University of Mannheim / Department of Economics

Working Paper Series

Education Policies and Taxation without Commitment

Sebastian Findeisen Dominik Sachs

Working Paper 14-16

May 2014

Education Policies and Taxation without Commitment *

Sebastian Findeisen
University of Mannheim

Dominik Sachs
CGS, University of Cologne

This version: May 27, 2014

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. Deviations from full commitment tend to make education policies more progressive, increasing the education subsidy for initially low skilled agents and decreasing it for initially high skilled agents. We provide suggestive cross-country correlations for this mechanism.

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

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

*Contact: findeisen@uni-mannheim.de, sachs@wiso.uni-koeln.de. We thank Friedrich Breyer, Leo Kaas, Fabrizio Zilibotti, Josef Zweimueller and seminar participants in Konstanz and Zurich for helpful comments and discussions. Sebastian Findeisen acknowledges the support of the University of Zurich (Forschungskredit of the University of Zurich, grant no. 53210603). Dominik Sachs' research is 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, as we will also exploit in our empirical section, the rate at which these different education levels are subsidized is very different.[3] We formalize this by allowing the government to set a nonlinear schedule of education subsidies. The income tax rate is linear and the revenue is redistributed lump-sum and used to finance education subsidies. We derive our results in a transparent and simple heterogeneous

[1]See 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.

[2]See 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 uncerainty.

[3]In fact, we will focus on two education levels in the theoretical analysis and think of primary education as exogenous. In the empirical part we look at cross-country correlations on relative public education expenditures on the tertiary level, relative to the primary and secondary level combined.

agent model with two types (Stiglitz 1982). Consistent with empirical evidence, individual wages are determined by both innate abilities and education levels.

We first consider the two polar cases where the government has full commitment to stick to tax promises and no commitment at all. Under full commitment the optimal income tax rate takes into account education incentives. The tax rate is smaller, when the effect of education on wages is large relative to the effect of innate abilities on wages for the initially high skilled. Intuitively, the more important the role of education for wages the more important taxes become to incentivize high-skilled agents to self-select into high education level. 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 as in 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.

Without any commitment, no tax promise of the government is credible and individuals rationally anticipate that the government re-optimizes after education is sunk. In line with previous results, this leads to excessive taxation and depresses human capital investment. An important result we find here concerns the design of education subsidies. We show that they tend to become more progressive when there is a commitment problem. The intuition is that a higher subsidy for low types and a lower subsidy for high types will compress the distribution of education. As education inequality is reduced, also the wage distribution in the next period is more compressed. And as wage inequality decreases, the redistributive government sets a lower tax rate in the second period. This lower tax rate will help to boost education incentives and helps to alleviate the commitment problem. This is consistent with the recent results from Farhi et al. (2012) who first detected a similar channel for nonlinear capital taxation.

We move on to study intermediate scenarios, where the government has some form of limited commitment, nesting the two polar cases. More concretely, the government can deviate from its announcements but this induces some output costs capturing the idea of a reputational loss.[4] The forward looking government wants to avoid costly deviation and announces policies respectively. Labor income taxes are still designed to take into account their effect on education incentives. However, the strength of the effect is decreasing the more severe the commitment problem is because too low tax promises lack credibility. Education policies become more progressive the more severe the commitment problem if

[4]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 depressed investment of future generations.

the government is sufficiently redistributive towards low types. Another way to look at this result is to interpret the design of education and tax policies as a choice to engage in redistribution *ex-ante* through more progressive education subsidies – which decreases wage inequality – as opposed to engaging in redistribution *ex-post* through income taxation. With a lack of commitment, the government tries to weaken its own temptation to engage in costly ex-post redistribution by increasing the amount of ex-ante redistribution.

We conclude by providing suggestive evidence for the described mechanisms in the form of cross-country correlations. We proxy for commitment power using data from the World Bank's Worldwide Governance Indicators. Specifically, we use the variable Government Effectiveness capturing "...the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." (Kaufmann et al., 2010). Controlling for income, geographical variables and the overall share of government involvement in education, we find a robust correlation, indicating that countries with higher policy credibility employ more regressive subsidies.

As already mentioned, this paper is related to Farhi et al. (2012), who consider capital taxation without commitment. An important difference is that, in their framework, deviation could imply full redistribution, i.e. 100% marginal tax rates on capital. In our case, deviation is constrained to be less extreme as the government still has to respect labor supply responses and therefore cannot fully tax away human capital returns. For simplicity, we work with a two period model, as their pioneering work already proves the equivalence in the results between the dynamic repeated game and the simple two period version.

This paper also relates to the work on time inconsistency and education policies by Konrad (2001) and Andersson and Konrad (2003).[5] Konrad (2001) shows how the time inconsistency problem is alleviated by the presence of private information in an optimal taxation framework. In particular, he shows that the strong no-education result obtained in Boadway, Marceau, and Marchand (1996) no longer applies, as with private information some rents of education are still captured by individuals, preserving some incentives to invest in education.[6] In our framework, a similar logic applies as the government uses a linear tax income rate together with lump-sum transfers, in the spirit of a simple negative income tax system. This also preserves some incentives to invest in education, even in the complete absence of credible policy promises, as full equalization of incomes is not feasible. In contrast to Konrad (2001), we consider ex-ante heterogeneous individu-

[5]In 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.

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

als and thereby address the progressivity of education subsidies. Lastly, Andersson and Konrad (2003) investigate education policies chosen by extortionary governments lacking commitment and how migration and tax competition affect policies.[7]

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

In Section 2, we introduce the formal model and look at optimal income taxation with exogenous education as a benchmark case. In Section 3, we derive optimal policies for the full-commitment government before we look at the other extreme case where the government cannot commit at all in Section 4. We then look at the intermediate case of partial commitment in Section 5 before we present some suggestive empirical evidence in Section 6 and conclude in Section 7.

2 The Model Basics

In this section, we present the model basics and characterize optimal taxes with exogenous education, an important benchmark which helps to understand the further results with endogenous education.

2.1 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 ex-ante heterogeneous agents. The θ_1-type and the θ_2-type with $\theta_2 > \theta_1$. Their masses are $f(\theta_1)$ and $f(\theta_2)$ with $f(\theta_1) + f(\theta_2) = 1$ and θ is assumed to be private information. In Period 1, they make a monetary educational investment e. The wage w they earn in period two is a function of innate type and education, i.e. $w(\theta, e)$.

