Erschienen in: The Scandinavian Journal of Economics. 2018, 120(4), pp. 1075-1099
https://doi.org/10.1111/sjoe.12246

Education Policies and Taxation without Commitment*

Sebastian Findeisen

University of Mannheim, DE-68131 Mannheim, Germany
findeisen@uni-mannheim.de

Dominik Sachs

University of Munich, DE-80539 Munich, Germany
dominik.sachs@econ.lmu.de

Abstract

We study the implications of limited commitment on education and tax policies set by benevolent governments. Consistent with real-world practices, a government can decide to subsidize different levels of education at different rates. A lack of commitment, however, affects the optimal structure of education subsidies. The direction of the effect depends on how labor taxes are designed. 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.

Keywords: Inequality; time inconsistency

I. Introduction

Public finance economists have long recognized that the challenges involved in designing optimal education policies and income tax systems are closely related. Income taxation influences the incentive to invest in education.[1] Education subsidies and policies, in turn, influence the choice of an optimal income tax system as they directly affect both the level and the distribution of wages. Many papers have studied the designs of education and tax policies jointly from a normative perspective – see, for example, Bovenberg and Jacobs (2005) for a state-of-the-art treatment

---

*We thankFriedrich Breyer, Leo Kaas, Fabrizio Zilibotti, Josef Zweimueller, and the two anonymous referees. D. Sachs' research was partly funded by a post-doctorate fellowship of the Fritz Thyssen Foundation and the Cologne Graduate School in Management, Economics and Social Sciences.

[1] See, for example, Abramitzky and Lavy (2014) for quasi-experimental evidence or Trostel (1993) for a structural approach.

in a heterogeneous agent model.[2] This strand of literature assumes that individuals make human capital investment decisions rationally, reacting to incentives set by the tax code and by education subsidies. Importantly, it is assumed that the government fully commits to the income tax schedule that it has announced before decisions on education are made.

In their work, Boadway *et al.* (1996) draw attention to the issue of time consistency, in the spirit of Kydland and Prescott (1977), which is key to the design of optimal tax and education policies. If the government fails to credibly commit to tax policies at a time when individuals are making education decisions, this can dramatically reduce the incentives of young individuals to invest in human capital. In their framework, Boadway *et al.* (1996) show that this results in underinvestment, 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, a 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 rates at which these three education levels are subsidized differ greatly.

We formalize this by allowing the government to set a non-linear schedule of education subsidies. We derive our results with a transparent and simple heterogeneous agent model with a two-type Stiglitz (1982) model. 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 is dependent 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 non-linear 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 type are set such that

---

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

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 for the fiscal externality, education is distorted downwards at the margin to relax the incentive constraint of the high type.

In cases with commitment problems, the government's tax promises lack credibility and individuals rationally anticipate that the government might reoptimize after education is sunk. More concretely, the government can deviate from its original policies, but this induces some output costs, capturing the idea of a reputational loss. In equilibrium, however, deviation does not occur.

The optimal tax rate is also higher than in the case with full commitment. The reason for this 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.

Finally, education subsidies become more progressive than in the full-commitment benchmark case. The key to this is that, in the case of deviation, the government opts for a higher tax rate. A more unequal distribution of wages makes the deviation and the higher tax rate more attractive to the social planner. This is because the planner values redistribution from high earners to low earners, and the incentive to reoptimize – 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 levels and, ultimately, wages in the next period.

Non-Linear 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 non-linear taxation and we address the reasoning behind it in more detail in Section IV. 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 non-linear. Limited commitment to tax policies makes education policies more regressive when the government uses non-linear 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 by the government that is sticking to its promises. In the case of deviation, the government reoptimizes after having learned the type of each individual and redistributes with type-dependent, lump-sum taxes, thereby setting 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. A higher level of education for the low type hence lead to a higher output gain for the deviating government as compared to the non-deviating government. To make this deviation less attractive and to resist its own temptation to reoptimize, 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 either case and so there is no difference in labor supply between the two cases. As a consequence, the education distortion is still zero even 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 always makes a deviation less likely, whereas a more compressed human capital (wage) distribution can make a deviation either less or more likely, depending on the 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 labor income tax, which creates a labor supply distortion. We show that the effect of limited commitment on education policies depends on how labor taxes are set; with non-linear labor taxation one in fact obtains a reversal of the result that greater inequality worsens credibility, because greater inequality makes a deviation less tempting.

