



# Optimal linear income taxes and education subsidies under skill-biased technical change

Bas Jacobs<sup>1,2,4</sup> · Uwe Thuemmel<sup>3,4</sup>

Accepted: 21 July 2022

© Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

This paper studies how linear tax and education policy should optimally respond to skill-biased technical change (SBTC). SBTC affects optimal taxes and subsidies by changing (1) direct distributional benefits of each policy instrument, (2) indirect, general-equilibrium effects on wages, and (3) education distortions. Analytically, the effect of SBTC on these three components is shown to be ambiguous. In simulations for the US economy, SBTC makes the optimal tax system more progressive and lowers optimal education subsidies. This is because for both income taxes and education subsidies; their direct distributional effects become more important, which more than offsets the larger general-equilibrium effects and increased education distortions.

**Keywords** Human capital · General equilibrium · Optimal taxation · Education subsidies · Technological change

**JEL Classification** H2 · H5 · I2 · J2 · O3

---

✉ Bas Jacobs  
b.jacobs@vu.nl  
<https://jacobs73.home.xs4all.nl/>  
Uwe Thuemmel  
mail@uwethuemmel.com  
<http://uwethuemmel.com>

<sup>1</sup> Department of Economics, School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

<sup>2</sup> Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, The Netherlands

<sup>3</sup> Department of Economics, University of Zurich, Schoenberggasse 1, 8001 Zurich, Switzerland

<sup>4</sup> CESifo, Munich, Germany

## 1 Introduction

Skill-biased technical change (SBTC) has been an important driver of rising income inequality in many developed countries over the last decades (see, e.g., Van Reenen 2011). Skill-biased technology raises the relative demand for skilled workers. If relative demand grows faster than relative supply, the skill premium increases and so does income inequality.<sup>1</sup> The idea that income inequality is the result of the ‘race between education and technology’ dates back to Tinbergen (1975). He suggested that, in order to contain inequality, governments should increase investment in education, so as to increase the relative supply of skilled workers and win the race with technology. Goldin and Katz (2010, Ch.9, pp. 350-351) take up Tinbergen’s metaphor and argue that US policy should respond to SBTC by making the tax system more progressive and by increasing financial aid for higher education.

Despite the obvious relevance of SBTC to explain rising skill premia and wage inequality, very little analysis exists on the normative question whether it is a good idea to make tax systems more progressive or to stimulate investments in higher education in response to SBTC. Therefore, this paper studies how skill-biased technical change affects optimal linear taxes and education subsidies. We do so by extending the standard model of optimal linear income taxation of Sheshinski (1972) with endogenous skill formation and embed it in the ‘canonical model’ of SBTC, where high-skilled and low-skilled workers are imperfect substitutes in production (Katz and Murphy, 1992; Violante, 2008; Acemoglu and Autor, 2011).<sup>2</sup> In our model, individuals differ in their earning ability. They decide how much to work and whether to enroll in higher education. Only individuals with a sufficiently high ability become high-skilled, everyone else remains low-skilled. The wages of high-skilled and low-skilled workers are endogenously determined by relative demand and relative supply, where skill-biased technical change drives relative demand. An inequality-averse government maximizes social welfare by optimally setting linear income taxes and education subsidies. Our findings are the following.

We start our analysis by deriving optimal tax and education policies for *given skill bias*. As is usual in optimal taxation, optimal policies trade off the redistributive benefits against the efficiency costs of each policy instrument. The benefits consist of the direct distributional impact and the indirect redistributive impacts that originate from the general-equilibrium effects on wages. The linear income tax directly reduces income inequality, but also generates general-equilibrium effects on wages that raise pre-tax income differentials: by discouraging investment in education, the relative supply of skilled workers falls and the relative wage of skilled workers increases.<sup>3</sup> Moreover, education subsidies result in distributional losses,

<sup>1</sup> For the canonical model of SBTC see Katz and Murphy (1992), Violante (2008) and Acemoglu and Autor (2011).

<sup>2</sup> Dixit and Sandmo (1977) and Hellwig (1986) elaborate further on the optimal linear tax model.

<sup>3</sup> Although relative wages may also respond to relative changes in hours worked, this mechanism does not play a role in our model, since we assume that high-skilled and low-skilled workers have equal labor-supply elasticities. Hence, relative labor supply does not change in response to changing the linear tax rate. See also Jacobs (2012).

since high-skilled individuals have higher incomes than low-skilled individuals. However, as suggested by Tinbergen, these direct distributional losses can be countered by general-equilibrium effects on wages. By increasing the relative supply of skilled workers, the skill premium declines, and this reduces income inequality. For both the income tax and the education subsidy, the direct distributional impacts and indirect distributional impacts that result from general-equilibrium effects are traded off against distortions in labor supply and in investment in education.

We then analyze how optimal policy should respond to a *change in skill bias*. In doing so, we demonstrate that the policy recommendations of Tinbergen (1975) and Goldin and Katz (2010) need not be correct. Our analysis shows that the optimal policy response depends on the effect of SBTC on (1) direct distributional impacts, (2) general-equilibrium effects, and (3) distortions in education.<sup>4</sup> We derive analytically that the effect of SBTC on direct distributional impacts, general-equilibrium effects, and distortions in skill formation are *all* theoretically ambiguous.

To resolve these theoretical ambiguities, we quantify the impact of SBTC on optimal tax and education policy by calibrating our model to the US economy using data from the US Current Population Survey and empirical evidence on labor market responses to tax and education policy. Given that our model is stylized, these simulations should be taken with caution. Our main aim is to get a quantitative sense of the impacts of SBTC on the main determinants of optimal policy in our model: direct distributional effects, general-equilibrium effects, and tax distortions on education. We simulate the response of optimal taxes and education subsidies to a rise in skill bias that is in line with the observed increase in the skill premium between 1980 and 2016. Moreover, we show that education is optimally subsidized on a net basis before the shock in skill bias to exploit general-equilibrium effects on wages for income redistribution. Hence, investment in higher education is optimally distorted upwards. Our main finding is that the optimal income tax rate increases with SBTC, while the optimal education subsidy declines with SBTC.

To understand which mechanisms drive these policy responses to SBTC, we numerically decompose the impact of SBTC into the main theoretical determinants of tax and education policy: direct distributional impacts, indirect general-equilibrium effects, and education distortions. We find that the optimal tax rate increases because the direct distributional benefits of taxing income increase and the distortions of net subsidies on education become larger, which overturn the larger general-equilibrium effects of taxing labor income. The optimal education subsidy declines with SBTC, since both the direct distributional losses and the distortions of net subsidies on education increase more than the stronger indirect general-equilibrium effects of subsidizing education.

The main lesson of our paper is that the impact of SBTC on optimal tax and education policy is far from straightforward. Indeed, while our model is stylized, it conveys a number of important messages for thinking about the optimal policy response to SBTC. While SBTC typically raises income inequality and the skill premium,

<sup>4</sup> The distortions in labor supply are invariant to SBTC due to a presumed constant elasticity of labor supply.

thus calling for higher taxes and lower education subsidies for redistributive reasons, SBTC also affects the power of tax and education policy to generate general-equilibrium effects, which work in the opposite direction. Moreover, we show that it is not obvious how tax distortions in education change in response to SBTC; in our simulations, the distortions of net subsidies on education increase.