We impose three intuitive assumptions on the wage function $w(\theta, e)$. First, education is productive and raises wages $\frac{\partial w(\theta,e)}{\partial e} > 0$. Second education and innate ability are complements implying higher marginal returns to education for the higher innate type: $\frac{\partial w(\theta_2,e)}{\partial e} - \frac{\partial w(\theta_1,e)}{\partial e} > 0$. Finally, innate abilities positively influence wages for a given level of education: $w(\theta_2, e) - w(\theta_1, e) > 0$. None of these assumption are needed for most of the

[7] In 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.

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. To minimize the notational burden we often write all the variables not as a function of θ but with subscript instead. E.g. e_1 instead of $e_1(\theta_1)$ or w_2 instead of $w(e_2(\theta_2), \theta_2)$. The utility functions are $U^1 = c^1$ in period one and $U^2 = c^2 - \Psi(h)$ in period two, 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 $y_i = w_i h_i$. Further we assume no discounting and a zero interest rate for notational convenience.

We are considering redistributive linear taxation. That is, we are interested in the policies of a government that is interested in redistributing from the high type θ_2 to the low type θ_1 via linear taxes used to finance a lump-sum rebate such as in a negative income tax system. To capture this redistributive concern, we set the Pareto weights $\tilde{f}(\theta_1)$ and $\tilde{f}(\theta_2)$ such that $\frac{\tilde{f}(\theta_1)}{f(\theta_1)} > \frac{\tilde{f}(\theta_2)}{f(\theta_2)}$. 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. The education margin, however, can also be influenced by nonlinear education subsidies. 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.

2.2 Optimal Policies with Exogenous Education

Assume a one period setting where education levels e_1 and e_2 are exogenous. In that case, the only relevant margin for the government when choosing taxes is the labor-leisure margin. The problem of the government then simply is

$$\max_t \quad \tilde{f}(\theta_1)\Bigg((1-t)w_1 h(t,w_1) + T - \Psi[h(t,w_1)]\Bigg)$$
$$+\tilde{f}(\theta_2)\Bigg((1-t)w_2 h(t,w_2) + T - \Psi[h(t,w_2)]\Bigg) \tag{1}$$

subject to a government budget constraint

$$T = t\left(f(\theta_1)w_1 h(t,w_1) + (1-f(\theta_1))w_2 h(t,w_2)\right) \tag{2}$$

and optimal labor supply of the individuals

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

The government thus 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.[8] It is then easy to show that the optimal linear tax rate t^{ex}, in this case with exogenous human capital, satisfies

$$\frac{t^{ex}}{1-t^{ex}} = \frac{\left(\tilde{f}(\theta_1) - f(\theta_1)\right)\left(\frac{y_2-y_1}{\bar{y}}\right)}{\varepsilon}, \tag{3}$$

where $\bar{y}$ is average income $f(\theta_1)y_1 + f(\theta_2)y_2$. The optimal tax rate is increasing in redistributive preferences $\left(\tilde{f}(\theta_1) - f(\theta_1)\right)$, increasing in inequality measured by $\frac{y_2-y_1}{\bar{y}}$ and decreasing in the elasticity of labor supply. The formula (3) is a variation for the optimal linear tax rate of Sheshinski (1972).[9] We refrain from providing a formal proof for this simple case as it is nested in the following formulas with endogenous educational attainment.

[8]In addition, one must also respect a non-negativity constraint on consumption that might be binding for some "extreme" Pareto weights. In the following we assume that the Pareto weights which we are considering will never be that "extreme".

[9]See Stantcheva (2013) for a similar formula in a discrete type setting.

3 Optimal Policies with Full Commitment

3.1 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[10] subject to a government budget constraint and subject to behavioral responses of the individuals. Thus, formally we have:

$$\begin{aligned} \max_{t,S(\cdot)} \quad & \tilde{f}(\theta_1)\Bigg((1-t)w_1h(t,w_1)+T-\Psi[h(t,w_1)]-e_1+S(e_1)\Bigg) \\ & +\tilde{f}(\theta_2)\Bigg((1-t)w_2h(t,w_2)+T-\Psi[h(t,w_2)]-e_1+S(e_2)\Bigg) \end{aligned} \tag{4}$$

subject to a government budget constraint

$$T = t\;\Bigg(f_1w(e_1,\theta_1)h(t,w_1)+f_2w_2h(t,w_2)-f_1S(e_1)-f_2S(e_2)\Bigg)$$

and optimal individual behaviour

$$\forall\; i=1,2\;:\;(e_i,h_i)=\arg\max_{e,h}(1-t)w(e,\theta_i)h+T-\Psi(h)-e+S(e). \tag{5}$$

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_1, e_2, c_1^1, c_2^1 directly, where c_1^1 and c_2^1 denote first period consumption.[11] In that case, we can replace (5) by

$$h(t,w_i)=\arg\max_{h}(1-t)hw_i-\Psi(h) \tag{6}$$

[10]As individuals already reveal their type with their education decision, the government could actually also levy individualized lump sum taxes in period 2 as argued by Boadway, Marceau, and Marchand (1996). In the spirit of the *Ramsey approach*, we constrain the labor income taxes to be linear, however. Alternatively, we could overcome this kind of informational inconsistency by assuming uncertainty in education returns as Konrad (2001). We choose the case with certainty and constrained linear taxes, however, because the exposition becomes particularly clear and the result about the relation between commitment power and progressivity of education subsidies is more transparent.

[11]With linear utility, the timing of consumption across the first and the second period is not determined. This implies that there are alternative ways to state the problem. E.g., we dropped first period consumption in problem (4) without loss of generality.

and an incentive compatibility constraint[12]

$$\begin{aligned}&c_2^1 + (1-t)w_2 h(t,w_2) - \Psi(h(t,w_2))\\ \geq &c_1^1 + (1-t)w(e_1,\theta_2)h(t,w(e_1,\theta_2)) - \Psi(h(t,w(e_1,\theta_2))).\end{aligned} \tag{7}$$

Notice that in the incentive constraint (7) the deviation utility on the right-hand-side, the terms $w(e_1,\theta_2)$ and $h(t,w(e_1,\theta_2))$, show up. A deviating high-skilled agent receives the education level of the low skilled agent e_1. 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 will call this $w(e_1,\theta_2) = w^c$, with a c for counterfactual as in equilibrium by incentive compatibility the wage will never be observed. We call the associated hour choice $h(t,w(e_1\theta_2)) = h^c$ and associated income $y^c = h^c w^c$.

