Submission Number: PET11-11-00054

Education, Insurance and Optimal Dynamic Taxation

Sebastian Findeisen
Zurich, Yale

Abstract

We study optimal tax and educational policies in dynamic private information economies, in which ex-ante heterogeneous individuals make an educational investment early in their life. Wages are stochastic and drawn from a continuous distribution, which is conditional on, both, education and initial type. Education shifts the distribution of wages as suggested by recent empirical evidence. Moreover, also in line with evidence, marginal returns to education differ across agents. We characterize constrained-efficient allocations and propose two different integrated tax and education systems with real world characteristics that can implement the allocation. These policy instruments are also able to overcome capital market imperfections in educational investment markets. Optimal labor and savings taxes are history dependent and depend on past education investments.

Contact: sebastian.findeisen@uzh.ch and dominik.sachs@uni-konstanz.de. Findeisen acknowledges the hospitality of Yale University. We thankMichelle Rendall, IvánWerning, Chris Winter, Fabrizio Zilibotti, Josef Zweimüller and seminar participants in Gerzensee, Konstanz and of various seminars in Zurich for helpful comments.
Submitted: February 22, 2011.

Education, Insurance and Optimal Dynamic Taxation

Sebastian Findeisen
University of Zurich

Dominik Sachs
University of Konstanz*

February 15, 2011

Abstract

We study optimal tax and educational policies in dynamic private information economies, in which *ex-ante* heterogeneous individuals make an educational investment early in their life. Wages are stochastic and drawn from a continuous distribution, which is conditional on, both, education and initial type. Education shifts the distribution of wages as suggested by recent empirical evidence. Moreover, also in line with evidence, marginal returns to education differ across agents. We characterize constrained-efficient allocations and propose two different integrated tax and education systems with real world characteristics that can implement the allocation. These policy instruments are also able to overcome capital market imperfections in educational investment markets. Optimal labor and savings taxes are history dependent and depend on past education investments.

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

Keywords: Optimal dynamic taxation, Education, Implementation

*Contact: sebastian.findeisen@uzh.ch and dominik.sachs@uni-konstanz.de. Findeisen acknowledges the hospitality of Yale University. We thank Michelle Rendall, Iván Werning, Chris Winter, Fabrizio Zilibotti, Josef Zweimüller and seminar participants in Gerzensee, Konstanz and of various seminars in Zurich for helpful comments.

1 Introduction

A constant rise in wage and earnings inequality has taken place in many advanced economies in the last 30 years (Krueger, Perri, Pistaferri, and Violante 2010). The role for effective policies to, both, insure people against negative transitory shocks and redistribute to make up for permanent differences has increased simultaneously. Education is a primary determinant of earnings. A vast empirical literature has documented a wage premium for better educated workers, which has increased substantially around 1980 in the US (Katz and Autor 1999). Income inequality is, hence, linked to educational inequality. But not all of the observed rise in the college premium is necessarily caused by an increased demand for skills acquired in college – Taber (2001) finds that at least parts of the observed rise in the educational skill premium should be attributed to unobserved skill heterogeneity realized before college decisions are made. At the same time, individuals face very different returns to education.[1] To the extent that people are aware about their idiosyncratic expected returns, this implies that people have different incentives to invest into education. At the same time education decisions are undertaken under substantial uncertainty. Labor taxation and capital taxation may hence insure individuals, even if they are *ex-ante* identical. To the extent that private markets fail to offer insurance for wage risk, the government can be the best provider of offering insurance via public policies.

In this paper, we conduct the normative exercise of investigating how governments should optimally set education and taxation policies. To this end we construct a dynamic model, in which *ex-ante* heterogeneous individuals are born with different *innate abilities* and decide on their level of education early in their life. Later in their life individuals work and their reward to labor supply is a function of their realized *labor skill* level. There is risk and individuals face a distribution of skill levels, which depends on education acquired and *innate abilities*. We allow the distribution function to include different marginal returns to education across agents. Also, holding education constant, *innate abilities* affect the conditional distribution of skills directly. These assumptions are made to capture key stylized facts found by the empirical literature on education and subsequent labor market outcomes.

Both *innate abilities* and *labor skills* are private information. We characterize constrained Pareto optimal allocations in this economy. In any constrained Pareto optimum, we show that labor wedges are history dependent; individuals with the same labor skill face, in general, different distortions, unless they were also identical *ex-ante*. The extent to which educational decisions are distorted depends mainly on two forces. First,

[1] See, for example, Lemieux (2006) or Carniero and Heckman (2003).

implicit educational subsidies are used to offset any distortionary impact of the labor tax on educational decisions. This is equivalent to a first-best rule, if skills would be observable. Second, the planner distorts education to relax binding incentive constraints. If the desired redistribution of income is downward (rich to poor), we show that this imposes an implicit tax and downward distortion of education. The key is that this improves the equity-efficiency trade-off for the planner, in the same spirit as positive marginal labor tax rates relax incentive constraints in the standard model. Finally, a positive implicit tax on savings turns out to be optimal. This is a familiar and well-understood feature of dynamic private information with endogenous labor supply that also holds up in our setting with education as an additional choice variable for agents.

In dynamic settings like ours the straightforward mapping from wedges to taxes known from the static model, in general, breaks down. We propose a decentralized implementation of Pareto optima with a labor tax code, that directly conditions on educational decisions. The existence of separate schedules gives individuals the correct incentives to self-select into education. At the same time the different tax schedules insure agents against labor market risk. Since the risk properties of the different education levels differ, this naturally leads to different optimal tax schedules. During the education period, individuals receive grants, which are tied to their education level chosen. In the implementation agents reveal themselves through their education decisions early in their life and later through their labor supply. Various labor tax codes in developed countries like the US already feature a form of dependence on education, with the possibility of deductions for educational expenditures. However, most of these policies allow only for deductibles on expenditure in the same time period. The prescription of our theory goes further by showing that labor tax codes should condition on education for the whole working life.

Related Literature. Several previous papers have studied problems of optimal labor taxation with education decisions. A strand of papers has worked under the assumption of *ex-ante* homogeneity. Grochulski and Piskorski (2010) focus on the implications of human capital investment for capital taxation. Anderberg (2009) concentrates mainly on the question whether and how education should be distorted relative to a first-best allocation, which, under his modeling assumptions, depends on how education influences the variance of log wages. Our model explicitly stresses the importance of heterogeneity, already at the point when education decisions are made.

In a static setting, Bovenberg and Jacobs (2005) analyze how endogenous education alters the result of the Mirrleesian tax problem. Relatedly, Bohacek and Kapicka (2008) study a model in a dynamic environment, but operate under certainty. The explicit dependence of the labor tax code on education arises in our model to give proper incentives for heterogeneous agents to obtain education as well as to insure against wage risk, which depends on education. These aspects are not present in the previous contributions.

In contrast to most predecessors to our paper, we work with a continuum of agents.[2] This connects us to recent work, which aims to generalize well-understood labor tax formulas (Diamond (1998) and Saez (2001)) from the static Mirrlees model with exogenous skills to a dynamic, stochastic setting (see Golosov, Troshkin, and Tsyvinski (2009) or Findeisen (2011)). To make the dynamic incentive problem tractable, we employ a first-order approach, for which we provide necessary and sufficient conditions. Importantly, we rely on the fact that private information evolves sequentially in our economy, which avoids solving a multidimensional screening problem. The paper is, hence, related to work studying dynamic mechanisms; in particular, Pavan, Segal, and Toikka (2009), who study the validity and robustness of the Mirrleesian first-order approach to mechanisms for a wide and general class of models, or Courty and Li (2000), who study optimal dynamic pricing of a monopolist.

This paper is organized as follows. Section 2.1 contains the basics of the model as well as its empirical justification. After that the Laissez-Faire allocation is described in Section 2.2 and properties of the First-Best are discussed in Section 2.3. In Section 3, we discuss dynamic incentive compatibility and describe the properties of constrained efficient allocations. A decentralized implementation of such constrained efficient allocations is provided in Section 4 before Section 5 concludes.

2 Baseline Model

2.1 Heterogeneity, Education and Labor Market Outcomes

We begin with a brief discussion of recent evidence and stylized facts from the empirical literature on education and labor market outcomes and state how these guide our modeling assumptions. These findings underscore the importance of the effect of initial heterogeneity and risk on the returns to education. A series of papers have used factor structure models to estimate returns to education (see Cunha and Heckman (2007) for a survey). Importantly, these methods can identify whole distributions of returns (instead of first and second moments only), do counterfactual analysis, and distinguish between ex-ante and ex-post returns. First, Cunha and Heckman (2006) document considerable residual uncertainty over future returns at the time of the decision to go to college or enter the labor market directly after high school, even after controlling for heterogeneity. Second, for both groups, actual college and high-school graduates, the density of lifetime

[2] The article by Bovenberg and Jacobs (2005) is the notable exception.

earnings when going to college, lies to the right of the density when entering the labor after high school. Note that for high-school graduates the college density is estimated counterfactually and vice-versa for college graduates. Third, returns to education differ widely across individuals. Carniero and Heckman (2003) document that the return can differ by as much 19% points across individuals for one year of college. In this vein, the literature has documented a complementarity effect- both cognitive and non-cognitive ability, either acquired early during childhood or innate, increase the return of education. For example, Altonji and Dunn (1996) find that the return is higher for children whose parents are highly educated; these parents are likely also more able to transmit or instill their offspring with higher initial human capital. Finally, there is evidence of a direct effect of these early abilities on earnings. Taber (2001) presents findings suggesting 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.

