

Education and Optimal Dynamic Taxation

Sebastian Findeisen

University of Zurich

Dominik Sachs

University of Konstanz*

October 21, 2011

Abstract

We study optimal tax and educational policies in a dynamic private information economy, in which *ex-ante* heterogeneous individuals make an educational investment early in their life and face a stochastic wage distribution. We characterize labor and education wedges in this setting analytically and numerically, using a calibrated example. We present ways to implement the optimum. In one implementation there is a common labor income tax schedule, and a repayment schedule for government loans given out to agents during education. These repayment plans are contingent on loan size and income and capture the history dependence of the labor wedges. Applying the model to US-data and a binary education decision (graduating from college or not) we characterize optimal labor wedges for individuals without college degree and with college degree. The labor wedge of college graduates as a function of income lies first strictly above their counterparts from high-school, but this reverses at higher incomes. The loan repayment schedule is hump-shaped in income for college graduates.

JEL-classification: H21, H23, I21

Keywords: Optimal dynamic taxation, Education, Implementation

*Contact: sebastian.findeisen@uzh.ch and dominik.sachs@uni-konstanz.de. We acknowledge the hospitality of Yale and Stanford University. We thank Dan Anderberg, Friedrich Breyer, Carlos da Costa, Mike Golosov, Florian Scheuer, Dirk Schindler, Kjetil Storesletten, Iván Werning, Christoph Winter, Fabrizio Zilibotti, and seminar participants in Bloomington, Gerzensee, Konstanz, Stanford, Yale and Zurich for helpful comments. We also thank Stefan Voigt for valuable research assistance.

1 Introduction

A well-established link exists between income inequality and education. In a recent survey Acemoglu and Autor (2010) note that the college premium has been rising steadily in the US for the last 45 years with no immediate end in sight. This widening wage premium has contributed to an increase in the dispersion of the distribution of wages and income also in other advanced economies (Krueger, Perri, Pistaferri and Violante 2010). However, not all of the observed rise in the college premium is necessarily caused by an increased demand for skills acquired in college to the extent that unobserved skills influence wages and labor market outcomes in general. Taber (2001) finds that at least some of the observed rise in the educational skill premium should be attributed to unobserved skill heterogeneity realized *before* college decisions are made.¹ Furthermore, since heterogeneity is already realized at a very early stage in life, individuals face very different returns to education, resulting in very different incentives to invest into education² Finally, it is well known that returns to education are subject to substantial uncertainty.³

What are the implications of these empirical relationships for optimal policies? How can non-linear labor taxes be used to dampen inequality? Should information about one's education level be included in the tax code? How should education policies be set and how are incentives to invest into human capital affected by the tax code for *ex-ante heterogeneous agents*? What role do policies have to insure risk?

In this paper, we try to make progress on these important questions in a normative Mirrleesian framework, using both theory and numerical simulations. Although previous papers have studied optimal taxation problems with endogenous education, they have typically abstracted from either *ex-ante* heterogeneity or uncertainty, or imposed exogenous restrictions on the available policy instruments (e.g. only linear taxes). To address this issue, we build a more realistic setting including both *ex-ante* heterogeneity and uncertainty in educational returns. In addition, the set of policy instruments is endogenous, arising solely from the informational structure of the problem.

More formally, we study an environment characterized by private information, which evolves stochastically over time. *Ex-ante* heterogeneous individuals are born with different *innate abilities* θ and decide on their level of education z early in their life. After education, agents learn their skill level a that can be interpreted as wage,

¹Indirect evidence for the importance of unobserved skills comes from the strong persistence of within education group inequality.

²See, for example, Lemieux (2006) or Carniero and Heckman (2003). We give a brief review of key empirical findings guiding our modeling later in the paper.

³See, e.g., Cunha, Heckman and Navarro (2005).

they work and generate income.⁴ The planner cannot observe innate ability θ , work effort and the skill level a , so the resulting allocation must be incentive compatible at every point in time. However, we do assume that the planner can observe educational investment z .

Labor market risk is taken into account in our model by the fact that individuals face a distribution of wages before entering the labor market. Once they enter, they learn their type a , which is private information. When deriving analytical results, we try to impose minimal structure on the stochastic process and work with a general conditional distribution function $G(a|z, \theta)$. *Ex-ante* heterogeneity manifests itself in two ways: 1) Holding education constant, the distribution of skills tomorrow depends on innate ability directly; and 2) the marginal return to education varies with the innate type, i.e. the cross derivative of $G(a|z, \theta)$ with respect to θ and z is typically not zero.

We are interested in characterizing the properties of constrained Pareto optimal allocations and investigating decentralized implementations that can generate these optimal allocations as a consequence of agents' choices in markets. In the truthful mechanism we employ, agents first reveal their innate ability level first, and only at a later stage their labor market skill level. The second stage is a familiar Mirrlees problem, in which agents choose from consumption-income bundles. In the first stage, agents select from a menu of education levels and utility promises. Our analysis shows that education can be understood as a *screening device* to separate agents in an incentive compatible allocation early in life. The utility promises are tailored to the endogenous skill distributions, which are functions of innate type and education. To tackle the problem we employ a first-order approach simplifying the set of incentive constraints, and then verify ex-post whether the obtained allocations are indeed incentive compatible.

A result arising immediately from the environment is the dependency of labor wedges on education decisions; individuals with the same income face, in general, different marginal distortions, unless they were also identical *ex-ante*. Conceptually, we show that optimal implicit marginal labor tax rates in the dynamic case consist of the sum of two parts: The first is the standard marginal tax formula from the static Mirrlees model for the Utilitarian planner obtained in Mirrlees (1971) and Saez (2001). Conditional on the Pareto weight assigned to the innate type, the planner maximizes expected utility in our dynamic model. In the static model, he tailors tax rates to the skill distribution to optimally redistribute across skill levels. In the dynamic model he does the same to optimally insure different innate types against labor market risk.

⁴After agents enter the labor market we assume no more shocks to ability. This is to strengthen the focus on the education-taxation link. Recent papers which have characterized labor wedges in dynamic Mirrlees models with shocks occurring every period are Farhi and Werning (2011) and Golosov, Troshkin and Tsyvinski (2010).

Since the endogenous skill distribution is different for each innate type, there is one insurance part of labor taxation for every initial type. For the second effect and to fix ideas, consider the case of only two innate types θ_L and θ_H . We show that the marginal tax rate at some income level $y(a^*)$ for the low type θ_L depends positively on the difference: $(1 - G(a^*|z, \theta_H)) - (1 - G(a^*|z, \theta_L)) = Pr(a > a^*|z, \theta_H) - Pr(a > a^*|z, \theta_L)$. To look at a specific example, let us consider a first-order stochastic dominance ordering of the type space $G(a|z, \theta_H) \succeq_{FOSD} G(a|z, \theta_L)$, so the above term is never negative. The planner increases the marginal distortion for the low type proportional to the difference in order to deter a deviation from high types, who have a higher likelihood of being a higher type than a^* and therefore of being affected by this higher marginal distortion through a decrease in consumption. Differences in the conditional distribution are exploited in this way by the planner to effectively deal with the incentive constraints in the first period

The dependency of the tax schedules on characteristics other than income is related to the idea of *tagging* and income taxation, as first proposed by Akerlof (1978).⁵ Whereas our analysis shares the feature of this method in that the social planner uses additional information to tailor marginal tax rates for each educational level to the respective skill distribution, it differs substantially in the sense that education is not an immutable *tag*, but rather an endogenous variable. One might, therefore, label the dependency of tax distortions on education as *endogenous tagging*.⁶

Our theory illustrates how individual education decisions are distorted by optimal government intervention in constrained Pareto optimal allocations. This depends again on two forces. First, implicit educational subsidies are used to offset the distortionary impact of the labor tax on educational decisions. This is equivalent to a first-best rule in a sense that innate ability would be observable. Second, the planner distorts education to relax binding incentive constraints. If the desired redistribution of income is downward (high to low innate types), we show that this imposes an implicit tax and hence a 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 static model.

We numerically illustrate these forces, parameterizing the skill distributions as log-normal with a Pareto tail, following standard practice in the literature. We set the location parameter of the log-normal to be a function of innate skills and education, allowing for differing marginal returns to education across agents. We discipline our parameter choice by matching monetary education expenditures on college education in the US. Holding income constant, implicit marginal tax rates on labor income differ

⁵More recently tagging is investigated by Cremer, Gahvari and Lozachmeur (2010), Mankiw and Weinzierl (2010) as well as Weinzierl (2011b).

⁶A similar logic arises in the recent paper of Scheuer (2011) in a model with endogenous occupational choices.

by up to 21 % points across different education levels in our numerical example. These differences are mainly explained by tailoring marginal tax rates to the different skill distributions. We find only small second-order effects of the dynamic incentive effect described above. For the education wedge we find that it is mainly determined by the first-best rule and that the downward distortion to relax incentive constraints is of minor magnitude.

We propose two ways to decentralize the optimum. In one implementation, the government offers a non-linear schedule of grants to individuals. These grants are linked to education levels. Agents self-select into education and use the grants to finance tuition and consumption during education. After graduating they are free to choose their work hours/effort and pay a labor income tax conditioning on their education choice. Together with a savings tax these policies can achieve the optimum. Such history dependent tax schedules might be considered hard to implement in practice, since they might be perceived as violating horizontal equity by society. Dealing with that, we provide another implementation that is basically a reinterpretation of the first one: Agents are offered loans during their education period that are linked to education levels. When working, all individuals face the same *history - independent* tax schedule and in addition a loan repayment schedule that depends both on income and on the size of the loan.⁷ Since the government effectively provides liquidity to agents during their education, both implementation would also be applicable if agents were credit constrained.

Finally, we present a simpler application of the model, using a binary education decision, i.e. going to college or not. Importantly, this enables us to use estimates of factual as well as counterfactual income distributions from the labor economics literature (Cunha and Heckman 2007, 2008), both of which are needed as an input to the model. Additionally, by reducing the dimensionality of the problem, a binary education choice also significantly reduces the complexity and sophistication of the policy instruments, coming another step closer to practical policy recommendations. For high-school and college graduates optimal implicit labor tax rates are both U-shaped as in Diamond (1998) and Saez (2001). Marginal tax rates for college graduates are first higher then smaller, before both schedules converge to roughly the same top tax rate. Using the high-school labor wedge schedule as the common labor income tax schedule, the implied income-contingent repayment schedule for college graduates is hump-shaped. In expectation college graduates pay a small yearly premium over the actuarially-fair annuity value of the loan they take from the government.

Related Literature. Several previous papers have studied problems of optimal labor taxation and their relation to education decisions. One strand of inquiry has

⁷Income-contingent loan repayments are actually in place in Australia, Great Britain, Sweden, Norway and New Zealand.

worked under the assumption of *ex-ante homogeneity* and *risky* human capital. Grochulski and Piskorski (2010) focus on the implications of unobservable human capital investment for capital taxation. Anderberg (2009) emphasizes that the risk properties of human capital are crucial for the question whether and how education should be distorted relative to a first-best rule.⁸ Da Costa and Maestri (2007) show that human capital should always be encouraged in a second-best environment. In contrast, our model explicitly stresses the importance of heterogeneity, already at the point when education decisions are made.

In a static setting with *heterogeneity* but without *uncertainty*, Bovenberg and Jacobs (2005) analyze how endogenous education alters the result of the Mirrleesian tax problem and conclude that distortions on the education margin through income taxation should be offset by education subsidies. Relatedly, Bohacek and Kapicka (2008) study a model in a dynamic environment with certainty and obtain similar results.⁹ Some of the mentioned papers analyze unobservable human capital investment and some the observable case. Which assumption is appropriate depends on the exact research question since in reality some parts of human capital are observable (type of degree, student fees, years at college) and some parts are not (e.g. study behavior). One aim of our paper is to argue that in an social optimum, using available public information about human capital can act as a screening device; that is why we focus on modeling the *observable* part of human capital investment.

Bénabou (2002) considers the case of *heterogeneity* and *uncertainty*. His approach is different from ours in two ways: first, he does not solve for constrained Pareto optimal allocations, but instead focuses on given non-linear instruments: education subsidies and income taxes. Second, he does not consider the optimum using both instruments simultaneously, but rather compares the implications of education subsidies with those of progressive income taxes for growth and efficiency. Our approach is to derive the full optimum constrained only by informational asymmetries and then discuss policy instruments implementing these allocations in line with the New Dynamic Public Finance literature.