The government's problem now reads as:

$$\begin{aligned}\max_{c_1^1,c_2^1,t,e_1,e_2} \quad &\tilde{f}(\theta_1)\Bigg(c_1^1 + (1-t)w_1 h(t,w_1) + T - \Psi[h(t,w_1)]\Bigg)\\ &+\tilde{f}(\theta_2)\Bigg(c_2^1 + (1-t)w_2 h(t,w_2) + T - \Psi[h(t,w_2)]\Bigg)\end{aligned} \tag{8}$$

subject to a government budget constraint

$$T = t\,\Bigg(f_1 w_1 h(t,w_1) + f_2 w_2 h(t,w_2) - f_1(c_1^1+e_1) - f_2(c_2^1+e_2)\Bigg), \tag{9}$$

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.

The solution (8) can then be implemented with a nonlinear subsidy function $S(\cdot)$ that has to yield the desired consumption levels, i.e. $S(e_1) = c_1^1 + e_1$ and $S(e_2) = c_2^1 + e_2$. 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

[12]Since we assume $\tilde{f}(\theta_1) > f(\theta_1)$, we focus on downward redistributive taxation where only the incentive constraint of the θ_2-type is binding.

at e_1 and e_2. In these cases, we know that the first-order condition for education of an individual can be rearranged as:

$$(1 - S'(e(\theta_i))) = (1-t)\frac{\partial w_i}{\partial e} h_i \ \ \forall\, i = 1,2.$$

In the following, we will therefore be interested in

$$s(\theta_i) \equiv 1 - (1-t)\frac{\partial w_i}{\partial e} h_i \ \ \forall\, i = 1,2.$$

Having computed an optimal allocation, we can therefore infer the implicit marginal education subsidies $s(\theta_1)$ and $s(\theta_2)$ for this allocation. For simplicity, we will call $s(\theta_1)$ and $s(\theta_2)$ education subsidies in the remainder of this paper.[13] Note also that throughout this paper, we only characterize *marginal* subsidies and not *average* subsidies.

3.2 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*

$$\frac{t^f}{1-t^f} = \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left(\frac{y_2 - y_1}{\bar{y}}\right) - \eta\left(\frac{y_2 - y_2^c}{\bar{y}}\right)}{\epsilon},$$

where the multiplier satisfies $\eta = \tilde{f}(\theta_1) - f(\theta_1)$.

Proof. See Appendix A.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 $\eta\left(\frac{y_2 - y_2^c}{\bar{y}}\right)$ 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 $y_2 - y_2^c$. y_2^c is the income level that the high type θ_2 would attain when only taking the education level of the low type e_1. The difference, hence, captures the effect of a higher education level for the high type on her earnings. The more important the effect of

[13] As 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 $w(\theta, e)$ have to be made, see, e.g., Bovenberg and Jacobs (2005, p.2010) for a discussion of that in a similar framework.

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 $y_2 = y_2^c$. 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 $y_1 = y_2^c$, 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_1^* and e_2^* 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 $e_1 = e_1^*$ and $e_2 = e_2^*$, i.e. $t^f(e_1^*, e_2^*) < t^{ex}(e_1^*, e_2^*)$.*

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*

$$s^f(\theta_1) = t^f \frac{\partial w_1}{\partial e_1} h_1(1+\epsilon) - \frac{\eta}{f(\theta_1)}(1-t^f)\left[h_2^c \frac{\partial w_2^c}{\partial e_1} - h_1 \frac{\partial w_1}{\partial e_1}\right]$$

and

$$s^f(\theta_2) = t^f \frac{\partial w_2}{\partial e_2} h_2(1+\epsilon).$$

Proof. See Appendix A.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 (Bovenberg and Jacobs, 2005): 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 $\frac{\partial w_1}{\partial e_1}$ and in the income tax rate.

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 if this issue.[14] For the high type θ_2 only the fiscal externality part is present because a standard "no-distortion-at-the-top" result applies for the second part.

4 Optimal Policies without Commitment

We now characterize policies when the government has no commitment power and contrast them to the full-commitment results. We start backwards, looking at optimal tax policies, once education decisions are sunk.

4.1 The Problem in Period Two

The problem of the planner is basically equivalent to that of the planner in Section 2.2, as the distribution of wages is taken as exogenous. In particular, the same tax formula applies:

$$\frac{t^{nc}}{1-t^{nc}} = \frac{\left(\tilde{f}(\theta_1) - f(\theta_1)\right)\left(\frac{y_2-y_1}{\bar{y}}\right)}{\varepsilon}. \tag{10}$$

In the following, we write the optimal tax rate for the second period planner as a function of both education levels, so $t^{nc}(e_1, e_2)$.

[14]Maldonado (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.

4.2 The Problem in Period One

In the first period, the planner anticipates that he will set taxes according to (10). Therefore, in the first period, the problem reads as:

$$\max_{c_1,c_2,e_1,e_2} = \tilde{f}(\theta_1)\Bigg(c_1^1 + (1 - t^{nc}(e_1,e_2))w_1 h_1(t^{nc}(e_1,e_2),w_1)$$

$$+ T - \Psi[h_1(t^{nc}(e_1,e_2),w_1)]\Bigg)$$

$$+ \tilde{f}(\theta_2)\Bigg(c_2^1 + (1 - t^{nc}(e_1,e_2))w_2 h_2(t^{nc}(e_1,e_2),w_2)$$

$$+ T - \Psi[h_2(t^{nc}(e_1,e_2),w_2)]\Bigg)$$

subject to the government budget constraint

$$T = t^{nc}(e_1,e_2)\Bigg(f(\theta_1)w_1 h_1(t^{nc}(e_1,e_2),w_1)$$

$$+ (1 - f(\theta_1))w_2 h_2(t^{nc}(e_1,e_2),w_2)$$

$$- f(\theta_1)c_1^1 - f(\theta_2)c_2^1 - f(\theta_2)e_2 - f(\theta_1)e_1\Bigg) \tag{11}$$

and the incentive constraint

$$c_2^1 + (1 - t^{nc}(e_1,e_2))w_2 h_2 - \Psi(h_2) \geq c_1^1 + (1 - t^{nc}(e_1,e_2))w_2^c h_2^c - \Psi(h_2^c).$$

This problem is very similar to the problem in Section 3. The difference is that the government cannot choose t but instead takes into account how it will choose t in the future once education decisions are sunk. The Lagrangian and first-order conditions are stated in Appendix B. The following proposition shows how optimal optimal education subsidies are designed in a no-commitment economy.