Also found in the literature are related papers analyzing the dynamics and stability of redistributive policies, in particular 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 concerned with the rich dynamics that those papers characterize, but it does add education subsidies into the picture, which interact with redistributive labor tax policies in equilibrium.

This paper is also related to the work on time inconsistency and education policies by Konrad (2001) and Andersson and Konrad (2003).[3] 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.[4]

Lastly, Andersson and Konrad (2003) investigate education policies chosen by extortionary governments lacking commitment and assess how migration and tax competition affect policies.[5] We depart from these papers by placing our focus on non-linear education subsidies as used in the real world.

II. Environment

We consider a two-period model, in which *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, which we refer to as being of low-ability type 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 that 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 main assumptions on the wage function. First, education is productive and raises wages $\partial w_i(e)/\partial e > 0$ for $i = l, h$. Second, education and innate ability complement each other, implying higher marginal returns to education for the higher innate type: $\partial w_h(e)/\partial e - \partial w_l(e)/\partial e > 0$. Finally, innate abilities positively influence wages for a

[3]In a related paper, Pereira (2009) studies linear education subsidies and shows that these subsidies offset some of the excessive redistribution from income taxes in the case where the government lacks commitment.

[4]Poutvaara (2003) shows that redistribution without commitment might still involve a higher level of education than in the *laissez-faire* if the insurance effect of taxes is important.

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

given level of education: $w_h(e) - w_l(e) > 0$. None of these assumptions is required for most of the results we derive, in the sense that all formulas are valid if we deviate from them. The assumptions do, however, ease the understanding of the model and have strong empirical support. For example, Card (1999) provides a comprehensive literature review 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; and Taber (2001) and Hendricks and Schoellman (2014) suggest that much of the rise in the college premium can 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 represents hours worked. For simplicity, we assume that $\Psi(h) = h^{1+1/\varepsilon}/(1+1/\varepsilon)$, i.e., that Ψ exhibits a constant elasticity of labor supply ε. Income before tax is denoted by $y_i = w_i(e_i)h_i$ for $i = l, h$. Further, we assume no discounting and a zero interest 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 non-linear tax problem without commitment.

We are considering redistributive taxation. That is, we are interested in the policies of a government that wants to redistribute 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 $\tilde{f}_l/f_l > \tilde{f}_h/f_h$. Thus, the government's objective is $\sum_{i=l,h} \tilde{f}_i(U_i^1 + U_i^2)$.

When deciding on the optimal degree of redistribution, the government has to take into account that higher taxes will lower incentives to work and lower incentives to invest in education. Concerning the sophistication of policy instruments, we consider two scenarios. In Section III, we consider a planner that can use non-linear education subsidies but only has access to linear labor income taxes. The revenue from 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 was first studied by Sheshinski (1972). In Section IV, we assume that the government is only constrained by informational asymmetries in the tradition of the mechanism-design approach. This implies two changes. First, labor income tax rates can now be non-linear. Secondly, in the case of deviation there is no informational asymmetry. A deviating government has all the required information about types because types were revealed in the education period. A deviating government can therefore apply individualized lump-sum taxation.

III. Linear Tax Instruments

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

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. Here, t denotes the linear tax rate and T denotes the lump-sum transfer. The government's problem is then simply

$$\max_t \; \tilde{f}_l \left[(1-t)w_l(e_l)h[t, w_l(e_l)] + T - \Psi\{h[t, w_l(e_l)]\}\right] + \tilde{f}_h \left[(1-t)w_h(e_h)h[t, w_h(e_h)] + T - \Psi\{h[t, w_h(e_h)]\}\right], \qquad (1)$$

subject to a government budget constraint

$$T = t\,\{f_l w_l(e_l)h[t, w_l(e_l)] + f_h w_h h[t, w_h(e_h)]\} \qquad (2)$$

and optimal labor supply of the individuals

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

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 are affected. It is then easy to show that the optimal linear tax rate t^{ex}, in this case with exogenous human capital, satisfies

$$\frac{t^{ex}}{1-t^{ex}} = \frac{(\tilde{f}_l - f_l)[(y_h - y_l)/\bar{y}]}{\varepsilon}, \qquad (3)$$

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 $(y_h - y_l)/\bar{y}$ and decreasing in the elasticity of labor supply. Equation (3) is a variation for the optimal linear tax rate of Sheshinski (1972).[6] We do not provide a formal proof for this simple case here as it is nested in the following formulas with endogenous educational attainment.