Two recent contributions have analyzed optimal labor wedges in dynamic Mirrlees economies without endogenous education but with productivity shocks in every period, which we abstract from. Farhi and Werning (2011) characterize the evolution of the labor wedge *over time* analytically and numerically. Golosov, Troshkin and Tsyvinski (2010) derive and illustrate optimal labor wedges across agents in the case of no

⁸Focusing on linear policy instruments Anderberg and Andersson (2003) as well as Jacobs, Schindler and Yang (2010) also discuss the importance of the risk properties of human capital.

⁹Kapicka (2006) introduces non-observable endogenous human capital into a dynamic, non-stochastic Mirrlees model where taxes can only be conditioned on current income. He shows that marginal tax rates are lowered due to the education margin. Bovenberg and Jacobs (2005) consider observable and unobservable investment and provide an interesting discussion about different implications.

income effects and non-separable preferences. Our analysis on labor wedges focuses on differences across education levels in an optimal allocation and is, therefore, complementary.

To make the dynamic incentive problem tractable, we employ a first-order approach. Importantly, we rely on the fact that private information evolves sequentially in our economy, which avoids solving a multidimensional screening problem. This paper is, thus, related to work on dynamic mechanisms, in particular that of Pavan, Segal and Toikka (2011) and Kapicka (2010), who study the validity and robustness of the Mirrleesian first-order approach in dynamic economies, as well as that of Courty and Li (2000) who study optimal dynamic pricing of a monopolist.

Finally our paper builds on recent important work by Werning (2011), who shows that decentralized implementations of incentive compatible allocations, which are characterized by non-stochastic, non-linear capital taxes that may even be history independent, do exist.

This paper is organized as follows. Section 2 contains the basics of the model as well as its empirical foundation and a short characterization of the Laissez-Faire allocation. In Section 3, we investigate dynamic incentive compatibility and describe the major properties of constrained efficient allocations. Further on, we present quantitative results with a calibrated example. Decentralized implementations of constrained efficient allocations are provided in Section 4. Using data from Cunha and Heckman (2007, 2008) we apply our model to the case of a binary education decision in Section 5. Section 6 concludes.

2 Baseline Model

2.1 Heterogeneity, Education and Labor Market Outcomes: Some Empirical Facts

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 (2008) 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 earnings when going to college, lies to the right of the density when entering the labor after high school indicating a first-order stochastic dominance shift. 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 to be 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.

2.2 The Model

Individuals, whose mass is normalized to one, live for two periods. In the first period individuals acquire human capital and in the second period individuals work. The labor market ability in period two, however, is stochastic in period one. The distribution of labor market abilities depends on educational investment and initial type. We now formalize these ideas.

In period one individuals differ in innate ability θ , which can be interpreted as a one dimensional aggregate of (non-)cognitive skills and family background and is distributed in the interval $[\underline{\theta}, \bar{\theta}]$ according to $F(\theta)$. When individuals learn their type θ , which is private information, they make an educational investment. ¹⁰

In period two individuals draw their labor market ability a from a continuous conditional distribution $G(a|z, \theta)$, which depends on *innate* ability θ and education z , and has bounded support $[\underline{a}, \bar{a}]$. When individuals learn their type a they make a labor-leisure decision.

To sharpen a few analytical results, it turns out helpful to place some structure on the behavior of $G(a|z, \theta)$:

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

¹⁰Reducing education to a one-dimensional, monetary variable is admittedly a simplification of reality. Nevertheless, we think it is reasonable to downsize the complex education process like this since it captures the important fact that more education always requires more resources and makes the analysis tractable.

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

These assumption will not be needed to derive our main results, but help to illustrate important aspects of the model. Whenever an assumption is needed for a result, we refer to it. The first and the second one capture the notion that education and the innate ability level should both have a direct effect on labor market outcomes represented by a first-order stochastic dominance shift. The third one captures their interaction and respects the compelling evidence of complementarity between early ability and education.

2.3 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:

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

Individuals' utility functions are well-behaved – $u(\cdot)$ is assumed to be increasing, at least twice continuously differentiable and concave, and $\Psi(\cdot)$ is assumed to be increasing, at least twice continuously differentiable and convex. The parameter a is an individual's labor market skill, meaning individuals with a higher a need 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 short-sale or enforcement constraints and an exogenous gross return R . This defines the indirect utility function:

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

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. In our numerical simulations in Section 5 we

explicitly allow for the fact that different levels of education last a different number of periods.

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

Proposition 2.1. *The Laissez-Faire allocation has the following properties:*

(i) *The Euler Equation holds:*

$$u'(c_1(\theta)) = \beta R \int_{\underline{a}}^{\bar{a}} u'(c_2(\theta, a)) dG(a|z(\theta), \theta)$$

(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:*

$$u'(c_1(\theta)) = \beta \int_{\underline{a}}^{\bar{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial z} da$$

(iv) *If Assumptions 1-3 hold, educational investment is increasing in innate ability and savings are decreasing, i.e. $z'(\theta) > 0$ and $s'(\theta) < 0$.*

Proof. See Appendix A.1 □

Parts (i)-(iii) follow directly from the 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 if Assumptions 1-3 are fulfilled. The proof provides instructive intuition for this result. It is sufficient to show that the objective defined by (2.3) 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 : $\max_{s, z} U(s, z; \theta, a, \beta) = u(-s - z) + \beta \int_{\underline{a}}^{\bar{a}} v_2(a, s) g(a|z, \theta) da$. This objective is supermodular in credit taken $-s$, education z and type if and only if:

$$\frac{\partial^2 U(s, z; \theta, \cdot)}{\partial s \partial \theta} < 0 \tag{1}$$

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

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

In Appendix A.1 we show that all inequalities hold. Equations (1) and (2) imply that the return to savings is lower for higher θ types and with higher education, since expected labor skills are also higher. Equation (3) 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 neg-

ative number, leaves most of the results from Proposition 2.1 unaffected.¹¹ 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. *Suppose Assumptions 1 to 3 hold. If agents face borrowing constraints $s \geq \phi$, education is still monotone in type θ , i.e. $z'(\theta) > 0$ in the laissez-faire equilibrium.*

Proof. Above some threshold type, agents reach the borrowing limit and set s equal to ϕ . Of those agents higher types still face the greater returns to education because of the complementarity and therefore choose a higher level of z . \square

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 savings, labor income and education, respectively:

Savings wedge:

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

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 wedge.

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:

$$\tau_y(\theta, a) = 1 - \frac{\Psi' \left(\frac{y(\theta, a)}{a} \right) \frac{1}{a}}{u'(c_2(\theta, a))}$$

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

¹¹Surveying the literature, Carniero and Heckman (2003) conclude that short-term borrowing constraints seem to have only a very small effect on educational decisions.

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

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

In this case, a positive wedge corresponds to an upward distortion of education.

In the decentralized systems we later propose for the implementation of constrained Pareto efficient allocations, we will show, which of these *implicit taxes* on labor income, savings and education will equal *explicit marginal taxes*

3 Constrained Pareto Optimal Allocations

In this Section we characterize constrained Pareto efficient allocations, where ‘constrained’ refers to the government being unable to observe agents’ type θ in period one and a in period two. In Subsection 3.1, we show that the problem is tractable using a first-order approach. In addition we provide necessary as well as sufficient conditions for this approach to be valid. In Subsection 3.2, we analyze optimality conditions and its consequences for optimal policies. In Subsection 3.3, we show how the results are extended to a T -period framework and in Subsection 3.4, we explore the implications for optimal policies using numerical simulations.

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 $\{v_2(\theta, a)\}_{a \in [\underline{a}, \bar{a}]}$. In the second period the screening takes place over consumption levels $c_2(\theta, a)$ and labor supply $y(\theta, a)$. 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.¹²

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

¹²A similar first-order approach can be readily applied in 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).

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 $r(a)$:

$$u(c_2(\theta, a)) - \Psi \left[\frac{y(\theta, a)}{a} \right] \geq u(c_2(\theta, r(a))) - \Psi \left[\frac{y(\theta, r(a))}{a} \right] \quad \forall a, r(a).$$

Define the associated indirect utility function of the agents as:

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

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

$$\frac{\partial v_2(\theta, a)}{\partial a} = \Psi' \left(\frac{y(\theta, a)}{a} \right) \frac{y(\theta, a)}{a^2}, \quad (4)$$

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

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

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:

$$\begin{aligned} U(\theta, \theta) &= u(c_1(\theta)) + \beta \int_{\underline{a}}^{\bar{a}} v_2(\theta, a) dG(a|z(\theta), \theta) \\ &\geq u(c_1(r(\theta))) + \beta \int_{\underline{a}}^{\bar{a}} v_2(r(\theta), a) dG(a|z(r(\theta)), \theta) = U(\theta, r(\theta)), \quad \forall \theta, r(\theta), \end{aligned} \quad (5)$$

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

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

We proceed by replacing the set of inequality constraints defined by (5) by local envelope conditions analogous to the ones for the second period:

$$\frac{dV(\theta)}{d\theta} = \beta \int_{\underline{a}}^{\bar{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial \theta} da. \quad (6)$$

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 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 characterizing necessary conditions for local and then global incentive compatibility. The first-order condition for an optimal report, evaluated at the revealing strategy, must obey:

$$\frac{\partial U(\theta, \theta)}{\partial r(\theta)} = 0. \quad (7)$$

For a local maximum it is necessary that:

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

Differentiating equation (7) yields:

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

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

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

We summarize these findings in the following Lemma:

Lemma 3.1. *An allocation is incentive compatible only if:*

- (i) $\frac{dV(\theta)}{d\theta} = \beta \int_{\underline{a}}^{\bar{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial \theta} da$
- (ii) $\frac{\partial^2 U(\theta, \theta)}{\partial r(\theta) \partial \theta} = \int_{\underline{a}}^{\bar{a}} \frac{\partial v_2(\theta, a)}{\partial \theta} \frac{\partial g(a|z(\theta), \theta)}{\partial \theta} da + \frac{\partial z(\theta)}{\partial \theta} \int_{\underline{a}}^{\bar{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial z \partial \theta} da \geq 0,$
- (iii) $\frac{\partial v_2(\theta, a)}{\partial a} = \Psi' \left(\frac{y(\theta, a)}{a} \right) \frac{y(\theta, a)}{a^2},$
- (iv) $\frac{\partial y(\theta, a)}{\partial a} \geq 0.$

This Lemma provides necessary or localized conditions for any incentive compatible allocation. Its first part (i) can be conveniently included into any Lagrangian or optimal control problem. Part (ii) can be interpreted as follows: $\frac{\partial^2 U(\theta, \theta)}{\partial r(\theta) \partial \theta} \geq 0$ holds if the initial type θ and the report r are complements. The expression can be decomposed into two terms. The first term captures the idea that promised utilities should

be tailored to different distributions of different initial types; roughly spoken higher initial types should have relatively higher promised utilities in better states. The second term enters this condition due to the endogenous investment $z(\theta)$ and captures the idea that individuals with higher returns to education should receive a higher level of education. Parts (iii) and (iv) are the well-known envelope and monotonicity condition from the static Mirrlees problem, holding the innate type θ constant. Like in the standard problem parts (iii) and (iv) are necessary *and* sufficient, which can be shown by a standard proof.

As often done in screening problems, our strategy for solving the second-best problem is to work with a relaxed problem with only restrictions (4) and (6) being imposed and then check ex-post whether incentive compatibility is fulfilled. In the numerical explorations in Section 3.4 we find that incentive compatibility is always satisfied and therefore the first-order approach is valid for the considered primitives.

We now present a result that is interesting especially from a theoretical point of view.

Lemma 3.2. *Suppose Assumptions 2 and 3 hold, and conditions (i), (iii), (iv) of Lemma 3.1 are satisfied and further we have:*

$$(i) \frac{\partial y(\theta, a)}{\partial \theta} > 0,$$

$$(ii) \frac{\partial z(\theta)}{\partial \theta} > 0,$$

then the considered allocation is incentive compatible.