Our simulations demonstrate that SBTC calls for a more progressive income tax system, while the subsidy rate on investment in higher education should decline. Therefore, we show that the suggestions of Tinbergen (1975) and Goldin and Katz (2010) to promote investment in higher education to win the race against technology need not be correct. Although these authors are right to emphasize the larger benefits of exploiting general-equilibrium effects with SBTC, our analysis reveals that (at least) two other effects need to be taken into account as well to judge whether education subsidies should optimally increase: larger inequality between high-skilled and low-skilled workers and larger upward distortions in education. We show that these latter two effects quantitatively dominate over general-equilibrium effects.

The remainder of this paper proceeds as follows. Section 2 reviews the literature and outlines our contributions. Section 3 sets up the model. Section 4 analyzes optimal policy. Section 5 presents the simulations. Finally, Sect. 6 concludes. Proofs of all propositions, additional derivations, and background materials are contained in the Appendix, part of which is available online only.

## 2 Related literature

We analyze optimal linear income taxes and education subsidies in an extension of the optimal linear tax model due to Sheshinski (1972) with an endogenous education decision on the extensive margin and endogenous wage rates for high-skilled and low-skilled labor as in Roy (1951).<sup>5</sup> We merge this model with the canonical model of SBTC, which goes back to Katz and Murphy (1992), Violante (2008), and Acemoglu and Autor (2011). This allows us to analyze optimal linear education subsidies and to explore the consequences of SBTC for optimal policies. Our paper makes a number of contributions to four strands in the literature.

First, we contribute to the literature that analyzes optimal income taxes jointly with optimal education subsidies; see, for example, Bovenberg and Jacobs (2005), Maldonado (2008), Bohacek and Kapicka (2008), Anderberg (2009), Jacobs and Bovenberg (2011), and Stantcheva (2017). We contribute to these papers by analyzing optimal tax and education policies with education on the extensive margin rather than on the intensive margin. We find that education subsidies are employed to alleviate tax distortions on education, but do not fully eliminate all tax-induced distortions on education, as in Bovenberg and Jacobs (2005). The government likes to tax away infra-marginal rents in education to redistribute income from high-skilled

<sup>5</sup> See also Dixit and Sandmo (1977) and Hellwig (1986) for extensions and further analysis of the optimal linear tax model.

to low-skilled workers—*ceteris paribus*—see also Findeisen and Sachs (2016) and Colas et al. (2021).

Second, we contribute to the literature on optimal income taxation and education subsidies in the presence of general-equilibrium effects on the wage distribution. Dur and Teulings (2004) analyze optimal log-linear tax and education policies in an assignment model of the labor market.<sup>6</sup> Like Dur and Teulings (2004), we find that education might be subsidized on a net basis to exploit general-equilibrium effects for income redistribution. Jacobs (2012) analyzes optimal linear taxes and education subsidies in a two-type version of the model of Bovenberg and Jacobs (2005) and shows that optimal education subsidies are *not* employed to generate general-equilibrium effects so as to compress the wage distribution. The reason is that with education on the intensive margin, the general-equilibrium effect of education subsidies is identical to the general-equilibrium effect of income taxes. Hence, education subsidies have no distributional value added over income taxes, but generate additional distortions in education. Our model does not have this property, since we analyze education on the extensive margin. We add to this literature by analyzing the optimal response of income taxes and education subsidies to skill-biased technical change.

Third, this paper is most closely related to papers that study the response of optimal tax and education policies to technical change (Ales et al., 2015; Heathcote et al., 2014; Jacobs and Thuemmel, 2021; Loebbing, 2020).<sup>7</sup> Jacobs and Thuemmel (2021) study optimal nonlinear tax and education policies in a random participation model with education-dependent nonlinear taxes and individuals differing along two dimensions: earning ability and costs of education. As a result, the income distributions of high- and low-skilled workers are overlapping, since there can be high- and low-skilled workers at each income level. They find that general-equilibrium effects never determine optimal policy. Intuitively, any redistribution from high-skilled to low-skilled workers via general-equilibrium effects can be achieved as well with the education-dependent tax system, while the distortions of exploiting general-equilibrium effects on skill formation can be avoided. This paper adds to Jacobs and Thuemmel (2021) by showing how optimal policy should be set if the government can, realistically, not employ skill-dependent income tax rates. Moreover, doing so allows us to simplify the model structure considerably by allowing for one-dimensional heterogeneity and a non-overlapping wage distribution.<sup>8</sup> In this setting, the

<sup>6</sup> Krueger and Ludwig (2016) study optimal income taxation and education subsidies in a dynamic framework. Like this paper, they highlight the interaction between income taxes and education subsidies. Moreover, they also emphasize the role of education subsidies in exploiting general-equilibrium effects for income redistribution. Unlike this paper, they do not study the effect of SBTC on optimal policy.

<sup>7</sup> Related is also Heckman et al. (1998) who estimate structural dynamic OLG models with skill-specific human capital accumulation technologies and SBTC. Using the same model, Heckman et al. (1999) demonstrate that general-equilibrium effects on wages largely offset the initial impacts of tax and education policies. These papers do not analyze optimal tax and education policies like we do.

<sup>8</sup> In this respect, our focus on linear policies is not a fundamental constraint, but the absence of education-dependent taxes is. Intuitively, a linear tax system with education-dependent tax rates is sufficiently rich to achieve exactly the same income redistribution as general-equilibrium effects on the wage distribution. Hence, the government does not want to redistribute indirectly via general-equilibrium effects if this redistribution can also be obtained directly, as the former will distort education choices.

government can redistribute more income beyond what can be achieved with the income tax system by exploiting general-equilibrium effects on wages. For this reason, education may even be subsidized on a net basis, which can never occur in Jacobs and Thuemmel (2021).

Furthermore, we add to the analyses in Loebbing (2020), Heathcote et al. (2014), and Ales et al. (2015), who study the response of optimal taxes to technical change, but not how education policy should respond. Loebbing (2020) studies the interaction between optimal nonlinear income taxes and directed technical change. Heathcote et al. (2014) study the impact of SBTC on the optimal degree of tax progressivity using a parametric tax function in a model with endogenous human capital formation and imperfect substitutability of skills.<sup>9</sup> Ales et al. (2015) analyze how the nonlinear income tax should adjust to technical change in a task-based model of the labor market with exogenous human capital decisions.<sup>10</sup> In line with all these papers, we confirm that the tax system becomes more progressive in response to technical change. Our contribution is to also analyze optimal education policy jointly with optimal tax policy and show that SBTC reduces optimal education subsidies.

### 3 Model

This section presents our model consisting of individuals, firms, and a government. Utility-maximizing individuals supply labor on the intensive margin and optimally decide to become high-skilled or to remain low-skilled on the extensive margin. Profit-maximizing firms employ high-skilled and low-skilled labor, while facing SBTC. The government optimally sets progressive income taxes and education subsidies to maximize social welfare.