In view of the above discussion, we model the relationship between education and subsequent labor market outcomes as a stochastic process. Let a be an individual's *labor market ability*. We assume that the parameter a is drawn from a continuous conditional distribution $G(a|z, \theta)$, which depends on *innate* ability θ and education z, and has bounded support $[\underline{a}, \overline{a}]$. Innate ability θ can be interpreted as a one dimensional aggregate of (non-)cognitive skills and family background and is distributed in the interval$[\underline{\theta}, \overline{\theta}]$ according to $F(\theta)$. The timing is such that agents first learn their initial type θ and then decide on education z. We place the following assumptions on $G(a|z, \theta)$:

Assumption 1: $G(a|z', \theta) \succeq_{FOSD} G(a|z, \theta) \Leftrightarrow G(a|z', \theta) \leq G(a|z, \theta)$, for all $z < z'$.

Assumption 2: $G(a|z, \theta') \succeq_{FOSD} G(a|z, \theta) \Leftrightarrow G(a|z, \theta') \leq G(a|z, \theta)$, for all $\theta < \theta'$.

Assumption 3: $\frac{\partial^2 G(a|z_1, \theta)}{\partial\theta\partial z} \leq 0$.

First, an increase in education induces a first-order stochastic dominance shift in the conditional distribution of skills. This is consistent with the evidence presented in Cunha and Heckman (2006), if the actual and counterfactual distributions of high school and college graduates are interpreted as the information set agents decide on. Second, in line with the evidence of the effect of innate skills on labor market outcomes holding education constant, we assume a first-order stochastic dominance shift also for higher initial skill level θ. Third, respecting the compelling evidence of complementarity between early ability (θ) and education, marginal returns differ and are higher for higher initial type.

2.2 Laissez Faire Equilibrium

To lay out the basic properties of the model, we start with the characterization of the government intervention free laissez-faire equilibrium. In the second period, after agents have learned their labor market skill a, they choose labor supply, taking savings or private debt as given. This gives rise to the indirect utility function:

$$v_2(a, s(\theta)) = \max_{y,c_2} u\left(c_2\right) - \Psi\left(\frac{y}{a}\right) \text{ s.t. } c_2 = y + Rs(\theta). \tag{1}$$

Individuals' utility functions are well-behaved- $u(.)$ is assumed to be increasing, at least twice continuously differentiable and concave, and $\Psi(.)$ is assumed to be increasing, at least twice continuously differentiable and convex.[3] The parameter a is an individual's labor market skill, meaning a higher a needs to provide less labor effort to earn any income y.

In the first period, agents decide how much to invest into education, and make a consumption/saving decision. Agents have access to a risk-free one period bond market; we impose no shortsale or enforcement constraints and an exogenous gross return R. This defines the indirect utility function:

$$V(\theta) = \max_{s,z,c_1} u(c_1) + \beta \int_{\underline{a}}^{\overline{a}} v_2(a, s) g(a|z, \theta) da \text{ s.t. } c_1 + z = -s. \tag{2}$$

As already anticipated in the last section, we model the conditional distribution of skills $g(a|z, \theta)$ as being determined by an agent's education level z and her innate ability θ. Moreover, we focus on educational investment as a direct monetary cost. This is consistent with the idea that tuition fees and other monetary expenses are the most important factors on the cost side driving educational decisions. It is also in line with a foregone earnings interpretation, where more education delays labor market entry. z can be a sum of both factors.[4] Additionally, it is possible to model education as a direct effort cost without affecting any of the main results.

We now present the main properties of the equilibrium without government policies:

Proposition 2.1. *Independent of the size and distribution of initial wealth, the Laissez-Faire allocation has the following properties:*
(i) The Euler Equation holds:
$u'(c_1(\theta)) = \beta R \int_{\underline{a}}^{\overline{a}} u'(c_2(\theta, a)) dG(a|z(\theta), \theta)$

[3]All of our results can be generalized to the case of non-separable preferences.

[4]Different people have likely different opportunity costs in terms of foregone wages in the real world. At the same time, the number of years spent in education are small compared to the number of years being active in the labor market. So over a lifetime the differences in forgone earnings become very small, which is why we assume constant cost across types. This keeps the model tractable.

(ii) Labor supply is undistorted: $\Psi'\left(\frac{y(\theta,a)}{a}\right)\frac{1}{a} = u'(c_2(\theta))$.
(iii) The marginal cost of education is equalized to marginal benefits: $u'(c_1(\theta)) = \beta R \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta),\theta)}{\partial z} da$
(iv) Educational investment is increasing in innate ability and savings are decreasing, i.e. $z'(\theta) > 0$ *and* $s'(\theta) < 0$.

Proof. See Appendix. □

Parts (i)-(iii) follow directly from first-order conditions. They are unsurprising properties, stating that private marginal rates of substitution are equated to technical marginal rates of transformation on the labor, capital, and education market.

Part (iv) states that without government policies, education and savings are monotone in innate ability θ. The proof provides instructive intuition for this result. It is sufficient to show that the objective defined by (2) is supermodular in all choice variables and type θ (see Milgrom and Shannon (1994)). Plugging in the budget constraint gives the problem reduced to two choices s and z: $\max_{s,z} U(s, z; \theta, a, \beta) = u(-s - z) + \beta \int_{\underline{a}}^{\overline{a}} v_2(a, s) g(a|z, \theta) da$. This objective is supermodular in credit taken $-s$ and education z, if and only if:

$$\frac{\partial^2 U(s, z; \theta, .)}{\partial s \partial \theta} < 0 \tag{3}$$

$$\frac{\partial^2 U(s, z; \theta, .)}{\partial s \partial z} < 0 \tag{4}$$

$$\frac{\partial^2 U(s, z; \theta, .)}{\partial z \partial \theta} > 0. \tag{5}$$

In Appendix A.1 we show that all inequalities hold. Equations (3) and (4) imply that the return to savings is lower for higher θ types and with higher education, since expected labor skills are also higher. Equation (5) holds, since innate abilities and education are complementary to each other. Taken together the direct effects of being of higher type on credit and education are being reinforced by the relationship between the endogenous variables.

So far, we have assumed no limits on the ability of agents to borrow against future labor income. Imposing an ad-hoc constraint of the form $s \geq \phi$, where ϕ is some negative number, leaves most of the results from Proposition 2.1 unaffected.[5] Notably, constrained agents will not be able to smooth consumption intertemporally as much as desired. Still education levels will be increasing in type:

Corollary 2.2. *If agents face borrowing constraints* $s \geq \phi$*, education is still monotone in type* θ*, i.e.* $z'(\theta) > 0$.

[5] Surveying the literature, Carniero and Heckman (2003) conclude that short-term borrowing constraints seem to have only a very small effect on educational decisions.

Intuitively, above some threshold type, savings will be constant and equal to ϕ. Higher types still face higher marginal returns to education and property (iv) will hold. Moreover, the monotonicity of education of also carries over, if agents have heterogeneous initial wealth $w(\theta)$, as long as $w(\theta)$ is non-decreasing in θ.

Corollary 2.3. *If initial wealth is non-decreasing in θ, education is monotone in type θ, i.e. $z'(\theta) > 0$.*

Proof. See Appendix. □

Empirically, wealthier families tend to instill their children with more early human capital captured by θ and also directly with more financial resources (Carniero and Heckman (2003)), which gives the assumption plausibility. The empirical literature has also documented sorting into education, based on heterogeneous expected returns (Cunha and Heckman (2007)). The monotonicity of education in the laissez-faire equilibrium is consistent with that fact.

For later purposes when we analyze optimal allocations and the respective tax systems that can implement such allocations, it is useful to define three wedges. They are equal to implicit marginal tax rates on capital, labor income, and education, respectively:

Savings wedge:

$$\tau_s(\theta) = 1 - \frac{u'(c_1(\theta))}{\beta R \int_{\underline{a}}^{\overline{a}} u'(c_2(\theta, a)) g(a|z_i, \theta) da}$$

Importantly, like all wedges the intertemporal wedge is defined for any given allocation. It is the proportional adjustment needed in the rate of return to make the Euler equation hold for an agent θ, given the particular allocation. It follows that in any allocation, there are as many wedges as agents- one for each innate skill level. $\tau_s(\theta) > (<) 0$ implies a downward (upward) distortion of savings. The same is true for the following wedges.