Proof. See Appendix A.2. □

This Lemma implies that instead of directly ex-post verifying whether period one incentive compatibility is satisfied in an allocation, one can alternatively check these two simple monotonicity conditions. If they are fulfilled, then the allocation is incentive compatible. However, for both to be fulfilled is *sufficient* and not *necessary*. Whereas condition (ii) was always fulfilled in our numerical examples, condition (i) often was violated for low a ; we will comment on the reasons in Section 3.4 when we present numerical illustrations of the model.

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 our case the distribution of types tomorrow is endogenous, since education is a choice. In recent contributions, Kapicka (2010) as well as Pavan, Segal and Toikka (2011) investigate the robustness and validity of the Mirrleesian first-order approach in a large class of general dynamic environments.

3.2 Characterization

The planner maximizes

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

subject to (4), (6) and the resource constraint:

$$\int_{\underline{\theta}}^{\bar{\theta}} \left[c_1(\theta) - z(\theta) + \frac{1}{R} \int_{\underline{a}}^{\bar{a}} (c_2(\theta, a) - y(\theta, a)) dG(a|z(\theta), \theta) \right] dF(\theta) = 0$$

where R is the gross return on savings. We let the planner assign Pareto weights $\tilde{F}(\theta)$ to individuals, depending (solely) on their initial skill level. Any distribution of these weights, which we normalize to satisfy $\int_{\underline{\theta}}^{\bar{\theta}} \tilde{f}(\theta) d\theta = 1$, corresponds to one point on the Pareto frontier. λ_R denotes the multipliers on the resource constraint and $\eta(\theta)$ the multiplier function of the first-period envelope conditions. In Appendix A.3 the Lagrangian and the first-order conditions of the problem are stated. We now characterize the wedges of second-best allocations.

3.2.1 Savings Distortions

It turns 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 in dynamic Mirrleesian models.¹³ Some manipulations of the first-order conditions yield the following proposition:

Proposition 3.3. *In any constrained Pareto optimum, the inverse Euler equation holds:*

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

Proof. See Appendix A.4.1. □

Jensen's inequality then implies $\beta E[u'(c_2(\theta, a))] > u'(c_1(\theta))$ – the optimal allocation dictates a wedge between the intertemporal rate of substitution and transformation; savings are discouraged.

¹³Diamond 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.

3.2.2 Labor Distortions

The following proposition characterizes the optimal labor wedge.¹⁴

Proposition 3.4. *At any constrained Pareto optimum, labor wedges satisfy:*

$$\frac{\tau_y(\theta, a)}{1 - \tau_y(\theta, a)} = \frac{1 + \varepsilon^u(\theta, a)}{\varepsilon^c(\theta, a)} \frac{u'(c_2(\theta, a))}{ag(a|z(\theta), \theta)} [\mathcal{A}(\theta, a) + \mathcal{B}(\theta, a)],$$

where

$$\begin{aligned} \mathcal{A}(\theta, a) = & G(a|z(\theta), \theta) \left[\int_a^{\bar{a}} \frac{1}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) \right. \\ & \left. - \frac{1 - G(a|z(\theta), \theta)}{G(a|z(\theta), \theta)} \int_a^a \frac{1}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) \right] \end{aligned}$$

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

where $\varepsilon^u(\theta, a)$ ($\varepsilon^c(\theta, a)$) is the uncompensated (compensated) labor supply elasticity of type (θ, a) .

Proof. See Appendix A.4.2. □

Elasticities play a double role for the optimal labor wedge. On the one hand, a higher *compensated* elasticity increases the excess burden of labor distortions and is therefore inversely related to optimal labor wedges; on the other hand, a higher *uncompensated* elasticity translates into higher income inequality for a given skill distribution, making insurance more valuable and therefore tends to increase optimal labor wedges. 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 reflecting a deadweight loss argument. The term $u'(c_2(\theta, a))$ can be interpreted as capturing income effects— for individuals with low consumption income effects are stronger and therefore the disincentive effect of marginal tax rates is weakened.

Conceptually, the labor wedge consists of two parts. The first one $\mathcal{A}(\theta, a)$ is a variation of the optimal tax formula in the static Mirrlees case, with the difference that for each initial type θ there is one separate tax function $\mathcal{A}(\theta, \cdot)$. $\mathcal{A}(\theta, a)$ disappears, if agents are risk neutral, and therefore second period insurance is not a concern. With risk-aversion, however, optimal policies provide insurance against the labor market

¹⁴In a recent paper Golosov, Troshkin and Tsyvinski (2010) provide formulas for dynamic optimal labor wedges with exogenous human capital, connecting them to empirical observables in the spirit of the contributions of Diamond (1998) and Saez (2001) for the static Mirrlees model.

risk agents face. Education enters through its effect on the conditional distribution of skills. The interpretation is analogous to the case with a utilitarian planner in the static model. Fixing θ , the bigger the difference in the marginal costs of providing a utility increase to agents with skill above a relative to the corresponding marginal cost for lower skilled agents, the higher the tax rate at the skill pair (θ, a) . The term is equivalent to the tax formula from the standard static Mirrlees problem with utilitarian welfare weights and, as shown in Appendix A.4.2, it can be rewritten as in Saez (2001).

The second term $\mathcal{B}(\theta, a)$ is novel and 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. Fixing a , the implicit tax rate is proportional to $\frac{\partial[1-G(a|z(\theta), \theta)]}{\partial\theta}$, which measures the change in the probability of becoming a higher type than a . Higher initial types have, education constant, a higher probability of reaching a skill level above a . For two neighboring θ the planner adjusts the labor wedge of the lower type to deter a deviation in the first-period. The increase in the implicit marginal tax rate increases average taxes for all skills $a^* \geq a$ and makes mimicking unattractive for the higher type.¹⁵

A no-distortion at the top and bottom goes through, since $\mathcal{B}(\theta, \bar{a}) = \mathcal{B}(\theta, \underline{a}) = \mathcal{A}(\theta, \bar{a}) = \mathcal{A}(\theta, \underline{a}) = 0$.

3.2.3 Education Distortions

The following proposition characterizes optimal education policies.

Proposition 3.5. *At any constrained Pareto optimum, the education wedge is given by:*

$$\begin{aligned} \tau_z(\theta) = & \frac{1}{R} \int_{\underline{a}}^{\bar{a}} (y(\theta, a) - c_2(\theta, a)) \frac{\partial g(a|z(\theta), \theta)}{\partial z(\theta)} da \\ & + \frac{\beta\eta(\theta)}{\lambda_R f(\theta)} \int_{\underline{a}}^{\bar{a}} \frac{\partial v_2(\theta, a)}{\partial a} \frac{\partial^2 G(a|z(\theta), \theta)}{\partial z(\theta) \partial \theta} da. \end{aligned}$$

Proof. See Appendix A.4.3. □

The first term captures the marginal fiscal gain of an increase in education. Integrating the first line by part gives:

$$-\frac{1}{R} \int_{\underline{a}}^{\bar{a}} \left(\frac{\partial y(\theta, a)}{\partial a} - \frac{\partial c_2(\theta, a)}{\partial a} \right) \frac{\partial G(a|z(\theta), \theta)}{\partial z(\theta)} da$$

¹⁵In principal, $\eta(\theta)$ might be negative. For redistributive preferences, however, i.e. non-increasing Pareto weights, $\eta(\theta)$ is usually positive. In our numerical simulations that are based on utilitarian Pareto weights, $\eta(\theta)$ is always positive.

By incentive compatibility $\frac{\partial y(\theta, a)}{\partial a}$ and $\frac{\partial c_2(\theta, a)}{\partial a}$ are both positive. Under Assumption 1 the derivative $\frac{\partial G(a|z(\theta), \theta)}{\partial z(\theta)}$ is negative. A higher level education always increases the expected transfer from an individual to the government. Indeed, later we present an implementation of the optimum, in which $y(\theta, a) - c_2(\theta, a)$ is equal to the labor tax an individual pays, i.e. there exists a fiscal externality. The education wedge offsets the distortion arising from the labor wedge on the education margin.¹⁶

We now turn to the second term. Under Assumption 3 the cross-derivative $\frac{\partial^2 G(a|z(\theta), \theta)}{\partial z(\theta) \partial \theta}$ is negative and $\frac{\partial v_2(\theta, a)}{\partial a}$ positive everywhere by second period incentive compatibility. Further, for redistributive preferences $\eta(\theta)$ is typically positive. Then the second part of the education wedge is negative and acts as an implicit *tax on education*. By downward distorting education, the planner relaxes binding incentive constraints and can redistribute more effectively in line with 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, for this incentive term a zero at the top and at the bottom $(\underline{\theta}, \bar{\theta})$ of the innate ability distribution holds.

3.3 Extension to T Periods

For expositional convenience, we have worked with a two period model so far. We next extend the results to a life-cycle context. We assume no further shocks to one's labor skill level after education. This is to strengthen the focus on the education and taxation link, while it also keeps the problem tractable. It is also in line with recent results from Huggett, Ventura, and Yaron (2010), who find that heterogeneity realized at the age of 23 contributes more to variability in lifetime earnings than subsequent shocks.

Let T_e be the number of years education takes place and T_w the number of years an individual works, such that $T_w + T_e = T$. During education, so for the first T_e years in the life-cycle, consumption is constant and given by $c_e(\theta)$, since there is certainty within the education period; due to the same reasoning consumption during the work period, $c_w(\theta, a)$ is also constant. Further, it can be shown that the labor supply distortions are constant over time (Werning (2007)) and therefore $y_w(\theta, a)$ is constant.

However, whereas there is no intertemporal wedge within the years of education and within one's working life, the inverse Euler equation still governs the relationship between $c_e(\theta)$ and $c_w(\theta, a)$ generating the familiar savings distortion. Assuming $\beta R = 1$, labor wedges are simple extensions of the expressions in Proposition 3.3,

¹⁶A similar relationship holds in the static model of Bovenberg and Jacobs (2005).

with $c_w(\theta, a), c_e(\theta)$ replacing $c_2(\theta, a), c_1(\theta)$. The education wedge takes into account the time horizon, which is a direct determinant of the return of education:

$$\begin{aligned} \tau_z(\theta) = & \sum_{t=T_e+1}^T \beta^{t-1} \int_{\underline{a}}^{\bar{a}} (y(\theta, a) - c_w\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial z(\theta)} da \\ & + \frac{\sum_{t=T_e+1}^T \beta^{t-1} \eta(\theta)}{\lambda_R f(\theta)} \int_{\underline{a}}^{\bar{a}} \frac{\partial v_2(\theta, a)}{\partial a} \frac{\partial^2 G(a|z(\theta), \theta)}{\partial z(\theta) \partial \theta} da, \end{aligned}$$

with $\beta R = 1$.

3.4 Numerical Illustration

This Section analyzes the derived properties numerically. We are not looking to provide a definite calibration, but rather numerically illustrate the theory here. Later in Section 5, we apply our model to the case of binary education decisions, i.e. college vs. high school using estimated counterfactual distributions from the literature, which uses panel data of labor market outcomes. We run two different sets of simulations, since they complement each other in the sense that the numerical example in this Section illustrates the optimal wedges with many ex-ante types, whereas for the simulation in Section 5 with only two types we can make use of real world data.

For the numerical illustration we set $\tilde{f}(\theta) = f(\theta)$, i.e. we solve for the utilitarian optimum. We assume log utility

$$U(c, l) = \ln(c) - \frac{(y/a)^\sigma}{\sigma},$$

as e.g. in Farhi and Werning (2011) and Weinzierl (2011a), and set $\sigma = 3$, implying a Frisch elasticity of 0.5. We assume one period to be one year. Education lasts four years and the working life lasts 45 years in our numerical illustrations. The yearly risk-free rate is assumed to be 3% ($R=1.03$) and $\beta = 1/R$.

Following common practice in the optimal taxation literature, we assume that abilities are distributed log-normally with a right side Pareto tail. We specify the location parameter of the log-normal μ distribution to be a function of θ and z :

$$\mu(\theta, z) = \mu_B + b\theta^c + dz^e + f\theta z$$

The specification implies (i) $\frac{\partial \mu(\theta, z)}{\partial z} > 0$, (ii) $\frac{\partial \mu(\theta, z)}{\partial \theta} > 0$, (iii) $\frac{\partial^2 \mu(\theta, z)}{\partial \theta \partial z} > 0$ for positive parameters b, c, d, e and f . For the log-normal distribution, Levy (1973) shows that an increase in the location parameter μ , holding the scale parameter σ constant, implies a FOSD-shift. This FOSD-property carries over to a log-normal distribution extended by a Pareto tail in our reported simulation as well as in all unreported simulations.

