Optimal Taxation and Education Policy with Skill-Biased Technological Change

Bas Jacobs
Erasmus University Rotterdam, Tinbergen Institute and CESifo*

Uwe Thuemmel
Erasmus University Rotterdam, VU University Amsterdam and Tinbergen Institute†

February 2, 2015
Preliminary draft - Please do not cite or circulate

Abstract

This paper studies the effect of skill-biased technological change (SBTC) on optimal linear income taxation and education subsidies in a model with heterogeneous individuals that acquire human capital on the extensive margin and supply labor on the intensive margin. Optimal income taxes trade off distributional benefits against tax distortions on both labor supply and skill formation. Human capital is optimally taxed on a net basis as taxes on ability rents from infra-marginal skilled workers are traded off against distortions on skill formation of marginal skilled workers. Distributional benefits of income taxes and net taxes on education decrease in general equilibrium, since both policies raise pre-tax wage inequality. Although education is optimally taxed on a net basis, optimal education subsidies can be positive as they reduce tax distortions on skill formation. For a special case we show analytically that SBTC induces lower optimal education subsidies if skill-bias is low. In contrast, the optimal income tax increases at low levels of skill-bias and decreases at high levels of skill-bias. Using simulations, we find that SBTC leads to higher net-taxes on skill-formation. While in all our simulations subsidies decrease with SBTC, the effect on the income tax is ambiguous. Moreover, in the presence of general equilibrium effects SBTC leads to a decline in the college wage premium, which is at odds with the data. Accounting for the rising costs of higher education resolves the puzzle. Finally, we calibrate our model to US data in 1980 and 2005 and conduct a policy experiment. We find that while optimal income taxes are similar to those in the data, optimal education subsidies should be lower and transfers should be higher – resulting in a welfare gain.
Keywords: Human capital; General equilibrium; Optimal taxation; Education subsidies, Technological Change.
JEL-Codes: H2; H5; I2; J2; O3.

*Address: Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Phone: +31 10 408 1441. Fax: +31 10 4089166. E-mail: bjacobs@ese.eur.nl. Homepage: http://people.few.eur.nl/bjacobs.

†Address: Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Phone: +31 10 408 1441. Fax: +31 10 4089166. E-mail: u.thuemmel@tinbergen.nl.

1 Introduction

Skill-biased technological change (SBTC) is seen as one of the main drivers of widening earnings differentials all across the world (OECD, 2011). Dutch Nobel-prize winner Jan Tinbergen (1975) called this the race between education and technology.[1] Currently, this race is lost by education as the relative demand for skilled workers outstrips relative supply of skilled workers in many countries, most notably the US (see e.g. Goldin and Katz, 2010; Acemoglu, Gancia, and Zilibotti, 2011). How should redistributive governments respond to SBTC when designing their income tax and education policies? Following Tinbergen (1975), many authors have suggested to raise investment in human capital so as to let education win the race with technological change (see e.g. Teulings (2000)). However, this policy is controversial since education subsidies are regressive. The high-skilled are the ones benefitting from both, subsidies and skill bias in technology. Others have argued to make tax systems more progressive in response to widening income differentials, see for example Piketty (2014).

Although highly policy-relevant, the answer to the question "How should governments respond to SBTC in setting tax and education policies?"is not known. This paper aims to provide a first pass at the answering. To that end we develop a simple model of optimal linear income taxation and education subsidies, based on Mirrlees (1971); Sheshinski (1972); Dixit and Sandmo (1977); Bovenberg and Jacobs (2005). Individuals are heterogeneous in their ability, which determines their productivity per hour worked. Individuals choose on the extensive margin whether to become high-skilled or remain low-skilled. Only individuals with a sufficiently high ability will invest resources and become high-skilled, otherwise they remain low-skilled. Given their skill decision, individuals supply labor on the intensive margin. In contrast to Mirrlees (1971) and Bovenberg and Jacobs (2005), wage rates respond endogenously to the aggregate supplies of low- and high-skilled labor. We derive rules for optimal tax and education policies and we analyze how these policies should respond to skill-biased technological change using simulations. Our paper has three main contributions.

First, we demonstrate that optimal income taxes should be positive. We confirm the standard result that income taxes are increasing in the distributional benefits of taxation and decreasing in the distortionary costs of taxation.[2] Distortions increase in the elasticity of labor supply. Income taxes should also be lower if skill formation is more heavily distorted by income taxation. Since direct costs of investment in education are not deductible from the income tax, income taxation reduces investments in human capital (see also Jacobs, 2005, Bovenberg and Jacobs 2005). In addition, income taxation causes general-equilibrium effects on the wage structure by increasing pre-tax wage differentials between high- and low-skilled workers (see also Jacobs 2012). Hence, general-equilibrium effects reduce the distributional benefits of income taxation and optimal income taxes are lower if general-equilibrium effects are more important.

Second, we demonstrate that in the policy optimum skill formation should be taxed on a net basis. That is, education subsidies are not sufficient to eliminate all distortions of income taxation on skill formation. Intuitively, the government wishes to redistribute education rents from the high-ability (i.e. infra-marginal) workers to the low-ability workers. However, net taxes on education distort the skill choices of the marginal workers that are indifferent between becoming high-skilled or remaining low-skilled. The government thus

[1] Goldin and Katz (2010) have adopted Tinbergen's notion of the "race between education and technology".

[2] See also Sheshinski (1972) and Dixit and Sandmo (1976).

trades off taxing rents in education and distorting skill formation.[3] In addition, general-equilibrium effects on wages might again off-set some of the distributional gains of net taxes on education. The skill premium increases when education is taxed on a net basis as the supply of high-skilled workers diminishes relative to the supply of low-skilled workers. Although education is taxed on a net basis, optimal education subsidies are typically positive. Intuitively, education subsidies alleviate the tax distortions of income taxation on skill formation (Bovenberg and Jacobs, 2005).

Third, we analyze the effects of skill-biased technological change on the setting of optimal income taxes and education subsidies. We show theoretically that stronger skill bias has ambiguous effects on optimal taxes and optimal education policies. For the special case of linear technology and uniformly distributed ability we show that if the government can either adjust the income tax or the subsidy, but not both, SBTC leads to falling subsidies at low levels of skill bias. In contrast, the optimal income tax increases with SBTC at low levels of skill-bias and decreases at high levels of skill-bias.

In order to study the joint optimization of taxes and subsidies we use simulations. We find that SBTC leads to higher net-taxes on skill-formation. While in all our simulations subsidies decrease with SBTC, the effect on the income tax is ambiguous. Moreover, in the presence of general equilibrium effects SBTC leads to a decline in the college wage premium, which is at odds with the data. Accounting for the rising costs of higher education resolves the puzzle. Finally, we calibrate our model to US data in 1980 and 2005 and conduct a policy experiment. We find that while optimal income taxes are similar to those in the data, optimal education subsidies should be lower and transfers should be higher – resulting in a welfare gain.

The remainder of this paper is structured as follows. In the next section we shortly review related literature. In Section 3 we set up the model. In Section 4 we derive the optimal tax and education policies. In Section 5 we study the effect of SBTC on optimal policy analytically, focusing on a special case. In Section 6 we simulate the consequences of SBTC for optimal policies. In Section 7 we calibrate our model to US data and conduct a policy experiment. Finally, Section 6 concludes.

2 Related literature

Our research combines three strands of the literature: skill-biased technological change, optimal income taxation with general-equilibrium effects, and optimal education subsidies. We are not aware of other studies which analyze all three aspects in conjunction. However, there is a large literature which has studied each of the factors in isolation. Moreover, optimal income taxation and education subsidies have been studied together, and there is recent work on skill-biased technological change and optimal income taxation.

Skill-biased technological change The idea that skill-biased technological change is responsible for increasing inequality can be traced back to Tinbergen (1975) who wrote about the "race between education and technology". Katz and Murphy (1992) apply this supply-demand framework and find a strong relationship between the college wage premium and the rate of growth of the supply of college graduates in the US.[4] Autor,

[3]This contrasts with Bovenberg and Jacobs (2005) as they consider human capital investment on the intensive margin.

[4]Card and DiNardo (2002) claim that the theory of SBTC is not able to explain the stabilization in wage inequality in the US in the 1990s and the rise in inequality along other dimensions, such as gender and race. To explain these findings Autor, Katz, and Kearney (2008) suggest a more nuanced version of

Katz, and Kearney (2008) and Goldin and Katz (2010) also argue that the canonical model of SBTC does a good job in explaining the behavior of the relative supply of skills and the college wage premium over four decades. In a review of the book by Goldin and Katz, Acemoglu and Autor (2012) highlight the strengths of the canonical model, but also suggest a more nuanced version of SBTC which is developed in more detail in Acemoglu, Gancia, and Zilibotti (2011). Given its empirical success this paper builds upon the canonical model of SBTC by Katz and Murphy (1992).. Although this simple model misses some of the more realistic features of the task-based model by Acemoglu, Gancia, and Zilibotti (2011), it allows us to study the fundamental mechanisms of SBTC and its effect on optimal policy in a transparent way.

Skill-biased technological change and optimal income taxation How should income taxes be set when skill bias raises earnings inequality? A recent study by Ales, Bellofatto, and Wang (2014) focuses on the effect of technological change on non-linear income taxes in a task-based model of the labor market (see e.g. Acemoglu, Gancia, and Zilibotti, 2011). If technological change is such that low talents catch up in simple tasks, but fall behind in complex ones, general-equilibrium effects compress the wage distribution at the bottom and expand it at the top. The optimal policy response is then to redistribute less from middle incomes to the bottom, and more from high incomes to the middle. However, if the complementarity between talents and tasks increases, taxes become less distortionary and marginal tax rates at the bottom are optimally higher. Which of the two forces dominates is an empirical question. Ales, Bellofatto, and Wang use US data from the Current Population Survey (CPS) in the 1970s and the 2000s and argue that both cases can occur depending on the evolution of technology.

Heathcote, Storesletten, and Violante (2014) study a model that features endogenous human capital formation and imperfect substitutability of skills. They calibrate their model to the US economy and analyze the impact of skill-biased technological change on the optimal degree of tax progressivity. In the absence of general-equilibrium effects, SBTC raises tax progressivity. However, when general-equilibrium effects are present, tax progressivity remains modest, but higher than in the model without SBTC. Our model captures the same general-equilibrium effects on the setting of optimal income taxes. The result from these papers suggest that skill-bias generally raises optimal tax progression, a finding this paper confirms.

Optimal income taxation with general-equilibrium effects In the optimal-tax literature, general-equilibrium effects on the wage structure have been extensively analyzed. Allen (1982) wasone of the first to study the consequences of endogenous wages for the optimal setting of the linear income tax. He demonstrated that general-equilibrium effects affect the optimal income tax, depending on the strength of income effects in labor supply. Simulations by Stern (1976)demonstrated that these effects were typically weak, however. In our setting, income effects in labor supply are absent so general-equilibrium effects do not affect optimal taxes for that reason.

An influential and related study by Stiglitz (1982) explored a two-type model with endogenous wages to analyze optimal non-linear taxation. The optimal marginal tax rate for the high-ability type is found to be negative, whereas positive marginal taxes are levied on low-ability individuals. Intuitively, incentive-compatibility constraints are relaxed by subsidizing labor supply of high-ability individuals. By compressing wages,

SBTC in which computer technology replaces routine jobs and leads to a polarization of skill demand.

high-ability types need to work more hours to mimic a low-ability individual, which helps redistribution.[5]

Rothschild and Scheuer (2013) generalize the model by Stiglitz (1982) by combining optimal non-linear income taxation à la Mirrlees (1971) with a Roy-model (Roy, 1951) of endogenous sectoral choice. In contrast to Stiglitz (1982), types are continuous and two-dimensional, and sectoral choice is endogenous. Like Stiglitz, Rothschild and Scheuer find that general-equilibrium effects call for lower marginal tax rates at the top and higher marginal tax rates at the bottom of the income distribution. However, endogenous sectoral choice makes the effect less pronounced than in Stiglitz (1982).

Like Rothschild and Scheuer,we incorporate endogenous wages and endogenous sectoral choice, which we refer to as human capital formation. In contrast to Rothschild and Scheuer, we assume that ability is one- rather than two-dimensional. This allows us to analytically analyze the joint determination of optimal tax and education policies and to trace down the consequences of SBTC for the setting of optimal polices. Like in this entire literature, we demonstrate that the optimal tax rate and the optimal education subsidy are affected by general-equilibrium effects on wage structure. Moreover, our simulations also demonstrate that general-equilibrium effects have only little influence on optimal taxes an education subsidies. In contrast to both Stiglitz (1982) and Rothschild and Scheuer (2013) we focus on linear income taxation and leave the extension to non-linear income taxes for future research.

Optimal taxation and education subsidies How should optimal education policies be determined jointly with optimal income taxation? Bovenberg and Jacobs (2005) extend the optimal tax analysis of Mirrlees (1971) with endogenous skill formation. Investments in education determine earnings ability. Income taxation not only distorts labor supply but also the incentives to invest in human capital. Bovenberg and Jacobs demonstrate that optimal education subsidies ensure that human capital investment is optimally efficient, i.e. net taxes on education are zero. Hence, education subsidies eliminate all tax distortions on skill formation. Education taxes/subsidies are not directly used for redistribution because the government can organize the same income redistribution with the income tax without distorting human capital investment.[6] The combination of income taxes and education subsidies allows the government to tax ability rents at higher effective rates thereby reducing the distortions of redistributing income.

Jacobs (2012) extends Bovenberg and Jacobs by endogenizing wage rates. With linear instruments there is no redistributive role for education policy; their only role is to ensure efficiency in human capital formation. Intuitively, net taxes on education have the same distributional benefits as taxes on income, and this remains so in general equilibrium. Hence, the government prefers taxing income rather than education to avoid distortions in skill formation. Only when non-linear tax instruments are analyzed, optimal education policy exploits general-equilibrium effects for redistribution. As in Stiglitz (1982) the high-skilled receive a marginal education subsidy, whereas the low-skilled are marginally taxed. Simulations demonstrated that these effects are very modest, however.

A related study is Dur and Teulings (2001) who also explore optimal income taxes and

[5]This formulation assumes that all workers cannot switch between jobs. When this is allowed, general-equilibrium effects cease to affect optimal tax policy. See also Saez (2002).

[6]Jacobs and Bovenberg (2011) demonstrate that the efficiency result depends on the shape of the earnings function. Only when earnings functions are weakly separable between education – on the one hand – and ability and labor – on the other hand – should optimal education investments optimally be efficient under non-linear income taxation. The earnings function should be also homothetic in education when linear taxes are analyzed.

education subsidies in a task-based model of the labor market. They find that stimulating human capital formation might be optimal if the regressive distributional incidence of education subsidies is more than off-set by a more compressed wage distribution. This would allow the government to reduce distortionary income taxation. These effects are also present in our analysis. In contrast to the rest of the literature, these authors suggest that general-equilibrium effects in the labor market are very strong.

We extend Bovenberg and Jacobs (2005) and Jacobs (2012) by analyzing a discrete rather than continuous human capital decision, which gives rise to inframarginal rents from ability in education. As a consequence, the income tax ceases to have the same distributional benefits as a net tax on education. And we demonstrate that education should optimally be taxed on a net basis. Discrete education choices thus break the efficiency results of Bovenberg and Jacobs (2005). This is related to Lozachmeur, Pavan, and Gomes (2014) who consider optimal non-linear income taxes and endogenous sectoral choice. Sector-specific transfers – which are part of the non-linear income tax schedule – can be interpreted as a sector-specific subsidy. They also demonstrate that sectoral choice is optimally distorted for redistribution. Their work differs from ours in that general-equilibrium effects are absent.

Finally, related are also Gerritsen and Jacobs (2013) and Gerritsen and Jacobs (2014) who focus on the desirability of minimum wages as a redistributive tool in addition to income taxes. These authors develop very similar models with endogenous human capital and endogenous wages. The most important difference is that wage rates in our models are not fully determined by minimum-wage policy, but are allowed to change in response to taxation and education policy.

3 Model

3.1 Individuals

There is a continuum of individuals of unit mass. Each worker is endowed with ability $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$ with $0 < \underline{\theta} < \overline{\theta} \leq \infty$, where θ is a draw from distribution $F(\theta)$ with $f(\theta)$ as the corresponding density. The level of technology is denoted by A. Depending on her ability and the level of technology, each individual decides whether to remain low-skilled or to invest resources to become high-skilled.[7] We index individuals by their – endogenous – skill level $j \in \{L, H\}$, where L stands for the low-skilled and H for the high-skilled workers. All individual choices depend on ability θ and the level of technology A. Since our model is static, we do not explicitly indicate the dependence on A. In contrast, the level of ability matters for choices at any level of A, and we therefore use a subscript θ to highlight this dependence. Moreover, we use superscripts L and H to distinguish between low- and high-skilled individuals.

Individuals have identical preferences over consumption c_θ^j and labor supply l_θ, h_θ, where l_θ (h_θ) denotes low-skilled (high-skilled) labor supply. We assume that preferences are quasi-linear in consumption and iso-elastic:

$$U_\theta^L \equiv c_\theta^L - \frac{l_\theta^{1+1/\varepsilon}}{1+1/\varepsilon}, \quad \varepsilon > 0, \tag{1}$$

[7]One may view investment in human capital as investment in higher education. Alternatively one could also interpret our model as an occupational choice model where workers optimally select the sector in which to work.

$$U_\theta^H \equiv c_\theta^H - \frac{h_\theta^{1+1/\varepsilon}}{1+1/\varepsilon}, \quad \varepsilon > 0. \tag{2}$$

Here,ε is the wage-elasticity of labor supply, which is identical for both skill groups.[8] This utility function is employed in nearly the entire optimal-tax literature, because income effects and heterogeneous labor-supply elasticities substantially complicate the analysis when general-equilibrium effects on wages are present.[9].

A low-skilled worker supplies l_θ units of effective labor, earns wage rate per efficiency unit of labor w_L, and earns gross income $z_\theta^L \equiv w_L \theta l_\theta$. Similarly, a high-skilled worker supplies θh_θ units of effective labor, earns a wage rate w_H per unit of of effective labor, and earns gross income $z_\theta^H \equiv w_H \theta h_\theta$. Investment in human capital is a discrete choice to become high-skilled. Workers that become high-skilled need to invest a fixed amount of resources $p > 0$. Consumption c is the numéraire commodity and its price is normalized to unity.

The government levies linear taxes t on labor income and provides a non-individualized lump-sum transfer b. The tax system is progressive if both t and b are positive. In addition, individuals that decide to become high-skilled receive an education subsidy sp. We express s as a fraction of p such that the net costs of skill-formation are $(1-s)\,p$. We do not restrict the education subsidy to be positive, hence we allow for the possibility that high-skilled individuals may have to pay an education tax. Low- and high-skilled workers of type θ thus face the following budget constraints:

$$c_\theta^L = (1-t)\,w_L \theta l_\theta + b, \tag{3}$$

$$c_\theta^H = (1-t)\,w_H \theta h_\theta - (1-s)\,p + b. \tag{4}$$

The informational assumptions underlying the government's policy instruments are that the government only verifies aggregate labor earnings, but can neither condition tax rates on individual earnings nor on skill choices. However, by providing subsidies dependent on skill levels we do allow for a skill-dependent tax system. These assumptions are realistic, as income taxes are usually independent of education and subsidies on education are usually independent from labor earnings.[10]

Workers maximize utility by choosing consumption, labor supply and investment in human capital, taking wage rates and government policy instruments as given. Given skill choice $j \in \{L, H\}$, optimal labor supply is obtained by maximizing utility (1) or (2), subject to the budget constraint, (3) or (4). Labor supplies of low- and high-skilled individuals are given by

$$l_\theta = [(1-t)\,w_L \theta]^\varepsilon, \tag{5}$$

$$h_\theta = [(1-t)\,w_H \theta]^\varepsilon. \tag{6}$$

Labor supply increases in net earnings per hour $(1-t)\,w_j \theta$, and the more so if labor supply is more elastic (higher ε). Income taxation distorts labor supply downwards as it drives a wedge between the social rewards of labor supply $(w_j \theta)$ and the private rewards of labor supply $((1-t)\,w_j \theta)$.