Labor wedge: The labor wedge is nonzero, if an individual would like to work more or less at the intervention-free market price (which is her productivity level a). Formally the labor wedge reads as:

$$\tau_y(\theta, a) = 1 - \frac{\Psi'\left(\frac{y(\theta,a)}{a}\right)\frac{1}{a}}{u'(c_1(\theta, a))}$$

It again has to be evaluated for a given allocation and there exists exactly one labor wedge for every type vector (θ, a).

Educational wedge: The educational wedge is nonzero, if the individual would want to obtain more or less education if it could do that at the market price z. Formally it reads as

$$\tau_z(\theta) = 1 - \frac{\beta \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta),\theta)}{\partial z(\theta)} da}{u'(c_1(\theta))}. \tag{6}$$

In the implementation we later propose for constrained Pareto efficient allocations, we will show, which of these *implicit taxes* will equal *explicit marginal tax rates*.

2.3 Unconstrained Efficient Allocations

In an unconstrained Pareto optimum no information is private. The planner is able to observe an individual's innate ability θ as well as realized productivity a. We characterize optimal decision rules for the assignment of bundles of consumption levels, labor supply and education attainment to everybody. The only restriction the government has to meet is the aggregate present-value resource constraint, making sure the allocation is feasible.

We let the planner assign Pareto-weights $\tilde{f}(\theta)$ to individuals, depending (solely) on their initial skill level. Sometimes, we make use of the ratio $\frac{\tilde{f}(\theta)}{f(\theta)}$, which we refer to the as net Pareto-weight. Other times we denote cumulated Pareto-weights by the function $\tilde{F}(\theta)$. Any distribution, which we normalize to satisfy $\int_{\underline{\theta}}^{\overline{\theta}} \tilde{f}(\theta) d\theta = 1$, of these weights corresponds to one point on the Pareto frontier; i.e. as long as we do not impose any restrictions on $\{\tilde{f}(\theta)\}_{\theta \in [\underline{\theta},\overline{\theta}]}$ the first-order conditions of the social planner characterize the whole pareto frontier.

We can establish the following properties of a first-best equilibrium:

Proposition 2.4. *Any unconstrained Pareto optimal allocation satisfies:*
(i) Initial consumption levels are determined by the net Pareto-weights: $u'(c_1(\theta)) = \frac{f(\theta)}{\tilde{f}(\theta)} \lambda_R$.
(ii) There is full insurance and perfect intertemporal smoothing: $u'(c_1(\theta)) = R\beta u'(c_2(\theta, a)) = R\beta u'(c_2(\theta))$
(iii) Labor supply is undistorted: $\Psi' \left(\frac{y(\theta,a)}{a} \right) \frac{1}{a} = u'(c_2(\theta))$.

Proof. See Appendix A.2. □

The last two conditions are well-understood attributes of a Pareto optimum. Note that part (ii) states that future consumption is only a function of initial type, which allows us to write: $c_2(\theta, a) = c_2(\theta)$. The first part makes clear that net Pareto-weight are inversely related to marginal utilities in all periods. The above conditions also imply zero savings and labor wedges. In contrast, as can easily be shown from the first-order conditions spelled out in in the appendix, the educational first-best wedge is positive and given by:

$$\begin{aligned}\tau_z^{FB} =& 1 - \frac{\beta \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta),\theta)}{\partial z(\theta)} da}{u'(c_1(\theta))} \\ =& \frac{1}{R} \int_{\underline{a}}^{\overline{a}} (y(\theta, a) - c_2(\theta)) \frac{\partial g(a|z(\theta), \theta)}{\partial z(\theta)} da > 0. \end{aligned} \tag{7}$$

Intuitively, the marginal subsidy to education is just equal to the marginal expected increase in the lump sum tax $(y(\theta, a) - c_2(\theta))$ paid by the individual. This can be viewed as a fiscal externality of education, not taken into account by an agent in a decentralized allocation. Additionally, the first line can be integrated by parts:

$$\tau_z^{FB} = 1 + \frac{\beta \int_{\underline{a}}^{\overline{a}} \frac{\partial v_2(\theta,a)}{\partial a} \frac{\partial G(a|z(\theta),\theta)}{\partial z(\theta)} da}{u'(c_1(\theta))} > 1.$$

The second term is always positive, since indirect utility in a first best is *decreasing in ability* conditional on θ, and the FOSD shift of education.[6] The marginal subsidy to education is, hence, bigger than 100% across all initial types θ. Not only is there full redistribution of resources in the second period, conditional on innate ability, but agents actually lose from a good skill draw, further weakening incentives to invest in education.

It seems worthwhile to refer to the work of da Costa and Maestri (2007) and Anderberg (2009). In their modeling framework probabilities are exogenous and wages in the respective states are endogenous (i.e. increasing in education). Whereas in such a framework Proposition 2.4 holds, there is no negative educational wedge. In that framework, even if there is full redistribution, education pays at the margin since more education increases wages but does not alter the lump-sum taxes individuals have to pay in period 2.

3 Constrained Efficient Allocations

In this section we consider constrained Pareto efficient allocations, where 'constrained' refers to the government being unable to observe agents' type θ in period 1 and a in period 2. We show that the problem is tractable using a first-order approach. In addition we provide necessary and sufficient condition for this approach to be valid.

[6] $\frac{\partial v_2(\theta,a)}{\partial a} < 0$ follows from part (iii) of Proposition 2.4; there is full consumption insurance but the more able work more.

3.1 Incentive Compatibility

We cast the problem as a sequential, dynamic mechanism – agents report an initial type θ in the first period, and, after uncertainty has materialized, report their productivity a in the second period. The planner assigns initial consumption levels $c_1(\theta)$ and education levels $z(\theta)$ to individuals with innate ability θ. Moreover, with each report there comes a sequence of utility promises for the next period $\{v_2(\theta, a)\}_{a \in [\underline{a}, \overline{a}]}$. In the second period the screening takes place over consumption levels $c_2(\theta, a)$, and labor supply $y(\theta, a)$ and individuals self-select according to their history (θ, a). All these quantities define an allocation in the economy. Dynamic incentive compatibility is ensured backwards, so we start analyzing the problem from the second period.[7]

3.1.1 Second Period Incentive Compatibility

By the revelation principle, we can restrict attention to direct mechanisms. Suppose in the first period agents have made truthful reports $r(\theta) = \theta$, although this is not necessary and just simplifies the exposition. Conditions for this to be true are given in the next section. Conditional on this report, the second period incentive constraint must be met for any history of types (θ, a) and reporting strategy $r(a)$:

$$u\left(c_1\left(\theta, a\right)\right) - \Psi\left[\frac{y(\theta, a)}{a}\right] \geq u\left(c_1\left(\theta, r(a)\right)\right) - \Psi\left[\frac{y(\theta, r(a))}{a}\right] \qquad \forall a, r(a).$$

Define the associated indirect utility function of the agents as:

$$v_2(\theta, a) = \max_{r(a)} u\left(c_2\left(\theta, r(a)\right)\right) - \Psi\left[\frac{y(\theta, r(a))}{a}\right]. \tag{8}$$

Like in a standard Mirrleesian problem preferences satisfy single-crossing for given first-period reports. For global incentive compatibility it is, hence, sufficient that all local envelope conditions hold:

$$\frac{\partial v_2(\theta, a)}{\partial a} = \Psi'\left(\frac{y(\theta, a)}{a}\right)\frac{y(\theta, a)}{a^2}, \tag{9}$$

and the usual monotonicity condition, stating that $y(\theta, a)$ is non-decreasing in ability levels a, is satisfied:[8]

$$\frac{\partial y(\theta, a)}{\partial a} \geq 0 \tag{10}$$

[7] Our approach can be readily extended to the case of a discrete choice for education; i.e. the planner deciding which agents to send to college and which not. Incentive compatibility has then to be characterized using modified envelope theorems in the spirit of Milgrom and Segal (2002). The full proofs to deal with the arising difficulties because of the non-continuities are available upon request.

[8] As usual, we assume that these constraints do not bind, when we solve the Pareto program. We abstract, hence, from bunching issues.