Proposition 4. *In a no-commitment economy, education subsidies satisfy*

$$s^{nc}(\theta_1) = t^{nc}\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - \frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_1}(y_2 - y_2^c)$$

and

$$s^{nc}(\theta_2) = t^{nc}\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - \frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_2}(y_2 - y_2^c)$$

where $\frac{\partial t^{nc}}{\partial e_1} < 0$ *and* $\frac{\partial t^{nc}}{\partial e_2} > 0$ *and the multiplier satisfies* $\eta = \tilde{f}(\theta_1) - f(\theta_1)$.

Proof. See Appendix B.1 □

In comparison to the full-commitment case in Proposition 3, there is now an additional term in both formulas for the optimal education subsidy. In fact, for the low type, this additional term favors higher subsidies and for the high type, this additional term favors lower subsidies. Together this tends to make education policies more progressive.

The intuition behind this result is clear and simple. In the case of the low type the additional term is given by:

$$-\frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_1}(y_2 - y_2^c) > 0.$$

A higher education subsidy and a higher level of education for the low type will decrease the optimal tax rate chosen by the government in period two $\frac{\partial t^{nc}}{\partial e_1} < 0$. This will strengthen education incentives for both types. In other words, the government anticipates its temptation to set too high taxes in the second period. By compressing the distribution of education across the two agents, it can avoid some of the harmful spillover from too high taxes on the education margin. Consistent with that argument, there is a downward adjustment in the optimal subsidy for the high type

$$-\frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_2}(y_2 - y_2^c) < 0,$$

as a higher education level for the high type will tend to increase taxes because of higher income inequality.

5 Varying the Degree of Commitment

In the previous section we studied two polar cases. We now look at economies, where the degree of commitment power of the government is allowed to differ, nesting the two cases from the previous sections. This allows us to show that smoother versions of our previous results hold.

5.1 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:

$$\mathcal{W}^2_{PC}(e_1, e_2, t) \geq \mathcal{W}^2_{Dev}(e_1, e_2) - \kappa, \tag{12}$$

where $\mathcal{W}^2_{PC}(e_1, e_2, t)$ is second period welfare as a function of education levels for both types and the promised tax rate t, under the assumption that the government sticks to its promise. $\mathcal{W}^2_{Dev}(e_1, e_2)$ 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 2.2.

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 from Section 4. 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, which naturally achieves the highest welfare. In this section we focus on the intermediate cases where κ lies between zero and $\bar{\kappa}$.

5.2 Optimal Policies and Discussion

In comparison to the full-commitment problem in Section 3, the government has to respect the credibility constraint (12) 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 C. 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:*

$$\frac{t^{pc}}{1 - t^{pc}} = \frac{\left(\tilde{f}(\theta_1) - f(\theta_1)\right)\left(\frac{y_2 - y_1}{\bar{y}}\right) - \frac{\eta}{1+\zeta}\left(\frac{y_2 - y_2^c}{\bar{y}}\right)}{\epsilon},$$

Proof. See Appendix C.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:*

$$s^{pc}(\theta_1) = t^{pc}\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - \frac{\eta}{f(\theta_1)}(1-t^{pc})\left[h_2^c\frac{\partial w_2^c}{\partial e_1} - h_1\frac{\partial w_1}{\partial e_1}\right] + \frac{\zeta}{f(\theta_1)}\left(\frac{\partial \mathcal{W}_{PC}}{\partial e_1} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_1}\right)$$

where $\frac{\partial \mathcal{W}_{PC}}{\partial e_1} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_1} > 0$ *and*

$$s^{pc}(\theta_2) = t^{pc}\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) + \frac{\zeta}{f(\theta_2)}\left(\frac{\partial \mathcal{W}_{PC}}{\partial e_2} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_2}\right).$$

where $\frac{\partial \mathcal{W}_{PC}}{\partial e_2} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_2} < 0$.

Proof. See Appendix C.2 □

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

$$\frac{\zeta}{f(\theta_1)}\left(\frac{\partial \mathcal{W}^2_{PC}}{\partial e_1} - \frac{\partial \mathcal{W}^2_{Dev}}{\partial e_1}\right) > 0$$

and

$$\frac{\zeta}{f(\theta_2)}\left(\frac{\partial \mathcal{W}^2_{PC}}{\partial e_2} - \frac{\partial \mathcal{W}^2_{Dev}}{\partial e_2}\right) < 0$$

respectively. This implies that whenever there is a commitment problem, the marginal value of low level education goes up as it strengthens the credibility of tax promises. Relatedly, the marginal value of high level education goes down as it increases the temptation to renege on tax promises and increase the tax rate to redistribute. Taken together, a more compressed education distribution leads to a more compressed wage distribution, decreasing the value of an ex-ante harmful deviation of the government. With limited commitment, the government wants to avoid excessive ex-post redistribution, by engaging already in ex-ante redistribution through the use of education policies. The larger the commitment problem, the larger ζ and the stronger this effect on the progressivity of education subsidies. In the next subsection, we illustrate with a numerical example how a decrease in κ leads to more progressive education subsidies.

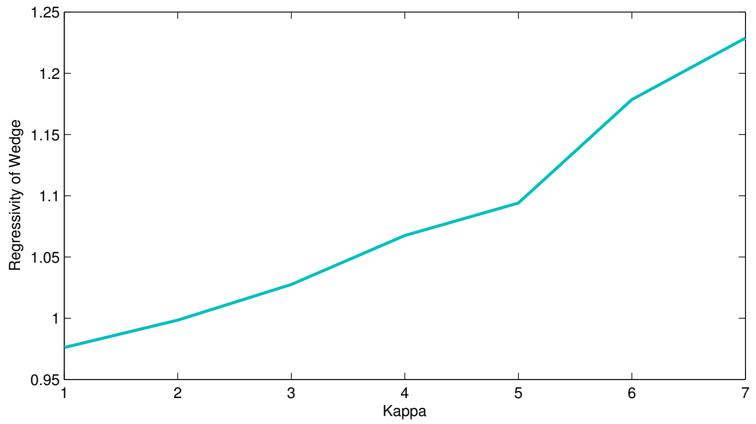

Figure 1: Regressivity of Education Subsidies

5.3 Numerical Illustration

We assume an equal mass of high and low types $f(\theta_1) = f(\theta_2) = 0.5$ and set the welfare weights to $\tilde{f}(\theta_1) = 0.9$ and $\tilde{f}(\theta_2) = 0.1$. We set the labor supply elasticity to 0.25. For wages, we assume that they are determined by a simple Cobb Douglas production function $w_i = \theta_i^{0.5} e_i^{0.5}$ with equal weights. There is a constant marginal cost of education. We start by assuming that the government has limited commitment power and the cost of reneging on tax promises (κ) is set at 5% of output calculated from the full-commitment economy.