#### 3.1 Individuals

There is a continuum of individuals of unit mass. Each worker is endowed with earnings ability  $\theta \in [\underline{\theta}, \bar{\theta}]$ , which is drawn from distribution  $F(\theta)$  with corresponding density  $f(\theta)$ . Individuals derive utility from consumption  $c$  and disutility from labor supply  $l$ . Individuals have identical quasi-linear preferences:

$$U(c, l) \equiv c - \frac{l^{1+1/\varepsilon}}{1 + 1/\varepsilon}, \quad \varepsilon > 0, \quad (1)$$

<sup>9</sup> The published version of Heathcote et al. (2014), Heathcote et al. (2017), no longer contains this extension.

<sup>10</sup> For task-based assignment models, see, e.g., Acemoglu and Autor (2011).

where  $\varepsilon$  is the constant wage elasticity of labor supply.<sup>11</sup> Consumption is the numéraire commodity and its price is normalized to unity.

In addition to optimally choosing consumption and labor supply, each individual makes a discrete choice whether to become high-skilled ( $H$ ) or to remain low-skilled ( $L$ ). We indicate an individual's education type by  $j \in \{L, H\}$  and define  $\mathbb{1}$  as an indicator function for being high-skilled:

$$\mathbb{1} \equiv \begin{cases} 1 & \text{if } j = H, \\ 0 & \text{if } j = L. \end{cases} \tag{2}$$

To become high-skilled, workers need to invest a fixed amount of resources  $p(\theta)$ , which captures expenses such as tuition fees, books, and the (monetary value of) effort. High-skilled individuals also forgo earnings as a low-skilled worker. The wage rate per efficiency unit of labor is denoted by  $w^j$ . Gross earnings are thus equal to  $z_\theta^j \equiv w^j \theta^j$ .

We model the direct costs of education as a weakly decreasing function of the worker's ability  $\theta$ :

$$p(\theta) \equiv \pi \theta^{-\psi}, \quad \pi \in (0, \infty), \quad \psi \in [0, \infty).$$

If  $\psi > 0$ , individuals with higher ability  $\theta$  have lower direct costs of education. Hence, more able students need to spend less on education, e.g., because they have lower costs of effort, lower tuition fees, require less tutoring, or obtain grants. Parameter  $\psi$  is only introduced to control the elasticity of enrollment in higher education in the simulations, which will be calibrated at empirically plausible values. However, in the theoretical derivations one can set  $\psi = 0$  without any loss of generality, such that all individuals face the same direct costs of education  $\pi$ .

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 the tax rate  $t$  and transfer  $b$  are positive.<sup>12</sup> In addition, high-skilled individuals receive a flat rate education subsidy  $s$  on total resources  $p(\theta)$  invested in education. 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. Workers of type  $\theta$  with education  $j$  thus face the following budget constraint:

$$c_\theta^j = (1 - t)z_\theta^j + b - (1 - s)p(\theta)\mathbb{1}. \tag{3}$$

The informational assumptions of our model are that individual ability  $\theta$  and labor effort  $l_\theta^j$  are not verifiable, but aggregate labor earnings  $\bar{z} \equiv \int_{\underline{\theta}}^{\bar{\theta}} z_\theta^j dF(\theta)$  and aggregate

<sup>11</sup> Since income effects are absent, compensated and uncompensated wage elasticities coincide. This utility function is employed in nearly the entire optimal tax literature with endogenous wages; see, e.g., Rothschild and Scheuer (2013) and Sachs et al. (2020). The reason is that income effects in labor supply and heterogeneous labor-supply elasticities substantially complicate the analysis if general-equilibrium effects on wages are present, see also Feldstein (1972), Allen (1982), and Jacobs (2012).

<sup>12</sup> We do not constrain the tax system to be progressive in the optimal tax program of the government, but in all our simulations the optimal tax system is indeed progressive.

gate education expenditures  $\int_{\underline{\theta}}^{\bar{\theta}} p(\theta) \mathbb{1} dF(\theta)$  are. Hence, the government can levy linear taxes on income and provide linear subsidies on education.<sup>13,14</sup> Importantly, the tax implementation does not exploit all information available to the government. In particular, we realistically assume that marginal tax rates are not conditioned on education choices, in contrast to Jacobs and Thuemmel (2021). Consequently, income taxes can no longer achieve the same income redistribution as reducing inequality via general-equilibrium effects on wage rates. Hence, exploiting general-equilibrium effects becomes socially desirable for income redistribution.

Workers maximize utility by choosing consumption, labor supply, and education, taking wage rates and government policy as given. For a given education choice, the first-order condition for maximizing utility in Eq. (1), subject to the budget constraint in Eq. (3), yields optimal labor supply for all workers of type  $\theta$  and education  $j$ :

$$l_{\theta}^j = [(1 - t)w^j\theta]^{\varepsilon}. \tag{4}$$

Labor supply increases in net earnings per hour  $(1 - t)w^j\theta$ , and more so if labor supply is more elastic (higher  $\varepsilon$ ). Income taxation distorts labor supply downward 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 the first-order condition in Eq. (4) into the utility function in Eq. (1), and using the budget constraint Eq. (3), the indirect utility function is obtained for all  $\theta$  and  $j$ :

$$V_{\theta}^j \equiv \frac{[(1 - t)w^j\theta]^{1+\varepsilon}}{1 + \varepsilon} + b - ((1 - s)p(\theta))\mathbb{1}. \tag{5}$$

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

$$\Theta = \left[ \frac{\pi(1 - s)(1 + \varepsilon)}{(1 - t)^{1+\varepsilon}((w^H)^{1+\varepsilon} - (w^L)^{1+\varepsilon})} \right]^{\frac{1}{1+\varepsilon+\psi}}. \tag{6}$$

All individuals with ability  $\theta < \Theta$  remain low-skilled, whereas all individuals with  $\theta \geq \Theta$  become high-skilled. A decrease in  $\Theta$  implies that more individuals become high-skilled. If  $w^H/w^L$  rises, more individuals invest in higher education. The same holds true for a decrease in the net cost of education  $(1 - s)\pi$ . The income tax potentially distorts the education decision, since the direct costs of education

<sup>13</sup> These informational assumptions imply that income taxes can be levied as proportional withholding taxes at the firm level and universities can collect education subsidies while proportionally reducing the costs of education to students.

<sup>14</sup> This implies that, although the government can subsidize education at a flat rate, it cannot infer individual ability  $\theta$  from aggregate investments in education.



are not tax-deductible, while the returns to education are taxed. Income taxes also reduce investment in education because they reduce labor supply, and thereby lower the ‘utilization rate’ of human capital. If labor supply would be exogenous ( $\varepsilon = 0$ ), and education subsidies would make all education expenses effectively deductible (i.e.,  $s = t$ ), education would be at its first-best level:  $\Theta = [\pi/(w^H - w^L)]^{1/(1+\varepsilon+\psi)}$  (see Jacobs 2005; Bovenberg and Jacobs 2005).