3.1.2 First Period Incentive Compatibility

Importantly, in the first period an agent takes into account the effect of her report about θ on future utility. First period incentive compatibility is ensured, if and only if the double continuum of weak inequalities holds:

$$\begin{aligned} U(\theta,\theta) &= u\left(c_1\left(\theta\right)\right) + \beta \int_{\underline{a}}^{\overline{a}} v_2(\theta,a) dG(a|z(\theta),\theta) \\ \geq & u\left(c_1\left(r(\theta)\right)\right) + \beta \int_{\underline{a}}^{\overline{a}} v_2(r(\theta),a) dG(a|z(r(\theta)),\theta) = U(\theta, r(\theta)), \qquad \forall \theta, r(\theta), \end{aligned} \tag{11}$$

where $U(\theta, r(\theta))$ is the expected utility of an individual of type θ reporting $r(\theta)$. The associated value function is:

$$V(\theta) = \max_{r(\theta)} u\left(c_1\left(r(\theta)\right)\right) + \beta \int_{\underline{a}}^{\overline{a}} v_2(r(\theta),a) dG(a|z(r(\theta)),\theta). \tag{12}$$

We proceed by replacing the set of inequality constraints defined by (11) by local envelope conditions analogous to the ones for the second period (9):

$$\frac{\partial V(\theta)}{\partial \theta} = \int_{\underline{a}}^{\overline{a}} v_2(\theta,a) \frac{\partial g(a|z(\theta),\theta)}{\partial \theta} da \tag{13}$$

Innate ability affects the indirect utility function *directly* through the serial correlation in types only. This *localization* of incentive constraints to make the problem tractable is clearly only valid, if they imply a maximum from the point of view of the agents for a truthful report and also imply *global* incentive compatibility. We now proceed by first characterizing necessary and sufficient conditions on primitives and endogenous variables for local and then global incentive compatibility. The first-order condition for an optimal report, evaluated at the revelation strategy, must obey:

$$\frac{\partial U(\theta,\theta)}{\partial r(\theta)} = 0 \tag{14}$$

For a local maximum it is necessary that:

$$\frac{\partial^2 U(\theta,\theta)}{\partial r(\theta)^2} \leq 0 \tag{15}$$

Differentiating equation (14) yields:

$$\frac{\partial^2 U(\theta,\theta)}{\partial r(\theta)^2} + \frac{\partial^2 U(\theta,\theta)}{\partial r(\theta) \partial \theta} = 0. \tag{16}$$

In any incentive compatible allocation it must, hence, hold that:

$$\frac{\partial^2 U(\theta,\theta)}{\partial r(\theta)\partial\theta} = \int_{\underline{a}}^{\overline{a}} \frac{\partial v_2(\theta,a)}{\partial\theta}\frac{\partial g(a|z(\theta),\theta)}{\partial\theta}da + \frac{\partial z(\theta)}{\partial\theta}\int_{\underline{a}}^{\overline{a}} v_2(\theta,a)\frac{\partial g(a|z(\theta),\theta)}{\partial z\partial\theta}da \geq 0 \tag{17}$$

We summarize these findings in the following lemma[9]:

Lemma 3.1. *An allocation is locally incentive compatible, if and only if:*
(i) $\frac{\partial V(\theta)}{\partial\theta} = \int_{\underline{a}}^{\overline{a}} v_2(\theta,a)\frac{\partial g(a|z(\theta),\theta)}{\partial\theta}da$
(ii) $\int_{\underline{a}}^{\overline{a}} \frac{\partial^2 v_2(\theta,a)}{\partial\theta\partial a}\frac{\partial G(a|z(\theta),\theta)}{\partial\theta}da + \frac{\partial z(\theta)}{\partial\theta}\int_{\underline{a}}^{\overline{a}} \frac{\partial v_2(\theta,a)}{\partial a}\frac{\partial G(a|z(\theta),\theta)}{\partial z\partial\theta}da \leq 0,$

where (ii) follows after integrating (17) by parts. The lemma provides necessary conditions for any incentive compatible allocation. Its first part (i) can be conveniently included into any Lagrangian or optimal control problem we later want to solve. Its second part carries an interesting intuition. The first term in the first integral measures how the skill premium $\frac{\partial v_2(\theta,a)}{\partial a} > 0$ is different for the marginally higher θ type. The planner has to tailor these skill premia to the marginal changes in the distribution of a. If these shifts are large, i.e. $|\frac{\partial G(a|z(\theta),\theta)}{\partial\theta}|$ is large then the effect of $\frac{\partial^2 v_2(\theta,a)}{\partial\theta\partial a}$. States which become more likely with higher innate ability must be rewarded accordingly in the next period, to avoid that an agent claims an allocation not designed for her. The second term in (ii) is always negative, if education is increasing in innate ability. Consider the case where the correlation of types across time is very weak- to be incentive compatible an allocation has then to offer increasing education levels in innate ability. Since education and innate ability are complements, higher skilled agents need to be offered a higher z. In the polar case, if ability is very strongly correlated across time, an allocation can be incentive compatible, even if education might be decreasing over some interval. We now show that there are less complicated *sufficient conditions* than the ones given in (ii) of the last lemma. In particular these are simple monotonicity conditions like (10), which are straightforward to check after a candidate allocation for a relaxed problem has been computed:

Lemma 3.2. *If an allocation satisfies:*
(i) $\frac{\partial V(\theta)}{\partial\theta} = \int_{\underline{a}}^{\overline{a}} v_2(\theta,a)\frac{\partial g(a|z(\theta),\theta)}{\partial\theta}da$
(ii) $\frac{\partial y(\theta,a)}{\partial\theta} \geq 0$ *and* $\frac{\partial z(\theta)}{\partial\theta} \geq 0,$
then the allocation is globally incentive compatible.

Proof. See Appendix A.3. □

[9]We again abstract from the issue of bunching in z.

Note that Lemma 3.2 extends the previous lemma from local to *global* incentive compatibility. Condition (ii) states that second period labor income should be non-decreasing in *innate ability* θ. This might be surprising, since θ does not affect indirect utility after productivity a is realized. In a decentralized allocation equation this condition will characterize an intuitive property of a constrained-efficient equilibrium. Agents with a higher initial skill will typically save less since expected labor productivity is also be higher for these agents. For the same realization of productivity a labor supply and income will, hence, be higher for these individuals, because of wealth effects. However, note these two "monotonicity conditions" are not directly comparable to the monotonicity condition on gross incomes in the static Mirrlees case. The latter is necessary for incentive compatibility, whereas the conditions in Lemma 3.2 are sufficient for incentive compatibility and might be violated over some interval of the skill set. Therefore, it is not valid to add the two monotonicity constraints to the second-best problem as this would reduce the set of incentive compatible allocations. As is standard in screening problems, our strategy for solving the second-best problem is to work with a relaxed problem with only restriction (13) imposed and then check ex-post whether the monotonicity constraints hold. In the numerical explorations in Section ? we find that these condition are always satisfied.[10]

Our results on dynamic incentive compatibility are related to previous work in the optimal non-linear pricing literature by Courty and Li (2000). They study optimal pricing schemes of a monopolist facing consumers with stochastic tastes. In a recent contribution, Pavan, Segal, and Toikka (2009) investigate the robustness and validity of the Mirrleesian first-order approach in very general dynamic environments.

3.2 Characterization

The planner maximizes

$$\int_{\underline{\theta}}^{\overline{\theta}} u(c_1(\theta))d\tilde{F}(\theta) + \beta \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} v_2(\theta, a)dG(a|z(\theta), \theta)d\tilde{F}(\theta) \tag{18}$$

subject to (9), (13) and the resource constraint

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[c_1(\theta) - z(\theta) + \int_{\underline{a}}^{\overline{a}} (c_2(\theta, a) - y(\theta, a)dG(a|z(\theta), \theta) \right] dF(\theta) = R \tag{19}$$

where R are some exogenous initial resources. In Appendix A.4 the Lagrangian and the first-order conditions of the problem are stated. In the

[10] See Farhi and Werning (2010) for a similar approach.

following we will characterize the wedges of the second-best efficient allocations.

3.2.1 Savings Distortions

We now derive a useful necessary condition of a constrained efficient optimum, concerning the agents' savings decision, from the education period to working life. It will turn out that the presence of education, which endogenously affects the probability distribution of tomorrow's skills, does not change the prescription of a positive intertemporal wedge, stemming from the optimality of the Inverse Euler equation, familiar from other dynamic Mirrleesian models.[11] Integrating the first-order condition for a utility promise over a yields[12]:

$$\left(\tilde{f}(\theta) - \eta'(\theta)\right)\beta - \frac{1}{R}\lambda_R E_{a|\theta}\left[\frac{1}{u'(c_2(\theta,a))}\right] f(\theta) - \int_{\underline{a}}^{\overline{a}} \mu'(\theta,a)da$$
$$-\beta\int_{\underline{a}}^{\overline{a}} \frac{\partial g(a|z(\theta),\theta)}{\partial\theta}\eta(\theta)da = 0$$

by the transversality condition for $\mu(\theta,a) : \mu(\theta,\underline{a}) = \mu(\theta,\overline{a}) = 0.$; further last term is zero. Rearranging the first-order condition for optimal first period consumption gives:

$$\lambda_R = u'(c_1(\theta))\frac{\tilde{f}(\theta)+\eta(\theta)}{f(\theta)}$$

Taking together yields:

$$\frac{1}{u'(c_1(\theta))} = \frac{1}{\beta R}\int_{\underline{a}}^{\overline{a}} \frac{1}{u'(c_2(\theta,a))}g(a|z(\theta),\theta)da = \frac{1}{\beta R}E_{a|\theta}\left[\frac{1}{u'(c_2(\theta,a))}\right], \tag{20}$$

and by Jensen's inequality: $\beta E\left[u'(c_2(\theta,a))\right] > u'(c_1(\theta))$- the optimal allocation dictates a wedge between the intertemporal rate of substitution and transformation, savings are hence discouraged.