The equilibrium tax rate in the partial-commitment case is $t^{pc} = 35.56\%$. For comparison it is $t^{fc} = 19.56\%$ in the full-commitment case. This illustrates the workings of the formula in Proposition 5, as the human capital effect on taxes is scaled down relative to the full-commitment benchmark, because the government lacks full credibility. A deviating government which would take the wage distribution as exogenously given (Section 2.2) would set a tax rate of 65.95%. Thus, although the government lacks commitment power, it still takes human capital investment incentives into account as it sets a significantly smaller tax rate (about 30 percentage points).

The main predictions from our analysis of education subsides concern the degree of the progressivity of education subsidies, see Proposition 6. In line with these predictions we find that the ratio of the subsidies $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ of the high relative to the low type is 0.99, as compared to 2.38 in the full-commitment case. A higher $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ ratio implies a more regressive incidence of subsidies.

Finally, we illustrate how the regressivity of education policies varies with the commitment technology. Figure 1 plots $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ against κ as it varies, measured in percentage points of output lost when reneging on tax promises. Moving from left to right, the commitment power of the government gradually increases. In line with the mechanism outlined above, a government with more commitment power can afford to set less progressive subsidies as its credibility increases.

6 Empirical Implications

The model predicts a more regressive incidence of subsidies when the ability of a government to commit is high. We now provide suggestive cross-country evidence for this. The estimates of this section should be interpreted as correlations only.

Data. As a measure for commitment power and the credibility of policy announcements we use data from the World Bank's Worldwide Governance Indicators database. We use the variable Government Effectiveness capturing "...the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." (Kaufmann et al., 2010). To proxy for the regressivity/progressivity of public education subsidies, we use the share of public education expenditures at each educational level relative to total public education expenditures. We then take the share spent on tertiary education relative to the total spending share on all lower levels of education (primary and secondary). The bigger the value of this variable, the more regressive is the incidence of public education expenditure, in the sense that more is spent on tertiary education relative to lower tier education. To construct the measure, we take data from the UNESCO on public educational expenditures across education levels. We also use GDP data, which we take from the World Bank database. We use the year 2008 as the most recent year with a reasonable number of observations. We are left with a sample of 54 countries, for which all the relevant data is available.

Results. For Figure 2, we regress our measure of the regressivity of public education on the government credibility index. The figure is an added-variable plot so that the regressively index is de-meaned. The correlation is positive and highly significant.[15] The coefficient in column one of Table 1 implies that a one standard deviation increase in policy credibility increases the regressivity of public education expenditures by 0.53 standard deviations. Next, we include continent dummies. Only exploiting the variation within continents increases the credibility coefficient. Adding the log of per capita GDP does not

[15] It is even stronger when excluding Lesotho and Cuba, which are two outliers with high regressivity but weak institutional commitment. On the other side, Singapore is an outlier with very strong policy credibility and a high incidence of regressivity.

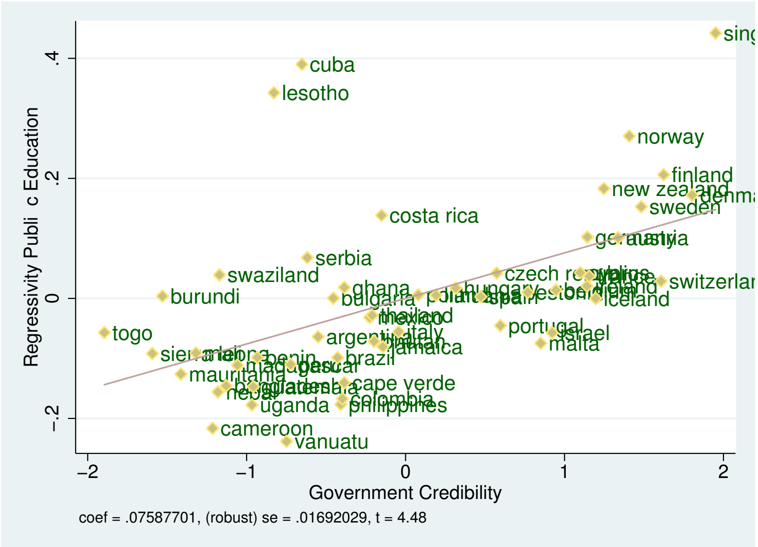

Figure 2: Government Education Expenditures and Policy Credibility

Correlation between regressivity in education expenditures, as proxy for subsidies, and government credibility. Corresponds to estimates from first column in Table 1. Since regressions include an intercept, the regressivtiy index on the Y-axis is measured relative to the cross-sectional sample mean.

Table 1: Credibility of Government and Education Regressivity

Dependent variable: Regressivity of Education Expenditure				
Policy Credibility	0.759***	1.223***	1.329**	1.160**
	(0.169)	(0.332)	(0.550)	(0.560)
Log GDP			- 0.020	- 0.021
			(0.030)	(0.032)
Total Education				0.031**
Expenditure Share				(0.013)
Continent Dummies	No	Yes	Yes	Yes
R-squared	0.290	0.352	0.356	0.490

Observations: 54. Year: 2008. List of countries see Appendix. Policy Credibility coefficient multiplied by 10. Robust errors. Last column based on 52 observation, since data on Bhutan and Uganda is missing.

affect the conclusion. Maybe surprisingly, income per capita seems not to be correlated with a more regressive incidence of public expenditure, as is seen in column three. The raw correlation between GDP per capita and our regressivity index is, however, positive and significant (0.49). But as the estimates indicate, this effect vanishes with continent dummies and controlling for government credibility. Finally, we control for the overall share of public education expenditures aggregated across all levels as a fraction of GDP. This approximates for the overall importance of the public sector in providing and paying for education. The main correlation concerning the effect of governmental policy credibility remains unaffected. As column four shows, countries in which the government has a relatively larger stake in education, tend to spend more on higher education.