Thus, given that the FOSD-property carries over, (i) and (ii) imply Assumptions 1 and 2. However, note that (iii) that does not necessarily imply Assumption 3 to be fulfilled.¹⁷ We set the scale parameter to $\sigma = 0.565$ and append a Pareto tail $a^P = 42.5$, following Mankiw, Weinzierl, and Yagan (2009). We chose the ‘thickness’ parameter equal to two for the tail.¹⁸

We work with ten different initial skill levels θ , which are uniformly distributed in $[0.1, 1]$. We assume constant marginal costs of education and choose them and the parameters μ_B, e, b, c such that with an approximation of the current tax and college subsidy system in the US the model roughly replicates per-capita expenditures on college education and, since the model produces ten endogenous skill distributions, the interval of the location parameters of the log-normal parts across these distributions contains the empirical one.¹⁹

Figure 1 plots four exemplary density functions for two levels of θ and z , highlighting how a higher initial type and a higher education level translate into a right shift of the distribution and a smaller peak.

Figure 2 shows the optimal labor wedges as a function of labor income. For every initial type θ the labor wedge schedule follows a U-shaped pattern in the beginning and then convergence to constant implicit marginal tax rates for high incomes. This pattern mirrors findings from the static Mirrlees model (Saez 2001 and Diamond 1998). Holding income constant, the labor wedge is first increasing in initial type θ . The difference becomes as large as 21% points between the highest and lowest θ . The reason is that higher initial types θ have skill densities to the right of their counterparts with lower θ for two reasons: the direct effect of θ on $G(a|\theta, z)$ and the indirect through z which is increasing in θ . Consequently, especially for the beginning of the skill distributions, higher labor wedges are less distorting for higher θ types, since there is a smaller mass of agents for higher θ types.²⁰ The planner levies higher marginal tax

¹⁷Formally, $\frac{\partial^2 G}{\partial \theta \partial z} = \frac{\partial^2 G}{\partial \mu \partial z} \frac{\partial \mu}{\partial \theta} + \frac{\partial G}{\partial \mu} \frac{\partial^2 \mu}{\partial \theta \partial z}$ might be positive, if $\frac{\partial^2 G}{\partial \mu \partial z}$ is sufficiently large.

¹⁸Saez (1999) shows that the thickness tail parameter of the income distribution converges to $\frac{\gamma}{1+\zeta^u}$, where ζ^u is the limiting uncompensated elasticity of labor supply and γ the tail parameter of the skill distribution. The utility function we use implies $\zeta^u = 0$. Since the empirical literature agrees on a Pareto tail parameter of two for earnings and wages, our choice for γ is a value of two.

¹⁹Following Gallipoli, Meghir and Violante (2011) we set the labor income tax to a flat rate of 27% and a lump sum transfer of one sixth of labor income per capita. We introduce a yearly education subsidy of 24%. With $\mu_B = 2, b = 0.7, c = 1, d = 0.3, e = 0.2, f = 0.01$ the model produces average education expenditures of about 29.57% of yearly income per education year (data: around 30.0%, Gallipoli, Meghir and Violante (2011)) and a distribution of realized $\mu(\theta, z)$ given the above parameters within the range $[2.21, 2.95]$ (data: 2.76, Mankiw, Weinzierl and Yagan (2009)). As a robustness check we try out alternative parameterizations giving a smaller range of $\mu(\theta, z)$'s, giving very similar qualitative results.

²⁰Since low income distortions are strongly increasing in θ , condition (i) of Corollary 3.2 is typically not fulfilled for low a .

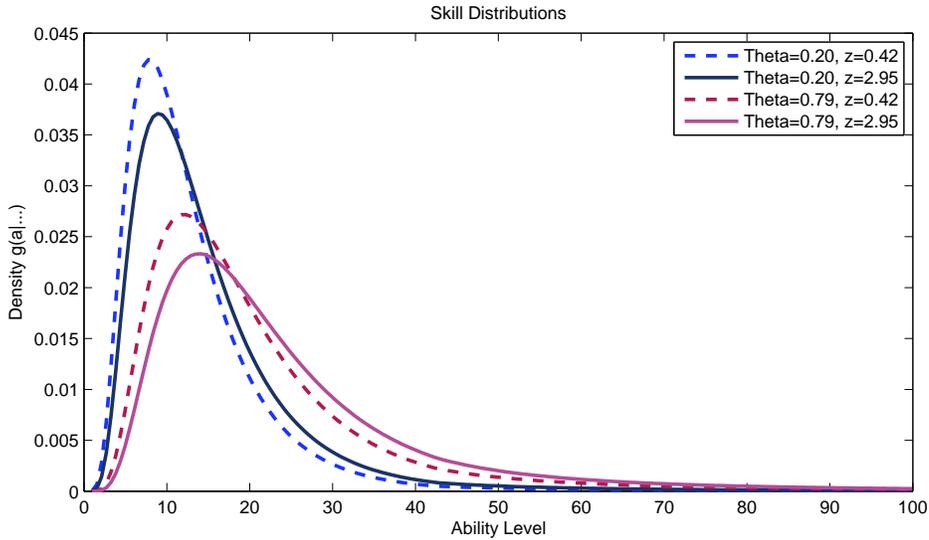


Figure 1: Skill Distributions

rates on higher θ types, raising tax revenues with less distortions. Since the ‘hump’ of the densities of a is shifted to the right for higher θ , labor wedges decrease in θ in regions where the mass is increasing in θ . For large incomes all labor wedges converge to a similar top rate because all distributions are characterized by a Pareto tail. The differences in the tails can become as big as 4% points, however.

We also distinguish between the insurance motive $\mathcal{A}(\theta, a)$ and the incentive effect $\mathcal{B}(\theta, a)$ of Proposition 3.3. We find a modest quantitative contribution of the incentive effect at all skill levels with an average of 2.56% percentage points. Figure 3 illustrates the decomposition for θ_5 , i.e. an intermediate innate ability level; we choose to illustrate the decomposition for this type because $\eta(\theta)$ is hump-shaped and therefore the dynamic incentive effect is most pronounced for intermediate θ -types. The incentive effect is particularly strong for low a ; this pattern holds for all θ -types. The reason is that $\frac{\partial G}{\partial \theta}$ is particularly strong for low a due to the properties of the log-normal distribution.

Figure 4 plots the education wedge across innate types and its decomposition into the fiscal part and the incentive part according to Proposition 3.5. The implicit subsidy to education is always positive and varies between 40% and 47%. The fiscal effect contributes quantitatively more than the incentive effect. The wedge also inherits its shape from the fiscal externality. The incentive effect is U-shaped and varies between 0 (no distortion at the top) and -2% points. Finally, we find an increasing savings wedge between 17% and 29% across initial types.

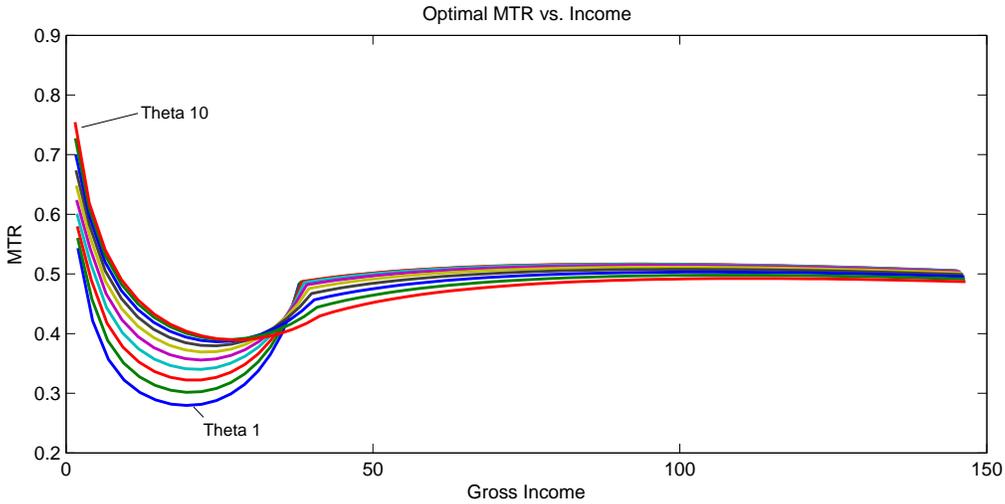


Figure 2: Labor Wedges

4 Implementation

So far we only considered a direct mechanism, in which individuals make reports about their realized type and the planner assigns bundles of consumption, labor supply and education levels as functions of the reports. The focus in the characterization of the optimal allocation was on wedges or *implicit* price distortions of the allocation. In this Section, we leave the abstract setting with no markets and explore two decentralized implementations of the constrained optima. In the first, individuals self-select into education and are offered government support conditioning on the education level chosen. When entering the labor market, income tax schedules will condition on education. The second implementation is closely related. Individuals also self-select into education; the difference is that the government support (conditional on educational choice) is not interpreted as a grant but rather as a loan. During their working life, individuals pay back an amount of the loan, which depends both, on their income and education level. In addition, there is an income tax schedule in place that conditions on income *only*. Both implementations are independent of the functioning of private credit markets since the government provides liquidity during education. We will first present the implementations for the two-period model and then generalize the results to the T period case.

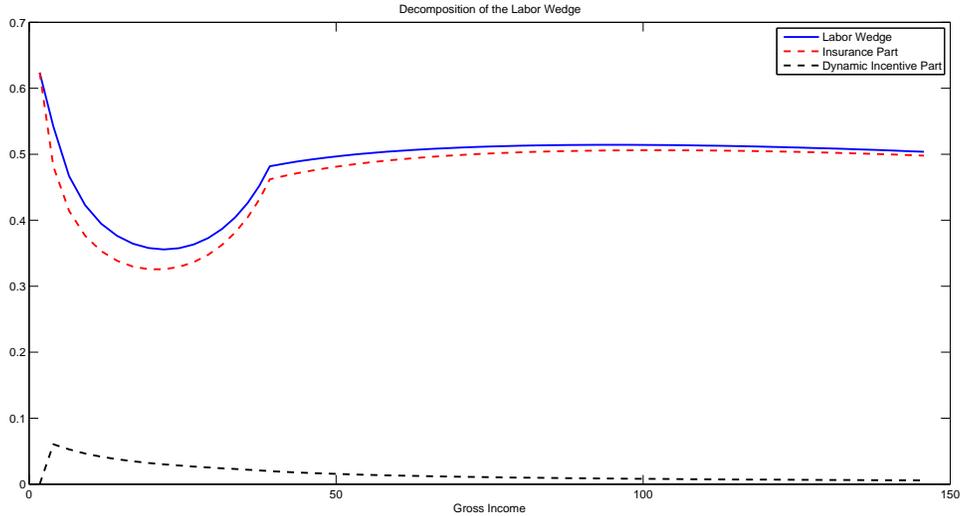
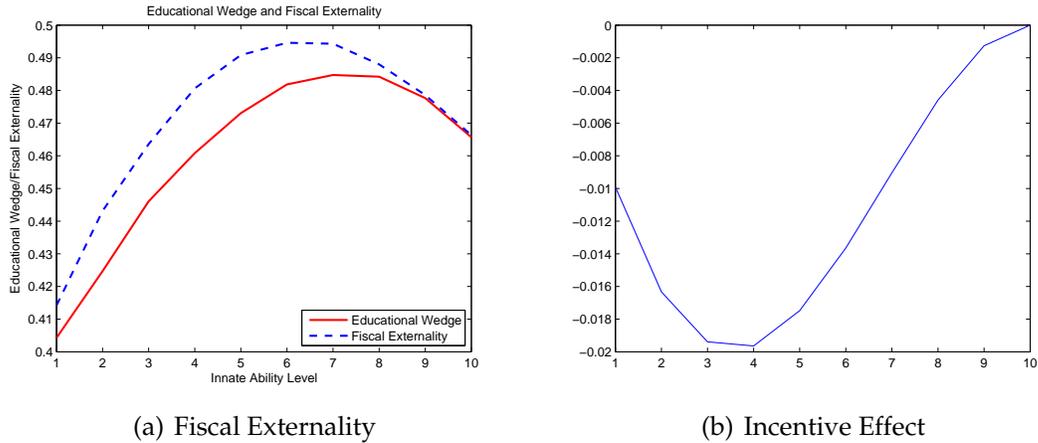


Figure 3: Decomposition of the Labor wedge



(a) Fiscal Externality

(b) Incentive Effect

Figure 4: Educational Wedge

4.1 Implementation 1: Student Grants and Income Taxes Conditioning on Education

4.1.1 Two Periods

The benevolent government, taking the role of the planner, offers a menu of student grants to the agents. These grants \mathcal{G} are conditional on education. In the second period, there is a tax schedule in place, which, importantly, does not only condition on earnings but also on educational investment. Further, savings taxes are high enough to make private savings disappear from the market; the definition of the savings tax builds on Werning (2011). We summarize this in the following proposition:

Proposition 4.1. *Any constrained Pareto optimal allocation in a two period economy can be implemented by a Grant schedule $\mathcal{G}(z)$, an education dependent income tax $T(y, z)$ and a savings tax $T^s(s)$, where*

- $\mathcal{G}(z(\theta)) = z(\theta) + c_1(\theta)$
- $T(y(\theta, a), z(\theta)) = y(\theta, a) - c_2(\theta, a)$
- $T^s(s)$ as defined in Appendix A.5.