[11] Diamond and Mirrlees (1978) and Rogerson (1985) were the first to derive it. In an important paper reviving the interest in the result Golosov, Kocherlakota, and Tsyvinski (2003) generalized it to a large class of dynamic environments, most importantly allowing for arbitrary skill processes. Much like the Atkinson and Stiglitz (1976) prescription of uniform commodity taxes, the robustness of a positive intertemporal wedge relies on the (weak) separability of consumption and work effort.

[12] The whole Pareto program and all optimality conditions are spelled out in the appendix. An alternative way, working equally well, to derive the Inverse Euler equation also in our context is a pertubation around an optimal allocation as pioneered in Rogerson (1985) and Golosov, Kocherlakota, and Tsyvinski (2003).

3.2.2 Labor Distortions

Rearranging the first-order condition for labor supply yields:

$$\frac{\Psi'\left(\frac{y(\theta,a)}{a}\right)}{au'(c_2(\theta,a))} = 1 - \frac{\mu(\theta,a)\left[\Psi''\left(\frac{y(\theta,a)}{a}\right)\frac{y(\theta,a)}{a^3} + \frac{1}{a^2}\Psi'\left(\frac{y(\theta,a)}{a}\right)\right]}{\frac{1}{R}\lambda_R g(a|z(\theta),\theta)f(\theta)}, \tag{21}$$

implying that labor supply is distorted downwards everywhere in the interior of the skill set as long as $\mu(\theta,a) > 0$. The opposite $\mu(\theta,a) < 0$ is unlikely but can be the case if the Pareto weight $\tilde{f}(\theta)$ is particularly increasing in θ, which implies that the state wants to redistribute in favor of the high types. As will be shown in section 4.1, where the labor wedges are interpreted as marginal labor income tax rates, the distortion can be decomposed into two parts

1. An insurance term: Conditional on being of type θ, individuals want to be insured against wage risk. This can be done by redistributive taxation in period 2; this effect always works in favor of positive marginal rates, i.e. downwards distortion of labor supply.
2. A redistribution term: Conditional on the Pareto weights $\{\tilde{f}(\theta)\}_{\theta\in[\underline{\theta},\overline{\theta}]}$ that the planner assigns to the different types, the planner wants to redistribute between different types.

3.2.3 Education Distortions

In a constrained Pareto optimum education decisions are distorted in the following way:

$$\begin{aligned}\tau_z^{SB} =& 1 - \frac{\beta\int_{\underline{a}}^{\overline{a}} v_2(\theta,a)\frac{\partial g(a|z(\theta),\theta)}{\partial z(\theta)}da}{u'(c_1(\theta))} \\ =& \frac{1}{R}\int_{\underline{a}}^{\overline{a}}\left(y(\theta,a) - c_2(\theta,a)\right)\frac{\partial g(a|z(\theta),\theta)}{\partial z(\theta)}da - \frac{\beta\eta(\theta)}{\lambda_R f(\theta)}\int_{\underline{a}}^{\overline{a}} v_2(\theta,a)\frac{\partial^2 g(a|z(\theta),\theta)}{\partial z(\theta)\partial\theta}da.\end{aligned}$$

Let $T(\theta,a) = y(\theta,a) - c_2(\theta,a)$; we will later show that the allocation can be decentralized with $T(\theta,a)$ being the total labor tax bill of an agent with history (θ,a).

To make further progress, integrate by parts to obtain:

$$\tau_z^{SB} = -\frac{1}{R}\int_{\underline{a}}^{\overline{a}}\frac{\partial T(\theta,a)}{\partial a}\frac{\partial G(a|z(\theta),\theta)}{\partial z(\theta)}da + \frac{\beta\eta(\theta)}{\lambda_R f(\theta)}\int_{\underline{a}}^{\overline{a}}\frac{\partial v_2(\theta,a)}{\partial a}\frac{\partial^2 G(a|z(\theta),\theta)}{\partial z(\theta)\partial\theta}da \tag{22}$$

The first part of the wedge is the same as in the unconstrained Pareto problem, see equation (7). A marginal increase in education triggers higher expected productivity and higher expected gross income next period. This will change expected labor tax liabilities. Again the educational wedge ensures that this spillover of the labor wedge on the educational margin is corrected for. As long as desired redistribution is not too strongly going in the direction from low to high types, which in turn might imply negative marginal tax rates over some range, this term will be positive. In contrast, the second part of the wedge has the opposite sign of $\eta(\theta)$.[13] As long as initial consumption $c_1(\theta)$ is weakly increasing in the initial type, however, which is elusive to prove theoretically on this level of generality, but emerges from all numerical exercises, the expression is *negative* calling for an implicit *tax on education*. By downward distorting education, the planner relaxes binding incentive constraints and can redistribute more effectively in line her preferences. This is a consequence of the complementarity assumption, stating that agents endowed with higher innate skills gain more from education at the margin. The bundle of a lower type, hence, becomes less attractive from the perspective of an agent, if education is downward distorted. Such an intuition is familiar from the standard static Mirrlees model concerning positive marginal income tax rates on the interior of skill set. Relatedly, by a transversality condition we can show that this second part of the educational distortion is zero at the top and at the bottom $(\underline{\theta}, \overline{\theta})$ of the innate ability distribution.

4 Implementation(s)

4.1 Student Loans and History Dependent Labor Taxes

So far we only considered direct mechanisms. In this section we explore a decentralized implementation. The benevolent government, taking the role of the planner, offers a menu of student grants to the agents. These grants L are conditional on education, which is chosen by the agents and observable. In the second period, there is a tax schedule in place, which, importantly, does not only condition on earnings but also on educational investment. To fix ideas, the budget constraint of an agent in both periods are given by:

$$c_1(\theta) + z(\theta) \leq L(z(\theta))$$
$$c_2(\theta, a) \leq y(\theta, a) - T(z(\theta), y(\theta, a))$$

What is obvious from this formulation, is that we implicitly have set savings using the private bond market to zero. Given that *all* agents in a constrained Pareto optimum will be savings constrained (see Section 3.2.1) and are actually eager to save, there seems to be a contradiction. Still setting

[13] Note $\frac{\partial v_2(\theta,a)}{\partial a}$ is positive everywhere by second period incentive compatibility and $\frac{\partial^2 G(a|z(\theta),\theta)}{\partial z(\theta)\partial\theta}$ negative by complementarity of innate skills and education.

savings to zero comes without any loss of generality. In fact, we can always find infinitely many savings tax schedules, which punish savings so much, that no agents wishes to deviate from the optimal allocation.Werning (2010) shows in a recent paper that there are even more degrees of freedom in the implementation. By a Ricardian equivalence argument, we can adjust $L(z(\theta))$ and $T(z(\theta), a)$ with lump-sum transfers and deductibles to arrive with a non-linear savings tax schedule, which produces non-zero private savings for every agent and the same allocation with the same distortion of consumption across periods. The full argument is found in Werning (2010).

Next, we investigate in more detail the forces behind optimal labor tax rates during adulthood.[14]

Proposition 4.1. *At any constrained Pareto optimum marginal labor income tax rates satisfy:*

$$\frac{T'_y(z(\theta), y(\theta, a))}{1 - T'_y(z(\theta), y(\theta, a))} = \frac{u'(c_2(\theta, a))}{\varepsilon^* a g(a|z(\theta), \theta)} \left[\mathcal{A}(\theta, a) + \mathcal{B}(\theta, a)\right],$$

where

$$\mathcal{A}(\theta, a) = \beta R G(a|z(\theta), \theta) \left[\frac{1}{u'(c_1(\theta))} - \frac{1}{\beta R G(a|z(\theta), \theta)} \int_{\underline{a}}^{a} \frac{1}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta)\right]$$

$$\begin{aligned}\mathcal{B}(Y(\theta, a)) &= -\frac{1}{f(\theta)\lambda_R} R\beta \frac{\partial G(a|z(\theta), \theta)}{\partial \theta} \eta(\theta) \\ &= \frac{1}{f(\theta)\lambda_R} \beta R \left|\frac{\partial G(a|z(\theta), \theta)}{\partial \theta}\right| \eta(\theta)\end{aligned}$$

and $\varepsilon^* = \frac{\Psi' \frac{1}{a}}{\Psi'' \frac{y}{a^2} + \Psi' \frac{1}{a}}$ *is proportional to the compensated labor supply elasticity.*