By substituting (5) into (1) and using (3), the indirect utility function for a low-skilled individual is found:

$$V_\theta^L \equiv b + \frac{[(1-t)\,w_L \theta]^{1+\varepsilon}}{1+\varepsilon}. \tag{7}$$

[8]Since income effects are absent, compensated and uncompensated wage elasticities coincide.

[9]See for example Jacobs (2012), Rothschild and Scheuer (2013).

[10]We aim to analyze linear and non-linear taxes that might differ by skill-type in a future version of this paper.

Similarly, substituting (6) into (2) and using (4), the indirect utility of high-skilled individual is obtained:

$$V_\theta^H \equiv b + \frac{[(1-t)\,w_H\theta]^{1+\varepsilon}}{1+\varepsilon} - (1-s)\,p. \tag{8}$$

A low-skilled individual chooses to invest in education so as to become high-skilled if and only if she derives higher utility from being high-skilled than low-skilled, i.e. when $V_\theta^H \geq V_\theta^L$. The critical level of ability Θ that separates the high-skilled from the low-skilled is determined by $V_\Theta^H = V_\Theta^L$, and is given by

$$\Theta = \left[\frac{(1-s)\,p\,(1+\varepsilon)}{(1-t)^{1+\varepsilon}\left(w_H^{1+\varepsilon} - w_L^{1+\varepsilon}\right)}\right]^{\frac{1}{1+\varepsilon}}. \tag{9}$$

All individuals with an ability $\theta < \Theta$ remain low-skilled, whereas individuals with $\theta \geq \Theta$ will become high-skilled. A decrease in Θ corresponds to more individuals becoming high-skilled. Individuals will only invest resources p in education when $w_H > w_L$, which is ensured by the Inada conditions on the production function – to be discussed below. Investment in human capital increases if the skill premium w_H/w_L is larger. In addition, skill formationincreases if the education subsidy s is larger as the net costs of education fall. Finally, taxes on income t distort investment in human capital because the direct costs of education are not deductible from the income tax, whereas the returns to education are taxed. The education choice is also distorted because income taxation reduces labor supply, and thereby lowers the 'utilization rate' of human capital. If labor supply would be exogenous ($\varepsilon = 0$) and subsidies on education make all education expenses effectively deductible ($s = t$) human capital investment is at first-best levels: $\Theta = \frac{p}{(w_H - w_L)}$ (see also Jacobs, 2005; Bovenberg and Jacobs, 2005).

3.2 Firms

A representative firm produces a homogeneous consumption good using low-skilled labor L and high-skilled labor H as inputs using a constant-returns-to-scale production technology:

$$\begin{aligned} Y(L^D, H^D), \quad Y_j(\cdot) &> 0, \quad Y_{jj}(\cdot) < 0, \quad Y_{LH}(\cdot) \geq 0, \\ \lim_{j\to 0} Y_j(\cdot) &= \infty, \quad \lim_{j\to\infty} Y_j(\cdot) = 0, \quad j = L, H. \end{aligned} \tag{10}$$

The subscript on the production function indicates a derivative. There are positive but diminishing marginal products to each labor input, and high- and low-skilled workers are co-operantfactors of production ($Y_{LH} \geq 0$). The production function $Y(\cdot)$ furthermore satisfies the Inada conditions; marginal products (and thus wages, see below) of each labor type go to zero (infinity) when input of this labor type goes to infinity (zero). High- and low-skilled workers are imperfect substitutes in production. We denote by $\sigma \equiv \frac{Y_L Y_H}{Y_{LH} Y}$ the elasticity of substitution between high-skilled and low-skilled labor. Within sectors, labor of different ability types θ are perfect substitutes.

The representative firm maximizes profits, taking wage rates on competitive labor markets as given. The first-order conditions are:

$$w_L = Y_L\left(L^D, H^D\right), \tag{11}$$

$$w_H = Y_H\left(L^D, H^D\right). \tag{12}$$

The marginal product of each labor input should equal its marginal cost. For later reference, we will denote by $\alpha \equiv HY_H(\cdot)/Y(\cdot)$ the share of total income that accrues to high-skilled labor.

3.3 General equilibrium

Labor-market clearing implies that aggregate effective labor supplies for for each skill type equal aggregate demands:

$$L = \int_{\underline{\theta}}^{\Theta} \theta l_\theta dF(\theta), \tag{13}$$

$$H = \int_{\Theta}^{\overline{\theta}} \theta h_\theta dF(\theta). \tag{14}$$

Goods-market clearing implies that total output Y equals aggregate demand for private consumption plus government consumption G: $Y = \int_{\underline{\theta}}^{\Theta} c_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} c_\theta^H dF(\theta) + G$.Since we have normalized the mass of individuals to one, average labor earnings $\bar{z}$ equals total income, which in turn equals output Y:

$$\overline{z} \equiv \int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} z_\theta^H dF(\theta) = Y. \tag{15}$$

3.4 Elasticities

Before deriving the optimal tax formulae it is instructive to derive the behavioral elasticities with respect to the income tax and education subsidies. Table 1 provides these elasticities. The derivations are given in Appendix A.

Elasticity of	Expression
Skill-formation	$\varepsilon_{\Theta,t} \equiv \frac{\partial \Theta}{\partial t}\frac{1-t}{\Theta} = \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right) > 0$
Low-skilled wage	$\varepsilon_{w_L,t} \equiv -\frac{\partial w_L}{\partial t}\frac{1-t}{w_L} = \left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right) > 0$
High-skilled wage	$\varepsilon_{w_H,t} \equiv -\frac{\partial w_H}{\partial t}\frac{1-t}{w_H} = -\left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right) < 0$
Low-skilled labor supply	$\varepsilon_{l,t} \equiv -\frac{\partial l_\theta}{\partial t}\frac{1-t}{l_\theta} = \varepsilon\left(1+\varepsilon_{w_H,t}\right) = \left(\frac{\sigma+\varepsilon+\delta\beta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon > 0$
High-skilled labor supply	$\varepsilon_{h,t} \equiv -\frac{\partial h_\theta}{\partial t}\frac{1-t}{h_\theta} = \varepsilon\left(1+\varepsilon_{w_L,t}\right) = \left(\frac{\sigma+\varepsilon+\delta(\beta-1)}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon > 0$

(a) Tax elasticities

Elasticity of	Expression
Skill-formation	$\varepsilon_{\Theta,s} \equiv -\frac{\partial \Theta}{\partial s}\frac{s}{\Theta} = \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho > 0$
Low-skilled wage	$\varepsilon_{l,s} \equiv \frac{\partial l_\theta}{\partial s}\frac{s}{l_\theta} = \left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon\rho > 0$
High-skilled wage	$\varepsilon_{h,s} \equiv \frac{\partial h_\theta}{\partial s}\frac{s}{h_\theta} = -\left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon\rho < 0$
Low-skilled labor supply	$\varepsilon_{w_L,s} \equiv \frac{\partial w_L}{\partial s}\frac{s}{w_L} = \left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho > 0$
High-skilled labor supply	$\varepsilon_{w_H,s} \equiv \frac{\partial w_H}{\partial s}\frac{s}{w_H} = -\left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho < 0$

(b) Subsidy elasticities

Table 1: Elasticities

$\beta \equiv \frac{w_H^{1+\varepsilon}}{w_H^{1+\varepsilon} - w_L^{1+\varepsilon}}$ is a measure for the skill-premium, $\delta \equiv \left(\frac{\Theta l_\Theta f(\Theta)}{L} + \frac{\Theta h_\Theta f(\Theta)}{H}\right)\Theta$ measures the importance of the marginal individual with ability Θ in aggregate effective labor supply, and $\rho \equiv \frac{s}{(1-s)(1+\varepsilon)} > 0$ captures the importance of education subsidies in the total direct costs of education.

From Table 1 follows that taxation depresses skill formation and low- and high-skilled labor supply. Education subsidies increase skill formation, boost low-skilled labor supply and reduce high-skilled labor supply. The effects of education subsidies on labor supply work only through general-equilibrium effects on wage rates.

Any policy that leads to a change in wage rates, indirectly affects skill formation and labor supply. If income taxation (education subsidies) reduces (increases) skill formation, the supply of high-skilled labor relative to low-skilled labor decreases (increases). As a result of more (less) skill formation, low-skilled wage rates increase (decrease), whereas high-skilled wage rates decrease (increase). General-equilibrium effects thus increase (decrease) low-skilled (high-skilled) labor supply. The strength of these general-equilibrium effects is captured by σ and ε. The more difficult it is to substitute skill types in production (lower σ), and the less elastic is labor supply (lower ε), the stronger will be general-equilibrium effects on wage rates, and the lower are the tax and subsidy elasticities of all policy variables. The effects of tax and education policy on skill formation and labor supply are therefore partially off-set by general-equilibrium effects in the wage structure.

In the absence of general-equilibrium effects on wages ($\sigma = \infty$) the elasticities are very simple. In fact, wage rates do not respond to changes in relative aggregate factor supplies ($\varepsilon_{w_j t} = \varepsilon_{w_j s} = 0$). Consequently, skill formation only responds to policy variables ($\varepsilon_{\Theta,t} = \varepsilon_{\Theta,s}/\rho = 1$), and labor supplies only respond to taxes, but not to subsidies ($\varepsilon_{j,t} = \varepsilon$, $\varepsilon_{j,s} = 0$).

4 Optimal policy

The government maximizes social welfare, which is a function Γ of indirect utilities of low- and high-skilled individuals:

$$\Gamma \equiv \int_{\underline{\theta}}^{\Theta} \gamma_\theta V_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta V_\theta^H dF(\theta), \tag{16}$$

where γ_θ denotes the Pareto weight of an individual with ability θ. The Pareto weights γ_θ are assumed to be strictly decreasing in θ. Moreover, we normalize the Pareto weights such that they sum to one: $\int_{\underline{\theta}}^{\overline{\theta}} \gamma_\theta dF(\theta) = 1$.

The government budget constraint states that total tax revenue equals spending on education subsidies s, non-individualized transfers b, and exogenous spending G:

$$t\left[\int_{\underline{\theta}}^{\Theta} w_L \theta l_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} w_H \theta h_\theta dF(\theta)\right] = sp\int_{\Theta}^{\overline{\theta}} dF(\theta) + b + G. \tag{17}$$

The government maximizes social welfare (16) by choosing the marginal tax rate t, the lump-sum transfer b, and the subsidy s, subject to (17).

4.1 Optimal transfers

The optimal lump-sum transfer b is derived in the next proposition.

Proposition 1. *Let the social welfare weight of an individual of type θ be defined as $g_\theta \equiv \frac{\gamma_\theta}{\eta}$, where η is the Lagrange multiplier on the government budget constraint, then the* *optimal lump-sum transfer ensures that*

$$\bar{g} \equiv \int_{\underline{\theta}}^{\overline{\theta}} g_\theta dF(\theta) = 1. \tag{18}$$

Proof. See Appendix B. □

The optimality condition equates the marginal costs of raising a transfer of one Euro (left-hand-side) with the marginal costs of doing so (right-hand-side)..

4.2 Tax wedges and distributional characteristics

In order to facilitate the optimal expressions for the tax rate t and the subsidy s, we introduce some additional notation. First, we define the net tax-wedge on skill formation Δ as:

$$\Delta \equiv t w_H \Theta h_\Theta - t w_L \Theta l_\Theta - sp. \tag{19}$$

Δ gives the marginal increase in government revenue when a marginally larger fraction of individuals become high-skilled. When $\Delta > 0$, education is taxed on a net basis. $tw_H\Theta h_\Theta$ gives the additional tax revenue when the marginal individual decides to become high-skilled. $tw_L\Theta l_\Theta$ gives the lost tax revenue as this individual no longer pays taxes as a low-skilled worker. The government also looses s in revenue if this individual invests in education. Δ then captures the total impact on the public budget of more skill formation.

We also employ Feldstein (1972)'s distributional characteristic of the income tax, which is defined as:

$$\xi \equiv \frac{\int_{\underline{\theta}}^{\Theta}(1-g_\theta)z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}}(1-g_\theta)z_\theta^H dF(\theta)}{\bar{z}\bar{g}} > 0. \tag{20}$$

ξ equals minus the normalized covariance between social welfare weights g_θ and labor earnings z_θ^j. Note that at the optimum we have $\bar{g} = 1$ from the first-order condition for b. ξ measures the social marginal value of more income redistribution via the income tax, expressed in monetary equivalents, as a fraction of taxed earnings. Marginal benefits of income taxation are positive, since the Pareto weights g_θ decline with ability θ. ξ lies between 0 and 1. ξ is larger if government has more redistributive social preferences. $\xi = 1$ for a Rawlsian/maxi-min social welfare function, which features Pareto weights $\gamma_{\underline{\theta}} = 1$ and $\gamma_\theta = 0$ for all $\theta > \underline{\theta}$. $\xi = 0$ for a utilitarian social welfare function with constant Pareto weights γ that do not depend on θ.[11] $\xi = 0$ as well if z_θ^j is equal for everyone so that the government is not interested in income redistribution. An alternative intuition for the distributional characteristic ξ is that it measures the social value raising an additional unit of revenue with the income tax. It gives the income-weighted average of the additional unit of revenue (the '1') minus the utility losses (g_θ) that raising this unit of revenue inflicts on tax payers.

Similarly, we define the distributional characteristic for education ζ as:

$$\zeta \equiv \frac{p\int_{\Theta}^{\overline{\theta}}(1-g_\theta)dF(\theta)}{\bar{g}} > 0. \tag{21}$$

Since education is a discrete choice, no income weights are applied as in the expression for ξ. From the first-order condition for b follows that $\zeta > 0$. Intuitively, the marginal

[11] Note that the absence of a redistributional preference in this case relies on a constant marginal utility of income at the individual level. In general, with non-constant private marginal utilities of income, also a utilitarian government has a preference for income redistribution, i.e. $\xi > 0$.

value of raising a unit of revenue for the government is lower than the utility loss inflicted letting the high-skilled tax payers pay a unit of income extra tax. This implies that taxing education generates distributional benefits.

Finally, we define the income-weighted social welfare weights of each skill group as

$$\bar{g}^L \equiv \frac{\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta)}{\int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta)}, \quad \bar{g}^H \equiv \frac{\int_{\Theta}^{\bar{\theta}} g_\theta z_\theta^H dF(\theta)}{\int_{\Theta}^{\bar{\theta}} z_\theta^H dF(\theta)}. \tag{22}$$

Armed with the additional notation, we are now able to state the main propositions of our paper.

4.3 Optimal linear income taxes

First, we provide the optimal income tax.

Proposition 2. *The optimal tax rate on labor income is determined by*

$$\frac{t}{(1-t)}\varepsilon + \frac{\Delta}{(1-t)} \frac{f(\Theta)\Theta}{\bar{z}} \varepsilon_{\Theta,t} = \xi - (\bar{g}^L - \bar{g}^H)\varepsilon_{GE}, \tag{23}$$

where $\varepsilon_{GE} \equiv (1-\alpha)\varepsilon_{w_L,t} = -\alpha\varepsilon_{wH,t} = \frac{\alpha(1-\alpha)\delta}{(\sigma+\varepsilon+\delta(\beta-\alpha)}$ *is the general-equilibrium elasticity.*

Proof. See Appendix B.1. □

The optimal income tax balances the distortions of income taxation on the left-hand side with its distributional benefits on the right-hand side. On the left-hand side, $\frac{t}{(1-t)}\varepsilon$ captures the marginal deadweight loss of distorting labor supply. The larger the wage elasticity of labor supply ε, the more do individuals reduce labor supply when tax rates increase. Moreover, labor-supply distortions increase more than proportionally with the level of taxation as captured by $\frac{t}{1-t}$. $\frac{\Delta}{(1-t)} \frac{f(\Theta)\Theta}{\bar{z}} \varepsilon_{\Theta,t}$ denotes the marginal deadweight loss of income taxation on human capital formation. A higher marginal tax rate discourages individuals from becoming high-skilled. The larger is the tax-elasticity of skill formation $\varepsilon_{\Theta,t} = \frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}$, the bigger are distortions of income taxation on skill formation. The higher the net tax wedge on human capital $\frac{\Delta}{1-t}$, the more income taxation distorts skill formation, and the lower should be the optimal tax rate be. $\frac{f(\Theta)\Theta}{\bar{z}}$ measures relative importance of tax distortions on the marginal graduate Θ. The higher is the mass of agents at Θ and the larger is their ability, the more important are tax distortions on human capital.

The right-hand side gives the distributional benefits of income taxation. The larger are the distributional benefit of higher income taxes – as captured by ξ – the higher should be the optimal tax rate. This is the standard term in optimal linear tax models. In addition, $(\bar{g}^L - \bar{g}^H)\varepsilon_{GE}$ captures distributional losses of general-equilibrium effects. As income taxation reduces skill formation, the supply of high-skilled labor falls relative to low-skilled labor. This raises high-skilled wages and lowers low-skilled wages. Consequently, social welfare declines since the income-weighted welfare weights of the low-skilled workers are larger than the income-weighted welfare weights of the high-skilled workers ($\bar{g}^L - \bar{g}^H > 0$). The direct gains of income redistribution are therefore reduced as income taxation leads to larger before-tax wage inequality. The general-equilibrium elasticity $\varepsilon_{GE} \equiv \frac{\alpha(1-\alpha)\delta}{(\sigma+\varepsilon+\delta(\beta-\alpha)}$ captures the strength of these general-equilibrium effects on the wage structure. A lower elasticity of substitution σ, and a lower labor-supply elasticity ε provoke stronger general-equilibrium responses that erode the distributional powers of income taxation. If the effective labor supply around the skill margin is relatively low compared to aggregate labor supply, i.e. $\delta \equiv \left(\frac{\Theta l_\Theta f(\Theta)}{L} + \frac{\Theta h_\Theta f(\Theta)}{H}\right)\Theta$ is small, general-equilibrium effects will not be important for setting optimal tax rates. In the absence of general equilibrium effects ($\sigma = \infty$), the general-equilibrium elasticity is zero ($\varepsilon_{GE} = 0$) and this effect drops out of the equation.

What is the effect of SBTC on optimal income taxes? SBTC is expected to raise the distributional benefits of income taxation ξ. Larger distributional gains call for higher income taxes. At the same time, the difference in welfare weights $\bar{g}^L - \bar{g}^H$ increases as inequality rises. Hence, general-equilibrium effects of income taxes generate larger distributional losses. If the direct distributional gains outweigh the losses due to general-equilibrium effects optimal taxes on income should increase – ceteris paribus. SBTC might also affect the elasticities of skill formation $\varepsilon_{\Theta,t}$ and the general-equilibrium elasticity ε_{GE}, but for small changes in SBTC we expect these effects to be less important. So we expect that SBTC raises optimal income tax rates provided that general-equilibrium effects are not too strong.

4.4 Optimal education subsidies

The optimal net tax on education follows in the next proposition.

Proposition 3. *The optimal net tax on skill formation is determined by*

$$\frac{\Delta}{1-t}\frac{\Theta f(\Theta)}{\bar{z}}\varepsilon_{\Theta,s} = \frac{1}{\bar{z}}\frac{s}{1-t}\zeta - \rho\left(\bar{g}^L - \bar{g}^H\right)\varepsilon_{GE} \tag{24}$$

Proof. See Appendix B.2. □

The formula gives the marginal distortions of taxing education on a net basis on the left-hand side and the distributional benefits of taxing education on the right-hand side. On the left-hand side, distortions increase if the optimal net tax on education $\frac{\Delta}{(1-t)}$ is larger. Recall that if $\Delta > 0$, human capital formation is taxed on a net basis. The $\frac{\Theta f(\Theta)}{\bar{z}}$-term is the same as in (23). It captures the economic importance of distorting the decision of the marginal graduate. $\varepsilon_{\Theta,s} = \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho$ is the elasticity of skill formation with respect to the subsidy on education. The larger the elasticity of skill formation, the lower should be the optimal net tax be on skill formation.