### 3.2 Firms

A representative firm produces the homogeneous consumption good by using aggregate low-skilled labor  $L$  and aggregate high-skilled labor  $H$  as inputs. Output  $Y$  is produced with a constant-returns-to-scale CES production technology:

$$Y(L, H, A) = B \left( \omega L^{\frac{\sigma-1}{\sigma}} + (1 - \omega)(AH)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{7}$$

$$A, B > 0, \quad \omega \in (0, 1), \quad \sigma > 1,$$

where  $\omega$  governs the income shares of low- and high-skilled workers,  $\sigma \equiv Y_H Y_L / (Y_{HL} Y)$  is the elasticity of substitution between low- and high-skilled labor, and skill bias is parameterized by  $A$ . We denote by  $\alpha \equiv H Y_H(\cdot) / Y(\cdot)$  the income share of high-skilled workers. We model technology like in the canonical model of SBTC. We assume that  $\sigma > 1$  to ensure that skill-biased technical change increases the relative wage of high-skilled to low-skilled workers (Acemoglu and Autor, 2011; Katz and Murphy, 1992; Violante, 2008). All theoretical results generalize to a general constant-returns-to-scale production technology that satisfies the Inada conditions and has an elasticity of substitution  $\sigma$  that is larger than unity, i.e.,  $\sigma > 1$  (see the Appendix).

The representative firm is competitive and maximizes profits by taking wage rates as given. The first-order conditions are:

$$w^L = Y_L(L, H, A), \tag{8}$$

$$w^H = Y_H(L, H, A). \tag{9}$$

In equilibrium, the marginal product of each labor input equals its marginal cost. Moreover, in equilibrium, wage rates  $w^L$  and  $w^H$  depend on skill bias  $A$ . To improve readability, we suppress arguments  $L, H,$  and  $A$  in the derivatives of the production function in the remainder of the paper.

Since we have normalized the mass of individuals to one, average (gross) labor earnings  $\bar{z}$  equals total income, which in turn equals output  $Y$ :

$$\bar{z} \equiv \int_{\underline{\theta}}^{\Theta} z_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} z_{\theta}^H dF(\theta) = Y. \tag{10}$$

### 3.3 Government

The government maximizes social welfare, which is given by

$$\int_{\underline{\theta}}^{\Theta} \Psi(V_{\theta}^L) dF(\theta) + \int_{\Theta}^{\bar{\theta}} \Psi(V_{\theta}^H) dF(\theta), \quad \Psi' > 0, \quad \Psi'' < 0, \quad (11)$$

where  $\Psi(\cdot)$  is a concave transformation of indirect utilities of low- and high-skilled workers. The government budget constraint states that total tax revenue equals spending on education subsidies, non-individualized transfers, and an exogenous government revenue requirement  $R$ :

$$t \left[ \int_{\underline{\theta}}^{\Theta} w^L \theta l_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} w^H \theta l_{\theta}^H dF(\theta) \right] = s \int_{\Theta}^{\bar{\theta}} p(\theta) dF(\theta) + b + R. \quad (12)$$

### 3.4 General equilibrium

In equilibrium, factor prices  $w^L$  and  $w^H$  ensure that labor markets and the goods market clear. Labor market clearing implies that aggregate effective labor supplies for each skill type equal aggregate demands:

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

$$H = \int_{\Theta}^{\bar{\theta}} \theta l_{\theta}^H dF(\theta). \quad (14)$$

Goods market clearing implies that total output  $Y$  equals aggregate demand for private consumption, education expenditures, and exogenous government spending  $R$ :

$$Y = \int_{\underline{\theta}}^{\Theta} c_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} (c_{\theta}^H + p(\theta)) dF(\theta) + R. \quad (15)$$

Due to the Inada conditions on the production technology, there will be a strictly positive mass of both high-skilled individuals and low-skilled individuals in general equilibrium (i.e.,  $0 < \Theta < \infty$ ) if  $\varepsilon > 0$  and  $0 \leq t < 1$ . Moreover, the skill premium will then always be positive, i.e.,  $w^H > w^L$ . That the equilibrium features a positive mass of low- and high-skilled workers jointly with  $w^H > w^L$  can be proven by contradiction. Suppose that there would be an equilibrium in which the high-skilled wage is lower than the low-skilled wage, i.e.,  $w^H < w^L$ . Then, from the expression for the optimal cutoff  $\Theta$  in Eq. (6) follows that nobody wants to become high-skilled ( $\Theta = 0$ ), since there are positive fixed costs of education ( $p(\theta) > 0$ ). However, if nobody wants to become high-skilled, then the wage of the high-skilled workers

goes to infinity in view of the Inada conditions on the production function in Eq. (7), i.e.,  $w^H \rightarrow \infty$ , which contradicts that  $w^H < w^L$ .

### 3.5 Behavioral elasticities

Before analyzing the optimal tax formulas, it is instructive to derive the general-equilibrium comparative statics of the model variables with respect to the income tax and education subsidy. Table 5 in Appendix A derives these behavioral elasticities.

The comparative statics of taxes and subsidies on behavior and wage rates can be summarized as follows. A higher income tax rate discourages both labor supply and investment in education. The latter because the direct costs of investment in human capital are not deductible from the income tax. The education subsidy boosts investment in education. Tax and education policy both affect the skill premium, i.e., the high-skilled wage relative to the low-skilled wage, by changing the relative supply of high-skilled to low-skilled labor. This occurs only via a change in investment in education, and not via changes in labor supply, since the labor-supply elasticity is the same for high-skilled and low-skilled workers. Larger income taxes raise the skill premium as fewer people become high-skilled. This implies that the adverse effect of taxation on high-skilled (low-skilled) labor supply is alleviated (exacerbated) by a rise in the skill premium. Similarly, education subsidies reduce the skill premium as more people become educated. As a result, education subsidies reduce high-skilled labor supply and increase low-skilled labor supply.

## 4 Optimal policy and SBTC

### 4.1 Optimal policy

The government maximizes social welfare in Eq. (11) by choosing the marginal income tax rate  $t$ , the lump-sum transfer  $b$ , and the education subsidy  $s$ , subject to the government budget constraint in Eq. (12). In order to interpret the expressions for the optimal tax rate  $t$  and the subsidy rate  $s$ , we introduce some additional notation.

First, we define the net tax wedge on skill formation  $\Delta$  as

$$\Delta \equiv tw^H\Theta l_{\Theta}^H - tw^L\Theta l_{\Theta}^L - sp(\Theta). \tag{16}$$

$\Delta$  gives the increase in government revenue if the marginal individual with ability  $\Theta$  decides to become high-skilled rather than remaining low-skilled. If  $\Delta > 0$ , education is taxed on a net basis.  $tw^H\Theta l_{\Theta}^H$  gives the additional tax revenue when the marginal individual with ability  $\Theta$  becomes high-skilled.  $tw^L\Theta l_{\Theta}^L$  gives the loss in tax revenue as this individual no longer pays taxes as a low-skilled worker. The government also loses  $sp(\Theta)$  in revenue due to subsidizing education of this individual.