[6] See Stantcheva (2014) for a similar formula in a discrete-type setting.

Optimal Policies with Full Commitment

The Government's Problem. We now consider the case where the educational decision is endogenous and the government can influence the agents' decisions by setting a non-linear subsidy schedule. Thus, the government chooses a (non-linear) subsidy function $S(e)$ and an income tax rate t subject to a government budget constraint and to the behavioral responses of individuals. Thus, formally we have

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

subject to a government budget constraint

$$T = t\,\{f_l w_l(e_l)h[t,w_l(e_l)] + f_h w_h(e_h)h[t,w_h(e_h)]\} - f_l S(e_l) - f_h S(e_h)$$

and optimal individual behavior

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

This problem has some similarities to the problem in Stiglitz (1982), where a non-linear 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 equation (5) by

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

and an incentive compatibility constraint[7]

$$c_h^1 + (1-t)w_h(e_h)h[t,w_h(e_h)] - \Psi\{h[t,w_h(e_h)]\} \geq c_l^1 + (1-t)w_h(e_l)h[t,w_h(e_l)] - \Psi\{h[t,w_h(e_l)]\}. \quad (7)$$

Notice that in the incentive constraint (7), there is a deviation utility on the right-hand side (i.e., the terms $w_h(e_l)$ and $h[t,w_h(e_l)]$). 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 by $h[t,w_h(e_l)] = h^c$ and the associated income by $y^c = h^c w_h(e_l)$, with a c for the counterfactual, as in equilibrium the high type will be truth-telling.

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

The government's problem can now be written as

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

subject to a government budget constraint

$$\begin{aligned} T \quad = \quad & t\left\{f_l w_l(e_l)h[t, w_l(e_l)] + f_h w_h(e_h)h[t, w_h(e_h)]\right\} \\ & - f_l(c_l^1 + e_l) - f_h(c_h^1 + e_h), \end{aligned} \tag{9}$$

and subject to equations (6) and (7), where we denote by η the Lagrangian multiplier of the incentive compatibility constraint. The Lagrangian and the first-order conditions are stated in Online Appendix A.1. Note that, in fact, c_l^1, c_h^2, and T are not pinned down uniquely. Because of the quasi-linearity of preferences, individuals are indifferent as to the best time 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.

Solution (8) can then be implemented with a non-linear 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 – equation (5) has to hold. Naturally, infinitely many non-linear subsidy schedules can implement the desired allocation, as in the non-linear tax problem with a two-type Stiglitz (1982) model. In the following, we are 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 the 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 what follows, therefore, we are 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 for the remainder of this paper.[8] Note

[8] As in the optimal taxation problem with discrete types, we can always pick a non-linear subsidy schedule such that the first-order conditions of an individual 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 this in a similar framework.

also that throughout this paper, we only characterize marginal subsidies and not average subsidies.

Optimal Tax and Education Policies. We start by characterizing the optimal linear income tax rate. As shown in the following proposition, the optimal linear tax rate is corrected by the endogeneity of education as compared to the optimal tax rate with exogenous education in (3).

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

$$\frac{t^f}{1-t^f} = \frac{(\tilde{f}_l - f_l)[(y_h - y_l)/\bar{y}] - \eta[(y_h - y^c)/\bar{y}]}{\varepsilon},$$

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

Proof: See Online Appendix A.1. □

The optimal tax rate for the case of endogenous education decisions also increases in income inequality and decreases in the level of the labor supply elasticity. As can be seen, there is an additional force given by $\eta[(y_h - y^c)/\bar{y}]$ in the numerator as compared to the case where education is taken to be exogenous. This decreases the optimal tax rate, and the bigger the difference $y_h - y^c$ the stronger the effect. y^c is the income level that the high type would attain if it received the education level of the low type e_l. Thus, the difference captures the effect of a high education level for the high type on earnings. The more significant the effect of education on earnings, the lower the tax rate tends to be. Consider the 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 1. *Let* e_l^* *and* e_h^* *be the solution to problem (8). Then the respective optimal linear tax rate is lower than the linear tax rate as defined by equation (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 $(y_h - y_l)/\bar{y} = (y_h - y_l)/(f_l y_l + f_h y_h)$ change by the same factor. Then the corollary follows directly. □

Income taxes are not the only tool at the government's disposal. Governments do rely on education subsidies to increase incentives to invest in education. We now characterize optimal education subsidies.