The distributional gains of net taxes on education are given on the right-hand side. Since $\zeta > 0$, taxing human capital yields net distributional benefits. The higher is the distributional gain of taxing education ζ, the more the government wishes to tax the high-skilled. $\rho(\bar{g}^L - \bar{g}^H)\varepsilon_{GE}$ captures again the adverse distributional consequences of taxing human capital on a net basis. When a positive tax wedge is levied on skill formation, skill formation falls, high-skilled wages rise and low-skilled wages fall. Consequently, general-equilibrium effects inflict larger welfare losses on the low-skilled than gains on the high-skilled, since the income-weighted welfare weights of the low-skilled are higher than that of the high-skilled ($\bar{g}^L - \bar{g}^H > 0$). The general-equilibrium elasticity $\varepsilon_{GE} \equiv \frac{\alpha(1-\alpha)\delta}{(\sigma+\varepsilon+\delta(\beta-\alpha)}$ captures the importance of these general-equilibrium effects. Since it is the same as above, it will not be discussed again. Like with income taxes, the indirect distributional losses due to general-equilibrium effects need to be weighed against the direct distributional gains for setting net taxes on skill formation.

Labor-market distortions do not play a role in determining the net tax on skill formation. The intuition is that net taxes on skill-formation cannot alleviate the labor-market distortions created by income taxation. Intuitively, by discouraging high-skilled labor supply, lower skill formation would alleviate labor-market distortions in the low-skilled segment of the labor market. However, by the same token, reducing skill formation also exacerbates labor-market distortions in the high-skilled segment of the labor market. These two effects exactly off-set each other.

What is the effect of SBTC on optimal net taxes on education? SBTC is expected to raise the distributional benefits of taxes on education ζ, which calls for net taxes on human capital. At the same time, the difference in welfare weights $\bar{g}^L - \bar{g}^H$ increases as inequality rises. Again, general-equilibrium effects of net education taxes generate distributional losses. However, as long as direct distributional benefits outweigh the indirect distributional losses due to general-equilibrium effects, optimal net taxes on human capital $\frac{\Delta}{1-t}$ should increase – ceteris paribus. Again, we expect that changes in the elasticities are of minor importance when SBTC is not too strong. We therefore conjecture that SBTC raises optimal net taxes on education.

We like to note that, even if income taxes t and net taxes on education Δ increase due to SBTC, it is not a priori clear how education subsidies should optimally respond to SBTC. The reason is that higher income taxes raise the net tax wedge on education $\Delta \equiv tw_H\Theta h_\Theta - tw_L\Theta l_\Theta - s$. It could be the case that the increase in taxation t is so strong that higher education subsidies s are needed to implement the optimal (higher) net tax wedge Δ on education.

4.5 Optimal income taxes with optimal education subsidies

In order to gain more intuition into the optimal setting of income taxes and education subsidies, we can substitute the optimal education tax into the expression of the optimal income tax, to find a simple expression for the optimal linear income tax.

Corollary 1. *If the net tax on education is optimized, the optimal income tax satisfies*

$$\varepsilon \frac{t^*}{1-t^*} = \xi - s^* \frac{\zeta}{\bar{z}\rho\,(1-t)}. \tag{25}$$

Proof. We can solve for the optimal structure of income taxes and net taxes on education by solving the optimality conditions for t, (23), and s, 24, for $\frac{t}{(1-t)}$ and $\frac{\Delta}{(1-t)}$. Then substitute (24) in (23). □

This expression sheds light on which instrument the government likes to use for income redistribution. It captures the relative advantage of using taxes on income relative to taxes on education to redistribute incomes. Optimal income taxes are larger when the equity efficiency trade-off in the labor market improves. That is, if there are higher distributional benefits ξ relative to tax distortions $\varepsilon\frac{t^*}{1-t^*}$. Moreover, optimal income taxes are lower if net taxes on education are a more powerful tool to combat inequality, i.e. when $s^*\frac{\zeta}{\bar{z}\rho(1-t)}$ is larger relative to ξ. We note that the distributional advantages of income taxes relative to education taxes do not depend on the general-equilibrium effects on the wage structure. We now turn to the effect of SBTC on optimal policy, and first focus on a special case.

5 Special case: Linear technology and uniform ability

We now focus on the case of linear technology. Moreover, we assume that ability is uniformly distributed on $[0, 1]$. We choose welfare weights

$$g_\theta \equiv \theta^{-r}, \ 0 < r < 1.$$

These simplifications allow us to solve the model in closed form. The production function is now given by

$$Y(L, AH) = L + AH,$$

with $A > 1$. We normalize the price of the output good to one. Taking the output price and wage rates as given, the firm's profit maximization problem then is

$$\max_{L,H} L + AH - w_L L - w_H H.$$

Ruling out corner solutions, the first-order conditions are given by

$$w_L = 1$$

and

$$w_H = A.$$

The ratio of wage rates is thus simply A. Linear technology rules out general equilibrium effects, hence a change in the supply of skills has no effect on the wage rates.

5.1 Threshold

The threshold is now given by

$$\Theta \equiv \left[\frac{(p-s)(1+\varepsilon)}{(1-t)^{1+\varepsilon}(A^{1+\varepsilon}-1)}\right]^{\frac{1}{1+\varepsilon}}.$$

Clearly, Θ decreases in A. The requirement of $0 \leq \Theta \leq 1$ imposes a lower bound $\underline{A}$ on A, such that $\Theta(\underline{A}) = 1$. Moreover, we note that

$$\lim_{A \to \infty} \Theta = 0,$$

hence if skill-bias becomes infinite, all individuals will work in the college sector.

5.2 Labor supplies

Labor supplies are given by

$$l_\theta = [(1-t) w_L \theta]^\varepsilon = [(1-t)\theta]^\varepsilon,$$

and

$$h_\theta = [(1-t) w_H \theta]^\varepsilon = [(1-t) A\theta]^\varepsilon.$$

Since wage rates are not affected by changes in labor supply, for a given θ, l_θ is now constant, while h_θ changes with skill-bias.

5.3 Labor market clearing

The conditions for labor market clearing are

$$L = \int_{\underline{\theta}}^{\Theta} \theta l_\theta dF(\theta) = (1-t)^\varepsilon \int_{\underline{\theta}}^{\Theta} \theta^{1+\varepsilon} dF(\theta)$$

and

$$H = A^\varepsilon (1-t)^\varepsilon \int_{\Theta}^{\overline{\theta}} \theta^{1+\varepsilon} dF(\theta).$$

Using that ability is uniformly distributed on the interval $[0,1]$, we obtain

$$L=(1-t)^{\varepsilon}\,\frac{\Theta^{2+\varepsilon}}{2+\varepsilon},$$

and

$$H=A^{\varepsilon}\,(1-t)^{\varepsilon}\,\frac{1-\Theta^{2+\varepsilon}}{2+\varepsilon}.$$

Substituting for Θ, we obtain aggregate labor supplies in terms of parameters only:

$$L=\frac{(1-t)^{\epsilon}\left(\frac{(\epsilon+1)(p-s)(1-t)^{-\epsilon-1}}{A^{\epsilon+1}-1}\right)^{\frac{1}{\epsilon+1}+1}}{\epsilon+2}$$

and

$$H=\frac{A^{\varepsilon}(1-t)^{\epsilon}\left(1-\left(\frac{(\epsilon+1)(p-s)(1-t)^{-\epsilon-1}}{A^{\epsilon+1}-1}\right)^{\frac{1}{\epsilon+1}+1}\right)}{\epsilon+2}.$$

5.4 Effect of SBTC on optimal policy

The effect of SBTC on the optimal income tax and the optimal education subsidy is ambiguous when both policies are optimized simultaneously. However, we can gain some insights into the effect of technical change on optimal policy when holding either the subsidy or the optimal tax rate fixed. Such an analysis is more than just a theoretical exercise. Often governments might be able to adjust only one policy at a time. We study how such an adjustment should be conducted optimally. First, we ask how the optimal income tax should be adjusted for a given level of the subsidy.

5.4.1 Lemma: Distributional characteristic

We first establish the following Lemma which will be used in subsequent proofs.

Lemma 1. *Under skill-biased technological change, the distributional characteristic ξ only indirectly depends on w_L and w_H via $\Theta(w_L, w_H)$ but not directly.*

Proof. See Appendix B.3. □

5.4.2 Optimal tax adjustment for given subsidy

In order to study the effect of SBTC on the optimal tax rate, we first analyze the effect of technological change on the different terms involved in the optimal tax rate formula. We derive that with linear technology and uniform ability the distributional characteristic of the income tax, ξ, first increases with skill-bias, and then decreases.

Lemma 2. *With linear technology, uniform ability, and a fixed subsidy, the distributional characteristic ξ first increases with SBTC and then falls.*

Proof. See Appendix B.4. □

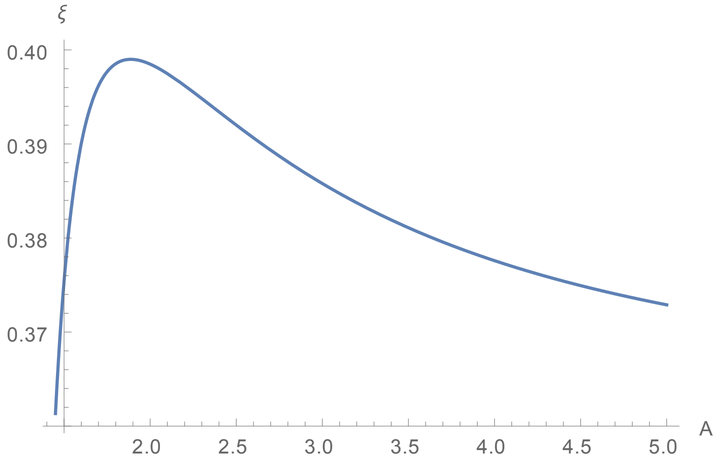
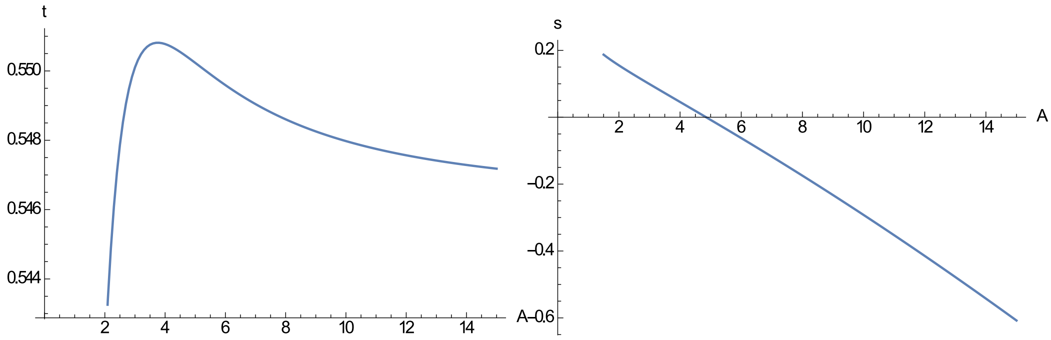


Figure 1: Behavior of ξ under SBTC - Uniformly distributed ability



(a) Optimal tax rate under SBTC with fixed subsidy - Linear technology and uniform ability

(b) Optimal subsidy under SBTC with fixed tax rate - Linear technology and uniform ability

Figure 2: Optimal tax rate and subsidy under SBTC when keeping the other variable fixed - Linear technology and uniform ability

We illustrate the behavior of ξ under SBTC in Figure 1. To gain intuition, we highlight again that ξ is a function of Θ only – and more importantly – varies with the share of high-skill income in total income. While this share is low, redistributing income from high-skilled individuals to low-skilled individuals raises welfare. However, once the share of high-skill income grows large, further redistribution of income lowers welfare.

Having analyzed the behavior of ξ under SBTC, we arrive at the following proposition.

Proposition 4. *With linear technology, uniform ability, and a fixed subsidy: the optimal tax rate*

- □ *increases with SBTC if skill-bias is low*
- □ *decreases with SBTC if skill-bias is large.*

Proof. See Appendix B.5. □

The behavior of the optimal tax rate under SBTC while keeping the subsidy fixed is shown in Figure 2a.

5.4.3 Optimal subsidy adjustment for given tax rate

If the government is able to adjust the subsidy, but needs to keep the income tax rate at its old level, it should optimally lower the subsidy if skill-bias is low:

Proposition 5. *With linear technology, uniformly distributed ability, and a fixed tax rate: the optimal subsidy*

- □ *decreases with SBTC if skill-bias is low*

Proof. ee Appendix B.6 □

If the government can simultaneously optimize the tax rate and the subsidy, the optimal behavior of tax rate and subsidy under SBTC is analytically ambiguous. In order to study optimal policy in this case, we now turn to simulations.

6 Simulation

Simulations are commonly used to study the behavior of optimal tax rates. Many simulation studies have assumed lognormal ability distributions. We follow the example of Tuomala (2010) and choose a lognormal ability distribution with parameters $\mu^\theta = 0.4$ and $\sigma^\theta = 0.39$. Like the canonical model of SBTC (see Acemoglu and Autor, 2011), we assume a constant-elasticity of substitution (CES) production function:

$$Y(L,H) = \left(\omega L^{\frac{\sigma-1}{\sigma}} + (1-\omega)(AH)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \quad \omega,\sigma > 0, \tag{26}$$

where ω is a parameter used to calibrate the income share of high-skilled income α. Skill-bias is parametrized by A and SBTC corresponds to an increase in A. To analytically and numerically trace down the consequences of skill-biased technological change we adopt a social welfare function with Pareto weights γ_θ that follow[12]

$$\gamma_\theta \equiv \theta^{-r},$$

where r captures the government's aversion to inequality. We choose a value of $r = 0.5$ which leads to income tax rates in a similar order of magnitude as in the US. Table 2 summarizes the parameters used. The compensated wage elasticity of labor-supply is set to 0.3, based on evidence for compensated labor-supply elasticities provided in Blundell and Macurdy (1999) and Meghir and Phillips (2008). The chosen parameters generate a variance of log-hourly wages which comes close to that reported by Heathcote, Perri, and Violante (2010) for the US.

In order to simulate SBTC, we increase the value of skill-bias starting from $A = 1.2$. For each level of skill-bias, the optimal policy is computed. We focus on two scenarios. In the first scenario we set the elasticity of substitution to $\sigma = 1.41$ as estimated by Katz and Murphy (1992), such that there are general equilibrium effects between high-skilled and low-skilled labor. In the second scenario, we choose $\sigma = 10000$ such that technology is effectively linear.

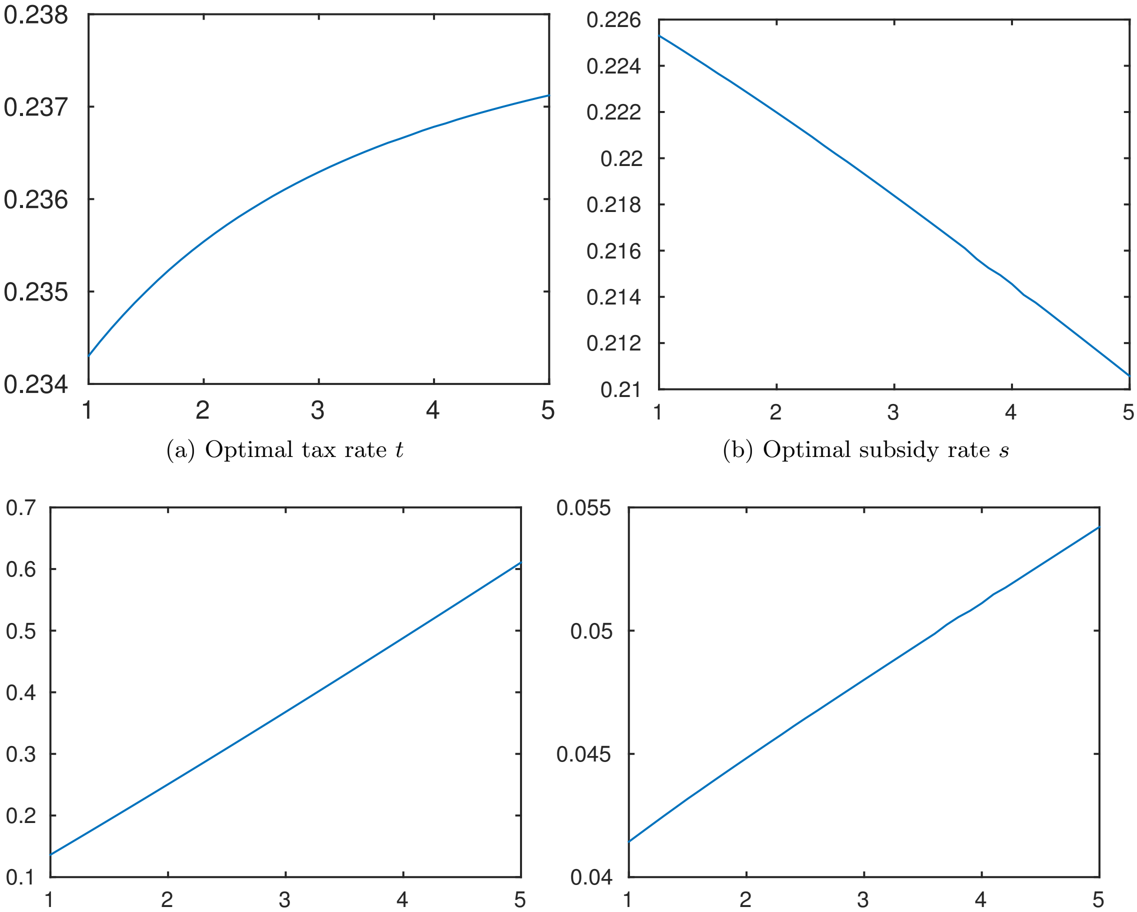
6.1 SBTC with general equilibrium effects

The results of the first scenario are displayed in Figure 3. We observe that the optimal tax

[12]When using a general social welfare function – a sum of a concave transformation of individual utilities – general-equilibrium effects make it very difficult to conduct comparative-statics analyses.