7 Conclusion

Optimal income tax and education policies depend on the degree of commitment power or policy credibility the government has. We build a transparent and simple heterogeneous agent model to understand the economic mechanisms involved. Individual wages are determined by both innate abilities and education levels. Without any commitment, the labor income tax does not take into account the incentives to acquire education. When some or full commitment is available, income tax rates are adjusted to incentivize education. The tax rate is smaller, when the effect of education on wages is large relative to the effect of innate abilities on wages for the initially high skilled. We allow the government to subsidize different levels of education at different rates. The main implication of limited commitment is that education policies become more progressive relative to the full-commitment benchmark: the government takes into account that a more compressed wage distribution limits its own temptation to tax excessively. By adjusting the distribution of education, the government effectively creates its own commitment device. This mirrors previous findings from Farhi, Sleet, Werning, and Yeltekin (2012) concerning the design of capital taxes. Using data on the credibility of policy announcements from the World Bank database, we find a positive and significant correlation between the degree of commitment power and how regressive education expenditures are across countries, consistent with the mechanism highlighted in this paper. This correlation is conditional on income and geographical controls.

A The Full-Commitment Planner

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

$$\begin{aligned}\mathcal{L} =& \tilde{f}(\theta_1)\Bigg(c_1^1 + (1-t)w(e_1,\theta_1)h(t,w_1) - \Psi[h(t,w_1)]\Bigg)\\ &+ \tilde{f}(\theta_2)\Bigg(c_2^1 + (1-t)w_2(e_2,\theta_2)h(t,w_2) - \Psi[h(t,w_2)]\Bigg)\\ &+ t\,\Bigg(f_1 w(e_1,\theta_1)h(t,w_1) + (1-f_1)w(e_2,\theta_2)h(t,w_2) - f_1(c_1^1+e_1) - f_2(c_2^1+e_2)\Bigg)\\ &+ \eta\left(c_2^1 + (1-t)w_2 h(t,w_2) - \Psi(h(t,w_2)) - c_1^1 + (1-t)y^c - \Psi(h^c)\right)\end{aligned}$$

The first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial c_1^1} = \tilde{f}_1 - f_1 - \eta = 0$$

$$\frac{\partial \mathcal{L}}{\partial c_2^1} = \tilde{f}_2 - f_2 + \eta = 0$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t} =& -\tilde{f}(\theta_1)y_1 - (1-\tilde{f}(\theta_1))y_2 + f(\theta_1)y_1 + (1-f(\theta_1))y_2 + tf(\theta_1)w_1\frac{\partial h_1}{\partial t}\\ &+ t(1-f(\theta_1))w_2\frac{\partial h_2}{\partial t} - \eta(w_2h_2 - w_2^c h_2^c) = 0\end{aligned} \tag{13}$$

$$\frac{\partial \mathcal{L}}{\partial e_1} = \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t)h_2^c\frac{\partial w_2^c}{\partial e_1}\right] = 0,$$

$$\frac{\partial \mathcal{L}}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2 = 0.$$

From the FOC for c_1^1, one directly obtains $\eta = \tilde{f}_1(\theta) - f_1(\theta)$.

A.1 Proof of Proposition 1

Two manipulations of (13) yield

$$\left(\tilde{f}_1 - f_1\right)(y_2 - y_1) - \frac{t}{1-t}\left(f_1 w_1 h_1 \frac{\partial h_1}{\partial 1-t}\frac{1-t}{h_1} + f_2 w_2 h_2 \frac{\partial h_2}{\partial 1-t}\frac{1-t}{h_2}\right) - \eta(y^c - y_1).$$

Now use $\varepsilon = \frac{\partial h_2}{\partial 1-t}\frac{1-t}{h_2} = \frac{\partial h_1}{\partial 1-t}\frac{1-t}{h_1}$ and $\bar{y} = f_1 y_1 + f_2 y_2$ and solve for $\frac{t}{1-t}$ to obtain the result.

A.2 Proposition 3

We start with the θ_2 type. Rewriting the FOC for e_2 yields

$$\tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \left(f_2 - \tilde{f}_2\right)(1-t)\frac{\partial w_2}{\partial e}h_2 = 0,$$

which yields

$$tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + f_2(1-t)\frac{\partial w_2}{\partial e}h_2 = 0.$$

This can be rewritten as

$$f_2(1-t)\frac{\partial w_2}{\partial e}h_2 = 1 - t\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon),$$

where the RHS is the definition of the implicit education subsidy for the θ_2-type.

Now we look at the θ_1-type. Rewriting the FOC for e_1 yields:

$$\begin{aligned}&\tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t)h_2^c\frac{\partial w_2^c}{\partial e_1}\right]\\ &+ \eta(1-t)h_1\frac{\partial w_1}{\partial e_1} - \eta(1-t)h_1\frac{\partial w_1}{\partial e_1} = 0.\end{aligned}$$

Now use $\eta = \tilde{f}_1 - f_1$ and obtain

$$tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta(1-t)\left[h_1\frac{\partial w_1}{\partial e_1} - h_2^c\frac{\partial w_2^c}{\partial e_1}\right] - f_1(1-t)h_1\frac{\partial w_1}{\partial e_1} = 0.$$

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

B The No-Commitment Planner

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

$$
\begin{aligned}
\mathcal{L} =& \tilde{f}(\theta_1)\Bigg(c_1^1 + (1-t^{nc}(e_1,e_2))w_1(e_1,\theta_1)h_1(t^{nc}(e_1,e_2),w_1) - \Psi[h_1(t^{nc}(e_1,e_2),w_1)]\Bigg)\\
&+\tilde{f}(\theta_2)\Bigg(c_2^1 + (1-t^{nc}(e_1,e_2))w_2(e_2,\theta_2)h_2(t^{nc}(e_1,e_2),w_2) - \Psi[h_2(t^{nc}(e_1,e_2),w_2)]\Bigg)\\
&+t^{nc}(e_1,e_2)\Bigg(f(\theta_1)w_1(e_1,\theta_1)h_1(t^{nc}(e_1,e_2),w_1)\\
&+(1-f(\theta_1))w_2(e_2,\theta_2)h_2(t^{nc}(e_1,e_2),w_2) - f(\theta_1)c_1^1 - f(\theta_2)c_2^1 - f(\theta_2)e_2 - f(\theta_1)e_1\Bigg)\\
&+\eta\Bigg(c_2^1 + (1-t^{nc}(e_1,e_2))w_2h_2 - \Psi(h_2) - c_1^1 + (1-t^{nc}(e_1,e_2))w_2^c h_2^c - \Psi(h_2^c)\Bigg)
\end{aligned}
$$