Proof. See Appendix A.5 □

Implementation of the savings wedges: The savings function $T^s(s)$ is prohibitively high such that all agents choose $s = 0$, hence in this implementation there are no private savings. However, as shown in Werning (2011) this comes without loss of generality: by a Ricardian equivalence argument, we can adjust $\mathcal{G}(z(\theta))$ and $T(y(\theta, a), z(\theta))$ 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 (2011).

Implementation of the labor wedges: Agents enter the second period with no savings as argued above. Their budget constraint is then: $T(y(\theta, a), z(\theta)) = y(\theta, a) - c_2(\theta, a)$. From the agents' optimality conditions for y and c_2 it follows that marginal tax rates $T_y(y(\theta, a), z(\theta))$ are equal to labor wedges $\tau_y(\theta, a)$ as characterized in Section 3.2.2.

Implementation of the education wedges: To fix ideas, the budget constraints of an agent in both periods are given by:

$$\begin{aligned} c_1(\theta) + z(\theta) &\leq \mathcal{G}(z(\theta)) \\ c_2(\theta, a) &\leq y(\theta, a) - T(z(\theta), y(\theta, a)), \end{aligned}$$

where we already imposed the zero savings. 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 non-linear grant schedule $\mathcal{G}(z)$, which depends on education chosen and 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:

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

A positive value of $\tau_z(\theta)$ encourages education at level θ . The incentive for agents to increase their educational attainment comes from: i) An increase in their grant measured by $\mathcal{G}'(z)^{21}$ and ii) an increase or decrease in their labor income tax burden for all states, i.e. $T_z(y(\theta, a), z(\theta))$.

4.1.2 T Periods

The possible extension to the life-cycle case is summarized as follows:

Corollary 4.2. *Any constrained Pareto optimal allocation in a T -period economy can be implemented by a Grant schedule $\mathcal{G}(z)$, an education dependent income tax $T_t(y^t, z)$ that conditions on the history of incomes and a savings tax $T^s(s)$.*

Proof. See Appendix A.6 □

The corollary is an application of insights from Werning (2007), who characterizes possible implementations in the dynamic deterministic Mirrlees problem. The problem here is similar to his model with the only difference that there is not only one tax schedule conditioning on the history of incomes y^t , but rather one for each θ -type or each education level z .

One might wonder why the optimum is not simply implementable with time independent schedules. The reason is that with such a tax system it might be possible for an individual to profit from tax arbitrage, if tax schedules are sufficiently concave. Instead of earning the assigned income y every period, it can then be favorable for an individual to earn $y - \epsilon$ today and $y + \epsilon$ tomorrow. Average gross income would be the same, however, due to concavity of $T(y)$ the tax burden would decrease; if this effect is strong enough, it might compensate for the higher disutility of labor. In reality such strategic behavior to exploit the non-linearity of the tax system seems unlikely to occur. The reason is that shifting labor income between periods in such a way is often very costly or simply infeasible due to adjustment costs and hours restrictions.²² Formally, let $\tilde{C}(y^*)$ denote the present value of total adjustment costs of an individual that chooses income history y^* , and if $y_{T_e+1}^* = y_{T_e+2}^* = \dots = y_T^*$, then $\tilde{C}(y^*) = 0$, so if labor supply is constant, there are no adjustment costs.

Definition For a given adjustment cost function $\tilde{C}(\cdot)$, an income tax schedule $T(y, z)$ is *tax arbitrage resistant*, if $V(y^*, c^*(y^*, T(y, z(\theta))), \tilde{C}(y^*)) \leq V(y^{truth}(\theta, a), c^{truth}(\theta, a)) \forall y^*$ and $\forall (\theta, a)$, where $c^*(y^*, T(y), \tilde{C}(y^*))$ is the optimal consumption sequence given y^* ,

²¹Theoretically it could be the case that \mathcal{G} is (partly) decreasing in z if $c_1(\theta)$ is sufficiently decreasing. However, this is rather unlikely and in all our numerical examples we have $c'_1(\theta) > 0$.

²²For a recent treatment of hours constraint in the public economics literature see Chetty, Friedman, Olsen and Pistaferri (2011); for the implications of adjustment costs on hours choices and the labor supply elasticity see Chetty (2011).

$T(y)$ and $V(\cdot, \cdot)$ is the respective (deterministic) working life utility conditional on the realization of a .

With this in mind, we make the following assumption stating that strategic deviations are never feasible for a worker because of adjustment costs:

Assumption 4: The adjustment cost function $\tilde{C}(\cdot)$ is such that the income tax schedules $T(y, z)$ with $T(y(\theta, a), z(\theta)) = y(\theta, a) - c_w(\theta, a)$ are tax arbitrage resistant.

This allows to implement constrained efficient allocation using a labor tax system that conditions on current income and one's education level only, summarized in the following corollary:

Corollary 4.3. *Assume that Assumption 4 holds. Then any constrained Pareto optimal allocation in a T -period economy can be implemented by a grant schedule $\mathcal{G}(z)$, an education dependent income tax $T(y, z)$ with $T(y(\theta, a), z(\theta)) = y(\theta, a) - c_w(\theta, a)$ and a savings tax $T^s(s)$.*

Proof. By Assumption 4 we know that individuals prefer $y^T(\theta, a)$ over any income sequence where y is not constant over time. By incentive compatibility we know that each individual prefers $y^T(\theta, a)$ over $y^T(\theta, a') \forall a', a$; combining this with the insights of Proposition 4.1 yields the result. \square

4.2 Implementation Two: Student Loans with Income-Contingent Repayment

4.2.1 Two Periods

The previous implementation required that people with the same income but different levels of education pay different taxes. In reality people might perceive this as a violation of horizontal equity concerns, which could hinder the political feasibility of such policies. In this light we now present a more appealing alternative implementation with only one labor income tax schedule and a repayment scheme of the education grant.²³ Technically, this can be seen as a simple reinterpretation of the previous implementation – we take the tax system of the $\underline{\theta}$ as the *common* labor income tax schedule and introduce an income-contingent repayment schedule, which conditions on the size of the loan.²⁴ Together both instruments are sufficient to replicate the optimal labor wedges. Formally we summarize this in the following proposition:

²³Diamond and Saez (2011) argue in a recent contribution that practical policy prescription from optimal tax models should not go against common hold normative views (horizontal equity for example) and limit complexity to a reasonable degree. The second implementation seems in line with these recommendations.

²⁴Related implementations are of course possible where the tax function of another θ -type can be the labor income tax schedule in place. In fact, somehow anything goes. The extreme case would just

Proposition 4.4. *A constrained Pareto optimal allocation in a two-period economy can be implemented by a Loan schedule $L(z)$, a Loan Repayment Schedule $\Gamma(y, L)$, an income tax $T(y)$ and a savings tax $T^s(s)$, where*

- $L(z(\theta)) = z(\theta) + c_1(\theta)$
- $\Gamma(y(\theta, a), L(z(\theta))) = c_2(\underline{\theta}, \tilde{a}(\underline{\theta}, y(\theta, a))) - c_2(\theta, a)$ if $y \in [y(\underline{\theta}, \underline{a}), y(\underline{\theta}, \bar{a})]$ and $\Gamma(y(\theta, a), L(z(\theta))) = y(\theta, a) - c_2(\theta, a)$ otherwise.
- $T(y) = y - c_2(\underline{\theta}, \tilde{a}(\underline{\theta}, y)) \forall y \in [y(\underline{\theta}, \underline{a}), y(\underline{\theta}, \bar{a})]$ and $T = 0$ otherwise.
- $T^s(s)$ as defined in Appendix A.5.

where $\tilde{a}(\theta, y)$ is the inverse of $y(\theta, \cdot)$ for a .

Proof. The budget constraint of an individual reads as:

$$\begin{aligned} c_1(\theta) + z(\theta) &\leq L(z(\theta)) \\ c_2(\theta, a) &\leq y(\theta, a) - T(y(\theta, a)) - \Gamma(y(\theta, a), L(z(\theta))), \end{aligned}$$

which is equivalent to the budget constraint in Implementation 1 since by definition $\mathcal{G}(z) = L(z) \forall z$ and $T(y, z) = T(y) + \Gamma(y, z) \forall y, z$. Hence it is a direct consequence of Proposition 4.1. \square

4.2.2 T Periods

We will now show to what extent these results carry over to the case of T -periods.

Corollary 4.5. *Any constrained Pareto optimal allocation in a T -period economy can be implemented by a Loan schedule $L(z)$, a Loan Repayment Schedule $\Gamma_t(y^t, L)$, an income tax $T_t(y^t)$ and a savings tax $T^s(s)$.*

In the same way as Corollary 4.2 follows from Proposition 4.1, this Corollary follows from Proposition 4.4, we therefore omit a formal proof. Now we show that using Assumption 4 the dependence of these policy instruments on the histories of incomes can be overcome.

Corollary 4.6. *Assume that Assumption 4 holds. Then any constrained Pareto optimal allocation in a T -period economy can be implemented by a Loan schedule $L(z)$, a Loan Repayment Schedule $\Gamma(y, L)$, that is constant over time, an income tax $T(y)$ that is constant over time and a savings tax $T^s(s)$.*

be to say that income taxes do not exist and all schedules that were interpreted as history dependent labor income schedules in implementation 1 can now be interpreted as repayment schedules. Taking the labor income tax schedule of the $\underline{\theta}$ -type, however, seems to be more natural in our view. Especially in our application of the theory in Section 5.

Again we omit a formal proof as the result follows from Corollary 4.5 in the same way as Corollary 4.3 follows from Corollary 4.2. The implementation is appealing, since there is only one labor tax schedule and a yearly income contingent repayment plan for each loan size.

5 An Application of the Model: College vs. High-School

We now present an empirically driven application of our model. We limit education to be a binary instead of a continuous choice. Agents either enter the labor market directly after high-school graduation or go to college before working. Additionally, we restrict the analysis to two different levels of innate ability.²⁵ These simplifications enable us to parameterize the model using factual and, importantly, estimated counterfactual earnings distributions from the empirical labor literature (Cunha and Heckman 2007, 2008). Further the simplification has the advantage that it is easy to incorporate foregone earnings as an implicit cost of education. As laid out in Proposition 3.4, these counterfactual terms contribute to optimal marginal tax rates by deterring deviations. The simplified model, arguably, also comes a step closer to more practical policy recommendations, since the optimum can be implemented with considerably less complexity; one non-linear labor income tax schedule and one non-linear repayment schedule for college loans will suffice in the application.

5.1 Parametrization

Agents live for 49 years after they graduate from high-school (age 16-65). Afterwards they enter the labor market directly, or decide to go to college and graduate in four years. We label the two innate types θ_{HS} and θ_{CO} and assume that it is *a priori* optimal that θ_{HS} chooses the lower educational attainment (high school) and that θ_{CO} chooses the higher educational attainment (college). We solve for the utilitarian optimum. The (binding) incentive constraint for type θ_{CO} , hence, reads as:

$$\begin{aligned} & \sum_{t=0}^3 \beta^t u(c_1(\theta_{CO})) + \sum_{t=4}^{48} \beta^t \int_{\underline{a}}^{\bar{a}} v_2(a, \theta_{CO}) g(a|CO, \theta_{CO}) da \\ &= \sum_{t=0}^{48} \beta^t \int_{\underline{a}}^{\bar{a}} v_2(a, \theta_{HS}) g(a|HS, \theta_{CO}), \end{aligned} \tag{8}$$

²⁵This strong assumption implies that all individuals with the same education level are ex-ante equal, which is clearly not realistic. It can be seen as an indispensable approximation, which we need to make in order to identify innate types from real world data.