Parameter	Value	Source
μ^θ	0.4	Tuomala (2010)
σ^θ	0.39	Tuomala (2010)
ω	0.45	fixed
ε	0.3	Blundell and Macurdy (1999); Meghir and Phillips (2008)
r	0.5	fixed

Table 2: Parameters for simulations



(a) Optimal tax rate t

(b) Optimal subsidy rate s

(c) Optimal transfer b

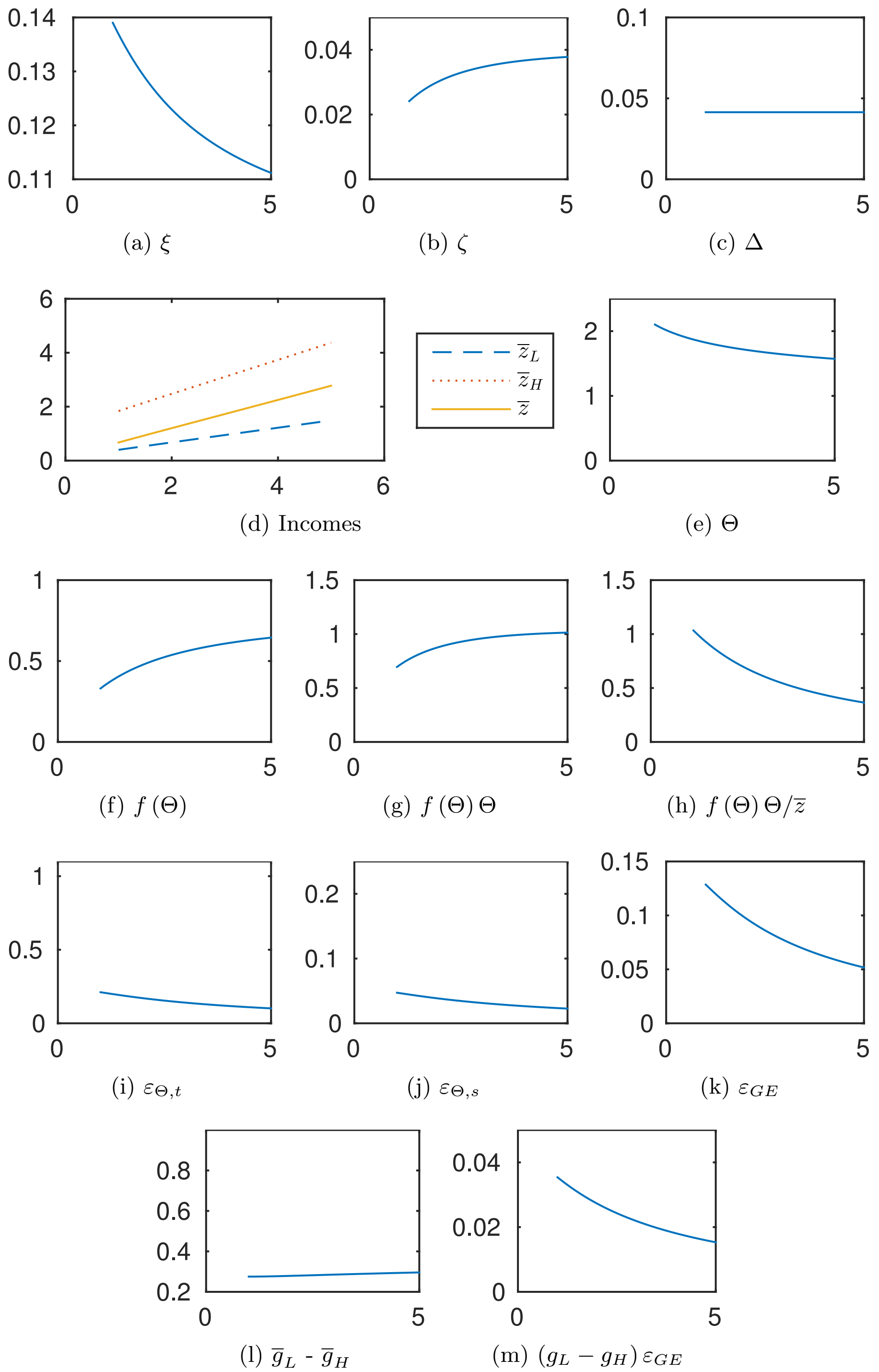
(d) Optimal net tax Δ

Figure 3: Lognormal ability, $\sigma = 1.41$, A on the horizontal axis

rate increases with SBTC, while the optimal subsidy falls (Panels 3a and 3b). The net tax on skill-formation increases as a consequence (Panel 3d). While optimal subsidies remain positive, it is thus optimal for redistribution to distort skill-formation more as skill-bias becomes more severe. Finally, as the economy grows, the optimal transfer increases (Panel 3c). In order to better understand what drives optimal policy, we conduct a ceteris paribus experiment in which we keep the tax rate and the subsidy at its optimal level for $A = 1.2$ – but do not re-optimize as we increase A. However, we make sure that labor markets clear by allowing L and H to adjust, and we enforce a balanced government budget by adjusting the transfer. We plot the different components of the first-order conditions in Figure 4.

We first focus on the components of condition (23) for the optimal tax rate. In Panel 4a, we observe that the distributional characteristic of the tax, ξ, falls with SBTC, which would call for a lower tax rate if all other elements of the first-order condition remained unchanged. However, this is not the case. While the net tax Δ stays virtually constant (Panel 3d), the threshold Θ falls (Panel 4e), and output $\overline{z}$ increases (Panel 4d). We also observe that the density at the threshold $f(\Theta)$ increases (Panel 4f). The weighted threshold $f(\Theta)\Theta$ increases as a consequence. However, this increase is compensated by the rise in $\overline{z}$, such that the term $f(\Theta)\Theta/\overline{z}$ drops. Since at the same time the tax elasticity of skill-formation, $\varepsilon_{\Theta,t}$, drops, a higher income tax rate imposes less distortions on skill-formation – which ceteris paribus induces an increase in the optimal tax rate. Finally, we consider the term $\left(\overline{g}^L - \overline{g}^H\right)\varepsilon_{GE}$. Although the difference in income-weighted welfare weights, $\overline{g}^L - \overline{g}^H$, increases with SBTC (Panel 4l), this is overcompensated by general equilibrium effects, ε_{GE}, becoming less important (Panel 4k) such that $\left(\overline{g}^L - \overline{g}^H\right)\varepsilon_{GE}$ drops (Panel 4m).

In order to understand the behavior of the optimal net tax Δ, we focus on the components in condition (24). The increase in the distributional characteristic of the subsidy, ζ, calls for a higher optimal net tax. Moreover, since the subsidy elasticity of skill-formation, $\varepsilon_{\Theta,s}$, falls (Panel 4i), Δ imposes less distortions on skill-formation, and Δ should be higher ceteris paribus. Finally, we note that ρ stays constant and $\rho\left(\overline{g}^L - \overline{g}^H\right)\varepsilon_{GE}$ thus falls, which also encourages an increase in the optimal net tax.



(a) ξ (b) ζ (c) Δ

(d) Incomes (e) Θ

(f) $f(\Theta)$ (g) $f(\Theta)\,\Theta$ (h) $f(\Theta)\,\Theta/\overline{z}$

(i) $\varepsilon_{\Theta,t}$ (j) $\varepsilon_{\Theta,s}$ (k) ε_{GE}

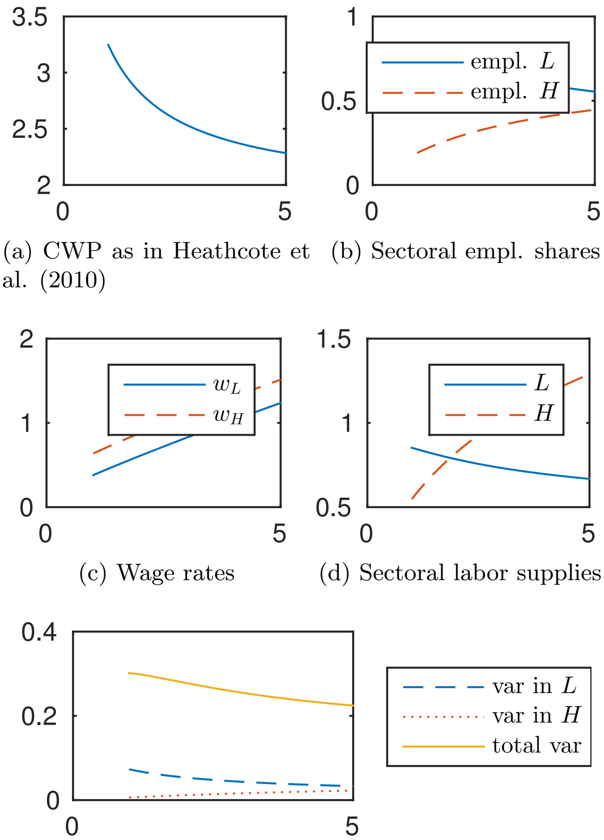
(l) $\overline{g}_L$ - $\overline{g}_H$ (m) $(g_L - g_H)\,\varepsilon_{GE}$

Figure 4: Lognormal ability, $\sigma = 1.41$, A on the horizontal axis - s and t kept constant

We now investigate whether our model is an appropriate description of the labor market. We do so by plotting the effect of SBTC on various labor market statistics in Figure 5 – again, holding s and t at their initial levels. Panel 5a displays the college wage premium measured as the ratio of average hourly wages of high and low-skilled individuals.[13] We find that SBTC reduces the college wage premium. This result is at odds with the data.

[13] Heathcote, Perri, and Violante (2010) use the same measure of the college wage premium.

In our model, SBTC leads to a strong inflow into the high-skill sector. Since there are complementarities between both types of labor, an increase in the supply and productivity of high-skilled labor also leads to a rise of the wage rate earned by the low skilled. These effects are so strong that w_L increases more than w_H (Panel 5c), which in turn leads to a fall in the college wage premium. The inflow into the high-skill sector in terms of employment and aggregate labor supply is illustrated in Panels 5b and 5d. In Panel 5e we plot the variance of log-hourly wages, by sector and overall. While SBTC leads to a slight increase in the variance of log-hourly wages in the college sector, the variance in the high-school sector, as well as the overall variance, falls. The drop in the overall variance is a consequence of two phenomena: the drop in the college wage premium and the fall of the variance within the high-school sector. This reduction in variance is similarly at odds with the data.



(a) CWP as in Heathcote et al. (2010)

(b) Sectoral empl. shares

(c) Wage rates

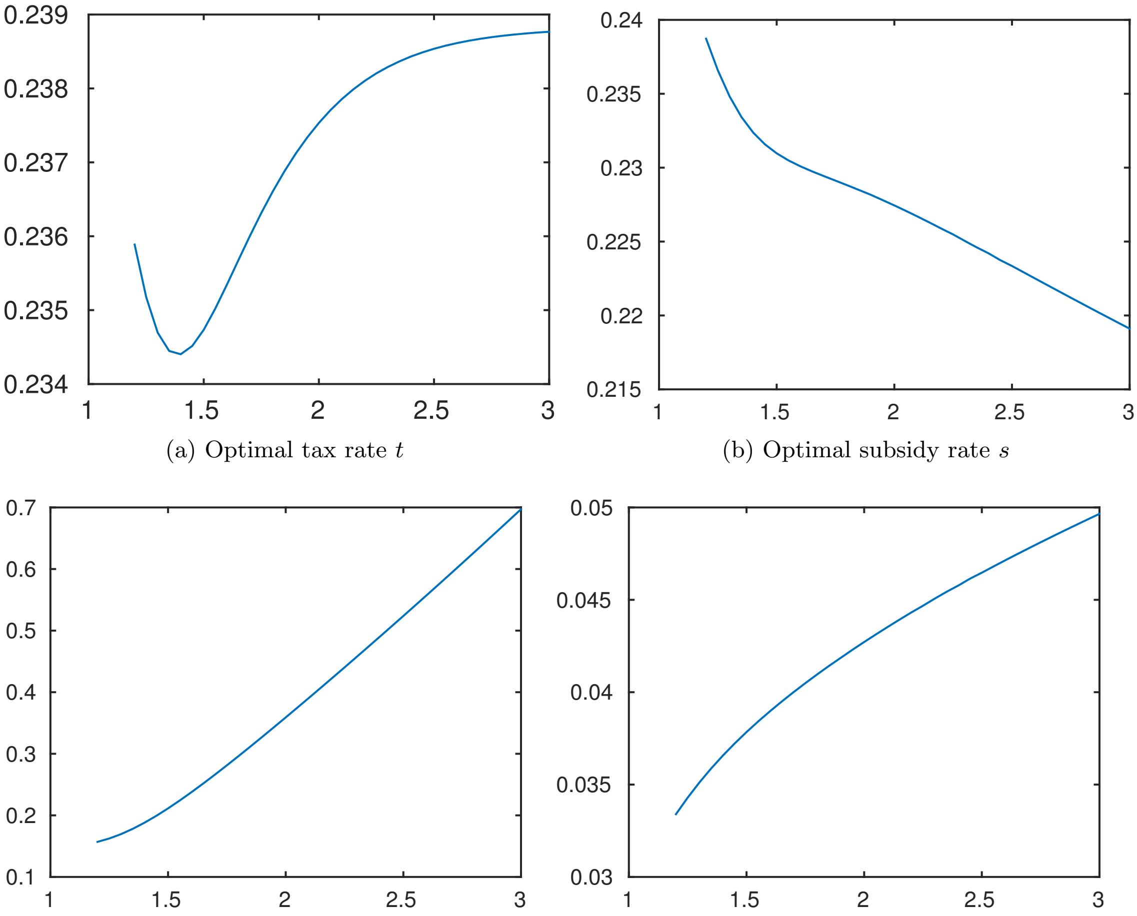
(d) Sectoral labor supplies

(e) Variance of log hourly wages

Figure 5: Labor market statistics - Lognormal ability, $\sigma = 1.41$, A on the horizontal axis - s and t kept constant

6.2 SBTC with linear technology

The results of the second scenario are plotted in Figure 6. Due to technology being almost linear, general equilibrium effects are now virtually absent. We observe that the optimal tax rate decreases for low values of skill bias, to then increase and stabilize (Panel 6a). In contrast, the optimal subsidy monotonically decreases with SBTC (Panel 6b). The combined effect on the net tax Δ is still such that SBTC leads it to increase (Panel 6d). As before, the optimal transfer rises (Panel 6c).



(a) Optimal tax rate t

(b) Optimal subsidy rate s

(c) Optimal transfer b

(d) Optimal net tax Δ

Figure 6: Lognormal ability, $\sigma = 10000$, A on the x-axis

To investigate the source of the changed pattern of the optimal tax rate, we again conduct a ceteris paribus analysis in which we keep s and t and constant; and plot the components of the first order conditions in Figure 7. As in the case with general equilibrium effects, the distributional characteristic ξ falls with SBTC (Panel 7a), which pushes towards lower tax rates. Moreover, as before Δ is virtually constant. Since general equilibrium effects are absent, the term $\left(\overline{g}^L - \overline{g}^H\right)\varepsilon_{GE}$ is now zero (Panel 7m). Moreover, the tax elasticity of skill-formation, $\varepsilon_{\Theta,t}$, now equals one (Panel 7i). Output $\overline{z}$ increases monotonically (Panel 7d). The threshold ability Θ falls as before, however, the effect is much stronger. This leads us to the source of the different behavior of the optimal tax rate: as Θ drops, $f(\Theta)$ becomes the mirror image of the ability density. With general equilibrium effects the inflow into college is dampened, and Θ remains to the right of the mode of the ability density. In contrast, Θ now crosses the mode, thus generating the hump-shaped pattern in Panel 7f). The distortions which a higher income tax imposes on skill-formation thus first increase with SBTC, thereby amplifying the drop in ξ, and inducing a fall in optimal tax rates. As skill-bias grows, distortions of skill-formation fall and approach zero, thereby overcompensating the drop in ξ, and inducing an increase in optimal tax rates.

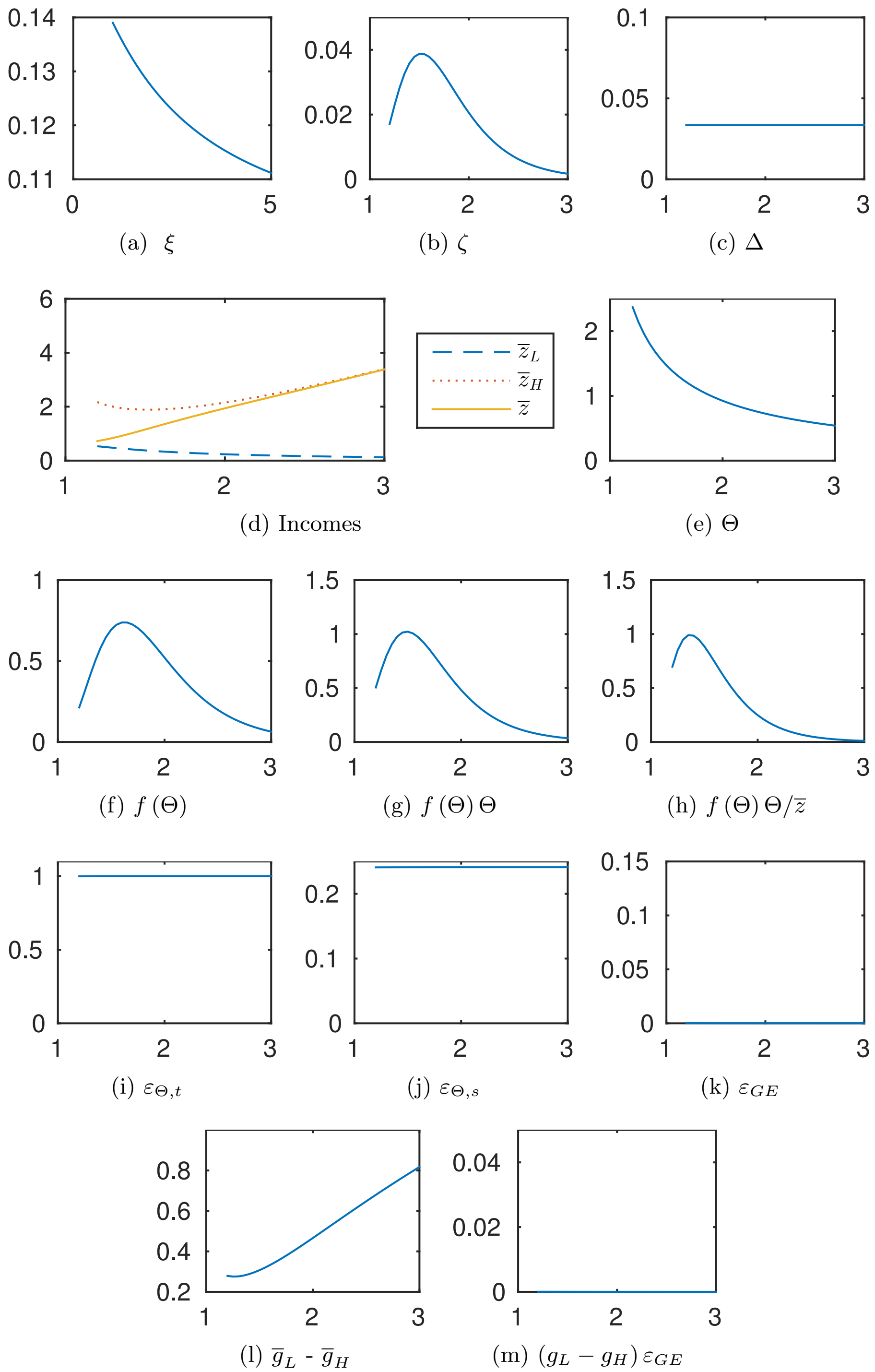
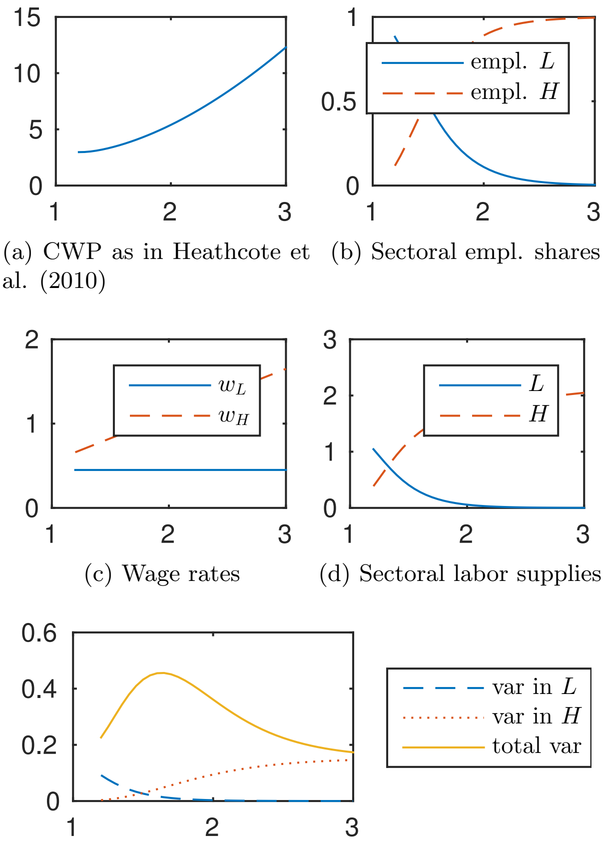


Figure 7: Lognormal ability, $\sigma = 1.41$, A on the horizontal axis - s and t kept constant

Next, we investigate how the model with linear technology performs with respect to core labor market outcomes. Keeping the ceteris paribus assumption, we plot key labor market statistics in Figure 8. The college wage premium shown in Panel 8a now increases with SBTC. The reason is plotted in Panel 8c: the wage rate of the high-skilled increases with skill-bias, whereas the wage rate of the low-skilled stays constant due to the absence of general equilibrium effects. The widening wage rate differential provides a strong incentive for skill-formation such that college enrollment responds much more strongly than under the first scenario (Panel 8b) – and so do sectoral labor supplies (Panel 8d). The total variance of log hourly wages displayed in Panel 8e increases at low levels of skill-bias when only a few individuals work in the high-skill sector. However, as more and more individuals become high-skilled, there is a tipping point at which the population becomes more homogeneous with respect to education – and with respect to log hourly wages – leading to a decrease in the variance.