For the first-order condition for e_1 and e_2, we know that their impact on Period 2 welfare via t is zero due to the envelope theorem. Thus the first-order conditions read as:

$$
\frac{\partial\mathcal{L}}{\partial e_1} = \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t^F)h_2^c\frac{\partial w_2^c}{\partial e_1}\right] - \eta\frac{\partial t}{\partial e_1}\left[w_2h_2 - w_c^c h_2^c\right] = 0,
$$

$$
\frac{\partial\mathcal{L}}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2 - \eta\frac{\partial t}{\partial e_2}\left[w_2h_2 - w_c^c h_2^c\right] = 0.
$$

B.1 Proof of Proposition 4

The optimal education subsidies can be obtained by some analytical manipulations almost equivalently as in Appendix A.2.

We now look at the derivatives of t with respect to e_1 and e_2. We know that:

$$
\frac{t}{1-t} = \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left[\frac{w_2h_2 - w_1h_1}{\overline{wh}}\right]}{\epsilon}.
$$

We now show that t is increasing in e_2 and decreasing in e_1. Define the implicit function:

$$F(e_1, e_2, t(e_1, e_2)) = \frac{t}{1-t} - \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left[\frac{w_2 h_2 - w_1 h_1}{\overline{wh}}\right]}{\epsilon} = 0.$$

As $F = 0$ for any (e_1, e_2), the derivatives of F w.r.t to e_1 and e_2 have to be zero as well. In general these derivates are characterized by

$$\frac{\partial F}{\partial t}\frac{\partial t}{\partial e_i} + \frac{\partial F}{\partial e_i} = 0$$

and therefore can reveal the sign of $\frac{\partial t}{\partial e_i}$. Spelling this out for e_1 yields

$$\frac{\partial t}{\partial e_1}\left[\frac{1}{(1-t)^2} - \frac{(\tilde{f}(\theta_1) - f(\theta_1))}{\epsilon \overline{y}^2}\left(\left(\frac{\partial y_2}{\partial t} - \frac{\partial y_1}{\partial t}\right)\overline{y} + \left(f(\theta_1)\frac{\partial y_1}{\partial t} + f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2 - y_1)\right)\right]$$
$$+ \frac{(\tilde{f}(\theta_1) - f(\theta_1))}{\epsilon \overline{y}^2}\frac{\partial y_1}{\partial e_1}((y_2 - y_1)f(\theta_1) + \overline{y}) = 0$$

and hence

$$\frac{\partial t}{\partial e_1}\left[\frac{1}{(1-t)^2} - \frac{(\tilde{f}(\theta_1) - f(\theta_1))}{\epsilon \overline{y}^2}\underbrace{\left(\left(\frac{\partial y_2}{\partial t} - \frac{\partial y_1}{\partial t}\right)\overline{y} + \left(f(\theta_1)\frac{\partial y_1}{\partial t} + f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2 - y_1)\right)}_{\equiv X}\right]$$
$$+ \frac{(\tilde{f}(\theta_1) - f(\theta_1))}{\epsilon \overline{y}^2}\frac{\partial y_1}{\partial e_1}y_2 = 0.$$

For X we obtain:

$$\begin{aligned}
&\left(\frac{\partial y_2}{\partial t} - \frac{\partial y_1}{\partial t}\right)\overline{y} - \left(f(\theta_1)\frac{\partial y_1}{\partial t} + f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2 - y_1) \\
&= \frac{\partial y_1}{\partial t}(-f(\theta_1)y_1 - (1 - f(\theta_1))y_2 - f(\theta_1)y_2 + f(\theta_1)y_1) + \frac{\partial y_2}{\partial t}(f(\theta_1)y_1 \\
&+ (1 - f(\theta_1))y_2 - (1 - f(\theta_1))y_2 + (1 - f(\theta_1))y_1)) \\
&= \frac{\partial y_1}{\partial t}(-y_2) + \frac{\partial y_2}{\partial t}y_1 = \frac{\partial y_1}{\partial 1-t}y_2 - \frac{\partial y_2}{\partial 1-t}y_1 = \varepsilon\frac{y_1 y_2}{1-t} - \varepsilon\frac{y_1 y_2}{1-t} = 0,
\end{aligned}$$

so it follows $\frac{\partial t}{\partial e_1} < 0$. Similar reasoning shows $\frac{\partial t}{\partial e_2} > 0$.

C The Partial-Commitment Planner

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

$$
\begin{aligned}
\mathcal{L} =& \tilde{f}(\theta_1)\Bigg(c_1^1 + (1-t)w(e_1,\theta_1)h(t,w_1) - \Psi[h(t,w_1)]\Bigg)\\
&+ \tilde{f}(\theta_2)\Bigg(c_2^1 + (1-t)w_2(e_2,\theta_2)h(t,w_2) - \Psi[h(t,w_2)]\Bigg)\\
&+ t\Bigg(f_1 w(e_1,\theta_1)h(t,w_1) + (1-f_1)w(e_2,\theta_2)h(t,w_2) - f_1(c_1^1+e_1) - f_2(c_2^1+e_2)\Bigg)\\
&+ \eta\left(c_2^1 + (1-t)w_2 h(t,w_2) - \Psi(h(t,w_2)) - c_1^1 + (1-t)y^c - \Psi(h^c)\right)\\
&+ \zeta\left(\mathcal{W}^2_{PC}(e_1,e_2,t) - \mathcal{W}^2_{Dev}(e_1,e_2) + \kappa,\right)
\end{aligned}
$$

where

$$
\begin{aligned}
&\tilde{f}_1(1-t)w_1h_1 - \Psi(h_1) + \tilde{f}_2(1-t)w_2h_2 -\\
&\left(\tilde{f}_1(1-t^d)w_1h_1(w_1,t^d) - \Psi(h_1(w_1,t^d)) + \tilde{f}_2(1-t)w_2h_2(w_2,t^d)\right)\\
&+ t\left(w_1h_1f_1 + w_2h_2f_2\right) - t^d\left(w_1h_1(w_1,t^d)f_1 + w_2h_2(w_2,t^d)f_2\right).
\end{aligned}
$$