Proof. See Appendix. □

As is well-understood, marginal tax rates are decreasing in the elasticity of labor supply with respect to the after-tax wage. Moreover, the weighted mass $ag(a|z(\theta), \theta)$ of agents whose labor supply is distorted by the tax is negatively related to the marginal tax. Next, consider the term in brackets of $\mathcal{A}(\theta, a)$. $\frac{1}{u'(c_1(\theta))}$ is equal to the marginal cost of raising lifetime utility of agents with initial type θ. $\frac{1}{G(a|z(\theta),\theta)} \int_{\underline{a}}^{a} \frac{1}{u'(c_2(\theta,a^*))} dG(a^*|z(\theta), \theta)$ is a truncated mean and the average marginal cost of raising utility for all types smaller than a. In a first best, the planner would equate these costs and

[14]See Findeisen (2011) for a more detailed discussion in a T period model with stochastic skills, but without endogenous education. In a recent paper Golosov, Troshkin, and Tsyvinski (2009) also provide formulas for dynamic optimal labor taxes, connecting them to empirical observables in the spirit of the contributions of Diamond (1998) and Saez (2001) for the static Mirrlees model.

smooth his own expenditures as well as utility levels across states and time. In constrained efficient allocations, incentive compatibility hinders full insurance, and the costs of providing utility varies with a. The planner can close this gap as much as possible by raising marginal labor tax rates, which hits all people with skill level bigger than a. Notice that $\mathcal{A}(\theta, a)$ disappears, if agents are risk neutral, and therefore second period insurance is not a concern. With risk-aversion the labor tax system provides insurance against the *innate and the educational* risk agents face. Education enters directly by influencing the need for insurance via the conditional distribution.

The other term $\mathcal{B}(Y(\theta, a))$ shows how labor tax rates are used to optimally supply dynamic incentives. In contrast to $\mathcal{A}(\theta, a)$ it is independent of risk-preferences, but vanishes with ex-ante homogeneous agents. $\left|\frac{\partial G(a|z(\theta),\theta)}{\partial \theta}\right|$ captures the informational advantage of the marginal type just above θ. The bigger the expression the more important a higher tax for type θ becomes to avoid the higher type claiming the allocation of θ. Note that education decisions enter by affecting the distribution and the marginal shifts. Intuitively, the planner offers labor tax systems to individuals, which give the correct incentives for revelation in the first period. In the appendix we show that $\eta(\theta)$ is equal to the cumulative mass of Pareto weights of lesser skilled agents, adjusted for incentive compatibility effects. For redistributive preferences, i.e. net Pareto weights non-increasing in θ, $\eta(\theta)$ will always be positive. Hence, the higher redistributive motives towards the low skilled or the lower the efficiency loss of redistributing, the higher are marginal tax rates.

Next, we discuss the instruments the government uses to implement the educational wedge, defined in Section 2.2. In contrast to the optimal labor *wedge*, which equals the optimal labor *tax*, there is no single policy instrument for which the education wedge equals the marginal distortion of the policy. Instead, the government uses two instruments: i) the nonlinear grant schedule $L(z)$, which depends on education chosen ii) the labor tax code in the second period. Using the agents' optimality conditions in the proposed implementation one can show that the wedge equals:

$$\tau_z(\theta) = L'(z) - \int_{\underline{a}}^{\overline{a}} \frac{u'(c_2(\theta, a))}{u'(c_1(\theta))} g(a|z(\theta), \theta) T_z(y(a), z) da$$

An increase in $\tau_z(\theta)$ encourages education at level θ. The incentive for agents to increase their educational attainment comes from: i) An increase in their grant measured by $L'(z)$ and ii) a deductible, reducing their labor income tax burden, if $T_z(y(a), z)$ is negative. Note that the gain from the deductible is an expected value weighted by the normalized shadow value of resources in each state.

The labor tax formulas can be generalized to a whole life cycle, if the realized skill level a remains constant over time. Arguably, the most important risks in life are the lottery of initial conditions (θ) and educational risk; for example, Huggett, Ventura, and Yaron (2010) find that differences across individuals at the age of 23 are more important than subsequent shocks received afterwards. This motivates the following proposition:

Proposition 4.2. *If individuals live for T periods and skills are fixed after period two and $\beta R = 1$, at any constrained Pareto optimum marginal labor income tax rates satisfy:*

$$\frac{T'_y(z(\theta), y(\theta, a))}{1 - T'_y(z(\theta), y(\theta, a))} = \frac{u'(c_c(\theta, a))}{\varepsilon^* a g(a|z(\theta), \theta)} \left[\mathcal{A}(\theta, a) + \mathcal{B}(\theta, a)\right],$$

where

$$\mathcal{A}(\theta, a) = G(a|z(\theta), \theta) \left[\frac{1}{u'(c_1(\theta))} - \frac{1}{G(a|z(\theta), \theta)} \int_{\underline{a}}^{a} \frac{1}{u'(c_c(\theta, a^*))} dG(a^*|z(\theta), \theta)\right]$$

$$\begin{aligned}\mathcal{B}(Y(\theta, a)) &= -\frac{1}{f(\theta)\lambda_R} \frac{\partial G(a|z(\theta), \theta)}{\partial \theta} \eta(\theta) \\ &= \frac{1}{f(\theta)\lambda_R} \left|\frac{\partial G(a|z(\theta), \theta)}{\partial \theta}\right| \eta(\theta)\end{aligned}$$

$c = 2, ..., T$ *and* $\varepsilon^* = \frac{\Psi' \frac{1}{a}}{\Psi'' \frac{y}{a^2} + \Psi' \frac{1}{a}}$ *is proportional to the compensated labor supply elasticity.*

Proof. See Appendix. □

To abstract from intertemporal effects, when patience dominates or the rate of return are sufficiently high or low, we set $\beta R = 1$. Consumption is constant for agents with a realized history (θ, a) and equal to $c_c(\theta, a)$ from period two onwards, and consequently so is labor supply and gross income $y_c(\theta, a)$. Individuals then face constant marginal tax rates and pay constant taxes during the rest of their lifetime. Ways to implement such an allocation, for example, via income averaging are presented in Werning (2007).

5 Conclusion

To be written.

A Appendix

A.1 Properties of the Laissez-Faire Allocation

To show part (iv), we show that:

$$\frac{\partial^2 U(s,z;\theta,.)}{\partial s \partial \theta} = -\beta \int_{\underline{a}}^{\overline{a}} \frac{\partial^2 v_2(a,s)}{\partial s \partial a} \frac{\partial G(a|z,\theta)}{\partial \theta} da \tag{23}$$

$$= -\beta \int_{\underline{a}}^{\overline{a}} \frac{\partial y(s,a)}{\partial s} \left[\Psi' \left(\frac{y(s,a)}{a} \right) \frac{1}{a^2} + \Psi'' \left(\frac{y(s,a)}{a} \right) \frac{y(s,a)}{a^2} \right] \frac{\partial G(a|z,\theta)}{\partial \theta} da < 0$$

$$\frac{\partial^2 U(s,z;\theta,.)}{\partial s \partial z} \tag{24}$$

$$= u''(c_1(\theta) - \beta \int_{\underline{a}}^{\overline{a}} \frac{\partial y(s,a)}{\partial s} \left[\Psi' \left(\frac{y(s,a)}{a} \right) \frac{1}{a^2} + \Psi'' \left(\frac{y(s,a)}{a} \right) \frac{y(s,a)}{a^2} \right] \frac{\partial G(a|z,\theta)}{\partial z} da < 0$$

$$\frac{\partial^2 U(s,z;\theta,.)}{\partial z \partial \theta} = -\beta \int_{\underline{a}}^{\overline{a}} \frac{\partial v_2(a,s)}{\partial a} \frac{\partial^2 G(a|z,\theta)}{\partial z \partial \theta} da > 0, \tag{25}$$

applying the envelope theorem and integrating by parts several times, as well as all three assumptions on the conditional distribution function. Also note that $\frac{\partial y(s,a)}{\partial s} < 0$, simply because of income effects.

To prove Corollary 2.3, note that (25) becomes:

$$\frac{\partial^2 U(s,z;\theta,.)}{\partial z \partial \theta} = -u''(c_1(\theta) w'(\theta) - \beta \int_{\underline{a}}^{\overline{a}} \frac{\partial v_2(a,s)}{\partial a} \frac{\partial^2 G(a|z,\theta)}{\partial z \partial \theta} da > 0,$$

under the assumption $w'(\theta) \geq 0$.