(a) CWP as in Heathcote et al. (2010)

(b) Sectoral empl. shares

(c) Wage rates

(d) Sectoral labor supplies

(e) Variance of log hourly wages

Figure 8: Labor market statistics - Lognormal ability, $\sigma = 1.41$, A on the horizontal axis - s and t kept constant

Based on our simulations we conclude that SBTC leads to an increase in the net-tax on skill-formation. General equilibrium effects matter for the qualitative behavior of the optimal tax rate, while the qualitative behavior of the optimal subsidy is not affected – it decreases in both scenarios. In the second scenario, the behavior of the college wage premium is more in line with the data. However, the assumption that low and high-skilled labor are perfect substitutes is implausible. We next turn to a factor which can resolve the puzzle of a decreasing college wage premium in the presence of general equilibrium effects: the rising costs of college.

6.3 The rising costs of college

In our simulations we have kept the costs of college constant while adjusting the degree of skill-bias. However, we observe that there has been a strong increase in tuition fees since the 1980s. Figure 9 plots the development of tuition fees including room and board for public and private 4-year colleges for the period 1980 to 2005, where we normalized the dollar amounts using the price level of the year 2000. Both, public and private tuition fees have more than doubled. Moreover, while attending private colleges was on average more than two times as expensive as attending public institutions in 1980, the gap has widened.

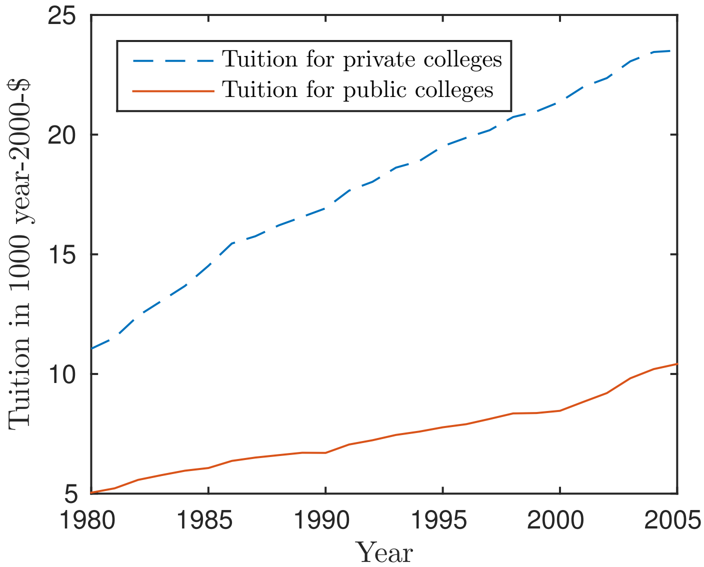
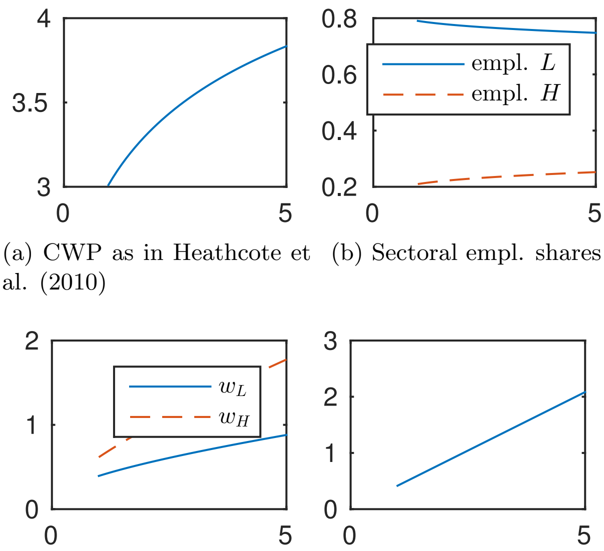


Figure 9: Tuition fees - Source: US National Center for Education Statistics

It has been shown empirically that an increase in tuition reduces college enrollment (Kane (2006); Deming and Dynarski (2009)). In our model, the same holds true: Θ increases in p. In the presence of general equilibrium effects, the reduction in the supply of college graduates increases the wage rate w_H. There is a second factor which pushes towards a higher wage rate w_H: individuals who decide to become skilled require a larger compensation the higher the costs of college. However, while a rise in the costs of college leads to an increase of the college wage premium, we do not observe a decline in the share of college graduates in the data. Neither SBTC nor an increase in tuition fees alone can thus generate the empirically observed patterns. However, the combination of the two mechanisms can generate both: an increase in the college wage premium and a rising share of college graduates. Allowing for the changing cost of college to enter the model is not in contradiction with the canonical model of SBTC. The canonical model takes the supply of college and high-school graduates as exogenously given – it is thus agnostic about mechanisms that interact with supply. However, the role of rising tuition rates for income inequality has been discussed in the literature. For example, Goldin and Katz (2010) partly attribute the slowing growth of the share of college graduates in the US to ever higher tuition fees – and link this development to rising inequality.

We next demonstrate that allowing for the cost of college to increase can indeed generate the empirically observed patterns. Figure 10 illustrates the behavior of the college wage premium, the employment shares, the wage rates and the costs of college for a version of the model in which general equilibrium effects are present ($\sigma = 1.41$) and in which the costs of college increase linearly with skill-bias. To not confound the effects of SBTC and the change in the tuition rate, we keep s and t fixed. The results are qualitatively in line with the data: both the college wage premium and the employment share of college graduates increase. Having shown that by accounting for the increasing costs of college



(a) CWP as in Heathcote et al. (2010) (b) Sectoral empl. shares

(c) Wage rates (d) Cost of college p

Figure 10: Lognormal ability, $\sigma = 1.41$, A on the horizontal axis - s and t kept constant

our model can generate the empirically observed patterns, we are now ready to calibrate our model in order to conduct a policy experiment.

7 Calibration and policy experiment

The goal of this section is to conduct a policy experiment in which we ask the question how the optimal income tax rate and education subsidy should have looked like – and should have changed – if the increase in income inequality has indeed been driven by SBTC. Moreover, we will investigate the potential welfare gains of choosing policy optimally.

7.1 Calibration

Most of the literature on SBTC has been concerned with developments in the US between 1980 and today. To let our model speak to these phenomena, we also choose to focus on US data after 1980. However, to not confound the effects of SBTC with those of the great recession, we limit ourselves to the period 1980 - 2005. Moreover, since our model is static, we do not consider it useful to analyze the entire transition between 1980 and 2005. Instead, we focus on the endpoints – 1980 and 2005 – which we consider as steady states. We calibrate our model as follows. First, we take values for the education subsidy, the tax rate and the costs of college in 1980 and 2005 from the data. We consider the difference between the tuition fees for private and public colleges as a first approximation of the education subsidy. We then normalize tuition for private colleges in 1980 to one – and express the subsidy as a fraction of these costs. The respective values are given in Table 3. Next, we focus on the economy in 1980 and choose the variance of ability and the level of skill-bias such that we match two moments in the data: the employment share of college graduates, and the variance of log hourly wages[14]. We find that a value of $\sigma_\theta = 0.3887$ and

[14] The variance of log hourly wages is taken from Heathcote, Perri, and Violante (2010). The fraction of college graduates in total employment is taken from http://www.census.gov/hhes/socdemo/education/data/cps/historical/index.html. All statistics are

$A = 1.2957$ let us match the respective moments perfectly. The value for σ_θ is surprisingly close to the one used in our simulations above – and thus close to the parameter used in Tuomala (2010). We summarize the parameters which we keep constant throughout the policy exercise in Table 4. We do not try to match the college wage premium. By construction, our model overestimates the college wage premium, as there is no overlap of the wage distributions attributed to high-school and college graduates. Even the worst college graduate earns more than the best high-school graduate – an artifact not found in the data. As a result, the college wage premium generated by our model is too high. If we were trying to match the college wage premium, the resulting variance of log hourly wages would be too low – inducing unrealistically low optimal taxes. In contrast, when targeting the employment share and the variance of log hourly wages, optimal tax rates come close to those observed in the data. In a final step, we choose the level of skill-bias in 2005 such that we come reasonably close to the moments in the data, while keeping the parameters of the ability distribution constant, but taking changes in policy and the costs of college into account. The outcomes of this exercise – including the calibrated values for skill-bias – are shown in Table 5. We observe that in order to bring the model close to the employment share of college graduates and the variance of log hourly wages in 2005, skill-bias is more than twice as large as in 1980. The moments are not matched perfectly, as a higher level of skill-bias would increase the share of college graduates, but would decrease the variation in wages. Although the college wage premium is largely overestimated by our model, the qualitative change between 1980 and 2005 goes in the right direction.

	1980	2005	Source
t	0.22	0.201	US Tax Policy Center
s	0.544	0.56	US National Center for Education Statistics
p	1	2.13	US National Center for Education Statistics

Table 3: Values taken from the data

Parameter	Value	
μ^θ	0.4	fixed
σ^θ	0.3887	calibrated to 1980 data
ω	0.45	fixed
σ	1.41	fixed
ε	0.3	fixed

Table 4: Parameters

for male workers.

	1980		2005	
	model	data	model	data
Level of skill bias A	1.2957		3.3584	
Variance of log hourly wages	0.3092	0.3092	0.3520	0.47
Share of college graduates	0.2090	0.21	0.2733	0.29
College premium as in Heathcote, Perri, and Violante (2010)	3.2105	1.44	3.2624	1.92

Table 5: Calibration outcomes

7.2 Policy experiment

Having calibrated our model, we can now ask the question what the optimal policy would have been in 1980 and 2005 – and what the potential welfare gains would be. In order to do so, we now take the parameters from Table 4 as well as the costs of college from Table 3 as given, and then optimize our model for the two levels of skill-bias which correspond to the years 1980 and 2005. The results are reported in Table 6 in which we compare the non-optimized model, in which the policy parameters are taken from the data, with the optimized model. We find that the marginal tax rate in the data is fairly close to the optimal value. Moreover, as in the data, the optimal tax rate decreases between 1980 and 2005. However, the drop in the optimal tax rate is less pronounced than the one observed in the data. When it comes to the optimal transfer and the optimal education subsidy, the differences between the data on the optimized model are more striking. The optimal education subsidy is less than half the empirical one. It is thus not optimal to subsidize the costs of college by more than 25%. Moreover, although there is a substantive increase in the costs of college between 1980 and 2005, the subsidy expressed as a fraction of the costs of college only increases marginally. However, we note that in absolute terms the optimal subsidy thus grows proportionately with the costs of college. As a result of the low optimal subsidy, the lump sum transfer received by every individual is much higher in the optimized model - which results in higher welfare. When comparing the labor market outcomes of the non-optimized with the optimized model, we find that the variance of log hourly wages should optimally be higher in 2005 than observed in the data, while the values in 1980 are fairly close to each other. To understand the source of this higher variance, it is indicative to compare the share of college graduates and the college premium between the non-optimized and the optimized model. By having a lower subsidy, the optimized model generates a higher share of college graduates than in the data in 1980, which corresponds to a lower college wage premium. In contrast, in 2005 the optimized model generates a lower share of college graduates than observed empirically. Due to general equilibrium effects, the wage rate of college graduates is therefore higher, which corresponds to a larger college wage premium.

Interestingly, the optimized model generates a larger increase in income inequality as measured by the dispersion of wages and the college wage premium than observed in the data – and this deviation from the data is welfare improving. The source of the rise in welfare is the larger lump sum transfer, which can be financed as less resources are spent on higher education subsidies. We highlight again that in our model there are two motives for having education subsidies – efficiency and redistribution. First, subsidies offset distortions on skill-formation which are induced by the income tax. Second, subsidies stimulate skill-formation which via general equilibrium effects compress the wage differential between college and high-school graduates and thus contribute to redistribution. However, neither the distortions on skill-formation nor the general equilibrium effects are strong enough to generate optimal subsidies as high as those observed in the data. Education subsidies are regressive in nature – and we observe too much re-distribution in the wrong direction in the data.

	1980		2005	
	not optimized	optimized	not optimized	optimized
b	0.0674	0.1662	0.2132	1.0064
s	0.5440	0.2238	0.5600	0.2244
t	0.2200	0.2334	0.2010	0.2321
Welfare (not yet in consumption equivalents)	0.4207	0.5171	0.9750	1.0064
Variance of log hourly wages	0.3092	0.3031	0.3520	0.3811
Share of college graduates	0.2090	0.2165	0.2733	0.2414
College premium as in Heathcote, Perri, and Violante (2010)	3.2105	3.1251	3.2624	3.6066

Table 6: Calibrated vs. optimized model

We acknowledge that our model is highly stylized and that there are other reasons for having education subsidies such as credit constraints and knowledge spillovers. However, in the absence of these factors the answer to SBTC is not to have ever increasing college subsidies. Still, rising costs of college justify increasing the total amount of subsidies, while the fraction of subsidies in total costs should stay roughly constant.

8 Conclusion

We study optimal linear income taxation and education subsidies in a model with heterogeneous individuals that acquire human capital on the extensive margin and supply labor on the intensive margin. We derive that optimal income taxes are positive. Income taxes are higher if distributional benefits increase and tax distortions on both labor supply and skill formation are lower.

Due to the discrete education choice, human capital is optimally taxed on a net basis. Net taxes on human capital redistribute ability rents from infra-marginal skilled workers to the low-skilled workers. However, this also distorts skill formation of the marginal skilled workers.

Distributional benefits of both income taxes and net taxes on education decrease in general equilibrium. Indeed, aggregate low-skilled labor supply increases and aggregate high-skilled labor supply decreases, which lowers low-skilled wages and raises high-skilled wages. As both policies raise pre-tax wage inequality, general-equilibrium effects cause distributional losses, which reduce their direct distributional gains. Although education is optimally taxed on a net basis, optimal education subsidies can be positive as they reduce tax distortions on skill formation. To what extent education is subsidized depends on how high income taxes and net education taxes should optimally be.

Focusing on the special case of uniformly distributed ability and linear technology, we study optimal policy responses to SBTC when the government can either adjust the tax rate (and the transfer) or the subsidy (and the transfer). We find that with SBTC the optimal tax rate increases at low levels of skill-bias, and decreases if skill-bias is high. In contrast, the optimal subsidy decreases at low levels of skill-bias. We proceed with simulations in order to study the effect of SBTC on policy when both the tax rate and the subsidy are optimized jointly. Using a log-normal ability distribution, we find that in the presence of general equilibrium effects the optimal income tax increases with SBTC, while the optimal subsidy falls – both contributing to an increase in the net tax on skill-formation. However, we also find that while SBTC increases the employment share of college graduates, the college wage premium actually falls. This result is at odds with the data.

In the case of linear technology the optimal tax rate first falls, to then increase and stabilize, while the optimal subsidy decreases throughout. The effect on the optimal net tax is as before. The reason for the different behavior of the tax rate originates in the behavior of the distortions on skill-formation. These distortions follow the reverse patter: they first increase with SBTC, inducing lower optimal tax rates, to then decrease and stabilize. With linear technology SBTC generates both an increase in the share of college graduates and a rising college wage premium. However, the assumption of college and high-school labor being perfect substitutes is implausible. We suggest an alternative explanation for the rising college wage premium: an increase in the costs of college. We demonstrate that the combination of SBTC and rising tuition fees can generate a rising college wage premium and an increasing share of college graduates while allowing for general equilibrium effects.

Finally, we calibrate our model to US data for the period 1980 and 2005 and conduct a policy experiment in which we ask the question how optimal policy should have looked like – and should have changed – if SBTC was the driving force behind rising inequality. We find that optimal tax rates generated by our model roughly coincide with those in the data. However, optimal subsidies should have been lower, both in 1980 and 2005. Moreover, the subsidy rate should have stayed roughly at the same level – a phenomenon actually observed in the data. Lower expenditures on subsidies result in higher transfers which are welfare improving. We conclude that the optimal policy response to SBTC is not to have ever increasing subsidy rates. In contrast, if the costs of college stay constant, subsidy rates should decrease. However, if the costs of college increase, subsidy rates might stay roughly constant – but are optimally lower than in the data. We acknowledge that our model is highly stylized. Especially the static nature might be restrictive. Optimal policy might respond differently to SBTC in an overlapping generations structure in which there are potential externalities of the education level of the old generation on economic choices of the young. Moreover, we highlight that in our model there is only one efficiency motive for having education subsidies: to offset distortions on skill-formation induced by the income tax. Our policy conclusions are likely to change in the presence of borowing constraints, especially if these interact with SBTC. A richer dynamic framework – potentially with borrowing constraints – is left to future research.

Appendix

A Derivation of elasticities

We define $\tilde{x} \equiv d \ln x/x$ as the relative change in variable x, with the exception of $\tilde{t} \equiv dt/(1-t)$. First, we log-linearize the labor-supply equations to obtain:

$$\tilde{h}_\theta = \varepsilon(\tilde{w}_H - \tilde{t}), \tag{27}$$

$$\tilde{l}_\theta = \varepsilon(\tilde{w}_L - \tilde{t}). \tag{28}$$

Next, we linearize the cut-off ability Θ to find:

$$\tilde{\Theta} = \tilde{t} - \rho\tilde{s} - \beta\tilde{w}_H - (1-\beta)\tilde{w}_L, \tag{29}$$

$$\rho \equiv \frac{s}{(1-s)(1+\varepsilon)}, \quad \beta \equiv \frac{w_H^{1+\varepsilon}}{w_H^{1+\varepsilon} - w_L^{1+\varepsilon}} > 1,$$

where ρ captures the importance of the subsidy in total costs of education, with a correction for ε to reduce notational burden later on. β is a measure for the skill premium in the initial equilibrium.

Next, we log-linearize the labor-market clearing conditions:

$$\tilde{H} = \varepsilon\left(\tilde{w}_H - \tilde{t}\right) - \delta_H\tilde{\Theta}, \quad \delta_H \equiv \frac{\Theta^2 h_\Theta f(\Theta)}{H}, \tag{30}$$

$$\tilde{L} = \varepsilon\left(\tilde{w}_L - \tilde{t}\right) + \delta_L\tilde{\Theta}, \quad \delta_L \equiv \frac{\Theta^2 l_\theta f(\Theta)}{L}. \tag{31}$$

Finally, we log-linearize the wage equations using the homogeneity of degree zero of the marginal product equations (i.e. $Y_{LL}L = -Y_{LH}H$ and $Y_{HH}H = -Y_{HL}L$) to find

$$\tilde{w}_H = \frac{(1-\alpha)}{\sigma}(\tilde{L} - \tilde{H}), \tag{32}$$

$$\tilde{w}_L = \frac{\alpha}{\sigma}(\tilde{H} - \tilde{L}), \tag{33}$$

$$\alpha \equiv \frac{HY_H(\cdot)}{Y(\cdot)}, \quad \frac{1}{\sigma} \equiv \frac{Y_{LH}(\cdot)Y(\cdot)}{Y_L(\cdot)Y_H(\cdot)}, \tag{34}$$

where α denotes the income share of the skilled worker in total output, and σ is the elasticity of substitution between low-skilled and high-skilled labor in production.

We now obtained a system of 7 equations – (27), (28), (29), (30), (31), (32), and (33) – in 7 unknowns $\left\{\tilde{h}_\theta, \tilde{l}_\theta, \tilde{\Theta}, \tilde{H}, \tilde{L}, \tilde{w}_H, \tilde{w}_L\right\}$. We solve this system for the relative changes in the tax rate $\tilde{t}$ and the subsidy rate $\tilde{s}$.

First, rewrite (30) and (31) by substituting out $\tilde{\Theta}$ using (29). Subtract both equations and define the composite variable $\delta \equiv \delta_L + \delta_H$ to find:

$$\tilde{H} - \tilde{L} = (\varepsilon + \beta\delta)\,\tilde{w}_H + (-\varepsilon + (1-\beta)\delta)\tilde{w}_L - \delta\tilde{t} + \delta\rho\tilde{s}. \tag{35}$$

Next, substitute $\tilde{w}_H$ from (32) and $\tilde{w}_L$ from (33) and solve for $\tilde{H} - \tilde{L}$ to find:

$$\tilde{H} - \tilde{L} = -\left(\frac{\sigma\delta}{\sigma + \varepsilon + \delta(\beta - \alpha)}\right)\tilde{t} + \left(\frac{\sigma\delta\rho}{\sigma + \varepsilon + \delta(\beta - \alpha)}\right)\tilde{s}. \tag{36}$$

The denominator is positive ($\sigma + \varepsilon + \delta(\beta - \alpha) > 0$) because $\beta > 1$, $\alpha < 1$ and all other parameters are positive.