For c_1 and c_2 we get the same FOC as in the full-commitment case. For t we get:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial t} =& (1+\zeta)\Bigg(-\tilde{f}(\theta_1)y_1 - (1-\tilde{f}(\theta_1))y_2 + f(\theta_1)y_1 + (1-f(\theta_1))y_2 + tf(\theta_1)w_1\frac{\partial h_1}{\partial t}\\
&+ t(1-f(\theta_1))w_2\frac{\partial h_2}{\partial t}\Bigg) - \eta(w_2h_2 - w_2^c h_2^c) = 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial e_1} =& \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t^F)h_2^c\frac{\partial w_2^c}{\partial e_1}\right]\\
&+ \zeta\frac{\partial w_1}{\partial e_1}\Bigg[\tilde{f}_1\left((1-t^{PC})h_1(e_1,t^{PC}) - (1-t^{Dev})h_1(e_1,t^{Dev})\right)\\
&+ f_1\left(t^{PC}h_i(e_1,t^{PC}) - t^{Dev}h_1(e_1,t^{Dev})\right)(1+\varepsilon_{h,w})\Bigg] = 0
\end{aligned}
$$

$$\frac{\partial \mathcal{L}}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2$$
$$+ \zeta\frac{\partial w_2}{\partial e_2}\Bigg[\tilde{f}_2\left((1-t^{PC})h_2(e_2,t^{PC}) - (1-t^{Dev})h_2(e_2,t^{Dev})\right)$$
$$+ f_2\left(t^{PC}h_2(e_2,t^{PC}) - t^{Dev}h_1(e_2,t^{Dev})\right)(1+\varepsilon_{h,w})\Bigg] = 0.$$

C.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; 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

C.2 Proof of Proposition 6

The FOC for e_1 and e_2 are equivalent to those in Appendix A.2 apart from the additional terms multiplied by ζ. The steps are, however, the same as in A.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:

$$\zeta\frac{\partial w_i}{\partial e_i}\Bigg[\frac{\tilde{f}_i}{f_i}\left((1-t^{PC})h_i(e_i,t^{PC}) - (1-t^{Dev})h_i(e_i,t^{Dev})\right)$$
$$+ \left(t^{PC}h_i(e_i,t^{PC}) - t^{Dev}h_i(e_i,t^{Dev})\right)(1+\varepsilon_{h,w})\Bigg]. \tag{14}$$

By assumption we have $\frac{\tilde{f}(\theta_1)}{f(\theta_1)} > 1$ and $\frac{\tilde{f}(\theta_2)}{f(\theta_2)} < 1$. 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 (14) can be rearranged as:

$$\zeta\frac{\partial w_i}{\partial e_i}\left(\left(h_i^{pc} - h_i^{dev}\right) - \left[t^{dev}h_i^{dev} - t^{pc}h_i^{pc}\right]\left[\frac{1+\varepsilon_{h,w}}{RF_i} - 1\right]\right). \tag{15}$$

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

$$\frac{h_i^{pc} - h_i^{dev}}{t^{dev}h_i^{dev} - t^{pc}h_i^{pc}} - \left[\frac{1+\varepsilon_{h,w}}{RF_i} - 1\right] \tag{16}$$

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 $h_i = (w_i(1-t))^\varepsilon$, (16) is $> (<) 0$ if:

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

which is >(<)0 whenever

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

We now have to show that $H(t) > H(t^d)$ for the low type and $H(t) < H(t^d)$ 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

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

which yields

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

As $RF_1 > 1$, we directly see that this condition is always fullfilled for the θ_1-type. Importantly, it is fulfilled for any $t > 0$ and therefore we know that $H(t^d) < H(t)$ for the low type.

How about our result for the θ_2-type? We need

$$RF_2 < \frac{(1-t)(1+\varepsilon) - t(1+\varepsilon)\varepsilon}{(1-t)(\varepsilon+1)} = 1 - \frac{t}{1-t}\varepsilon$$

and

$$RF_2 < \frac{(1-t)(1+\varepsilon) - t(1+\varepsilon)\varepsilon}{(1-t)(\varepsilon+1)} = 1 - \frac{t^d}{1-t^d}\varepsilon.$$

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

$$\frac{\tilde{f}_2}{f_2} < 1 - (f_2 - \tilde{f}_2) \underbrace{\frac{y_2 - y_1}{\bar{y}}}_{\equiv A}.$$

This is equivalent to

$$\tilde{f}_2(1 - f_2 A) < f_2(1 - f_2 A).$$

Thus, whenever $1 - f_2 A > 0$, we have our result. Term A can be written as

$$A = \frac{w_2^{1+\varepsilon}(1-t)^\varepsilon - w_1^{1+\varepsilon}(1-t)^\varepsilon}{f_1 w_1^{1+\varepsilon}(1-t)^\varepsilon + f_2 w_2^{1+\varepsilon}(1-t)^\varepsilon} = \frac{w_2^{1+\varepsilon} - w_1^{1+\varepsilon}}{f_1 w_1^{1+\varepsilon} + f_2 w_2^{1+\varepsilon}}.$$

$f_2 A < 1$ therefore implies

$$f_2 w_2^{1+\varepsilon} - f_2 w_1^{1+\varepsilon} < f_1 w_1^{1+\varepsilon} + f_2 w_2^{1+\varepsilon}$$

and hence

$$-f_2 w_1^{1+\varepsilon} < f_1 w_1^{1+\varepsilon}$$

which is always fulfilled.

C.3 List of Countries in the Empirical Part

Argentina, Austria, Bangladesh, Belgium, Benin, Bhutan, Brazil, Bulgaria, Burundi, Cameroon, Cape Verde, Colombia, Costa Rica, Cuba, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ghana, Guatemala, Hungary, Iceland, Ireland, Israel, Italy, Jamaica, Lesotho, Lithuania, Madagascar, Mali, Malta, Mauritania, Mexico, Nepal, New Zealand, Norway, Peru, Philippines, Poland, Portugal, Serbia, Sierra Leone, Singapore, Spain, Swaziland, Sweden, Switzerland, Thailand, Togo, Uganda.

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.

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.

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.

KAUFMANN, D., A. KRAAY, AND M. MASTRUZZI (2010): "The worldwide governance indicators: methodology and analytical issues," *World Bank policy research working paper Nr. 5430*.

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.

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.

SHESHINSKI, E. (1972): "The optimal linear income-tax," *The Review of Economic Studies*, 39(3), 297–302.

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.