Proposition 2. *In a full-commitment economy, education subsidies satisfy*

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

and

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

Proof: See Online Appendix A.1. □

Looking at the education subsidy for the low type, one can see that there are two parts to it. The first term reflects the fiscal externality effect of individuals' education decisions: these impose an externality on the government budget, as individuals with a higher education level pay higher taxes. The government internalizes this fiscal externality by subsidizing education in a Pigouvian way. As the formula reveals, the more elastic the labor supply, the larger the subsidy. In addition, the more individuals' working hours are affected by wage increases, the greater the fiscal externality of the government's budget. On a related note, the education subsidy increases in the marginal return of education $\partial w_l(e_l)/\partial e_l$ and in the income tax rate. Notice that even if the labor supply is not distorted ($\varepsilon = 0$), the education decision is still 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 complement each other. The marginal return to education increases with innate ability. As the government's policies are redistributive, there is a movement towards reducing education subsidies, as they tend to benefit the initially high types. Maldonado (2008) shows that in the case of a correlation between educational investment and innate ability, education should be taxed. See also Jacobs and Bovenberg (2011) for a discussion of this issue.[9] For the high type, only the fiscal externality part is present

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

because a standard “no-distortion-at-the-top” result applies for the second part.

Optimal Policies with Lack of Commitment

We now look at economies where the degree of government commitment is allowed to differ, which comprises the case of full commitment of the previous section as a special case.

Costs of Deviation and Commitment. Following Farhi *et al.* (2012), we introduce the 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 of reputational loss. Farhi *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 reduced investment in future generations.

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

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

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

These types of deviation costs make it possible to flexibly capture different levels of limited commitment. At one extreme, when κ is zero, there is no way for the government to credibly commit and we arrive at the no-commitment case. This case is also studied in an earlier version of this paper (Findeisen and Sachs, 2014). At the other extreme, when κ is above some positive threshold $\bar{\kappa} > 0$, all tax promises are fully credible and we arrive at the full-commitment solution, 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 the deviating government is implementing. It can be shown that a deviating government would set the tax rate according to the same rule as in equation (3). The reason for this is that for a deviating government education incentives are considered sunk. This is summarized in the following lemma.

Lemma 1. *A deviating government takes education levels as exogenous and therefore sets the linear tax rate according to*

$$\frac{t^{dev}}{1-t^{dev}} = \frac{(\tilde{f}_l - f_l)[(y_h - y_l)/\bar{y}]}{\varepsilon}. \tag{11}$$

Optimal Policies and Discussion. In comparison to the full-commitment problem, 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 Online Appendix A.2. The following proposition shows the optimal income tax rate for this case.

Proposition 3. *In a partial-commitment economy, the optimal linear tax rate satisfies*

$$\frac{t^{pc}}{1-t^{pc}} = \frac{(\tilde{f}_l - f_l)[(y_h - y_l)/\bar{y}] - [\eta/(1+\zeta)][(y_h - y^c)/\bar{y}]}{\varepsilon}. \tag{12}$$

Proof: See Online Appendix A.2. □

We can see how this case comprises 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 $\kappa > \bar{\kappa}$), then 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 to complement education subsidies. This effect is now scaled down by $1/(1+\zeta)$. 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 4. *In a partial-commitment economy, education subsidies satisfy*

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

where

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

and

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

where

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

Proof: See Online Appendix A.2. □

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

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

and

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

This implies that whenever there is a commitment problem, the marginal value of education subsidies for the low type increase as this strengthens the credibility of the tax promises. The marginal benefit of high-level education subsidies decreases, as a higher level of education for the high type increases the government's temptation to renege on tax promises and increase the tax rate in order to redistribute.

The reasoning behind this result lies in the difference between the tax rates on the equilibrium path and in case of deviation. On the equilibrium path, the tax rate is set lower than during deviation. The deviating planner will take education as sunk and set a tax rate similar to that in the problem with exogenous education (11). This makes education for the low type unambiguously more attractive on the equilibrium path than in the deviation case, and is therefore a driving force for higher education subsidies for the low type. This is the case because the low type works more hours on the equilibrium path (as the tax rate is lower), which directly increases the benefits of better education. Also, the low type keeps a share $(1-t)$ of their earnings, whereas the share t is divided between the high type and the low type through the payment of the lump-sum transfer T. Consumption of the low type is valued more highly. Because $(1-t)$ is higher on the equilibrium path, education of the low type is also valued more highly through this channel.