Substituting (36) in (32) and (33) yields the linearized high-skilled and low-skilled wage rates:

$$\tilde{w}_H = \left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\tilde{t} - \left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho\tilde{s}, \tag{37}$$

$$\tilde{w}_L = -\left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\tilde{t} + \left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho\tilde{s}. \tag{38}$$

Substituting (37) and (38) in (29) gives:

$$\tilde{\Theta} = \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\tilde{t} - \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho\tilde{s}. \tag{39}$$

Substitution of (37) and (38) in (27) and (28) yields the linearized labor supplies:

$$\tilde{h} = -\left(\frac{\sigma+\varepsilon+\delta(\beta-1)}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon\tilde{t} - \left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon\rho\tilde{s}, \tag{40}$$

$$\tilde{l} = -\left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon\tilde{t} + \left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon\rho\tilde{s}. \tag{41}$$

Finally, the linearized equations for aggregate high-skilled and low-skilled labor are given by:

$$\begin{aligned}\tilde{H} = &-\left[\left(1-\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon + \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\delta_H\right]\tilde{t} \\ &+\left[\left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon + \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\delta_H\right]\rho\tilde{s}.\end{aligned} \tag{42}$$

$$\begin{aligned}\tilde{L} = &-\left[\left(1+\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon - \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\delta_L\right]\tilde{t} \\ &+\left[\left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon - \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\delta_L\right]\rho\tilde{s}.\end{aligned} \tag{43}$$

We can now find explicit expressions for the tax elasticities by setting $\tilde{s} = \tilde{A} = 0$ and defining

$$\varepsilon_{\Theta,t} \equiv \frac{\partial\Theta}{\partial t}\frac{1-t}{\Theta} = \frac{\tilde{\Theta}}{\tilde{t}} = \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right) > 0, \tag{44}$$

$$\varepsilon_{w_L,t} \equiv -\frac{\partial w_L}{\partial t}\frac{1-t}{w_L} = \frac{\tilde{w}_L}{\tilde{t}} = \left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right) > 0, \tag{45}$$

$$\varepsilon_{w_H,t} \equiv -\frac{\partial w_H}{\partial t}\frac{1-t}{w_H} = \frac{\tilde{w}_H}{\tilde{t}} = -\left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right) < 0. \tag{46}$$

$$\varepsilon_{l,t} \equiv -\frac{\partial l_\theta}{\partial t}\frac{1-t}{l_\theta} = \frac{\tilde{l}}{\tilde{t}} = \varepsilon\left(1+\varepsilon_{w_H,t}\right) = \left(\frac{\sigma+\varepsilon+\delta\beta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon > 0, \tag{47}$$

$$\varepsilon_{h,t} \equiv -\frac{\partial h_\theta}{\partial t}\frac{1-t}{h_\theta} = \frac{\tilde{h}}{\tilde{t}} = \varepsilon\left(1+\varepsilon_{w_L,t}\right) = \left(\frac{\sigma+\varepsilon+\delta(\beta-1)}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon > 0. \tag{48}$$

Similarly, we obtain the subsidy elasticities by setting $\tilde{t} = \tilde{A} = 0$ and defining

$$\varepsilon_{\Theta,s} \equiv -\frac{\partial \Theta}{\partial s}\frac{s}{\Theta} = \frac{\tilde{\Theta}}{\tilde{s}} = \left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho > 0, \tag{49}$$

$$\varepsilon_{w_L,s} \equiv \frac{\partial w_L}{\partial s}\frac{s}{w_L} = \frac{\tilde{w}_L}{\tilde{s}} = \left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho > 0, \tag{50}$$

$$\varepsilon_{w_H,s} \equiv \frac{\partial w_H}{\partial s}\frac{s}{w_H} = -\frac{\tilde{w}_H}{\tilde{s}} = \left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho < 0, \tag{51}$$

$$\varepsilon_{l,s} \equiv \frac{\partial l_\theta}{\partial s}\frac{s}{l_\theta} = \frac{\tilde{l}}{\tilde{s}} = \varepsilon\varepsilon_{w_L,s} = \left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon\rho > 0, \tag{52}$$

$$\varepsilon_{h,s} \equiv \frac{\partial h_\theta}{\partial s}\frac{s}{h_\theta} = \frac{\tilde{h}}{\tilde{s}} = \varepsilon\varepsilon_{w_H,s} = -\left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\varepsilon\rho < 0. \tag{53}$$

We note that in the case of ABTC, $\varepsilon^{ABTC}_{\Theta,t} = \frac{1}{A}\varepsilon_{\Theta,t}$, and $\varepsilon^{ABTC}_{\Theta,s} = \frac{1}{A}\varepsilon_{\Theta,s}$.

B Optimal policy

Introducing η as the Lagrange multiplier on the government budget constraint, we can formulate the Lagrangian for deriving optimal policy as:

$$\begin{aligned}\max_{b,t,s} \mathcal{L} \equiv & \int_{\underline{\theta}}^{\Theta} \gamma_\theta V^L_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta V^H_\theta dF(\theta) \\ & +\eta\left[\int_{\underline{\theta}}^{\Theta} t w_L \theta l_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} (t w_H \theta h_\theta - sp)\, dF(\theta) - b - G\right],\end{aligned} \tag{54}$$

Necessary, first-order conditions for an optimum are given by:

$$\frac{\partial \mathcal{L}}{\partial b} = \int_{\underline{\theta}}^{\Theta} \gamma_\theta \frac{\partial V^L_\theta}{\partial b} dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta \frac{\partial V^H_\theta}{\partial b} dF(\theta) - \eta = 0, \tag{55}$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t} = & \int_{\underline{\theta}}^{\Theta} \gamma_\theta \frac{\partial V^L_\theta}{\partial t} dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta \frac{\partial V^H_\theta}{\partial t} dF(\theta) + \eta\left[\int_{\underline{\theta}}^{\Theta} w_L \theta l_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} w_H \theta h dF(\theta)\right] \\ & + \eta\left[\int_{\underline{\theta}}^{\Theta} t w_L \theta \frac{\partial l_\theta}{\partial t} dF(\theta) + \int_{\Theta}^{\overline{\theta}} t w_H \theta \frac{\partial h_\theta}{\partial t} dF(\theta)\right] \\ & + \eta\left[\int_{\underline{\theta}}^{\Theta} t \frac{\partial w_L}{\partial t} \theta l_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} t \frac{\partial w_H}{\partial t} \theta h_\theta dF(\theta)\right] \\ & + \underbrace{\left[\gamma^L_\Theta V^L_\Theta - \gamma^H_\Theta V^H_\Theta\right]}_{=0} f(\Theta)\frac{\partial \Theta}{\partial t} - \eta\left[t w_H \Theta h_\Theta - t w_L \Theta l_\Theta - sp\right] f(\Theta)\frac{\partial \Theta}{\partial t} = 0,\end{aligned} \tag{56}$$

$$\frac{\partial \mathcal{L}}{\partial s} = \int_{\underline{\theta}}^{\Theta} \gamma_\theta \frac{\partial V_\theta^L}{\partial s} dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta \frac{\partial V_\theta^H}{\partial s} dF(\theta) - \eta \left[\int_{\Theta}^{\overline{\theta}} dF(\theta) \right] \tag{57}$$
$$+ \eta \left[\int_{\underline{\theta}}^{\Theta} t w_L \theta \frac{\partial l_\theta}{\partial s} dF(\theta) + \int_{\Theta}^{\overline{\theta}} t w_H \theta \frac{\partial h_\theta}{\partial s} dF(\theta) \right]$$
$$+ \eta \left[\int_{\underline{\theta}}^{\Theta} t \frac{\partial w_L}{\partial s} \theta l_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} t \frac{\partial w_H}{\partial s} \theta h_\theta dF(\theta) \right]$$
$$+ \underbrace{\left[\gamma_\Theta^L V_\Theta^L - \gamma_\Theta^H V_\Theta^H \right]}_{=0} f(\Theta) \frac{\partial \Theta}{\partial s} - \eta \left[t w_H \Theta h_\Theta - t w_L \Theta l_\Theta - sp \right] f(\Theta) \frac{\partial \Theta}{\partial s} = 0.$$

Note that $V_\Theta^L = V_\Theta^H$ because the marginal graduate Θ is indifferent between being high-skilled or low-skilled.

Next, use Roy's lemma to derive that

$$\frac{\partial V_\theta^i}{\partial b} = 1, \tag{58}$$
$$\frac{\partial V_\theta^H}{\partial t} = -\theta w_H h_\theta + (1-t)\, \theta h_\theta \frac{\partial w_H}{\partial t}, \tag{59}$$
$$\frac{\partial V_\theta^L}{\partial t} = -\theta w_L l_\theta + (1-t)\, \theta l_\theta \frac{\partial w_L}{\partial t}, \tag{60}$$
$$\frac{\partial V_\theta^H}{\partial s} = p + (1-t)\, \theta h_\theta \frac{\partial w_H}{\partial s}, \tag{61}$$
$$\frac{\partial V_\theta^L}{\partial s} = (1-t)\, \theta l_\theta \frac{\partial w_L}{\partial s}. \tag{62}$$

Recall that the net tax wedge on skill formation is defined as $\Delta \equiv t w_H \Theta h_\Theta - t w_L \Theta l_\Theta - s$. And we define $g_\theta \equiv \gamma_\theta / \eta$ as the social welfare weight of individual θ, where g_θ gives the monetized value of providing this individual with an additional Euro. Therefore, we can simplify the first-order conditions as:

$$\frac{\partial \mathcal{L}}{\partial b} = 0 : \int_{\underline{\theta}}^{\Theta} \frac{\gamma_\theta}{\eta} dF(\theta) + \int_{\Theta}^{\overline{\theta}} \frac{\gamma_\theta}{\eta} dF(\theta) = \int_{\underline{\theta}}^{\Theta} g_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} g_\theta dF(\theta) = 1. \tag{63}$$

$$\frac{\partial \mathcal{L}}{\partial t} = \int_{\underline{\theta}}^{\Theta} \gamma_\theta \left(-\theta w_L l_\theta + (1-t)\, \theta l_\theta \frac{\partial w_L}{\partial t} \right) dF(\theta) \tag{64}$$
$$+ \int_{\Theta}^{\overline{\theta}} \gamma_\theta \left(-\theta w_H h_\theta + (1-t)\, \theta h_\theta \frac{\partial w_H}{\partial t} \right) dF(\theta)$$
$$+ \eta \left[\int_{\underline{\theta}}^{\Theta} w_L \theta l_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} w_H \theta h_\theta dF(\theta) \right]$$
$$+ \eta \left[\int_{\underline{\theta}}^{\Theta} t w_L \theta \frac{\partial l_\theta}{\partial t} dF(\theta) + \int_{\Theta}^{\overline{\theta}} t w_H \theta \frac{\partial h_\theta}{\partial t} dF(\theta) \right]$$
$$+ \eta \left[\int_{\underline{\theta}}^{\Theta} t \frac{\partial w_L}{\partial t} \theta l_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} t \frac{\partial w_H}{\partial t} \theta h_\theta dF(\theta) \right] - \eta \frac{\Delta}{1-t} f(\Theta)\, \Theta \frac{\partial \Theta}{\partial t} \frac{1-t}{\Theta} = 0,$$

$$\frac{\partial \mathcal{L}}{\partial s} = \int_{\underline{\theta}}^{\Theta} \gamma_\theta \left((1-t)\,\theta l_\theta \frac{\partial w_L}{\partial s} \right) dF(\theta) \tag{65}$$

$$+ \int_{\Theta}^{\overline{\theta}} \gamma_\theta \left(1 + (1-t)\,\theta h_\theta \frac{\partial w_H}{\partial s} \right) dF(\theta)$$

$$- \eta \left[\int_{\Theta}^{\overline{\theta}} dF(\theta) \right] + \eta \left[\int_{\underline{\theta}}^{\Theta} t w_L \theta \frac{\partial l_\theta}{\partial s} dF(\theta) + \int_{\Theta}^{\overline{\theta}} t w_H \theta \frac{\partial h_\theta}{\partial s} dF(\theta) \right]$$

$$+ \eta \left[\int_{\underline{\theta}}^{\Theta} t \frac{\partial w_L}{\partial s} \theta l_\theta dF(\theta) + \int_{\Theta}^{\overline{\theta}} t \frac{\partial w_H}{\partial s} \theta h_\theta dF(\theta) \right] - \eta \frac{\Delta}{s} \Theta f(\Theta) \frac{\partial \Theta}{\partial s} \frac{s}{\Theta} = 0.$$

We will simplify the first-order conditions for t and s in a number of steps.

B.1 Optimal income tax

Rewrite the first-order condition for t using the definitions for $z_\theta^L \equiv w_L \theta l_\theta$ and $z_\theta^H \equiv w_H \theta h_\theta$ to find:

$$\frac{\partial \mathcal{L}}{\partial t} = - \left[\int_{\underline{\theta}}^{\Theta} \gamma_\theta z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta z_\theta^H dF(\theta) \right] + \eta \left[\int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} z_\theta^H dF(\theta) \right] \tag{66}$$

$$+ \frac{t}{1-t} \eta \left[\int_{\underline{\theta}}^{\Theta} z_\theta^L \frac{\partial l_\theta}{\partial t} \frac{1-t}{l_\theta} dF(\theta) + \int_{\Theta}^{\overline{\theta}} z_\theta^H \frac{\partial h_\theta}{\partial t} \frac{1-t}{h_\theta} dF(\theta) \right]$$

$$+ \int_{\underline{\theta}}^{\Theta} \left[\gamma_\theta + \eta \frac{t}{1-t} \right] z_\theta^L \frac{\partial w_L}{\partial t} \frac{1-t}{w_L} dF(\theta) + \int_{\Theta}^{\overline{\theta}} \left[\gamma_\theta + \eta \frac{t}{1-t} \right] z_\theta^H \frac{\partial w_H}{\partial t} \frac{1-t}{w_H} dF(\theta)$$

$$- \eta \frac{\Delta}{1-t} f(\Theta)\, \Theta \frac{\partial \Theta}{\partial t} \frac{1-t}{\Theta} = 0.$$

And, simplify the first-order condition for t using the definitions of elasticities:

$$\frac{\partial \mathcal{L}}{\partial t} = - \left[\int_{\underline{\theta}}^{\Theta} \gamma_\theta z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta z_\theta^H dF(\theta) \right] + \eta \left[\int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} z_\theta^H dF(\theta) \right] \tag{67}$$

$$- \frac{t}{1-t} \eta \left[\int_{\underline{\theta}}^{\Theta} z_\theta^L \varepsilon_{l,t} dF(\theta) + \int_{\Theta}^{\overline{\theta}} z_\theta^H \varepsilon_{h,t} dF(\theta) \right] - \int_{\underline{\theta}}^{\Theta} \left[\gamma_\theta + \eta \frac{t}{1-t} \right] z_\theta^L \varepsilon_{w_L,t} dF(\theta)$$

$$- \int_{\Theta}^{\overline{\theta}} \left[\gamma_\theta + \eta \frac{t}{1-t} \right] z_\theta^H \varepsilon_{w_H,t} dF(\theta) - \eta \frac{\Delta}{1-t} f(\Theta)\, \Theta \varepsilon_{\Theta,t} = 0.$$

Important to note here is that all elasticities are independent of θ (they do depend on Θ, however). Hence, they can all be taken out of the integral signs. Next, we define average incomes of the low- and high-skilled

$$\bar{z}^L \equiv \int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta), \quad \bar{z}^H \equiv \int_{\Theta}^{\overline{\theta}} z_\theta^H dF(\theta). \tag{68}$$

Dividing (67) by η and substituting for the definitions, we obtain

$$- \left[\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} g_\theta z_\theta^H dF(\theta) \right] + \bar{z}^L + \bar{z}^H - \frac{t}{1-t} \left[\varepsilon_{l,t} \bar{z}^L + \varepsilon_{h,t} \bar{z}^H \right] \tag{69}$$

$$- \varepsilon_{w_L,t} \int_{\underline{\theta}}^{\Theta} \left[g_\theta + \frac{t}{1-t} \right] z_\theta^L dF(\theta) - \varepsilon_{w_H,t} \int_{\Theta}^{\overline{\theta}} \left[g_\theta + \frac{t}{1-t} \right] z_\theta^H dF(\theta)$$

$$- \frac{\Delta}{1-t} f(\Theta)\, \Theta \varepsilon_{\Theta,t} = 0.$$

Next, define the distributional characteristic of labor income as:

$$\xi \equiv 1 - \frac{\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} g_\theta z_\theta^H dF(\theta)}{[\bar{z}^L + \bar{z}^H] \int_{\underline{\theta}}^{\overline{\theta}} g_\theta dF(\theta)}. \tag{70}$$

Note also that $\bar{z} = \bar{z}^L + \bar{z}^H$ and $w_L L = \bar{z}^L$ and $w_H H = \bar{z}^H$ so that we can write for the income shares:

$$\alpha = \frac{\bar{z}^H}{\bar{z}^L + \bar{z}^H}, \quad 1 - \alpha = \frac{\bar{z}^L}{\bar{z}^L + \bar{z}^H} \tag{71}$$

Hence, the optimal income tax expression can be written as

$$\begin{aligned} \xi = & \frac{t}{1-t}\left[(1-\alpha)(\varepsilon_{l,t} + \varepsilon_{w_L,t}) + \alpha(\varepsilon_{h,t} + \varepsilon_{w_H,t})\right] + \frac{\Delta}{1-t}\frac{f(\Theta)\Theta}{\bar{z}}\varepsilon_{\Theta,t} \\ & + \varepsilon_{w_L,t}\frac{\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta)}{[\bar{z}^L + \bar{z}^H]} + \varepsilon_{w_H,t}\frac{\int_{\Theta}^{\overline{\theta}} g_\theta z_\theta^H dF(\theta)}{[\bar{z}^L + \bar{z}^H]}. \end{aligned} \tag{72}$$

Use substitute income-weighted social welfare weights of each skill group: $\bar{g}^L \equiv \int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta)/\bar{z}^L$ and $\bar{g}^H \equiv \int_{\Theta}^{\overline{\theta}} g_\theta z_\theta^H dF(\theta)/\bar{z}^H$ so that the optimal-tax expression is given by :

$$\begin{aligned} & \frac{t}{1-t}\left[(1-\alpha)(\varepsilon_{l,t} + \varepsilon_{w_L,t}) + \alpha(\varepsilon_{h,t} + \varepsilon_{w_H,t})\right] + \frac{\Delta}{(1-t)}\frac{f(\Theta)\Theta}{\bar{z}}\varepsilon_{\Theta,t} \\ & = \xi - \varepsilon_{w_H,t}\alpha\bar{g}^H - \varepsilon_{w_L,t}(1-\alpha)\bar{g}^L. \end{aligned} \tag{73}$$

Finally, substitute for the elasticities to find:

$$\frac{t}{(1-t)}\varepsilon + \frac{\Delta}{(1-t)}\frac{f(\Theta)\Theta}{\bar{z}}\left(\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right) = \xi - \frac{(1-\alpha)\alpha\delta}{(\sigma+\varepsilon+\delta(\beta-\alpha))}(\bar{g}^L - \bar{g}^H). \tag{74}$$

B.2 Optimal net tax on education

Using similar steps as a above we will rewrite the optimal education subsidy using the definitions for $z_\theta^L \equiv w_L \theta l_\theta$ and $z_\theta^H \equiv w_H \theta h_\theta$ to find:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} = & \int_{\underline{\theta}}^{\Theta} \gamma_\theta \left(\frac{(1-t)}{s} z_\theta^L \frac{\partial w_L}{\partial s}\frac{s}{w_L}\right) dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta \left(p + \frac{(1-t)}{s} z_\theta^H \frac{\partial w_H}{\partial s}\frac{s}{w_H}\right) dF(\theta) \\ & - \eta\left[p\int_{\Theta}^{\overline{\theta}} dF(\theta)\right] + \eta\left[\int_{\underline{\theta}}^{\Theta} \frac{t}{s} z_\theta^L \frac{\partial l_\theta}{\partial s}\frac{s}{l_\theta} dF(\theta) + \int_{\Theta}^{\overline{\theta}} \frac{t}{s} z_\theta^H \frac{\partial h_\theta}{\partial s}\frac{s}{h_\theta} dF(\theta)\right] \\ & + \eta\left[\int_{\underline{\theta}}^{\Theta} \frac{t}{s}\frac{\partial w_L}{\partial s}\frac{s}{w_L} z_\theta^L dF(\theta) + \int_{\Theta}^{\overline{\theta}} \frac{t}{s}\frac{\partial w_H}{\partial s}\frac{s}{w_H} z_\theta^H dF(\theta)\right] - \eta\frac{\Delta}{s}\Theta f(\Theta)\frac{\partial\Theta}{\partial s}\frac{s}{\Theta} = 0. \end{aligned} \tag{75}$$