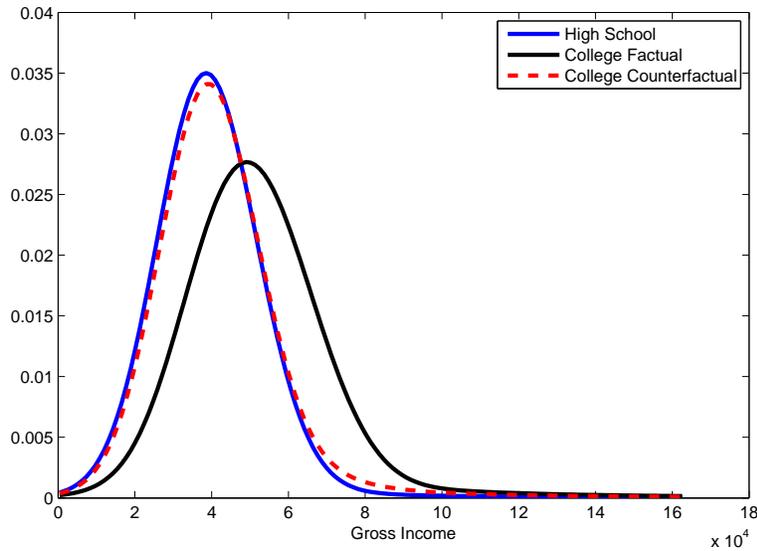


Figure 5: Earnings Distributions

where $g(a|CO, \theta_{CO})$ is the factual ability distribution of college graduates and $g(a|HS, \theta_{CO})$ is the counterfactual ability distribution; i.e. the distribution that would result if they entered the labor market directly. Note that college types now have to be compensated for their foregone labor earnings, the implicit cost of college education. We have abstracted from this in the full model for simplicity in Sections 2-4, but add it here for realism. To get the ability distributions, we take the factual and counterfactual earnings distributions for high-school graduates plotted in Cunha and Heckman (2007) in Figures 1 and 2.²⁶ Since they consider the present value of lifetime earnings (18-65), we take a 47 years annuity with the same present value, i.e. we take something similar to average annual earnings. After using a Kernel smoother (bandwidth=5000), we append a Pareto tail at earnings of \$86,000 and truncate the earnings distributions at \$350,000. The resulting relevant distributions are illustrated (for earnings up to \$150,000) in Figure 5.²⁷ Given a (linear) approximation of the real world tax system we calibrate the implied skill distributions as input for the model from the optimality conditions of the agents.²⁸ The share of high school and college types are set to 64.19% and 35.81%, respectively, as reported in Cunha and Heckman

²⁶We used the software GetData Graph Digitizer to read out the data from the graphs.

²⁷The counterfactual earnings distribution for high school graduates is not represented since it is only relevant for constrained Pareto optimal policies if the incentive constraint of type θ_{HS} is binding, which is only true if the planner wants to redistribute in favor of the more able types θ_{CO} ; i.e. not in the here considered utilitarian case.

²⁸Saez (2001) has pioneered the approach to calibrate skills distributions from actual income distributions. We employ the same approximation as in the calibration of the full model in Section 3.4. This has been found to give a good fit to the US tax system by Jaquet, Lehmann, and Von der Linden (2011).

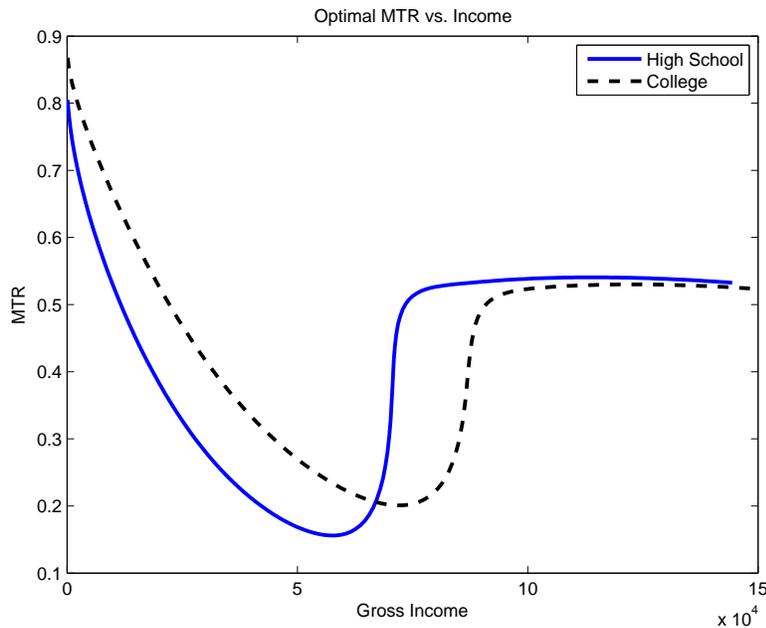


Figure 6: Optimal Marginal Tax Rates

(2008). Following Gallipoli, Meghir and Violante (2011) we set the annual monetary cost of college education to \$7,900.²⁹ The flow utility function, the yearly discount factor and the interest rate take the same values as in Section 3.4.

5.2 Results

5.2.1 Two Income Tax Schedules

Figure 6 shows the optimal implicit marginal tax rates as a function of yearly income up to \$150,000 for high school and college graduates. By Corollary 4.3 there exists an implementation, in which these schedules are equal to explicit labor income tax schedules. In combination with Figure 5, one can see in which way tax distortions are tailored to the different skill/earnings distributions. Since the hump of skill distribution is shifted to the right for college graduates, labor wedges are shifted to the right, too. Further, marginal labor distortions are very high for low income earners with college degree since the mass of those is very small compared to the mass of those with college degree and high earnings. At income levels of around \$100,000 both systems converge to roughly the same top marginal tax rate of 53%. The incentive effect (part \mathcal{B} of the optimal tax formula in Proposition 3.4) contributes very little to the optimal

²⁹According to 'Education Digest' and 'Student Financing of Undergraduate Education:1999-2000', published by the National Centre for Education Statistics (NCES), the average yearly college tuition was 7.900 US-\$ in the year 2000.

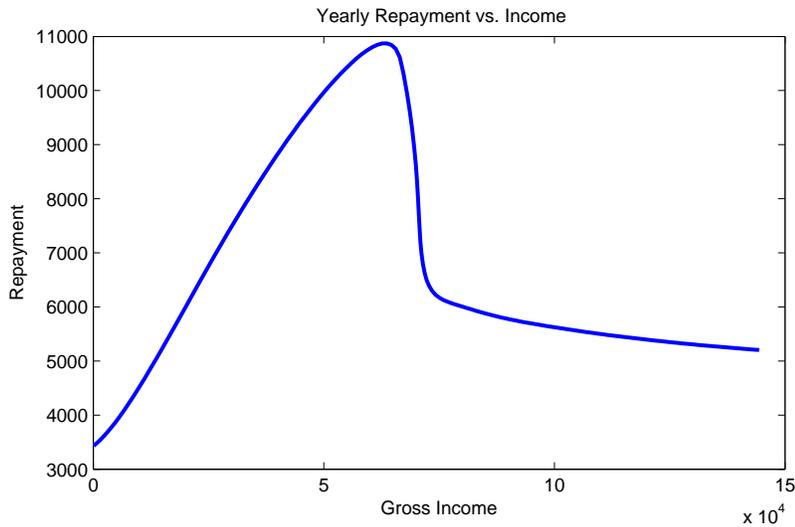


Figure 7: Yearly Repayment

high school tax schedule. Ignoring it would yield optimal tax rates, which differ by 0.17% percentage points on average (weighted by income and density). This is consistent with Figure 5 showing that college types only have a slight distributional advantage over high school types, conditional on entering the labor market directly after high school. So, although the model requires counterfactual distribution to include the influence of dynamic incentive effects deterring deviations, these effects seem to have a negligible impact on optimal marginal tax rates. This suggests that even without estimates of distributions for counterfactual education choices, meaningful tax simulations can be conducted.

5.2.2 One Income Tax Schedule and Repayments for College Graduates

By Corollary 4.6 one can interpret the marginal tax rates of the θ_{HS} -type(s) in Figure 6 as marginal tax rates of an income tax schedule common to all individuals. The marginal tax rates for the θ_{CO} -type in Figure 6 can, instead, also be interpreted as the aggregate of marginal distortion through the income tax and an income-dependent loan repayment schedule. The loan repayment schedule yielding the correct aggregate marginal distortions is illustrated in Figure 7. Repayment first increases with income and then drastically decreases before it converges.³⁰ One might wonder why the schedule is not increasing in y as an insurance device against labor market risk. The reason is that the labor income tax schedule also acts as an insurance device and actually their interaction yields optimal insurance. The common income tax schedule

³⁰It still decreases slightly in the graph, however, this is due to the truncation of the distribution made in our simulation. If the distributions were unbounded, the loan repayment would converge as long as the term $B(Y)$ in the labor wedge formula is negligible, which is the case in our simulations.

is tailored to optimally insure the θ_{HS} -type against labor market risk. The repayment schedule for the college grants corrects this schedule to fit to the skill distribution of the θ_{CO} -type(s). For very low incomes marginal distortions on labor supply should be higher for the θ_{CO} -type(s) for them being very unlikely to be that unskilled; as a consequence that repayment schedule is increasing for low incomes. At incomes of around \$60,000 marginal tax rates drastically increase due to the properties of the skill distribution of the θ_{HS} -type(s), which does not fit with the skill distribution of the θ_{CO} -types since their mass at that income level is still very high. As a consequence, the repayment schedule has to counteract the income tax schedule and therefore drastically decrease in income in this area. For higher incomes where labor distortions of the θ_{CO} -type(s) and the θ_{HS} -type(s) converge to the same value, the repayment schedule does not have to counteract the income tax schedule in any way and therefore becomes flat.³¹

5.2.3 Average Tax Burdens

We now turn to optimal average tax distortions across education levels. Working in the implementation with one tax schedule and an income contingent repayment plan, the average tax rate for college graduates reflects the average labor income tax rate plus repayment divided by gross income. Figure 8(a) illustrates these average tax rates for incomes above \$20,000. That average taxes are higher for θ_{CO} -types at every income level is not a surprise since the repayment is positive for each income level. In this context the standard concept of an average tax rate may be misleading, however, since college graduates receive transfers to finance their consumption and educational expenses when in college. A concept that better indicates the average burden a θ_{CO} -type has to bear, is a corrected average tax rate defined by $(T(y) - L_a)/y$ where L_a is the annuity payment, if individuals had to repay the loan at the market interest rate, and $T(y)$ is the labor tax paid at income y . Figure 8(b) displays these corrected average tax rates. For incomes up to roughly 45,000 US-\$ effective average tax rates are lower for college graduates; intuitively college graduates receive insurance against bad labor market outcomes by repaying only little amounts. Then for incomes up to roughly \$80,000 θ_{CO} -types pay higher corrected average tax rates. For higher incomes the pattern reverses. The graph might even suggest that college

³¹This last result hinges on the assumption that the tail of the skill distributions is characterized by the same Pareto parameter. Estimating the Pareto parameter separately for the income distribution of individuals with college degree and without might provide interesting insights for the optimal design of such repayment schedules. If, e.g., the Pareto parameter is smaller for college graduates this might call for a higher top tax rate for college graduates than for individuals with a college degree (see Saez 2001), which could be implemented as a repayment schedule that is linearly increasing for high incomes.