A.2 First Best Policies: Pareto Problem and Proof of Proposition 2.4

The Lagrangian reads as:

$$\begin{aligned}
\max_{v_0(\theta), v_2(\theta,a), y_1(\theta,a), c_1(\theta)} & \int_{\underline{\theta}}^{\overline{\theta}} u(c_1(\theta)) d\tilde{F}(\theta) \\
& + \beta \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} v_2(\theta,a) dG(a|z(\theta),\theta) d\tilde{F}(\theta) \\
& + \frac{1}{R} \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} y(\theta,a) dG(a|z(\theta),\theta) dF(\theta) \\
& - \frac{1}{R} \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} u^{-1} \left[v_2(\theta,a) + \Psi\left(y(\theta,a)/a\right) \right] dG(a|z(\theta),\theta) dF(\theta) \\
& - \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \left(c_1(\theta) + z(\theta) \right) dF(\theta)
\end{aligned}$$

With first-order conditions:

$$\tilde{f}(\theta) - \lambda_R f(\theta) \frac{1}{u'(c_1(\theta))} = 0 \tag{c_{1FB}}$$

$$g(a|\theta)\tilde{f}(\theta)\beta - \lambda_R g(a|\theta) f(\theta) \frac{1}{u'(c_2(\theta,a))} = 0, \tag{v_{1FB}}$$

$$1 = \frac{1}{u'(c_2(\theta,a))} \Psi' \left(\frac{y(\theta,a)}{a} \right) \frac{1}{a} \tag{y_{FB}}$$

$$\tilde{f}(\theta)\beta \int_{\underline{a}}^{\overline{a}} v_2(\theta,a) \frac{\partial g(a|z(\theta),\theta)}{\partial z(\theta)} da + \frac{1}{R} \lambda_R f(\theta) \int_{\underline{a}}^{\overline{a}} (y(\theta,a) - c_2(\theta,a)) \frac{\partial g(a|z(\theta),\theta)}{\partial z(\theta)} da = \lambda_R f(\theta) \tag{z_{FB}}$$

Proposition 2.4 directly follows from these first-order conditions.

A.3 Proof of Lemma 3.2

First, we show that $\frac{\partial y(\theta,a)}{\partial \theta} \geq 0$ and $\frac{\partial z(\theta)}{\partial \theta} \geq 0$ are sufficient for local incentive compatibility. Part (ii) of Lemma 3.1 clearly holds, if $\frac{\partial^2 v_2(\theta,a)}{\partial \theta \partial a} \geq 0$ and $\frac{\partial z(\theta)}{\partial \theta} \geq 0$. Using (9):

$$\frac{\partial^2 v_2(\theta,a)}{\partial \theta \partial a} = \frac{\partial y(\theta,a)}{\partial \theta} \left[\Psi'' \left(\frac{y(\theta,a)}{a} \right) \frac{1}{a} + \frac{1}{a^2} \Psi' \left(\frac{y(\theta,a)}{a} \right) \right], \tag{26}$$

which is non-negative, if and only if $\frac{\partial y(\theta,a)}{\partial \theta} \geq 0$.

Next, we investigate global incentive compatibility. For type θ truth-telling is a global maximum if:

$$\frac{\partial U(\theta, r(\theta))}{\partial r(\theta)} > (<) 0 \ \text{ if } \ r(\theta) < (>) \theta. \tag{27}$$

We now show that this always holds if $\frac{\partial z(\theta)}{\partial \theta} \geq 0$ and $\frac{\partial y(\theta,a)}{\partial \theta} \geq 0$. Note that for an individual of type $r(\theta)$ truth-telling to be optimal requires:

$$\frac{\partial U(r(\theta), r(\theta))}{\partial r(\theta)} = \theta.$$

Thus we can rewrite condition (27) as:

$$\frac{\partial U(\theta, r(\theta))}{\partial r(\theta)} - \frac{\partial U(r(\theta), r(\theta)}{\partial r(\theta)} > (<) 0 \ \text{ if } \ r(\theta) < (>) \theta. \tag{28}$$

The LHS can be rewritten as:

$$\beta \int_{\underline{a}}^{\overline{a}} \frac{\partial v_2(r(\theta),a)}{\partial r(\theta)} \left[g(a|z(r(\theta)),\theta) - g(a|z(r(\theta)), r(\theta)) \right] +$$

$$\beta \int_{\underline{a}}^{\overline{a}} v_2(r(\theta), a) \left[\frac{\partial g(a|z(r(\theta)), \theta)}{\partial z(r(\theta))} - \frac{\partial g(a|z(r(\theta)), r(\theta))}{\partial z(r(\theta))} \right] \frac{\partial z(r(\theta))}{\partial r(\theta)} da. \quad (29)$$

Integrating by parts one can rewrite this as:

$$-\beta \int_{\underline{a}}^{\overline{a}} \frac{\partial^2 v_2(r(\theta), a)}{\partial r(\theta) \partial a} \left[G(a|z(r(\theta)), \theta) - G(a|z(r(\theta)), r(\theta)) \right]$$

$$-\beta \int_{\underline{a}}^{\overline{a}} \frac{\partial v_2(r(\theta), a)}{\partial a} \left[\frac{\partial G(a|z(r(\theta)), \theta)}{\partial z(r(\theta))} - \frac{\partial G(a|z(r(\theta)), r(\theta))}{\partial z(r(\theta))} \right] \frac{\partial z(r(\theta))}{\partial r(\theta)} da. \quad (30)$$

This expression has the sign as $(\theta - r(\theta))$, if $\frac{\partial^2 v_2(r(\theta),a)}{\partial r(\theta)\partial a}$, $\frac{\partial v_2(r(\theta),a)}{\partial a}$ and $\frac{\partial z(r(\theta))}{\partial r(\theta)}$ are non-negative, using Assumptions ? and ?. The set of admissible strategies $r(\theta)$ is the same as the set of possible θ types. So incentive compatibility is guaranteed, if $\frac{\partial^2 v_2(\theta,a)}{\partial\theta\partial a}$, $\frac{\partial v_2(\theta,a)}{\partial a}$ and $\frac{\partial z(\theta)}{\partial \theta}$ are non-negative. This is true, if $\frac{\partial z(\theta)}{\partial\theta} \geq 0$ and $\frac{\partial y(\theta,a)}{\partial\theta} \geq 0$, using (26) and (9).

A.4 Second Best Policies: Pareto Problem and Optimality Conditions

$$\begin{aligned}
\max_{c_1(\theta), v_2(\theta,a), z(\theta), y(\theta,a)} & \int_{\underline{\theta}}^{\overline{\theta}} u(c_1(\theta)) d\tilde{F}(\theta) \\
& + \beta \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) dG(a|z(\theta), \theta) d\tilde{F}(\theta) \\
& + \frac{1}{R} \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} y(\theta, a) dG(a|z(\theta), \theta) dF(\theta) \\
& - \frac{1}{R} \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} u^{-1} \left[v_2(\theta, a) + \Psi\left(y(\theta, a)/a \right) \right] dG(a|z(\theta), \theta) dF(\theta) \\
& - \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \left(c_1(\theta) + z(\theta) \right) dF(\theta) \\
& - \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} \left(\mu'(\theta, a) v_2(\theta, a) + \mu(\theta, a) \Psi' \left(\frac{y(\theta, a)}{a} \right) \frac{y(\theta, a)}{a^2} \right) da d\theta \\
& - \int_{\underline{\theta}}^{\overline{\theta}} \eta'(\theta) \left[u(c_1(\theta)) + \beta \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) dG(a|z(\theta)) da \right] d\theta \\
& - \beta \int_{\underline{\theta}}^{\overline{\theta}} \eta(\theta) \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial \theta} da d\theta
\end{aligned}$$

With first-order conditions:

$$u'(c_1(\theta))(\tilde{f}(\theta) - \eta'(\theta)) - \lambda_R f(\theta) = 0 \qquad (c_{1SB})$$

$$\left(\tilde{f}(\theta) - \eta'(\theta)\right) \beta g(a|z(\theta), \theta) - \lambda_R \frac{1}{R} \frac{1}{u'(c_2(\theta, a))} g(a|z(\theta), \theta) f(\theta) - \mu'(\theta, a)$$
$$- \beta \frac{\partial g(a|z(\theta), \theta)}{\partial \theta} \eta(\theta) = 0 \qquad (v_{2SB})$$

$$\frac{1}{R} \lambda_R g(a|z(\theta), \theta) f(\theta) - \mu(\theta, a) \left[\Psi'' \left(\frac{y(\theta, a)}{a} \right) \frac{y(\theta, a)}{a^3} + \frac{1}{a^2} \Psi' \left(\frac{y(\theta, a)}{a} \right) \right]$$
$$- \frac{1}{R} \lambda_R g(a|z(\theta), \theta) f(\theta) \frac{\Psi' \left(\frac{y(\theta,a)}{a} \right)}{a u'(c_2(\theta, a))} = 0, \qquad (y_{SB})$$

$$\tilde{f}(\theta) \beta \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial z(\theta)} da + \frac{1}{R} \lambda_R f(\theta) \int_{\underline{a}}^{\overline{a}} \frac{\partial g(a|z(\theta), \theta)}{\partial z(\theta)} \left(y(\theta, a) - c_2(\theta, a) \right) da$$
$$- \eta'(\theta) \beta \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial z(\theta)} da - \beta \eta(\theta) \int_{\underline{a}}^{\overline{a}} v_2(\theta, a) \frac{\partial^2 g(a|z(\theta), \theta)}{\partial z(\theta) \partial \theta} da - \lambda_R f(\theta) = 0$$
$$(z_{SB})$$