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

We now present a complementary result that we discussed initially in the introduction. Note that the difference between the tax rates off and on the equilibrium path captures the planner's incentives to deviate in a single number. The former is given by equation (11) and the latter by equation (12). The incentive to deviate is stronger the larger $[\eta/(1+\zeta)][(y_h - y^c)/\bar{y}]$. The difference $y_h - y^c$ in turn increases with the wage difference. Thus, a more equal wage distribution makes this term smaller and therefore the tax rates on and off equilibrium are more similar. In other words, a more equal wage distribution renders the deviation less attractive.

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

IV. Non-Linear Labor Taxes

We now turn to policies that are only constrained by informational asymmetries. The difference between this section and the previous section is twofold. On the one hand, the government can tax income at a non-linear rate. The second aspect concerns the deviation: if the government deviates from its announced policy path and reoptimizes, it no longer faces an information problem because individuals have revealed their types 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 the implications as compared to the case with linear taxes.

Optimal Policies with Exogenous Education

As a benchmark, we look at the case with only one period and exogenous levels of education. Individuals differ in their wages w_h and w_l. The government maximizes the usual social objective $\sum_{i=l,h} f_i[c_i - \Psi(h_i)]$. Therefore, 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 was first studied extensively by Stiglitz (1982). As we show in Online Appendix B.1, optimal marginal tax rates in this case satisfy

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

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

with $\eta = \tilde{f}_l - f_l$. The low type faces a positive distortion that increases with redistributive preferences (i.e., $\tilde{f}_l - f_l$), inequality and the labor supply elasticity. The latter two are captured by

$$\left\{\Psi'(h_l) - \Psi'\left[\frac{h_l w_l(e_l)}{w_h(e_h)}\right]\frac{w_l(e_l)}{w_h(e_h)}\right\}.$$

This term increases with $w_h(e_h)/w_l(e_l)$ (hence the inequality) and decreases with the labor supply elasticity – the more convex Ψ, the higher the labor supply elastcity.

The result in equation (15) is similar to the result obtained for the optimal linear tax rate (3) in Section III. The key difference is that the 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 (e.g., Mirrlees, 1971; Stiglitz, 1982; Saez, 2001). This might seems counterintuitive at first. However, it does not take into account the average tax rate that the high-productivity type pays. Inframarginal tax rates can be quite high. The reason the marginal tax rate is zero is that it only applies to the very last (marginal) unit of income and would therefore not in fact raise any revenue.[10] However, it would distort the high type's labor supply decisions. Note that this has already been applied to education subsidies in Proposition 2.

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

Optimal Policies with Full Commitment

We now turn to the case with two periods, where again educational investments are made in the first period, but this time with non-linear labor income taxes. In this case, the government's problem is written 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 following proposition characterizes optimal policies.

Proposition 5. *In the full-commitment case with non-linear labor taxes, the optimal allocation has the following properties. (a) First- and second-period consumption is not pinned down; only* $c_h^1 + c_h^2$ *and* $c_l^1 + c_l^2$ *are pinned down. (b) Labor wedges are still characterized by equations* (15) *and* (16). *(c) Optimal education subsidies are given by*

$$s_l^f = \tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l} + \frac{\eta}{\lambda f_l} \Psi' h_l \frac{\partial [w_l(e_l)/w_h(e_l)]}{\partial e_l}$$

and $s_h^f = 0$*, with* $\eta = \tilde{f}_l - f_l$*.*

Proof: See Online Appendix B.2. □

First of all, consumption levels are not pinned down because individuals are indifferent as to whether they consume in period 1 or period 2. This is related to the previous discussion. Importantly, this changes when we consider lack of commitment; see the discussion after Proposition 6.

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

Third, for education subsidies we find similar results for the low type as we do 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 movement 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 present because of the usual no-distortion-at-the-top result, as described in the previous section.

Optimal Policies with Lack of Commitment

We now study the limited commitment case. If the planner deviates from the announced tax plan, κ units of output 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 Online 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 the announced tax policies once 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 the 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 reoptimization stage, the planner is able to identify individuals and no longer faces an informational constraint. The planner can hence assign zero labor wedges for each type: $\tau_h^{dev} = \tau_h^{dev} = 0$. 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 6. *In the partial commitment case with non-linear labor taxes, the optimal allocation has the following properties. (a) Consumption for the low type across time is again indeterminate; for the high type, all consumption is front-loaded to period 1. (b) Labor wedges are still characterized by equations* (15) *and* (16). *(c) Education wedges are*

$$\begin{aligned} s_l^{dev} &= \tau_h h_l \frac{\partial w_l(e_l)}{\partial e_l} + \frac{\eta}{f_l} \Psi' h_l \underbrace{\frac{\partial [w_l(e_l)/w_h(e_l)]}{\partial e_l}}_{<0} \\ &\quad + \underbrace{\zeta \frac{\tilde{f}_l}{f_l} \frac{\partial w_l(e_l)}{\partial e_l} (h_l - h_l^{dev})}_{<0} \end{aligned}$$

and $s_h^{dev} = 0$, *with* $\eta = \tilde{f}_l - f_l$.