Second, we derive a measure for the distributional benefits of taxing income. In particular, let the social welfare weight of an individual of type  $\theta$  be defined as  $g_{\theta} \equiv \Psi'(V_{\theta})/\eta$ , where  $\eta$  is the Lagrange multiplier on the government budget

constraint—see below. Following Feldstein (1972), we define the distributional characteristic  $\xi$  of the income tax as

$$\xi \equiv \frac{\int_{\theta}^{\Theta} (1 - g_{\theta}) z_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} (1 - g_{\theta}) z_{\theta}^H dF(\theta)}{\bar{z}\bar{g}} > 0. \quad (17)$$

$\xi$  equals minus the normalized covariance between social welfare weights  $g_{\theta}$  and labor earnings  $z_{\theta}^j$ .  $\xi$  measures the social marginal value of income redistribution via the income tax, expressed in monetary equivalents, as a fraction of taxed earnings. Marginal distributional benefits of income taxation are positive, since the social welfare weights  $g_{\theta}$  decline with ability  $\theta$ . We have  $0 \leq \xi \leq 1$ , where  $\xi$  is larger if the government has stronger redistributive social preferences. For a Rawlsian/maxi-min social welfare function, which features  $\Psi'_{\theta} = 1/f(\theta) \gg 1$  and  $\Psi'_{\theta} = 0$  for all  $\theta > \underline{\theta}$ , we obtain  $\xi = 1$  if the lowest ability is zero ( $\underline{\theta} = 0$ ). In contrast, for a utilitarian social welfare function with constant weights  $\bar{\Psi}' = 1$ , we obtain  $\xi = 0$ .<sup>15</sup> We also derive that  $\xi = 0$  if  $z_{\theta}^j$  is equal for everyone so that the government is not interested in income redistribution with the income tax.

An alternative intuition for the distributional characteristic  $\xi$  is that it measures the social value of 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 ( $g_{\theta}$ ) that raising this unit of revenue inflicts on tax payers.

Third, we similarly derive a measure for the distributional benefits of taxing education:

$$\zeta \equiv \int_{\Theta}^{\bar{\theta}} \theta^{-\psi} (1 - g_{\theta}) dF(\theta) \geq 0. \quad (18)$$

$\zeta$  captures the marginal benefits of income redistribution from the high-skilled to the low-skilled via a higher net tax on education (lower net education subsidy). It gives the value of an additional unit of revenue (the '1') minus the utility losses ( $g_{\theta}$ ) that raising this unit of revenue inflicts on high-skilled tax payers.

In contrast to the expression for  $\xi$ , the distributional benefits in  $\zeta$  are not divided by average earnings and the average welfare weight of the high-skilled, since the education choice is discrete.<sup>16</sup> However, the distributional benefits  $\zeta$  are scaled with the cost of education by the term  $\theta^{-\psi}$  because the costs of education decline with  $\theta$ , and more so if  $\psi$  is larger. If the costs of education are larger for individuals with a lower ability  $\theta$ , and every individual receives a linear subsidy on total costs, low-ability individuals receive higher education subsidies in absolute amounts. If the

<sup>15</sup> The absence of a redistributive 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.,  $\xi > 0$ .

<sup>16</sup> This is similar to results in models of optimal participation taxes; see, e.g., Diamond (1980) or Saez (2002).

costs of education are the same for each individual, we have that  $\psi = 0$ , and the distributional characteristic  $\zeta$  only depends on the social welfare weights  $g_\theta$ .

Fourth, we define the income-weighted social welfare weights of each education group as

$$\tilde{g}^L \equiv \frac{\int_{\underline{\theta}}^{\Theta} g_\theta z_\theta^L dF(\theta)}{\int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta)} > \tilde{g}^H \equiv \frac{\int_{\Theta}^{\bar{\theta}} g_\theta z_\theta^H dF(\theta)}{\int_{\Theta}^{\bar{\theta}} z_\theta^H dF(\theta)}. \tag{19}$$

The social welfare weights of the low-skilled are on average higher than the social welfare weights of the high-skilled, since the social welfare weights continuously decline in income.

Fifth, we define the ‘general-equilibrium elasticity’  $\epsilon_{GE}$  as

$$\begin{aligned} \epsilon_{GE} &\equiv \frac{\alpha(1-\alpha)\zeta\delta}{\sigma + \epsilon + \zeta\delta(\beta - \alpha)}, \quad \zeta \equiv \frac{1 + \epsilon}{1 + \epsilon + \psi}, \quad \beta \equiv \frac{1}{1 - (w^L/w^H)^{1+\epsilon}}, \\ \delta &\equiv \left( \frac{\Theta l_\Theta^L f(\Theta)}{L} + \frac{\Theta l_\Theta^H f(\Theta)}{H} \right) \Theta, \end{aligned} \tag{20}$$

where  $\zeta$  is a parameter combination of labor-supply and education elasticities ( $\epsilon$  and  $\psi$ ),  $\beta$  is a measure of the inverse skill premium, and  $\delta$  is a measure for effective labor supply around the skill margin  $\Theta$ . The general-equilibrium elasticity  $\epsilon_{GE}$  measures the response of the relative wage  $w^H/w^L$  in general equilibrium to the relative change in  $H/L$  in general equilibrium, taking into account simultaneous changes in relative demand ( $\sigma$ ) and relative supply ( $\epsilon$  and  $\psi$ ). This term captures the quantitative importance of general-equilibrium effects of tax and education policies on the relative wages of high-skilled and low-skilled workers. The general-equilibrium elasticity  $\epsilon_{GE}$  decreases if labor-supply and education responses are more elastic ( $\psi$  and  $\epsilon$  higher) and there is a larger mass of labor around the skill cutoff ( $\delta$  higher). In that case, the quantities of high-skilled labor relative to low-skilled labor respond more strongly to tax policy changes, leaving less room for relative wage effects to clear the labor markets for high-skilled and low-skilled labor. The general-equilibrium elasticity also increases if the skill premium is higher (inverse skill premium  $\beta$  is lower) and it is ambiguous in the share of high-skilled labor income ( $\alpha$ ).<sup>17</sup>

Armed with the additional notation, we are able to state the conditions for optimal policy in the next proposition.

<sup>17</sup> These comparative statics follow from differentiating the general-equilibrium elasticity with respect to these variables.

**Proposition 1** *The optimal lump-sum transfer, income tax, and net tax on education are determined by*

$$\bar{g} \equiv \int_{\theta}^{\bar{\theta}} g_{\theta} dF(\theta) = 1, \quad (21)$$

$$\frac{t}{1-t} \varepsilon + \frac{\Delta}{(1-t)\bar{z}} \Theta f(\Theta) \varepsilon_{\Theta,t} = \xi - (\tilde{g}^L - \tilde{g}^H) \varepsilon_{GE}, \quad (22)$$

$$\frac{\Delta}{(1-t)\bar{z}} \Theta f(\Theta) \varepsilon_{\Theta,s} = \frac{s\pi}{(1-t)\bar{z}} \zeta - \rho(\tilde{g}^L - \tilde{g}^H) \varepsilon_{GE}, \quad (23)$$

where  $\varepsilon_{\Theta,t} \equiv \frac{\partial \Theta}{\partial t} \frac{1-t}{\Theta}$  is the elasticity of  $\Theta$  with respect to the net-of-tax rate  $1-t$ ,  $\varepsilon_{\Theta,s} \equiv -\frac{\partial \Theta}{\partial s} \frac{s}{\Theta}$  is the elasticity of  $\Theta$  with respect to the subsidy rate  $s$ , and  $\rho \equiv \frac{s}{(1-s)(1+\varepsilon)} > 0$  captures the importance of education subsidies in the total direct costs of education.

