. ZILIBOTTI (2012): "Rotten Parents and Disciplined Children: A Politico-Economic Theory of Public Expenditure and Debt," *Econometrica*, 80(6), 2785–2803.
- STANTCHEVA, S. (2013): "Optimal Income Taxation with Adverse Selection in the Labor Market," *Review of Economic Studies*, forthcoming.
- STIGLITZ, J. E. (1982): "Self-selection and Pareto efficient taxation," *Journal of Public Economics*, 17(2), 213–240.
- TABER, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?," *Review of Economic Studies*, 68(3), 665–691.
- TROSTEL, P. A. (1993): "The effect of taxation on human capital," *Journal of Political Economy*, pp. 327–350.