Proof: See Online Appendix B.3. □

A first interesting result is that consumption of the high type is completely front-loaded. In general, the individual is indifferent as to the best time to consume. As a result of the government's temptation, however, it is better to give the high type all consumption in the first period so that these resources cannot be used for excessive redistribution in the case of a deviation, which makes the deviation less attractive.[11] In contrast, for the low type, 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. Note also 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 those of Section III with linear labor taxes. Here, with non-linear labor taxes, the lack of commitment leads to a lower education subsidy for the low type and a constant zero subsidy for 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? The key factor 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, unlike for the case with linear labor taxes, this additional redistribution is not carried out with more progressive taxes but with type-dependent, lump-sum taxes. Labor distortions are also zero. Therefore, labor supply is at its first-best efficient level for the low type in the deviation case. Thus, a higher level of education for the low type is more highly valued in the deviation case, which makes deviation more tempting at the margin. For the high type, this motivation is not there because the high type faces a zero labor supply distortion both on the equilibrium path and in the deviation case.[12]

---

[11] Note that this was not the case for linear taxation. The reason for this is that, for 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 a deviation case, the planner could then also pay out a lower lump-sum transfer T. But this decrease of T due to front-loading is the same both on and off the equilibrium path and therefore does not relax or tighten the credibility constraint in the linear tax case.

[12] Note that for 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 non-linear tax case, these effects are not present because the planner is not constrained on how to use the additional resources through education.

Taken together, lack of commitment leads to more regressive education subsidies. For the low type, lack of commitment creates a driving force towards a lower education subsidy. For the high type, the education subsidy remains at zero. This result of more regressive subsidies increases with ζ (i.e., the severity of the commitment problem).

V. 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 straightforward: at the stage when education decisions are sunk, the decision of whether to reoptimize 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 the labor taxes. First, we 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 increases with wage inequality. The government tends to make education subsidies

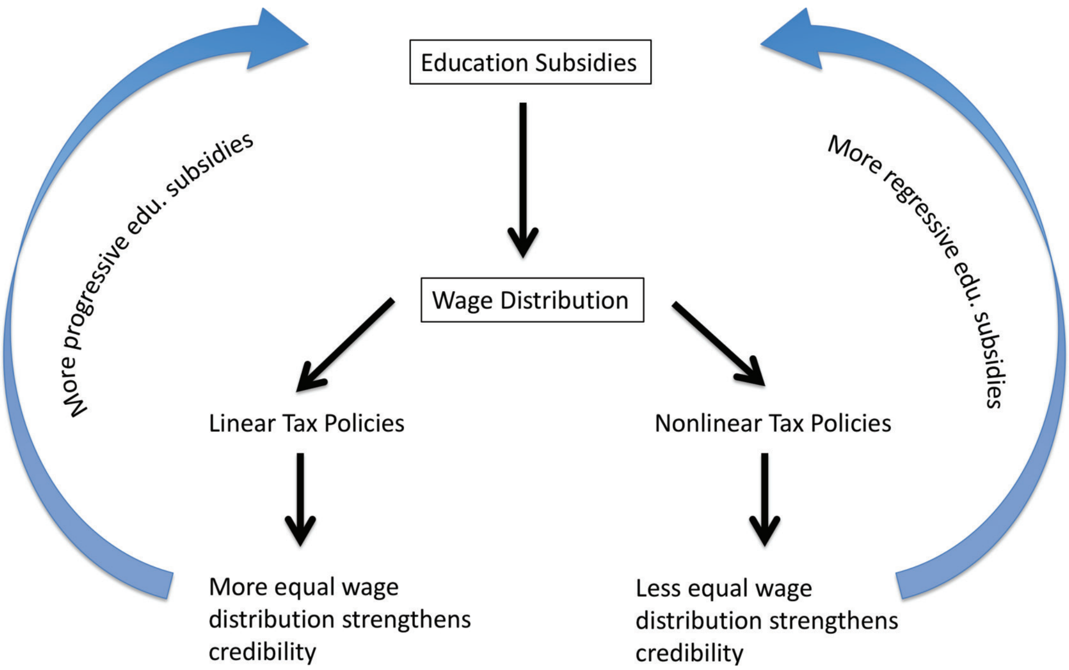

Fig. 1. Link between education subsidies and policy credibility

more progressive, relative to the full-commitment case, in order to achieve a more equal distribution of wages.