Simplify the first-order condition for s using the definitions of the subsidy elasticities:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} = & \int_{\underline{\theta}}^{\Theta} \gamma_\theta \left(\frac{(1-t)}{s} z_\theta^L \varepsilon_{w_L,s}\right) dF(\theta) + \int_{\Theta}^{\overline{\theta}} \gamma_\theta \left(p + \frac{(1-t)}{s} z_\theta^H \varepsilon_{w_H,s}\right) dF(\theta) \\ & - \eta p \int_{\Theta}^{\overline{\theta}} dF(\theta) + \eta\left[\frac{t}{s}(\varepsilon_{l,s} + \varepsilon_{w_L,s})\bar{z}^L + \frac{t}{s}(\varepsilon_{h,s} + \varepsilon_{w_H,s})\bar{z}^H\right] - \eta\frac{\Delta}{s}\Theta f(\Theta)\varepsilon_{\Theta,s} = 0. \end{aligned} \tag{76}$$

All elasticities are independent from θ (they do depend on Θ). Hence, they can be taken out of the integral signs. Consequently, we get after dividing by η and multiplication with $s/(1-t)$:

$$\begin{aligned}&\varepsilon_{w_L,s}\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta)+\varepsilon_{w_H,s}\int_{\Theta}^{\overline{\theta}} g_\theta z_\theta^H dF(\theta)-\frac{s}{1-t}p\int_{\Theta}^{\overline{\theta}}(1-g_\theta)dF(\theta)\\&+\frac{t}{1-t}\varepsilon_{l,s}\bar{z}^L+\frac{t}{1-t}\varepsilon_{h,s}\bar{z}^H+\frac{t}{1-t}\varepsilon_{w_L,s}\bar{z}^L+\frac{t}{1-t}\varepsilon_{w_H,s}\bar{z}^H-\frac{\Delta}{1-t}\Theta f(\Theta)\,\varepsilon_{\Theta,s}=0.\end{aligned}\tag{77}$$

Divide by $\bar{z}$, use $\bar{g}^L\equiv\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta)/\bar{z}^L$ and $\bar{g}^H\equiv\int_{\Theta}^{\overline{\theta}} g_\theta z_\theta^H dF(\theta)/\bar{z}^H$ and the definition of α to write

$$\begin{aligned}&\varepsilon_{w_L,s}(1-\alpha)\bar{g}^L+\varepsilon_{w_H,s}\alpha\bar{g}^H-\frac{1}{\bar{z}}\frac{s}{1-t}p\int_{\Theta}^{\overline{\theta}}(1-g_\theta)dF(\theta)\\&+\frac{t}{1-t}\varepsilon_{l,s}(1-\alpha)+\frac{t}{1-t}\varepsilon_{h,s}\alpha+\frac{t}{1-t}\varepsilon_{w_L,s}(1-\alpha)\\&+\frac{t}{1-t}\varepsilon_{w_H,s}(\alpha)+\frac{1}{\bar{z}}\frac{\Delta}{1-t}\Theta f(\Theta)\,\varepsilon_{\Theta,s}=0.\end{aligned}\tag{78}$$

Collect terms to arrive at

$$\begin{aligned}&\varepsilon_{w_L,s}(1-\alpha)\bar{g}^L+\varepsilon_{w_H,s}\alpha\bar{g}^H-\frac{1}{\bar{z}}\frac{s}{1-t}p\int_{\Theta}^{\overline{\theta}}(1-g_\theta)dF(\theta)\\&+\frac{t}{1-t}(1-\alpha)(\varepsilon_{l,s}+\varepsilon_{w_L,s})+\frac{t}{1-t}\alpha(\varepsilon_{h,s}+\varepsilon_{w_H,s})+\frac{1}{\bar{z}}\frac{\Delta}{1-t}\Theta f(\Theta)\,\varepsilon_{\Theta,s}=0.\end{aligned}\tag{79}$$

Now, we use the elasticities to substitute

$$\begin{aligned}\left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho(1-\alpha)\bar{g}^L-\left(\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho\alpha\bar{g}^H=\\\left(\frac{\alpha(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho\left(\bar{g}^L-\bar{g}^H\right),\end{aligned}\tag{80}$$

and

$$\begin{aligned}(1-\alpha)(\varepsilon_{l,s}+\varepsilon_{w_L,s})=\\(1-\alpha)\left(\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\varepsilon\rho+\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\rho\right)=\\(1-\alpha)(1+\varepsilon)\frac{\alpha\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\rho,\end{aligned}\tag{81}$$

and

$$\begin{aligned}\alpha(\varepsilon_{h,s}+\varepsilon_{w_H,s})=\\\alpha\left(-\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\varepsilon\rho-\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\rho\right)=\\-\alpha(1+\varepsilon)\frac{(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}.\end{aligned}\tag{82}$$

Thus

$$\frac{t}{1-t}\left(1-\alpha\right)\left(\varepsilon_{l,s}+\varepsilon_{w_L,s}\right)+\frac{t}{1-t}\alpha\left(\varepsilon_{h,s}+\varepsilon_{w_H,s}\right)=0. \tag{83}$$

The condition then simplifies to

$$\left(\frac{\alpha\left(1-\alpha\right)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho\left(\overline{g}^{L}-\overline{g}^{H}\right)-\frac{1}{\overline{z}}\frac{s}{1-t}p\int_{\Theta}^{\overline{\theta}}(1-g_{\theta})dF(\theta) \tag{84}$$
$$+\frac{1}{\overline{z}}\frac{\Delta}{1-t}\Theta f\left(\Theta\right)\varepsilon_{\Theta,s}=0.$$

Substituting for $\varepsilon_{\Theta,s}$ leads to

$$\left(\frac{\alpha\left(1-\alpha\right)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)}\right)\rho\left(\overline{g}^{L}-\overline{g}^{H}\right)-\frac{1}{\overline{z}}\frac{s}{1-t}p\int_{\Theta}^{\overline{\theta}}(1-g_{\theta})dF(\theta) \tag{85}$$
$$+\frac{\Delta}{1-t}\frac{\Theta f\left(\Theta\right)}{\overline{z}}\frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)}\rho=0.$$

And writing it in a more compact form again, using

$$\varepsilon_{GE}\equiv(1-\alpha)\varepsilon_{w_L,t}=-\alpha\varepsilon_{wH,t}=\frac{\alpha(1-\alpha)\delta}{(\sigma+\varepsilon+\delta(\beta-\alpha)},$$

we obtain

$$\varepsilon_{GE}\left(\overline{g}^{L}-\overline{g}^{H}\right)-\frac{1}{\overline{z}}\frac{s}{1-t}p\int_{\Theta}^{\overline{\theta}}(1-g_{\theta})dF(\theta) \tag{86}$$
$$+\frac{\Delta}{1-t}\frac{\Theta f\left(\Theta\right)}{\overline{z}}\varepsilon_{\Theta,s}=0.$$

which can be rearranged to

$$\frac{\Delta}{1-t}\frac{\Theta f\left(\Theta\right)}{\overline{z}}\varepsilon_{\Theta,s}=\frac{1}{\overline{z}}\frac{s}{1-t}p\int_{\Theta}^{\overline{\theta}}(1-g_{\theta})dF(\theta)-\rho\left(\overline{g}^{L}-\overline{g}^{H}\right)\varepsilon_{GE}.$$

We now define the distributional characteristic of the education subsidy as

$$\zeta\equiv p\int_{\Theta}^{\overline{\theta}}(1-g_{\theta})dF(\theta),$$

and obtain

$$\frac{\Delta}{1-t}\frac{\Theta f\left(\Theta\right)}{\overline{z}}\varepsilon_{\Theta,s}=\frac{1}{\overline{z}}\frac{s}{1-t}\zeta-\rho\left(\overline{g}^{L}-\overline{g}^{H}\right)\varepsilon_{GE}.$$

B.3 ξ only a function of Θ

Proof. We define the distributional characteristic of the low-skill sector, ξ_L, as minus the normalized covariance between welfare weights and incomes in the low-skill sector

$$\xi_L\equiv-\frac{\text{cov}\left[g_{\theta}^{L},z_{\theta}^{L}\right]}{\tilde{g}^{L}\tilde{z}^{L}}=-\frac{w_L\text{cov}\left[g_{\theta}^{L},h_{\theta}^{L}\theta\right]}{\tilde{g}^{L}\tilde{z}^{L}},$$

where $\tilde{g}^{L}$ and $\tilde{z}^{L}$

$$\tilde{g}^{L}\equiv\frac{1}{F\left(\Theta\right)}\int_{\underline{\theta}}^{\Theta}g_{\theta}dF\left(\theta\right),$$

$$\tilde{z}^L \equiv \frac{1}{F(\Theta)} \int_{\underline{\theta}}^{\Theta} z_\theta dF(\theta) = \frac{1}{F(\Theta)} w_L \int_{\underline{\theta}}^{\Theta} h_\theta^L \theta dF(\theta).$$

The distributional characteristic ξ_L is thus not directly dependent of w_L, and is a function of Θ only. We thus write $\xi_L(\Theta)$. Analogously , we define the distributional characteristic of the high-skill sector as

$$\xi_H = -\frac{\operatorname{cov}\left[g_\theta^H, z_\theta^H\right]}{\tilde{g}^H \tilde{z}^H} = -\frac{w_H \operatorname{cov}\left[g_\theta^H, h_\theta^H \theta\right]}{\tilde{g}^H \tilde{z}^H},$$

with

$$\tilde{g}^H \equiv \frac{1}{1-F(\Theta)} \int_{\Theta}^{\overline{\theta}} g_\theta dF(\theta),$$

$$\tilde{z}^L \equiv \frac{1}{1-F(\Theta)} \int_{\Theta}^{\overline{\theta}} z_\theta dF(\theta) = \frac{1}{F(\Theta)} w_H \int_{\Theta}^{\overline{\theta}} h_\theta^H \theta dF(\theta),$$

and we can thus write $\xi_H(\Theta)$.

Next, we rewrite $\xi_L(\Theta)$ as

$$\begin{aligned}
\xi_L(\Theta) = -\frac{\operatorname{cov}\left[g_\theta^L, z_\theta^L\right]}{\tilde{g}^L \tilde{z}^L} &= -\frac{\frac{1}{F(\Theta)} \int_{\underline{\theta}}^{\Theta} \left(g_\theta - \tilde{g}^L\right)\left(z_\theta - \tilde{z}^L\right) dF(\theta)}{\tilde{g}^L \tilde{z}^L} \\
&= -\frac{1}{F(\Theta)} \frac{1}{\tilde{g}^L \tilde{z}^L} \left(\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta dF(\theta) - \tilde{g}^L \int_{\underline{\theta}}^{\Theta} z_\theta dF(\theta) \right. \\
&\quad \left. - \tilde{z}^L \int_{\underline{\theta}}^{\Theta} g_\theta dF(\theta) + F(\Theta) \tilde{g}^L \tilde{z}^L \right) \\
&= -\frac{1}{\tilde{g}^L \tilde{z}^L} \left(\frac{1}{F(\Theta)} \int_{\underline{\theta}}^{\Theta} g_\theta z_\theta dF(\theta) - \tilde{g}^L \tilde{z}^L - \tilde{z}^L \tilde{g}^L + \tilde{g}^L \tilde{z}^L \right) \\
&= -\frac{1}{\tilde{g}^L \tilde{z}^L} \left(\frac{1}{F(\Theta)} \int_{\underline{\theta}}^{\Theta} g_\theta z_\theta dF(\theta) - \tilde{g}^L \tilde{z}^L \right) \\
&= 1 - \frac{1}{F(\Theta)} \frac{\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta dF(\theta)}{\tilde{g}^L \tilde{z}^L}.
\end{aligned}$$

If we now define

$$\overline{g}^L \equiv \int_{\underline{\theta}}^{\Theta} g_\theta dF(\theta)$$

and

$$\overline{z}^L \equiv \int_{\underline{\theta}}^{\Theta} z_\theta dF(\theta),$$

we have

$$\tilde{g}^L = \frac{1}{F(\Theta)} \overline{g}^L, \ \tilde{z}^L = \frac{1}{F(\Theta)} \overline{z}^L.$$

And thus

$$\xi_L(\Theta) = -\frac{\operatorname{cov}\left[g_\theta^L, z_\theta^L\right]}{\tilde{g}^L \tilde{z}^L} = 1 - F(\Theta) \frac{\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta dF(\theta)}{\overline{g}^L \overline{z}^L}.$$

Similarly, we can rewrite

$$\xi_H(\Theta) = 1 - (1 - F(\Theta)) \frac{\int_{\Theta}^{\overline{\theta}} g_\theta z_\theta dF(\theta)}{\overline{g}^H \overline{z}^H}.$$

Using the expressions for $\xi_L(\Theta)$ and $\xi_H(\Theta)$, we obtain

$$\begin{aligned}
\frac{1}{F(\Theta)}\frac{\tilde{g}^L\tilde{z}^L}{\overline{gz}}\xi_L(\Theta)+\frac{1}{1-F(\Theta)}\frac{\tilde{g}^H\tilde{z}^H}{\overline{gz}}\xi_H(\Theta) =& \frac{1}{F(\Theta)}\frac{\tilde{g}^L\tilde{z}^L}{\overline{gz}}+\frac{1}{1-F(\Theta)}\frac{\tilde{g}^H\tilde{z}^H}{\overline{gz}} \\
&-\frac{\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta dF(\theta)+\int_{\Theta}^{\overline{\theta}} g_\theta z_\theta dF(\theta)}{\overline{gz}} \\
=& \frac{1}{F(\Theta)}\frac{\tilde{g}^L\tilde{z}^L}{\overline{gz}}+\frac{1}{1-F(\Theta)}\frac{\tilde{g}^H\tilde{z}^H}{\overline{gz}}+\xi-1
\end{aligned}$$

and thus

$$\begin{aligned}
\xi &= 1+\frac{1}{F(\Theta)}\frac{\tilde{g}^L\tilde{z}^L}{\overline{gz}}(\xi_L(\Theta)-1)+\frac{1}{1-F(\Theta)}\frac{\tilde{g}^H\tilde{z}^H}{\overline{gz}}(\xi_H(\Theta)-1) \\
&= 1-\frac{1}{F(\Theta)}\frac{\tilde{g}^L\tilde{z}^L}{\overline{gz}}(1-\xi_L(\Theta))-\frac{1}{1-F(\Theta)}\frac{\tilde{g}^H\tilde{z}^H}{\overline{gz}}(1-\xi_H(\Theta)).
\end{aligned}$$

All expressions involved are independent of w_L and w_H and we therefore have that ξ is a function of Θ only. □

B.4 Behavior of ξ under SBTC

The distributional characteristic ξ is given by

$$\xi = 1-\frac{\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta)+\int_{\Theta}^{\overline{\theta}} g_\theta z_\theta^H dF(\theta)}{\overline{z}\overline{g}}.$$

The expressions for incomes are now

$$z_\theta^L = l_\theta\theta w_L = (1-t)^\varepsilon\theta^{1+\varepsilon},$$

and

$$z_\theta^H = h_\theta\theta w_H = (1-t)^\varepsilon A^{1+\varepsilon}\theta^{1+\varepsilon}.$$

We can then write

$$\begin{aligned}
\xi &= 1-\frac{(1-t)^\varepsilon\int_{\underline{\theta}}^{\Theta}\theta^{1-r+\varepsilon}dF(\theta)+(1-t)^\varepsilon A^{1+\varepsilon}\int_{\Theta}^{\overline{\theta}}\theta^{1-r+\varepsilon}dF(\theta)}{\overline{z}\overline{g}} \\
&= 1-\frac{(1-t)^\varepsilon\left[\frac{1}{2-r+\varepsilon}\theta^{2-r+\varepsilon}\right]_0^\Theta+(1-t)^\varepsilon A^{1+\varepsilon}\left[\frac{1}{2-r+\varepsilon}\theta^{2-r+\varepsilon}\right]_\Theta^1}{\overline{z}\overline{g}} \\
&= 1-\frac{(1-t)^\varepsilon\frac{\Theta^{2-r+\varepsilon}}{2-r+\varepsilon}+(1-t)^\varepsilon A^{1+\varepsilon}\frac{1-\Theta^{2-r+\varepsilon}}{2-r+\varepsilon}}{\overline{z}\overline{g}}
\end{aligned}$$

Moreover, we use that

$$\overline{g} = \int_0^1\theta^{-r}dF(\theta) = \frac{1}{1-r}.$$

We first establish that ξ takes on the same values when either everyone chooses to stay low-skilled ($\Theta = 1$), or chooses to become high-skilled ($\Theta = 0$).

$$\xi_{|\Theta=1} = 1-\frac{\int_0^1 g_\theta z_\theta^L dF(\theta)}{\overline{z}_{|\Theta=1}\overline{g}}.$$

Where substituting for

$$\overline{z}_{|\Theta=1} = \int_0^1 z_\theta^L dF(\theta) = (1-t)^\varepsilon \int_0^1 \theta^{1+\varepsilon} dF(\theta) = (1-t)^\varepsilon \frac{1}{2+\varepsilon}$$

and using

$$\int_0^1 g_\theta z_\theta^L dF(\theta) = (1-t)^\varepsilon \int_0^1 \theta^{1+\varepsilon-r} dF(\theta) = (1-t)^\varepsilon \frac{1}{2+\varepsilon-r}$$

leads to

$$\xi_{|\Theta=1} = 1-(1-r)\frac{(1-t)^\varepsilon}{(1-t)^\varepsilon}\frac{2+\varepsilon}{2+\varepsilon-r} = 1-(1-r)\frac{2+\varepsilon}{2+\varepsilon-r}.$$

Next, we note that at the other extreme of everyone being high-skilled, i.e. at the limit of $A \to \infty$ and thus $\Theta = 0$, we have

$$\xi_{|\Theta=0} = 1 - \frac{\int_0^1 g_\theta z_\theta^H dF(\theta)}{\overline{z}_{|\Theta=1}\overline{g}},$$

where substituting for

$$\overline{z}_{|\Theta=0} = \int_0^1 z_\theta^H dF(\theta) = (1-t)^\varepsilon A^{1+\varepsilon} \int_0^1 \theta^{1+\varepsilon} dF(\theta) = (1-t)^\varepsilon A^{1+\varepsilon}\frac{1}{2+\varepsilon},$$

and using

$$\int_0^1 g_\theta z_\theta^H dF(\theta) = (1-t)^\varepsilon A^{1+\varepsilon} \int_0^1 \theta^{1+\varepsilon-r} dF(\theta) = (1-t)^\varepsilon A^{1+\varepsilon}\frac{1}{2+\varepsilon-r},$$

leads to

$$\xi_{|\Theta=0} = 1-(1-r)\frac{(1-t)^\varepsilon A^{1+\varepsilon}}{(1-t)^\varepsilon A^{1+\varepsilon}}\frac{2+\varepsilon}{2+\varepsilon-r} = 1-(1-r)\frac{2+\varepsilon}{2+\varepsilon-r} = \xi_{|\Theta=1}.$$

The distributional characteristic thus takes on the same value in both extreme cases. Next, we investigate how ξ changes with SBTC.