A.5 Proof of Proposition 4.1 and Proposition 4.2

A.5.1 Proposition 4.1

Rewriting (y_{SB}):

$$\lambda_R g(a|z(\theta), \theta) f(\theta) \left[1 - \frac{\Psi' \left(\frac{y(\theta,a)}{a} \right)}{a u'(c_1(\theta, a))} \right]$$
$$- \mu(\theta, a) \left[\Psi'' \left(\frac{y(\theta, a)}{a} \right) \frac{y(\theta, a)}{a^3} + \frac{1}{a^2} \Psi' \left(\frac{y(\theta, a)}{a} \right) \right] = 0.$$

Dividing by $\frac{\Psi'}{au'}$ and $\lambda_R g(a|z, \theta) f(\theta)$ and using the FOC conditions of the individual in the second period, i.e. $u'(1 - T') = \Psi' \frac{1}{a}$ yields

$$\frac{T'_\theta(Y(\theta, a))}{1 - T'_\theta(Y(\theta, a))} = \frac{\mu(\theta, a)}{\lambda_R g(a|z(\theta), \theta) f(\theta) a} \left[\frac{\Psi'' \frac{y}{a^2} + \Psi' \frac{1}{a}}{\frac{\Psi'}{au'}} \right],$$

which can be written as

$$\frac{T'_\theta(Y(\theta, a))}{1 - T'_\theta(Y(\theta, a))} = u' \cdot \frac{\mu(\theta, a)}{\lambda_R g(a|z(\theta), \theta) f(\theta) a} \frac{1}{\varepsilon^*},$$

where $\varepsilon^* = \frac{\Psi' \frac{1}{a}}{\Psi'' \frac{y}{a^2} + \Psi' \frac{1}{a}}$ is proportional to the compensated labor supply elasticity, see, e.g., Atkinson and Stiglitz (1980, p418).

The multiplier $\mu(\theta, a)$ can be obtained using (v_{2SB}) and (c_{1SB}):

$$\mu(\theta,a) = \frac{\lambda_R f(\theta)}{u'(c_1(\theta))}\beta G(a|z(\theta),\theta) - \frac{\lambda_R}{R} f(\theta) \int_{\underline{a}}^{a} \frac{1}{u'(c_2(\theta,a))} dG(a|z(\theta),\theta)$$

$$-\beta \frac{\partial G(a|z(\theta),\theta)}{\partial \theta} \eta(\theta^*),$$

yielding:

$$\frac{T'_\theta(Y(\theta,a))}{1 - T'_\theta(Y(\theta,a))} = \frac{u'(c_2(\theta,a))}{\varepsilon^* a g(a|z(\theta),\theta)} \left[\mathcal{A}(Y(\theta,a)) + \mathcal{B}(Y(\theta,a))\right]$$

where

$$\mathcal{A}(Y(\theta,a)) = \frac{\beta G(a|z(\theta),\theta)}{u'(c_1(\theta))} - \frac{1}{R}\int_{\underline{a}}^{a} \frac{1}{u'(c_2(\theta,a))} dG(a|z(\theta),\theta)$$

$$\mathcal{B}(Y(\theta,a)) = -\frac{1}{f(\theta)\lambda_R} \beta R \frac{\partial G(a|z(\theta),\theta)}{\partial \theta} \eta(\theta).$$

From (c_{1SB}), $\eta(\theta)$ is given by:

$$\eta(\theta) = \tilde{F}(\theta) - \lambda_R \int_{\underline{\theta}}^{\theta} \frac{1}{u'(c_1(\theta))} f(\theta) d\theta.$$

The direct benefit of raising utils for agents with skill lower than θ is $\tilde{F}(\theta)$. The monetary cost is $\int_{\underline{\theta}}^{\theta} \frac{1}{u'(c_1(\theta))} f(\theta) d\theta$, transformed into utils by λ_R.

A.5.2 Proposition 4.2

The fact that consumption and labor wedges will be constant in a deterministic Mirrleesian economy is proofed in Werning (2007) and generalizes to our setting with a continuum of skills. It follows that with $\beta R = 1$, flow utility is constant from period 2 onwards. The lifetime utility of an agent with draws (θ, a) from the second period onwards is hence:$V_2(\theta,a) = (T-1)v_c(\theta,a)$, with $v_c(\theta,a) = u(c_c(\theta,a)) - \Psi\left(\frac{y_c(\theta,a)}{a}\right)$. The envelope condition, necessary for incentive compatibility, becomes (analogous to (9)):

$$\frac{\partial V_2(\theta,a)}{\partial a} = \sum_{t=2}^{T} \Psi'\left(\frac{y(\theta,a)}{a}\right) \frac{y(\theta,a)}{a^2}. \tag{31}$$

Constant consumption $c_c(\theta, a)$ can be written as:

$$c_c(\theta, a) = u^{-1}\left[v_c(\theta, a) + \Psi\left(\frac{y_c(\theta, a)}{a}\right)\right] = c_c(\theta, a) = u^{-1}\left[\frac{V_2}{T-1} + \Psi\left(\frac{y_c(\theta, a)}{a}\right)\right] \tag{32}$$

Writing up a Lagrangian with (31) as the appropriate incentive constraint and (32) inserted into the present value budget constraint, taking first order conditions and combining with optimality conditions for the household, like in the last two appendices, directly delivers the result.

References

ALTONJI, J., AND T. DUNN (1996): "The effects of family characteristics on the return to education," *The review of economics and statistics*, 78(4), 692–704.

ANDERBERG, D. (2009): "Optimal policy and the risk properties of human capital reconsidered," *Journal of Public Economics*, 93(9-10), 1017–1026.

ATKINSON, A., AND J. STIGLITZ (1976): "The design of tax structure: direct versus indirect taxation," *Journal of Public Economics*, 6(1-2), 55–75.

BOHACEK, R., AND M. KAPICKA (2008): "Optimal human capital policies," *Journal of Monetary Economics*, 55(1), 1–16.

BOVENBERG, L., AND B. JACOBS (2005): "Redistribution and education subsidies are Siamese twins," *Journal of Public Economics*, 89(11-12), 2005–2035.

CARNIERO, P., AND J. HECKMAN (2003): "Human capital policy," *Inequality in America: What Role for Human Capital Policies*.

COURTY, P., AND H. LI (2000): "Sequential Screening," *The Review of Economic Studies*, 67(4), 697–717.

CUNHA, F., AND J. HECKMAN (2006): "The Evolution of Labor Earnings Risk in the US Economy," (665).

CUNHA, F., AND J. J. HECKMAN (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 14(6), 870–893.

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.

DIAMOND, P. (1998): "Optimal income taxation: An example with a U-shaped pattern of optimal marginal tax rates," *American Economic Review*, 88(1), 83–95.

DIAMOND, P. A., AND J. A. MIRRLEES (1978): "A model of social insurance with variable retirement," *Journal of Public Economics*, 10(3), 295–336.

FARHI, E., AND I. WERNING (2010): "Insurance and taxation over the life cycle," .

FINDEISEN, S. (2011): "Pareto Optimal Labor and Savings Taxation Under Uncertainty," *Working Paper, Zurich*.

GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2009): "Optimal Dynamic Taxes," .

GROCHULSKI, B., AND T. PISKORSKI (2010): "Risky human capital and deferred capital income taxation," *Journal of Economic Theory*, 145(3), 908–943.

HUGGETT, M., G. VENTURA, AND A. YARON (2010): "Sources of Lifetime Inequality," *American Economic Review*, forthcoming.

KATZ, L. F., AND D. H. AUTOR (1999): "Changes in the wage structure and earnings inequality," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3 of *Handbook of Labor Economics*, chap. 26, pp. 1463–1555. Elsevier.

KRUEGER, D., F. PERRI, L. PISTAFERRI, AND G. L. VIOLANTE (2010): "Cross Sectional Facts for Macroeconomists," *Review of Economic Dynamics*, 13(1), 1–14.

LEMIEUX, T. (2006): "Postsecondary Education and Increasing Wage Inequality," *American Economic Review*, 96(2), 195–199.

MILGROM, P., AND I. SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70(2), 583–601.

MILGROM, P., AND C. SHANNON (1994): "Monotone comparative statics," *Econometrica*, 62(1), 157–180.

PAVAN, A., I. SEGAL, AND J. TOIKKA (2009): "Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization, and Informational Disclosure," *Mimeo*.

ROGERSON, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

SAEZ, E. (2001): "Using elasticities to derive optimal income tax rates," *Review of Economic Studies*, 68(1), 205–229.

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.

WERNING, I. (2007): "Optimal Fiscal Policy with Redistribution," *The Quarterly Journal of Economics*, 122(3), 925–967.

——— (2010): "Nonlinear Capital Taxation," *Working Paper, MIT*.