**Proof** See Appendix B. □

We interpret each optimality condition in Proposition 1 in the following subsections and relate our results to earlier findings in the literature.

#### 4.1.1 Optimal transfer $b$

The optimality condition for the lump-sum transfer  $b$  in Eq. (21) equates the average social marginal benefit of giving all individuals one euro more in transfers (left-hand side) to the marginal costs of doing so (right-hand side). This expression is standard in optimal linear tax models; see also Sheshinski (1972), Dixit and Sandmo (1977), and Hellwig (1986).<sup>18</sup>

#### 4.1.2 Optimal income tax $t$

The optimal income tax in Eq. (22) equates the total marginal distortions of income taxation on the left-hand side with its distributional benefits on the right-hand side, for any value of the optimal education subsidy, including the optimal level.<sup>19</sup> 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  $\varepsilon$ , the more income taxes distort labor supply.  $\frac{\Delta}{(1-t)\bar{z}} \Theta f(\Theta) \varepsilon_{\Theta,t}$  denotes the marginal distortion of the education decision due

<sup>18</sup> The inverse of  $\bar{g}$  is the marginal cost of public funds. At the tax optimum, the marginal cost of public funds equals one. Intuitively, the marginal social value of public resources is equal to the marginal social value of private resources if the transfer is optimized. See also Jacobs (2018).

<sup>19</sup> Hence, the expression also describes the policy optimum if the education subsidy would not be available to the policy maker. The expression for the optimal income tax does assume, however, that transfers are optimized.

to the income tax. A higher marginal tax rate discourages individuals from becoming high-skilled. The larger is the elasticity  $\varepsilon_{\Theta,t}$ , the larger are the distortions of income taxation on education. The higher the net tax wedge on education (in terms of net income)  $\Delta/((1-t)\bar{z})$ , the more income taxation distorts education, and the lower should the optimal tax rate be. If education subsidies are higher, they counter the distortions of income taxes on education by lowering  $\Delta$ , and optimal income taxes will be set higher—*ceteris paribus*. Hence, education subsidies allow for more progressive income taxes by alleviating the distortions on skill formation, as in Bovenberg and Jacobs (2005).  $\Theta f(\Theta)$  measures the ‘size of the tax base’ at the marginal graduate  $\Theta$ . The higher the mass of individuals  $f(\Theta)$  and the larger their ability  $\Theta$ , the more important are tax distortions in education.

The right-hand side of Eq. (22) gives the distributional benefits of income taxes. The larger are the marginal distributional benefits of income taxes—*as captured by  $\xi$* —the higher should be the optimal tax rate. This is the standard term in optimal linear tax models; see also Sheshinski (1972), Dixit and Sandmo (1977), and Hellwig (1986). In addition,  $(\tilde{g}^L - \tilde{g}^H)\varepsilon_{GE} > 0$  captures the distributional losses of income taxes due to general-equilibrium effects on the wage structure. Income taxation reduces skill formation. Hence, the supply of high-skilled labor falls relative to low-skilled labor. This raises high-skilled wages and depresses low-skilled wages. Consequently, before-tax inequality goes up and social welfare declines, since the income-weighted social welfare weights of the low-skilled workers are larger than the income-weighted social welfare weights of the high-skilled workers ( $\tilde{g}^L > \tilde{g}^H$ ). The direct gains of income redistribution ( $\xi$ ) are therefore reduced by indirect, redistributional losses due general-equilibrium effects on the wage distribution  $(\tilde{g}^L - \tilde{g}^H)\varepsilon_{GE}$ . The general-equilibrium elasticity  $\varepsilon_{GE}$  captures the strength of these general-equilibrium effects of income taxes. A lower elasticity of substitution  $\sigma$ , a lower labor-supply elasticity  $\varepsilon$ , and a lower education elasticity  $\psi$  provoke stronger general-equilibrium responses that erode the distributional powers of income taxation. Intuitively, if quantities adjust only little, relative wages will need to adjust relatively more to clear labor markets.

In the absence of general-equilibrium effects ( $\sigma = \infty$ ), the general-equilibrium elasticity is zero ( $\varepsilon_{GE} = 0$ ). In that case, also education decisions are (required to be) exogenous ( $\varepsilon_{\Theta,t} = 0$ ); see Appendix A.2. Consequently, the standard linear income tax in the absence of general-equilibrium effects and human capital distortions is obtained:  $\varepsilon t/(1-t) = \xi$ . See also Sheshinski (1972), Dixit and Sandmo (1977), and Hellwig (1986).

As a special case, we can also derive the optimal income tax without education subsidies. This allows us to relate the optimal income tax to Feldstein (1972), Allen (1982), and Jacobs (2012), who also study optimal linear income taxes with general-equilibrium effects on wages. The optimal income tax in the absence of education policy can be found by setting  $s = 0$  in Eq. (22):

$$\frac{t}{1-t}\varepsilon + \frac{t}{1-t}\left(\frac{w^H\Theta l_{\Theta}^H - w^L\Theta l_{\Theta}^L}{\bar{z}}\right)\Theta f(\Theta)\varepsilon_{\Theta,t} = \xi - (\tilde{g}^L - \tilde{g}^H)\varepsilon_{GE}. \quad (24)$$

We find that optimal linear income taxes are determined by general-equilibrium effects on wages. Our economic mechanism is different than in Feldstein (1972), Allen (1982), and Jacobs (2012). In all these papers, general-equilibrium effects depend on differences in (uncompensated) wage elasticities of labor supply between high-skilled and low-skilled workers. In particular, if high-skilled workers have the largest uncompensated wage elasticity of labor supply, then linear income taxes depress labor supply of high-skilled workers more than that of low-skilled workers, and this generates general-equilibrium effects on wages, which results in larger before-tax income inequality. Optimal income taxes are lowered accordingly.<sup>20</sup> High- and low-skilled individuals can have different uncompensated labor-supply elasticities due to differences in income elasticities or compensated elasticities. However, this mechanism is not relevant here, since we assume no income effects and compensated wage elasticities of labor supply are equal for both skill types. Indeed, the relative supply of skilled labor does not change due to changes in relative hours worked, but due to endogenous education choices. Income taxes unambiguously generate larger pre-tax income inequality due to general-equilibrium effects, since they reduce skill formation. This contrasts to the contributions that abstract from an endogenous education decision on the extensive margin.

If education subsidies are constrained to be zero, the optimality condition for optimal income taxes in Eq. (24) does not fundamentally change compared to the condition for optimal income taxes with non-zero, and potentially optimal, education subsidies in Eq. (22). The main difference is that the net tax wedge on education is now unambiguously positive, i.e.,  $\Delta \equiv tw^H\Theta_\Theta^H - tw^L\Theta_\Theta^L > 0$ . Since direct costs of education are not subsidized, income taxes distort skill formation, besides labor supply. This additional tax distortion lowers optimal income taxes below levels that would be obtained in case skill formation would be exogenous, i.e., where  $\varepsilon_{\Theta,t} = 0$ ; see also Jacobs (2005).