Differentiating ξ with respect to A yields

$$\begin{aligned}\frac{\partial \xi}{\partial A} = \frac{1}{(2+\varepsilon-r)\bar{z}(A)^2}(r-1)\Theta(A)^{-r} \\ \cdot\Bigg(\Theta^{\varepsilon+2}\left(\left(A^{\epsilon+1}-1\right)(1-t)^\epsilon\bar{z}' - (\epsilon+1)\bar{z}(A-At)^\epsilon\right) \\ +\Theta(-r+\epsilon+2)\bar{z}\Theta'\left((-(t-1)\Theta)^\epsilon - A(-A(t-1)\Theta)^\epsilon\right) \\ +\Theta^r(A-At)^\epsilon\left((\epsilon+1)\bar{z} - A\bar{z}'\right)\Bigg).\end{aligned}$$

The sign of this expression is determined by

$$\begin{aligned}\Lambda \equiv \overline{z}'\left(\Theta^r A^{1+\varepsilon} - \Theta^{\varepsilon+2}\left(A^{\epsilon+1}-1\right)\right) \\ +\overline{z}\left((1+\varepsilon)A^\varepsilon\left(\Theta^{\varepsilon+2}-\Theta^r\right) + (2+\varepsilon-r)\Theta'\Theta^{1+\varepsilon}\left(A^{1+\varepsilon}-1\right)\right).\end{aligned} \tag{87}$$

We evaluate this expression at $\Theta = 1$ in order to study the behavior of ξ under SBTC starting from the extreme case in which all individuals choose to be low-skilled. We obtain

$$\Lambda_{|\Theta=1} \equiv \overline{z}' + \overline{z}\left((2+\varepsilon-r)\Theta'\left(A^{1+\varepsilon}-1\right)\right)$$

We now use

$$\frac{\Theta'}{\Theta} = \frac{A^{\varepsilon}}{1-A^{1+\varepsilon}}$$

and hence

$$\Theta' = \frac{A^{\varepsilon}}{1-A^{1+\varepsilon}}\Theta$$

to arrive at

$$\Lambda_{|\Theta=1} \equiv \overline{z}' - \overline{z}\left((2+\varepsilon-r)A^{\varepsilon}\right).$$

We also have

$$\frac{\overline{z}'}{\overline{z}} = \frac{\Theta^{1+\varepsilon}\left((\epsilon+2)\left(A^{\epsilon+1}-1\right)\Theta' + (\epsilon+1)\Theta A^{\epsilon}\right) - (1+\varepsilon)A^{\epsilon}}{\left(A^{\epsilon+1}-1\right)\Theta^{2+\varepsilon} - A^{\epsilon+1}}$$

Evaluating the expression at $\Theta = 1$ yields

$$\begin{aligned}
\frac{\overline{z}'}{\overline{z}}|_{\Theta=1} =& \frac{\Theta^{1+\varepsilon}\left((\epsilon+2)\left(A^{\epsilon+1}-1\right)\Theta' + (\epsilon+1)\Theta A^{\epsilon}\right) - (1+\varepsilon)A^{\epsilon}}{\left(A^{\epsilon+1}-1\right)\Theta^{2+\varepsilon} - A^{\epsilon+1}} \\
=& \frac{(\epsilon+2)\left(A^{\epsilon+1}-1\right)\Theta' + (\epsilon+1)A^{\epsilon} - (1+\varepsilon)A^{\epsilon}}{\left(A^{\epsilon+1}-1\right) - A^{\epsilon+1}} \\
=& -\left((\epsilon+2)\left(A^{\epsilon+1}-1\right)\Theta' + (\epsilon+1)A^{\epsilon} - (1+\varepsilon)A^{\epsilon}\right)
\end{aligned}$$

Substituting for Θ' leads to

$$\begin{aligned}
\frac{\overline{z}'}{\overline{z}}|_{\Theta=1} =& -\left((\epsilon+2)\left(-A^{\varepsilon}\right) + (\epsilon+1)A^{\epsilon} - (1+\varepsilon)A^{\epsilon}\right) \\
=& (\epsilon+2)A^{\varepsilon} - (1+\varepsilon)A^{\epsilon} + (1+\varepsilon)A^{\epsilon} \\
=& A^{\varepsilon}\left(2+\varepsilon\right).
\end{aligned}$$

The condition for $\Lambda_{|\Theta=1} > 0$ thus translates into

$$\begin{aligned}
A^{\varepsilon}\left(2+\varepsilon\right) > A^{\varepsilon}\left(2+\varepsilon-r\right) \Leftrightarrow \\
r > 0,
\end{aligned}$$

which is true by assumption.

Next, note that evaluating the sign of $\partial\xi/\partial A$ at $\Theta = 0$ yields

$$\Lambda_{|\Theta=0} = 0.$$

We thus established that at low levels of skill-bias such that $\Theta = 1$, ξ increases, while at high levels of skill-bias ($\Theta \rightarrow 10$), ξ asymptotically approaches $\xi_{|\Theta=0}$. Moreover, since $\xi_{|\Theta=1} = \xi_{|\Theta=0}$, and since ξ is continuous in A, the result follows.

B.5 Optimal tax rate adjustment under SBTC with fixed subsidy

The behavior of ξ under SBTC has been established in Section 5.4.1. The optimal tax rate formula also involves a term which corresponds to the distortions of skill-formation. We now analyze the effect of SBTC on this term:

$$\frac{\Delta}{(1-t)}\frac{f(\Theta)\Theta}{\bar{z}}\varepsilon_{\Theta,t},$$

with Δ being the net-tax on skill-formation, defined as

$$\Delta \equiv tw_H\Theta h_\Theta - tw_L\Theta l_\Theta - s.$$

With linear technology, we have

$$\begin{aligned}\Delta =& tA\Theta\left[(1-t)A\Theta\right]^\varepsilon - t\Theta\left[(1-t)\Theta\right]^\varepsilon - s\\ =& t\left((1-t)^\varepsilon\Theta^{1+\varepsilon}\left(A^{1+\varepsilon}-1\right)\right) - s.\end{aligned}$$

We then have

$$\frac{\partial\Delta}{\partial A} = t(\epsilon+1)(1-t)^\epsilon\Theta^\varepsilon\left(\left(A^{\epsilon+1}-1\right)\Theta' + \Theta A^\epsilon\right)$$

The sign of this expression is determined by

$$\left(A^{\epsilon+1}-1\right)\Theta' + \Theta A^\epsilon.$$

The condition for $\frac{\partial\Delta}{\partial A} > 0$ thus can be written as

$$\begin{aligned}A^{\epsilon+1} - 1 <& -\frac{\Theta A^\varepsilon}{\Theta'},\\ A^{\varepsilon+1} <& 1 - \frac{\Theta}{\Theta'}A^\varepsilon,\end{aligned}$$

where we take into account that $\Theta' < 0$. We now use that in general

$$\frac{\Theta'}{\Theta} = \frac{w_H^\varepsilon w_H'}{w_L - w_H^{1+\varepsilon}}$$

and thus in the case of linear technology

$$\frac{\Theta'}{\Theta} = \frac{A^\varepsilon}{1-A^{1+\varepsilon}}.$$

We thus have that with linear technology $\frac{\partial\Delta}{\partial A} = 0$, hence the net tax on skill-formation does not change with skill-bias.

Next, we focus on

$$\frac{f(\Theta)\Theta}{\bar{z}}. \tag{88}$$

With uniform ability $f(\Theta) = 1$. The effect of SBTC on this term is thus

$$\frac{\Theta'\bar{z} - \Theta\bar{z}'}{\bar{z}^2} = \frac{A^\varepsilon}{1-A^{1+\varepsilon}}\frac{1}{\bar{z}^2}\Theta\left(\bar{z} - \bar{z}'\right). \tag{89}$$

The sign of this expression is determined by

$$\bar{z} - \bar{z}'$$

$$\frac{(1-t)^{\epsilon}\left(\left(1-(A+1)A^{\epsilon}\right)\Theta^{\varepsilon+2}+(A-\epsilon-1)A^{\epsilon}\right)}{\epsilon+2}$$
$$=\frac{(1-t)^{\epsilon}\left(\left(1-(A+1)A^{\epsilon}\right)\Theta^{\varepsilon+2}+(A-\epsilon-1)A^{\epsilon}\right)}{\epsilon+2}$$

The sign is determined by

$$\left(1-A^{1+\varepsilon}+A^{\epsilon}\right)\Theta^{\varepsilon+2}+\left(A^{1+\varepsilon}-\epsilon A^{\varepsilon}-A^{\varepsilon}\right)$$

The expression in 89 is positive if

$$\left(1-A^{1+\varepsilon}+A^{\epsilon}\right)\Theta^{\varepsilon+2}+\left(A^{1+\varepsilon}-\epsilon A^{\varepsilon}-A^{\varepsilon}\right)<0 \Leftrightarrow$$
$$\Theta^{\varepsilon+2}<\frac{\left(\epsilon A^{\varepsilon}-A^{1+\varepsilon}+A^{\varepsilon}\right)}{\left(1-A^{1+\varepsilon}+A^{\epsilon}\right)}.$$

The LHS is upper bounded by 1, while the RHS is unbounded. The condition will hold for A large enough. For low values of A the condition will not hold, though. We conclude that SBTC first leads to a decrease in (88), but as skill-bias continues to rise, (88) will increase as well. Finally, we note that with linear technology

$$\varepsilon_{\Theta,t}\equiv\frac{\partial\Theta}{\partial t}\frac{1-t}{\Theta}=1,$$

hence a 1% increase in the tax rate decreases college attendance by 1%, irrespective of the amount of skill-bias. We therefore have

$$\begin{aligned}\frac{\partial}{\partial A}\left(\frac{\Delta}{(1-t)}\frac{f\left(\Theta\right)\Theta}{\overline{z}}\varepsilon_{\Theta,t}\right)=&\frac{\Delta}{(1-t)}\varepsilon_{\Theta,t}\frac{\partial}{\partial A}\left(\frac{\Theta}{\overline{z}}\right)\\=&\frac{\Delta}{(1-t)}\varepsilon_{\Theta,t}\frac{A^{\varepsilon}}{1-A^{1+\varepsilon}}\frac{1}{\overline{z}^{2}}\Theta\left(\overline{z}-\overline{z}'\right),\end{aligned}$$

where the condition for this expression to be positive is

$$\Theta^{\varepsilon+2}<\frac{\left(\epsilon A^{\varepsilon}-A^{1+\varepsilon}+A^{\varepsilon}\right)}{\left(1-A^{1+\varepsilon}+A^{\epsilon}\right)},$$

which is satisfied for large values of A. We conclude that at low values of skill-bias, and increase in A dampens the distortions due to the net-tax on skill-formation, while these distortions are amplified as skill-bias grows large.

There are thus two forces which push towards a higher marginal tax rate if SBTC happens at modest levels of skill-bias, while at high levels of skill-bias both forces push towards a lower marginal tax rate.

B.6 Optimal subsidy adjustment under SBTC with fixed tax rate

Proof. With linear technology the condition for the optimal subsidy in simplifies to

$$\Delta\Theta\varepsilon_{\Theta,s}=s\zeta. \tag{90}$$

We next show that if $\Theta=1$, $s>0$: At $\Theta=1$ we have $\zeta=0$, and the condition in (90) simplifies to

$$\Delta\Theta\varepsilon_{\Theta,s}=0\Leftrightarrow$$
$$\frac{s\left(t\left(A^{\epsilon+1}-1\right)(1-t)^{\epsilon}-s\right)}{(\epsilon+1)(p-s)}=0.$$

Solving for s yields

$$s_{|\Theta=1} = t\left(A^{\epsilon+1}-1\right)(1-t)^{\epsilon} > 0,$$

hence at the extreme where no individual attends college, the optimal education subsidy is positive.

We next consider the effect of SBTC on 90. We have $\partial\Delta/\partial A = 0$, $\partial\Theta/\partial A < 0$, $\partial\varepsilon_{\Theta,s}/\partial A = 0$. The LHS of (90) thus decreases with SBTC. The effect of SBTC on the distributional characteristic ζ is ambiguous. With uniform ability we have

$$\zeta = \Theta\left(\Theta^{-r}-1\right),$$

and thus

$$\frac{\partial\zeta}{\partial A} = \frac{\partial\Theta}{\partial A}\left((1-r)\Theta^{-r}-1\right).$$

However, we can sign this derivative at the extremes of $\Theta = 1$ and $\Theta \to 0$. We obtain

$$\frac{\partial\zeta}{\partial A}_{|\Theta=1} = \frac{\partial\Theta}{\partial A}(-r) > 0,$$

and

$$\lim_{\Theta\to 0}\frac{\partial\zeta}{\partial A} = -\infty < 0.$$

To gain intuition, recall that ζ measures the distributional benefits of a marginally higher *net tax* on education.For high values of Θ, few individuals are in the college sector; moreover, these individuals receive a low welfare weight. Having a marginally higher net tax on education is thus beneficial. However, these distributional benefits start to fall as soon as the first individual with a welfare weight above average enters the college sector. Moreover, the decrease is rapid as the individuals with very low ability receive a high welfare weight – and taxing them additionally when they enter the college sector decreases welfare.

Using that ζ increases under SBTC for high levels of Θ means that the RHS of (90) increases. We now ask the question how the subsidy needs to adjust in order to restore equality of (90). We have $\partial\Delta/\partial s < 0$, $\partial\Theta/\partial s < 0$, and $\partial\varepsilon_{\Theta,s}/\partial s > 0$. The impact of a change in s on ζ works via Θ; and since an increase in s lowers Θ (just as an increase in A lowers Θ), we have that $\partial\zeta/\partial s$ is positive for low values of A (Θ close to 1), while it is negative for large values of A (Θ close to 0).

We consider a situation at which $\Theta = 1$; and conjecture that in order to restore equilibrium the optimal subsidy needs to drop, thereby reducing the RHS and increasing the LHS of 90. In order to prove that this indeed is the case, we assume to the contrary, that s needs to increase, such as to raise both the LHS and the RHS, but the LHS more strongly. We show that this implies a contradiction. We evaluate the derivative of the LHS with respect to s of 90 at $\Theta = 1$:

$$\frac{\partial\left(\Delta\Theta\varepsilon_{\Theta,s}\right)}{\partial s}_{|\Theta=1} = \frac{\underbrace{t\left(A^{\epsilon+1}-1\right)(1-t)^{\epsilon}}_{>0}\left(\underbrace{t(\epsilon+1)\left(A^{\epsilon+1}-1\right)(1-t)^{\epsilon}\Theta'(s)}_{<0}\right)}{(\epsilon+1)\left(\underbrace{p-t\left(A^{\epsilon+1}-1\right)(1-t)^{\epsilon}}_{?}\right)}.$$

The numerator of this expression is negative. If we can show that the denominator is positive it follows that an increase in s at $\Theta = 1$ lowers the LHS of 90, and thus moves us away from equilibrium. In order to determine the sign of the denominator, we solve

$$\Theta = 1 \Leftrightarrow$$

$$\underline{A} = \left(\frac{(1-t)^{\epsilon}(t\epsilon+1)}{p(\epsilon+1)+(t\epsilon+1)(1-t)^{\epsilon}} \right)^{-\frac{1}{\epsilon+1}}.$$

We thus derived the level of skill-bias $\underline{A}$ at which no individual finds it beneficial to enter the college sector, and thus $\Theta = 1$. Substituting $\underline{A}$ for A in $p - t\left(A^{\epsilon+1}-1\right)(1-t)^{\epsilon}$ yields

$$p - t\left(\underline{A}^{\epsilon+1}-1\right)(1-t)^{\epsilon} = \frac{p\,(1-t)}{t\epsilon+1} > 0.$$

We thus derived that at $\Theta = 1$ an increase in s leads to a contradiction. As a consequence, s has to drop in order to restore equilibrium. □

References

Acemoglu, D., and D. Autor (2011): "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. Volume 4, Part B, pp. 1043–1171. Elsevier.

——— (2012): "What Does Human Capital Do? A Review of Goldin and Katz's The Race between Education and Technology," *Journal of Economic Literature*, 50(2), 426–463.

Acemoglu, D., G. Gancia, and F. Zilibotti (2011): "Offshoring and directed technical change," Discussion paper, mimeo.

Ales, L., A. A. Bellofatto, and J. J. Wang (2014): "Taxing Atlas: Using Firm Data to Derive Optimal Income Tax Rates?," Discussion paper.

Allen, F. (1982): "Optimal linear income taxation with general equilibrium effects on wages," *Journal of Public Economics*, 17(2), 135–143.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008): "Trends in U.S. Wage Inequality: Revising the Revisionists," *Review of Economics and Statistics*, 90(2), 300–323.

Blundell, R., and T. Macurdy (1999): "Chapter 27 Labor supply: A review of alternative approaches," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. Volume 3, Part A, pp. 1559–1695. Elsevier.

Bovenberg, A. L., and B. Jacobs (2005): "Redistribution and education subsidies are Siamese twins," *Journal of Public Economics*, 89(11–12), 2005–2035.

Card, D., and J. E. DiNardo (2002): "Skill-biased technological change and rising wage inequality: Some problems and puzzles," *Journal of Labor Economics*, 20(4), 733–783.

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

Dixit, A., and A. Sandmo (1977): "Some Simplified Formulae for Optimal Income Taxation," *The Scandinavian Journal of Economics*, 79(4), 417.

Dur, R., and C. N. Teulings (2001): "Education and Efficient Redistribution," SSRN Scholarly Paper ID 288946, Social Science Research Network, Rochester, NY.

Feldstein, M. S. (1972): "Distributional equity and the optimal structure of public prices," *The American Economic Review*, p. 32–36.

Gerritsen, A., and B. Jacobs (2013): "Optimal minimum wages and optimal redistribution in competitive labor markets with endogenous skill formation," Discussion paper, Erasmus University Rotterdam.

——— (2014): "Is a Minimum Wage an Appropriate Instrument for Redistribution?," Discussion paper, CESifo.

Goldin, C., and L. F. Katz (2010): *The Race between Education and Technology.* Belknap Press of Harvard University Press.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic Dynamics*, 13(1), 15–51.

Heathcote, J., K. Storesletten, and G. L. Violante (2014): "Optimal Tax Progressivity: An Analytical Framework," Working Paper 19899, National Bureau of Economic Research.

Jacobs, B. (2005): "Optimal Income Taxation with Endogenous Human Capital," *Journal of Public Economic Theory*, 7(2), 295–315.

——— (2012): "Optimal redistributive tax and education policies in general equilibrium," *International Tax and Public Finance*, pp. 1–26.

Jacobs, B., and A. L. Bovenberg (2011): "Optimal Taxation of Human Capital and the Earnings Function," *Journal of Public Economic Theory*, 13(6), 957–971.

Kane, T. J. (2006): "Chapter 23 Public Intervention in Post-Secondary Education," in *Handbook of the Economics of Education*, ed. by E. Hanushek, and F. Welch, vol. Volume 2, pp. 1369–1401. Elsevier.

Katz, L. F., and K. M. Murphy (1992): "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107(1), 35–78.

Lozachmeur, J., A. Pavan, and R. Gomes (2014): "Differential Taxation and Occupational Choice," Taxation Theory Conference Cologne.

Meghir, C., and D. Phillips (2008): "Labour supply and taxes," http://www.ifs.org.uk/publications/4166.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38(2), 175–208.

OECD (2011): "Divided We Stand," .

Piketty, T. (2014): *Capital in the Twenty-First Century.* Harvard University Press, Cambridge Massachusetts.

Rothschild, C., and F. Scheuer (2013): "Redistributive Taxation in the Roy Model," *The Quarterly Journal of Economics*, 128(2), 623–668.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Sheshinski, E. (1972): "The Optimal Linear Income-tax," *The Review of Economic Studies*, 39(3), 297–302.

Stern, N. H. (1976): "On the specification of models of optimum income taxation," *Journal of Public Economics*, 6(1–2), 123–162.

Stiglitz, J. E. (1982): "Self-selection and Pareto efficient taxation," *Journal of Public Economics*, 17(2), 213–240.

Teulings, C. N. (2000): "Bridging the Gap between 'Joe Sixpack' and 'Bill Gates': On the Efficiency of Institutions for Redistribution," *De Economist*, 148(5), 603–624.

Tinbergen, J. (1975): *Income distribution: Analysis and policies.* North-Holland Pub. Co. ; American Elsevier, Amsterdam; New York.

Tuomala, M. (2010): "On optimal non-linear income taxation: numerical results revisited," *International Tax and Public Finance*, 17(3), 259–270.