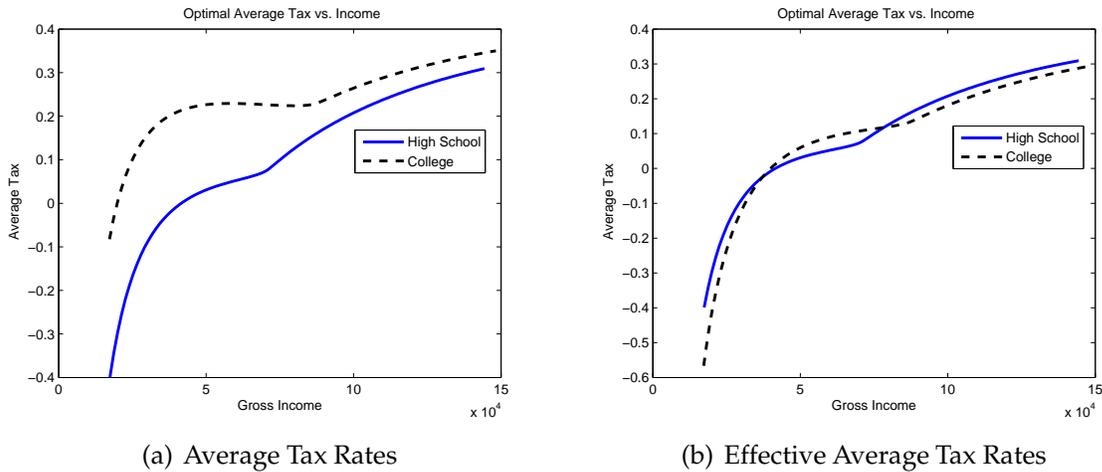


Figure 8: Average Tax Rates

types pay lower effective average tax rates in expectation, and hence repay less than the actuarially fair annuity every year on average. However, since a big mass of θ_{CO} -types earns income between \$45,000 and \$85,000 the opposite is true: The expected yearly repayment is \$563 above the annuity. The provision of loans by the government to college students, hence, serves two roles: 1) It circumvents possible credit market frictions by providing liquidity through income contingent loans. 2) On average, it increases the implicit interest on that loans to be able to redistribute from θ_{CO} to θ_{HS} -types.

5.3 Education As a Tag?

We have shown that in constrained Pareto optima labor wedges condition on education. This feature of optimal policies bears resemblance with a simple *tagging policy*, first put forward by Akerlof (1978) and more recently investigated by Cremer Gahvari, and Lozachmeur (2010), Mankiw and Weinzierl (2010) as well as Weinzierl (2011). Moreover, the simulations of optimal policies in the previous Section as well as Section 3.4 suggested that incentive effects contribute quantitatively very little to optimal labor wedges. This might suggest that self-selection into education, does not constrain the planner's problem significantly, and ignoring the incentive constraint (8) for the θ_{CO} -type and simply condition the labor tax code on education would result in similar optimal policies.

We will now argue why this is misleading. We, therefore, compute an allocation with such a simple tagging policy.³² Figure 9 illustrates the resulting effective average tax rates. Effective average taxes for college graduates are higher everywhere,

³²The only difference is that now the planner does not take (8) into account.

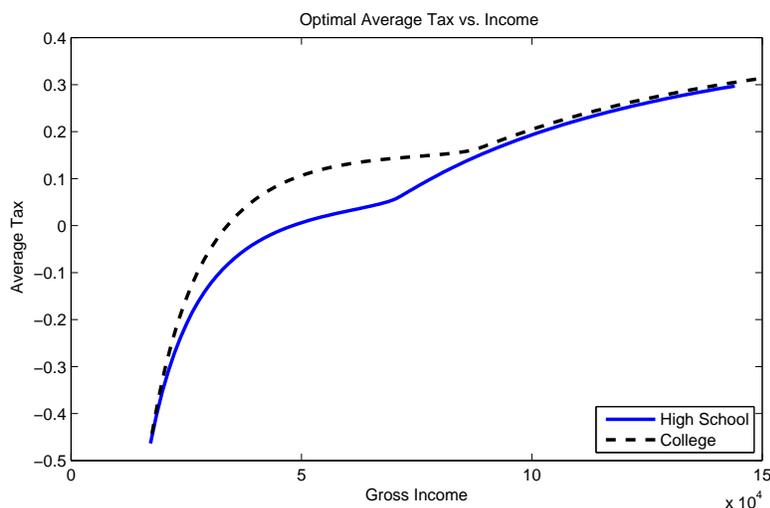


Figure 9: Effective Average Tax Rates Under Simple Tagging

indicating that a simple tag policy results in an allocation giving strong incentives for high innate types to deviate, i.e. entering the labor market directly. This becomes more stark when considering the expected yearly premium over the actuarially fair annuity, which is now \$4102 – about a sevenfold increase in comparison with the constrained optimum taking incentive effects into account. In contrast, labor wedges stay almost unaffected in comparison to the constrained optimum, consistent with the small quantitative effect of the incentive term we found in the previous Sections.³³

6 Conclusion

In this paper, we have studied the implications of endogenous education decisions on optimal tax policies in a general dynamic environment with heterogeneous agents and uncertainty. Putting no exogenous restrictions on available instruments, we were particularly interested which policies with real world appeal could implement the optimum. An attractive way to decentralize Pareto optimal allocations is to have the government support students to finance consumption and tuition during education. During their working life students pay back these loans, contingent on income and loan size. We have put the model to work illustrating our policy recommendations for the case of a binary education choice: graduating from college or entering the labor market directly after high-school. This enabled us to use empirical income distributions to parameterize the model and interpret it with respect to results from the static

³³A similar logic applies in the model of Scheuer (2011) with endogenous occupational choices. In his model, individuals choose to become an entrepreneur or worker. So although the optimal tax system in his model conditions on occupations, this is different from a pure tag because of endogenous sorting into occupations.

literature (Saez 2001). The labor wedge of college graduates as a function of income lies first strictly above their counterparts from high-school, but this reverses at higher incomes. The optimal repayment plan is hump-shaped and college graduates pay a slightly positive premium over the actuarially fair annuity in expectation.

We have abstracted from some several elements, which can be tackled in future work. For example, the planner had full commitment to honor all utility promises and deliver by history dependent distortions. Relaxing this assumption by incorporating political economy aspects could yield interesting insights. There could be an additional force to have a decreasing education subsidy; this keeps inequality in check and thereby strengthens credibility, which in turn mitigates the detrimental effects of the lack of commitment on human capital investment. This would counteract the role of increasing subsidies to offset the effect of progressive taxes, we found dominating in our numerical section (Figure 4). This idea that current policies influence future credibility has recently been put forward by Farhi, Sleet, Werning, and Yeltekin (2011) for the case of capital taxation.

Finally, our simulation have shown that counterfactual distribution have a quantitatively small contribution to optimal marginal tax rates. So even without the availability of counterfactual earnings distribution, meaningful tax simulation can be conducted for other countries than the US.

A Appendix

A.1 Properties of the Laissez-Faire Allocation

To show part (iv), we show that:

$$\frac{\partial^2 U(s, z; \theta, \cdot)}{\partial s \partial \theta} = -\beta \int_{\underline{a}}^{\bar{a}} \frac{\partial^2 v_2(a, s)}{\partial s \partial a} \frac{\partial G(a|z, \theta)}{\partial \theta} da \quad (9)$$

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

$$\frac{\partial^2 U(s, z; \theta, \cdot)}{\partial s \partial z} \quad (10)$$

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

$$\frac{\partial^2 U(s, z; \theta, \cdot)}{\partial z \partial \theta} = -\beta \int_{\underline{a}}^{\bar{a}} \frac{\partial v_2(a, s)}{\partial a} \frac{\partial^2 G(a|z, \theta)}{\partial z \partial \theta} da > 0, \quad (11)$$

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.

A.2 Proof of Lemma 3.2

Consider some admissible reporting strategy $r(\theta) = \theta'$.

$$\begin{aligned} \frac{\partial U(\theta, \theta')}{\partial r(\theta)} &= u_c(c_1(\theta')) \frac{\partial c_1(\theta')}{\partial r(\theta)} + \beta \int_{\underline{a}}^{\bar{a}} \frac{\partial v_2(\theta', a)}{\partial r(\theta)} g(a|z(\theta'), \theta) da \\ &\quad + \frac{\partial z(\theta')}{\partial r(\theta)} \int_{\underline{a}}^{\bar{a}} v_2(\theta', a) \frac{\partial g(a|z(\theta'), \theta)}{\partial z(\theta')} da \end{aligned}$$

and

$$\begin{aligned} 0 = \frac{\partial U(\theta', \theta')}{\partial r(\theta)} &= u_c(c_1(\theta')) \frac{\partial c_1(\theta')}{\partial r(\theta)} + \beta \int_{\underline{a}}^{\bar{a}} \frac{\partial v_2(\theta', a)}{\partial r(\theta)} g(a|z(\theta'), \theta') da \\ &\quad + \frac{\partial z(\theta')}{\partial r(\theta)} \int_{\underline{a}}^{\bar{a}} v_2(\theta', a) \frac{\partial g(a|z(\theta'), \theta')}{\partial z(\theta')} da \end{aligned}$$

Subtracting from one another gives:

$$\begin{aligned} \frac{\partial U(\theta, \theta')}{\partial r(\theta)} &= \beta \int_{\underline{a}}^{\bar{a}} \left[\frac{\partial v_2(\theta', a)}{\partial r(\theta)} (g(a|z(\theta'), \theta) - g(a|z(\theta'), \theta')) \right. \\ &\quad \left. + \frac{\partial z(\theta')}{\partial r(\theta)} v_2(\theta', a) \left(\frac{\partial g(a|z(\theta'), \theta)}{\partial z(\theta')} - \frac{\partial g(a|z(\theta'), \theta')}{\partial z(\theta')} \right) \right] da. \end{aligned}$$

We are now looking when this last expression always has the same sign as the difference $(\theta - \theta')$, which is clearly sufficient for global incentive compatibility. For $(\theta - \theta') > 0$, by Assumption 2, the first line is positive, if $\frac{\partial v_2(\theta', a)}{\partial r(\theta)}$ or equivalently $\frac{\partial v_2(\theta, a)}{\partial \theta}$ in a truthful mechanism is increasing in a . This can be shown to be equivalent to $\frac{\partial y(\theta, a)}{\partial \theta} > 0$ using the envelope theorem. The second line is positive if $\frac{\partial z(\theta')}{\partial r(\theta)} > 0$ or equivalently $\frac{\partial z(\theta)}{\partial \theta} > 0$ in a truthful mechanism. The last condition (iii) is a routine exercise and a proof can be found, for example in Salanié (2003).

A.3 Second Best Policies: Pareto Problem and Optimality Conditions

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

With first-order conditions:

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

$$\begin{aligned}
& (\tilde{f}(\theta) - \eta'(\theta)) \beta g(a|z(\theta), \theta) - \lambda_R \frac{1}{R} \frac{1}{u'(c_2(\theta, a))} g(a|z(\theta), \theta) f(\theta) - \mu'(\theta, a) \\
& - \beta \frac{\partial g(a|z(\theta), \theta)}{\partial \theta} \eta(\theta) = 0 \quad (v_{2SB})
\end{aligned}$$

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

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

A.4 Proofs of Propositions 3.3-3.5

A.4.1 Proposition 3.3

Combining equations (c_{1SB}) and (v_{2SB}) and integrating directly gives the result.

A.4.2 Proposition 3.4

Rewriting (y_{SB}) :

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

Dividing by $\frac{\Psi'}{a u'}$ and $\lambda_R g(a|z, \theta) f(\theta)$ and using the definition of the labor wedge, i.e. $u'(1 - \tau_y) = \Psi' \frac{1}{a}$ yields

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

which can be written as

$$\frac{\tau_y(\theta, a)}{1 - \tau(\theta, a)} = R \cdot \frac{\hat{\mu}(\theta, a)}{\lambda_R g(a|z(\theta), \theta) f(\theta) a} \frac{1 + \varepsilon_u(\theta, a)}{\varepsilon_c(\theta, a)},$$

where $\frac{\Psi' \frac{1}{a}}{\Psi'' \frac{y}{a^2} + \Psi' \frac{1}{a}} = \frac{1 + \varepsilon_u(\theta, a)}{\varepsilon_c(\theta, a)}$ can be shown by simple algebra, see Saez (2001, p.227).

In particular, with the isoelastic specification used in the computations $\frac{(y/a)^\sigma}{\sigma}$ one can verify that this term is equal to $\frac{1}{\sigma}$.