#### 4.1.3 Optimal net tax on education $\Delta$

The optimality condition for education subsidies is given in Eq. (23). The left-hand side gives the marginal distortions of taxing education on a net basis, while the right-hand side gives the distributional benefits of doing so, for any value of the income tax rate, including the optimal level.<sup>21</sup> If  $\Delta > 0$ , education is taxed on a net basis. The optimal education subsidy  $s$  follows from the net tax on education  $\Delta \equiv tw^H\Theta_\Theta^H - tw^L\Theta_\Theta^L - sp(\Theta)$ . Education is distorted downwards more if the optimal net tax on education  $\frac{\Delta}{(1-t)\bar{z}}$  is larger. Distortions on education are larger (higher  $\Delta$ ) if the income tax  $t$  is set at a higher level—*ceteris paribus*.  $\Theta f(\Theta)$  is the same as

<sup>20</sup> The reverse is also true: if low-skilled individuals have the highest uncompensated wage elasticity of labor supply, then income taxes depress labor supply of low-skilled workers more than that of high-skilled workers, which results in smaller before-tax wage differentials, and income taxes are optimally increased for that reason.

<sup>21</sup> Hence, the expression also describes the policy optimum if the income tax would not be available to the policy maker. The expression for the optimal education subsidy does assume, however, that transfers are optimized.



in Eq. (22). It captures the size of the tax base at the marginal graduate  $\Theta$ . The larger the education elasticity  $\varepsilon_{\Theta,s}$  with respect to the subsidy rate  $s$ , the more skill formation responds to net taxes on education, and the lower should be the optimal net tax on education.

For given distributional benefits of net taxes on education on the right-hand side of Eq. (23), and for a given elasticity of education on the left-hand side of Eq. (23), the optimal subsidy  $s$  on education rises if the income tax rate  $t$  increases, so as to keep the net tax  $\Delta$  constant. These results are similar to Bovenberg and Jacobs (2005) who show that education subsidies should increase if income taxes are higher so as to alleviate the distortions of the income tax on skill formation—*ceteris paribus*.<sup>22</sup>

Note that there is no impact of education subsidies on labor-supply distortions. Intuitively, a marginally higher education subsidy does not directly affect labor supply on the intensive margin. However, the subsidy does affect labor supply indirectly via changes in the wage distribution.

The distributional gains of net taxes on education are given on the right-hand side of Eq. (23). Since  $\zeta > 0$ , taxing education yields net distributional benefits. The higher is the distributional gain of taxing education  $\zeta$ , the higher is the net tax education—*ceteris paribus*. In contrast to Bovenberg and Jacobs (2005), it is generally not optimal to set the education subsidy exactly equal to the tax rate (i.e.,  $s = t$ ) to obtain a zero net tax on education (i.e.,  $\Delta = 0$ ). Since investment in education generates infra-marginal rents for all but the marginally skilled individual, the government likes to tax these rents and redistribute income from high-skilled to low-skilled workers. This finding is in line with Findeisen and Sachs (2016) and Colas et al. (2021), who also analyze optimal education policies with discrete education choices.<sup>23</sup>

Furthermore, lower net taxes (or even net subsidies) on education generate general-equilibrium effects on wages that are captured by  $\rho(\tilde{g}^L - \tilde{g}^H)\varepsilon_{GE}$ . Lower taxes (or higher subsidies) give distributional gains, since pre-tax income inequality declines. Social welfare then increases, since the income-weighted social welfare weights of the low-skilled are higher than those of the high-skilled ( $\tilde{g}^L > \tilde{g}^H$ ). The general-equilibrium elasticity  $\varepsilon_{GE}$  captures the strength of general-equilibrium effects. If general-equilibrium effects are sufficiently strong, education may even be subsidized on a net basis rather than taxed on a net basis (i.e.,  $\Delta < 0$ ), which is in fact the case in our baseline simulation below. This finding confirms Dur and Teulings (2004) who analyze optimal log-linear tax and education policies in an assignment model of the labor market and find that optimal education subsidies may need to be positive. In the absence of general-equilibrium effects ( $\sigma = \infty$ ), the general-equilibrium elasticity is zero ( $\varepsilon_{GE} = 0$ ), and education subsidies are not deployed to exploit general-equilibrium effects for income redistribution.

<sup>22</sup> See also Maldonado (2008), Bohacek and Kapicka (2008), Anderberg (2009), Jacobs and Bovenberg (2011), and Stantcheva (2017).

<sup>23</sup> Related is Gomes et al. (2018) who show that it is optimal to distort occupational choice in a two-sector model if optimal income taxes cannot be conditioned on occupation as in our model.

The finding that education may be subsidized on a net basis contrasts with Jacobs (2012), who also analyzes optimal linear taxes and education subsidies with general-equilibrium effects. However, he models education on the intensive rather than the extensive margin, as in Bovenberg and Jacobs (2005). Education subsidies should then not be employed to generate general-equilibrium effects, because the general-equilibrium effect of linear education subsidies is identical to the general-equilibrium effect of linear income taxes. Hence, education subsidies have no distributional value added over income taxes, but only generate additional distortions in education.

If the government does not have access to income taxes at all ( $t = 0$ ), then Eq. (23) reduces to

$$-s \frac{\pi}{\bar{z}} \Theta^{1-\psi} f(\Theta) \varepsilon_{\Theta,s} = \frac{s\pi}{\bar{z}} \zeta - \rho(\bar{g}^L - \bar{g}^H) \varepsilon_{GE}. \quad (25)$$

In this case, like in the general case, optimal subsidies on education remain ambiguous in sign. On the one hand, direct income redistribution, as captured by the first term on the right-hand side, calls for a tax on education, while on the other hand general-equilibrium effects, as captured by the second term on the right-hand side, call for a subsidy on education.

Our findings also differ from Jacobs and Thuemmel (2021). They analyze optimal nonlinear income taxes that can be conditioned on skill type in a random participation model with education-dependent nonlinear taxes and individuals differing along two dimensions: earning ability and costs of education. They find that education is always taxed on a net basis, in contrast to this paper. In their framework, general-equilibrium effects do *not* enter optimal policy rules for both income taxes and education subsidies. The reason is that any redistribution from high-skilled to low-skilled workers via a general-equilibrium effects can be achieved as well with the income tax system, while the distortions in education can be avoided. Our analysis shows that tax and education policies should exploit general-equilibrium effects on the wage distribution in the realistic case that tax schedules cannot be conditioned on education. By generating general-equilibrium effects on wages, the government can redistribute more income beyond what can be achieved with the income tax system alone.<sup>24</sup>

## 4.2 Effects of SBTC on optimal policy

To understand the mechanisms behind the optimal policy response to SBTC, we study the comparative statics of the optimal policy rules with respect to SBTC. SBTC affects optimal policy through three channels: (1) distributional benefits, (2) education distortions, and (3) general-equilibrium effects. We do not report the

<sup>24</sup> Furthermore, we should note that it is not the linearity of the tax schedule that drives our results. If we would allow for skill-dependent linear tax rates, general-equilibrium effects will also not be exploited for income redistribution, because skill-dependent linear taxes can achieve exactly the same amount of income redistribution. The reason is that wage rates are linear prices so that linear tax rates are sufficient to achieve the same income redistribution as general-equilibrium effects.

effect of SBTC on labor-supply distortions, since the marginal excess burden of income taxes ( $\epsilon t/(1 - t)$ ) is not affected by SBTC because the labor-supply elasticity  $\epsilon$  is the same for all individuals.