In contrast, in the non-linear labor taxation case, where private information about types leads to deviations from the first best, a more unequal wage distribution strengthens credibility. The payoff of a deviation then decreases with wage inequality. This is because in the deviation case informational frictions disappear and the government can set zero labor wedges and implement an efficient level of labor supply. This implies that 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) improve the commitment problem and make tax promises set *ex ante* more credible. Thus, the government implements more regressive education subsidies to achieve a more unequal wage distribution.

Differences from Capital Taxation and Limited Commitment

The results from the model with linear labor taxes are reminiscent of the results of Farhi *et al.* (2012). They show that a more equal distribution of 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 lower wealth inequality are more credible as they reduce the chance of excessive wealth distribution later on. Human capital, by contrast, is only taxed indirectly through labor income tax. What we show is that policies leading to greater wage inequality can be more or less credible, depending on the pattern of labor distortions.

Other Public Spending

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

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

where $u_i(g)$ is the type-dependent utility for the public good. Because 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 non-linear tax case, the optimal level of the public good would be the same in both the second- and first-best cases. Consequently, it would be the same on and off the equilibrium path and would not alter the commitment problem in any way. How might the situation differ for linear taxes? Unless there was a strong preference for the public good and one ended up in a corner solution, where all tax revenue was used for g and $T = 0$, there would be no influence on deviation incentives. The reason for this is that, if not in a corner solution, a deviating government will use all its 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.[13]

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 tax rates when undertaking human capital investments. This is in line with results for a constant tax rate, where the government sets lower taxes when taking education incentives into account.[14] Abramitzky and Lavy (2014) provide quasi-experimental evidence on the negative effect of redistributive taxation on education investment. More structural and model-based approaches, such as the classic work by Trostel (1993), have also unearthed the major effects of income taxation on human capital investment. A second operative margin in the theory is the fact that individuals' education decisions are influenced by education subsidies. There is a large body of empirical literature estimating this for college enrolment, surveyed by Kane (2006) and Deming and Dynarski (2009). The consensus is that the behavioral response is of a sizable magnitude. 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 on how lack of commitment influences education policies, depending on how labor taxes are designed. Therefore, we leave a detailed investigation of these issues for further research.[15]

---

[13]Certainly, one could think of other ways to model the public good such that it interacted with labor supply, for example. However, given the standard way of modeling public goods in political economics or macroeconomics as additively separable (e.g., Song *et al.*, 2012), our results would be unaffected by the presence of public goods.

[14]In the non-linear case, optimal tax rates are described by the same formula both taking and not taking into account endogenous human capital investment; see Section IV.

[15]In an earlier version of this paper, we provided suggestive cross-national evidence for a more regressive incidence of education subsidies when a government's ability to commit is high.

VI. Conclusion

In this study, we have built a simple model that shows how education and tax policies are affected when the government lacks full commitment. It is assumed that individuals are born with different levels of innate abilities, undertake human capital investments early in life and make labor supply decisions later. While we also characterize labor tax policies in detail, our main results concern the design of education policies. The impact of commitment problems on education subsidies depends on the labor tax system: education subsidies become more progressive if labor tax rates are linear but more regressive if the labor tax system is non-linear, in the spirit of Mirrlees (1971). 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. In the case with physical capital, wealth can always be taxed and redistributed directly by the government. Human capital, by contrast, is taxed indirectly only by labor income tax, which also creates a labor supply distortion. This interaction with labor tax means that the results depend on what labor tax systems are available to the government. Further work might integrate wealth and human capital accumulation together with capital taxation and education policies into one model; we leave this for a future study.

References

Abramitzky, R. and Lavy, V. (2014), How Responsive is Investment in Schooling to Changes in Redistributive Policies and in Returns?, *Econometrica 82*, 1241–1272.

Anderberg, D. (2009), Optimal Policy and the Risk Properties of Human Capital Reconsidered, *Journal of Public Economics 93*, 1017–1026.