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

$$\begin{aligned}
\mu(\theta, a) &= \frac{\lambda_R f(\theta)}{u'(c_1(\theta))} \beta G(a|z(\theta), \theta) - \frac{\lambda_R}{R} f(\theta) \int_{\underline{a}}^a \frac{1}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) \\
&\quad - \beta \frac{\partial G(a|z(\theta), \theta)}{\partial \theta} \eta(\theta^*),
\end{aligned}$$

yielding:

$$\frac{\tau_y(\theta, a)}{1 - \tau(\theta, a)} = \frac{1 + \varepsilon_u(\theta, a)}{\varepsilon_c(\theta, a)} \frac{u'(c_2(\theta, a))}{ag(a|z(\theta), \theta)} [\mathcal{A}(\theta, a) + \mathcal{B}(\theta, a)]$$

where

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

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

Using the inverse Euler equation, the term $\mathcal{A}(\theta, a)$ can be written as in Proposition 3.3.

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 .

Relation to the formula of Saez (2001)

The insurance part of the labor wedge can be expressed as in Saez (2001), for our case with separable preferences. This relation applies if agents do not differ ex-ante. By some abuse of notation, then $\mathcal{B}(\theta, a) = 0$ and for $\mathcal{A}(\theta, a)$, using the inverse Euler equation, we obtain

$$\begin{aligned} \mathcal{A}(\theta, a) &= \int_{\underline{a}}^{\bar{a}} \frac{G(a|z(\theta), \theta)}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) - \int_{\underline{a}}^a \frac{1}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) \\ &= \int_{\underline{a}}^{\bar{a}} \frac{G(a|z(\theta), \theta)}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) - \int_{\underline{a}}^{\bar{a}} \frac{1}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) \\ &\quad + \int_a^{\bar{a}} \frac{1}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) \\ &= \int_a^{\bar{a}} \frac{1}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) - \int_{\underline{a}}^{\bar{a}} \frac{1 - G(a|z(\theta), \theta)}{u'(c_2(\theta, a^*))} dG(a^*|z(\theta), \theta) \end{aligned}$$

where the second equality follows from the transversality condition. This term can be expressed as in Saez (2001) as shown by Mankiw, Weinzierl and Yagan (2009) in their online appendix.

A.4.3 Proposition 3.5

Plugging c_{1SB} into z_{SB} gives:

$$\begin{aligned} & \frac{\lambda_R f(\theta)}{u'(c_1(\theta))} \beta \int_{\underline{a}}^{\bar{a}} v_2(\theta, a) \frac{\partial g(a|z(\theta), \theta)}{\partial z(\theta)} da + \frac{1}{R} \lambda_R f(\theta) \int_{\underline{a}}^{\bar{a}} \frac{\partial g(a|z(\theta), \theta)}{\partial z(\theta)} (y(\theta, a) - c_2(\theta, a)) da \\ & - \beta \eta(\theta) \int_{\underline{a}}^{\bar{a}} v_2(\theta, a) \frac{\partial^2 g(a|z(\theta), \theta)}{\partial z(\theta) \partial \theta} da - \lambda_R f(\theta) = 0 \end{aligned}$$

Proposition 3.5 directly follows.

A.5 Proof of Proposition 4.1

Starting from a direct mechanism we show in four steps that optimal allocations can indeed be implemented with the policy instruments as defined in Proposition A.5.

Step 1: Introduce savings

Imagine the constrained efficient allocation is implemented by a direct mechanism. From that point on, assume that individuals could freely save s . With savings tax $T^s(s, r_1)$, the budget constraints read as

$$\tilde{c}_1(r_1) + s = c_1(r_1) \quad (12)$$

$$\tilde{c}_2(r_1, r_2) = c_2(r_1, r_2) + Rs - T^s(s, r_1) \quad (13)$$

Define the optimal report r_2 about a , for a given report r_1 about θ , a given savings tax schedule $T^s(s, r_1)$ and a given level of savings s :

$$r_2^*(a, r_1, s, T^s) = \arg \max_{r_2} u [c_2(r_1, r_2) + Rs - T^s(s, r_1)] - \psi \left(\frac{y(r_1, r_2)}{a} \right)$$

Then the optimal report in period 1, for a given level of savings and a given savings tax schedule $T^s(s, r_1)$, is defined by

$$\begin{aligned} r_1^*(\theta, s, T^s(r_1, s)) &= \arg \max_{r_1} u(c_1(r_1) - s) \\ &+ \beta \int_{\underline{a}}^{\bar{a}} u [c_2(r_1, r_2^*) + Rs - T^s(s, r_1)] - \psi \left(\frac{y(r_1, r_2^*)}{a} \right) dG(a|z(r_1), \theta) \end{aligned} \quad (14)$$

Then define a hypothetical tax schedule $T^*(r_1, s, \theta)$ for each θ implicitly by

$$V(\theta) = V(\theta, s, r_1^*, T^*(r_1, s, \theta)) \quad \forall s.$$

This hypothetical tax schedule would make individuals of type θ indifferent between truth telling and the optimal joint deviation for any s . It is hypothetical since it does not only depend on the report r_1 , which is observable but also on the unobservable type θ . However, we know that for each θ such a tax schedule exists. Therefore taking the upper envelope over these functions, we define $\hat{T}(s, r_1)$ by

$$\hat{T}(s, r_1) = \sup_{\theta} T^*(r_1, s, \theta) \quad (15)$$

Lemma A.1. *A constrained efficient allocation can be implemented by a direct mechanism extended by a savings choice and history dependent savings tax schedules $\hat{T}^s(s, r_1)$.*

Step 2: Make the savings tax history independent

A simple way to make the savings tax history-independent is to take the upper envelope of all function $T^s(s, r_1)$, i.e.

$$T^s(s) = \sup_{r_1} \hat{T}^s(s, r_1). \quad (16)$$

Lemma A.2. *A constrained efficient allocation can be implemented by a direct mechanism extended by a savings choice and history independent savings tax schedules $T^s(s)$.*

Step 3: Allow for labor-leisure decisions

To get closer to a decentralized implementation now assume the following extended direct mechanism.

1. Individuals report $r(\theta)$
2. They get assigned 'income to consume' $c_1(\theta)$
3. They face the savings tax schedule $T^s(s)$ and save $s(\theta) = 0$
4. In period two, instead of directly revealing their type, individuals of type θ face an income tax schedule that is defined by

$$T(y(\theta, a), \theta) = y(\theta, a) - c_2(\theta, a) \quad \forall a$$

By the same arguments as in the standard Mirrlees model it follows that this extended direct mechanism can also implement the constrained efficient allocations.

We can summarize this in the following lemma.

Lemma A.3. *A constrained efficient allocation can be implemented by a direct mechanism in the first period extended by a savings choice and a history independent differentiable savings tax schedule $\tilde{T}^s(s)$ and a history dependent labor income tax schedule $T(Y, z)$ in period two.*

Step 4: Complete Decentralization - allow for educational investment

1. Individuals buy (or tell the government that they want to buy) $z(\theta)$ units of education
2. They get assigned a student loan $\mathcal{G}(z(\theta)) = c_1(\theta) + z(\theta)$ (and are obliged to actually buy $z(\theta)$ units of education)
3. They face the savings tax schedule $T^s(s)$ and save $s(\theta) = 0$
4. in period two, instead of directly revealing their type, individuals of type θ face an income tax schedule that is defined by

$$T(y(\theta, a), z(\theta)) = y(\theta, a) - c_2(\theta, a) \forall a$$

Since the mechanism in step 4 is just a reformulation of the mechanism in step 3 this directly leads us to Proposition 4.1. \square

A.6 Proof of Proposition 4.2

Once θ is revealed, it is a direct consequence of Werning (2007, Appendix IV) that there exists a tax system $T_t(y^t)$ for every θ that leads to truth telling of a conditional on θ being revealed. It then directly follows that a capital tax as defined in steps 2 of Appendix A.5 combined with a grant schedule $\mathcal{G}(z)$ with $\mathcal{G}(z(\theta)) = z(\theta) + c_e(\theta) \sum_{t=0}^{T_e} \beta^t$ implements the optimum.

A.7 Proof of Proposition 4.3

By Assumption 5 we know that individuals prefer $y^T(\theta, a)$ over any income sequence where y is not constant over time. By incentive compatibility we know that each individual prefers $y^T(\theta, a)$ over $y^T(\theta, a') \forall a', a$. Then it follows that a history independent tax schedule $T(y, z)$ in every period with $T(y(\theta, a), z(\theta)) = y(\theta, a) - c_w(\theta, a)$ implements the allocation.

References

- ACEMOGLU, D., AND D. AUTOR (2010): "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics Vol. 4 (forthcoming)*.
- AKERLOF, G. (1978): "The economics of " tagging" as applied to the optimal income tax, welfare programs, and manpower planning," *The American economic review*, 68(1), 8–19.

- 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.
- ANDERBERG, D., AND F. ANDERSSON (2003): "Investments in Human Capital, Wage Uncertainty, and Public Policy," *Journal of Public Economics*, 87(7-8), 1521 – 1537.
- ATKINSON, A., AND J. STIGLITZ (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6(1-2), 55–75.
- (1980): *Lectures on public economics*. McGraw-Hill Book Co.
- BÉNABOU, R. (2002): "Tax and Education Policy in a Heterogenous-Agent Economy: What of Levels of Redistribution Maximize Growth and Efficiency?," *Econometrica*, 70(2), 481–517.
- 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*.
- CHETTY, R. (2011): "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica* (forthcoming).
- CHETTY, R., J. FRIEDMAN, T. OLSEN, AND L. PISTAFERRI (2011): "Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics*.
- COURTY, P., AND H. LI (2000): "Sequential Screening," *The Review of Economic Studies*, 67(4), 697–717.
- CREMER, H., F. GAHVARI, AND J. LOZACHMEUR (2010): "Tagging and Income Taxation: Theory and an Application," *American Economic Journal: Economic Policy*, 2(1), 31–50.
- CUNHA, F., AND J. HECKMAN (2008): "The Evolution of Labor Earnings Risk in the US Economy," *Working Paper, University of Chicago*.
- CUNHA, F., J. HECKMAN, AND S. NAVARRO (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57(2), 191.

- 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., C. SLEET, I. WERNING, AND S. YELTEKIN (2011): "Nonlinear Capital Taxation Without Commitment," *The Review of Economic Studies* (forthcoming).
- FARHI, E., AND I. WERNING (2011): "Insurance and Taxation over the Life Cycle," *Working Paper, NBER*.
- GALLIPOLI, G., C. MEGHIR, AND G. VIOLANTE (2011): "Equilibrium effects of education policies: A quantitative evaluation," *Working Paper*.
- 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 (2010): "Optimal Dynamic Taxes," *Working Paper, Yale University*.
- 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.
- JACOBS, B., D. SCHINDLER, AND H. YANG (2010): "Optimal Taxation of Risky Human Capital," *Scandinavian Journal of Economics* (forthcoming).
- KAPICKA, M. (2006): "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping," *Review of Economic Dynamics*, 9(4), 612–639.
- (2010): "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach," *Working Paper*.
- 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.
- LEVY, H. (1973): "Stochastic Dominance among Log-Normal Prospects," *International Economic Review*, 14(3), pp. 601–614.
- MANKIW, N., AND M. WEINZIERL (2010): "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, 2(1), 155–176.
- MANKIW, N., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23(4), 147–174.
- 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 (2011): "Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization, and Informational Disclosure," *Working Paper*.
- ROGERSON, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.
- SAEZ, E. (1999): "Essays on the Economics of Income Taxation," *Dissertation, Massachusetts Institute of Technology*.
- (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.
- SALANIÉ, B. (2003): *The Economics of Taxation*. The MIT Press.
- SCHEUER, F. (2011): "Entrepreneurial taxation, occupational choice, and credit market frictions," *Working Paper, Stanford University*.
- 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.
- WEINZIERL, M. (2011a): "The Surprising Power of Age-Dependent Taxes," *The Review of Economic Studies* (forthcoming).
- (2011b): "Why Do We Redistribute So Much But Tag So Little? Normative Diversity and Optimal Taxation," *Working Paper, Harvard Business School*.
- WERNING, I. (2007): "Optimal Fiscal Policy with Redistribution," *The Quarterly Journal of Economics*, 122(3), 925–967.

—— (2011): “Nonlinear Capital Taxation,” *Working Paper, MIT*.