Andersson, F. and Konrad, K. A. (2003), Human Capital Investment and Globalization in Extortionary States, *Journal of Public Economics 87*, 1539–1555.

Boadway, R., Marceau, N., and Marchand, M. (1996), Investment in Education and the Time Inconsistency of Redistributive Tax Policy. *Economica 63*, 171–189.

Bovenberg, L. and Jacobs, B. (2005), Redistribution and Education Subsidies are Siamese Twins, *Journal of Public Economics 89*, 2005–2035.

Card, D. (1999), The Causal Effect of Education on Earnings, in O. C. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Volume 3A, Elsevier, Amsterdam, 1801–1863.

Carneiro, P. and Heckman, J. (2005), Human Capital Policy, in B. M. Friedman (ed.), *Inequality in America: What Role for Human Capital Policies?*, MIT Press, Cambridge, MA, 77–240.

Da Costa, C. E. and Maestri, L. J. (2007), The Risk Properties of Human Capital and the Design of Government Policies, *European Economic Review 51*, 695–713.

Deming, D. and Dynarski, S. (2009), Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor, National Bureau of Economic Research (NBER) Working Paper 15387.

Farhi, E., Sleet, C., Werning, I., and Yeltekin, S. (2012), Non-linear Capital Taxation Without Commitment, *Review of Economic Studies 79*, 1469–1493.

Findeisen, S. and Sachs, D. (2014), Education Policies and Taxation without Commitment, Mannheim Working Paper 14-16.

Hassler, J., Rodríguez Mora, J. V., Storesletten, K., and Zilibotti, A. (2003), The Survival of the Welfare State, *American Economic Review 93* (1), 87–112.

Hassler, J., Krusell, P., Storesletten, K., and Zilibotti, F. (2005), The Dynamics of Government, *Journal of Monetary Economics 52*, 1331–1358.

Hendricks, L. and Schoellman, T. (2014), Student Abilities during the Expansion of US Education, 1950–2000, *Journal of Monetary Economics 63*, 19–36.

Jacobs, B. and Bovenberg, A. (2011), Optimal Taxation of Human Capital and the Earnings Function, *Journal of Public Economic Theory 13*, 957–971.

Kane, T. J. (2006), Public Intervention in Post-secondary Education, in E. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Volume 2, Elsevier, Amsterdam, 1369–1401.

Konrad, K. A. (2001), Privacy and Time-consistent Optimal Labor Income Taxation, *Journal of Public Economics 79*, 503–519.

Kydland, F. E. and Prescott, E. C. (1977), Rules Rather than Discretion: The Inconsistency of Optimal Plans, *Journal of Political Economy 85*, 473–491.

Lemieux, T. (2006), Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? *American Economic Review 96* (3), 461–498.

Maldonado, D. (2008), Education Policies and Optimal Taxation, *International Tax and Public Finance 15*, 131–143.

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

Pereira, J. (2009), Essays on Time-Consistent Fiscal Policy, PhD Thesis, European University Institute, Firenze, Italy.

Poutvaara, P. (2003), Investment in Education and Redistributive Taxation without Precommitment, *FinanzArchiv (Public Finance Analysis) 59*, 177–188.

Poutvaara, P. (2006), On the Political Economy of Social Security and Public Education, *Journal of Population Economics 19*, 345–365.

Poutvaara, P. (2011), The Expansion of Higher Education and Time-consistent Taxation, *European Journal of Political Economy 27*, 257–267.

Richter, W. F. (2009), Taxing Education in Ramsey's Tradition, *Journal of Public Economics 93*, 1254–1260.

Saez, E. (2001), Using Elasticities to Derive Optimal Income Tax Rates, *Review of Economic Studies 68*, 205–229.

Sheshinski, E. (1972), The Optimal Linear Income-tax, *Review of Economic Studies 39*, 297–302.

Song, Z., Storesletten, K., and Zilibotti, F. (2012), Rotten Parents and Disciplined Children: A Politico-economic Theory of Public Expenditure and Debt, *Econometrica 80*, 2785–2803.

Stantcheva, S. (2014), Optimal Income Taxation with Adverse Selection in the Labor Market, *Review of Economic Studies 81*, 1296–1329.

Stiglitz, J. E. (1982), Self-selection and Pareto Efficient Taxation, *Journal of Public Economics 17*, 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*, 665–691.

Trostel, P. A. (1993), The Effect of Taxation on Human Capital, *Journal of Political Economy 101*, 327–350.