We analytically derive how an increase in skill bias affects the terms in the formula for the optimal income tax rate in Eq. (22) and in the formula for the optimal subsidy rate in Eq. (23). Online Appendix A contains the formal derivations and more detailed explanations. Table 1 summarizes the analytical comparative statics and shows that the impact of SBTC on *all* elements of the expressions for optimal income taxes and optimal education subsidies in Proposition 1 is theoretically ambiguous. To gain a better understanding of the sign and quantitative size of these effects, we proceed by numerically analyzing the impact of SBTC on optimal policy. Table 1 also summarizes the outcomes of our simulations of the impact of SBTC on optimal policy rules, to which we turn next.

## 5 Simulation

In this section, we simulate the consequences of SBTC for optimal tax and education policy. To do so, we first calibrate the model to the US economy. Then, we analyze the comparative statics of SBTC on optimal policy and uncover the channels through which SBTC affects optimal tax and education policy. Given the ambiguous theoretical effects of SBTC on optimal policy, the purpose of the simulations is to better understand how optimal policy should respond to SBTC in a reasonably quantified model.

### 5.1 Calibration

We calibrate our model to the US economy using data from the US Current Population Survey.<sup>25</sup> We choose 1980 as the base year for the calibration, since evidence of SBTC emerges around that time. The final year is 2016. To compute levels and changes in the skill premium and the share of high-skilled workers in the data, we classify individuals with at least a college degree as high-skilled and all other individuals as low-skilled. The share of high-skilled workers in the working population was 24% in 1980 and 47% in 2016. We define the skill premium as average hourly earnings of high-skilled workers relative to average hourly earnings of low-skilled workers minus one:

$$\text{skill premium} \equiv \frac{w^H}{w^L} \frac{\frac{1}{1-F(\Theta)} \int_{\Theta}^{\bar{\theta}} \theta dF(\theta)}{\frac{1}{F(\Theta)} \int_{\underline{\theta}}^{\Theta} \theta dF(\theta)} - 1. \tag{26}$$

In the data, the skill premium changed from 47% in 1980 to 83% in 2016, which is an increase of 76%.

<sup>25</sup> Details of the data and our sample are discussed in Appendix C.

**Table 1** Effect of SBTC on determinants of optimal tax and subsidy rate

	Distributional benefits	Education distortions	General-equilibrium effects
Comparative statics of the optimal tax rate			
Term in Eq. (22)	$\xi$	$\frac{\Delta}{(1-t)\bar{\varepsilon}} \Theta f(\Theta) \varepsilon_{\Theta,t}$	$(\bar{g}^L - \bar{g}^H) \varepsilon_{GE}$
Direction	$\left\{ \begin{array}{ll} \text{analytical} & \uparrow\downarrow \\ \text{numerical} & \uparrow \end{array} \right.$	$\begin{array}{l} \uparrow\downarrow \\ \downarrow \end{array}$	$\begin{array}{l} \uparrow\downarrow \\ \uparrow \end{array}$
Comparative statics of the optimal subsidy rate			
Term in Eq. (23)	$\frac{s\pi}{(1-t)\bar{\varepsilon}} \zeta$	$\frac{\Delta}{(1-t)\bar{\varepsilon}} \Theta f(\Theta) \varepsilon_{\Theta,s}$	$\rho(\bar{g}^L - \bar{g}^H) \varepsilon_{GE}$
Direction	$\left\{ \begin{array}{ll} \text{analytical} & \uparrow\downarrow \\ \text{numerical} & \uparrow \end{array} \right.$	$\begin{array}{l} \uparrow\downarrow \\ \downarrow \end{array}$	$\begin{array}{l} \uparrow\downarrow \\ \uparrow \end{array}$

Derivations for the analytical comparative statics are provided in Appendix A. The details of the numerical comparative statics are given in Table 4

We set the compensated wage elasticity of labor supply to  $\varepsilon = 0.3$ , based on evidence reported in the surveys of Blundell and MaCurdy (1999) and Meghir and Phillips (2010). Although estimated uncompensated labor-supply elasticities are typically lower, we use a higher value to approximate the compensated labor-supply elasticity. Moreover, in our model,  $\varepsilon$  can also be interpreted as the elasticity of taxable income, encompassing more intensive margins, such as tax avoidance and evasion. The empirical literature reports figures in the range of 0.15 to 0.40 for the elasticity of taxable income; see the survey by Saez et al. (2012). We study alternative values for the elasticity  $\varepsilon$  in the robustness checks. For the ability distribution  $F(\theta)$ , we choose a log-normal distribution with a Pareto tail.<sup>26</sup>

The production technology is modeled according to the production function in Eq. (7). We set the elasticity of substitution between skilled and unskilled workers to  $\sigma = 2.9$ , following Acemoglu and Autor (2012).<sup>27</sup> We normalize the level of skill bias in 1980 to  $A_{1980} = 1$ . SBTC between 1980 and 2016 then corresponds to an increase from  $A_{1980}$  to  $A_{2016}$ , while keeping all other parameters in the production function constant.

We adopt a social welfare function with a constant elasticity of inequality aversion  $\phi > 0$ :

$$\Psi(V_\theta) = \begin{cases} \frac{V_\theta^{1-\phi}}{1-\phi}, & \phi \neq 1 \\ \ln(V_\theta), & \phi = 1 \end{cases} \tag{27}$$

<sup>26</sup> See Reed and Jorgensen (2004) for a discussion of the log-normal Pareto distribution. We use the single Pareto log-normal which originates from their eq. (23) when  $\beta \rightarrow \infty$ .

<sup>27</sup> Katz and Murphy (1992) have estimated that  $\sigma = 1.41$  for the period 1963–1987. Acemoglu and Autor (2012) argue that for the period up until 2008, a value of  $\sigma = 2.9$  fits the data better.

$\phi$  captures the government's desire for income redistribution.  $\phi = 0$  corresponds to a utilitarian social welfare function, whereas for  $\phi \rightarrow \infty$  the social welfare function converges to a Rawlsian social welfare function.<sup>28</sup> In the simulations, we assume  $\phi = 1$  as this leads to optimal marginal tax rates that are in a similar range as the marginal tax rates observed in the data. We also consider other parameters for  $\phi$  in the robustness checks.

We calibrate the model taking the tax rate, the transfer, and the education subsidy for 1980 and 2016 as given. The marginal tax rate in 1980 was on average  $t = 34.1\%$ , while it was  $t = 27.5\%$  in 2016 (National Bureau for Economic Research, 2021). The transfer  $b$  is pinned down by the average tax rate, which was  $18.1\%$  in 1980 and  $15.8\%$  in 2016. The subsidy rate is set at  $s = 47\%$  for 1980 (Gumport et al., 1997) and at  $s = 35\%$  for 2016 (OECD, 2018).<sup>29</sup> The subsidy rate corresponds to the share of government spending in total spending on higher education. At the calibrated equilibrium, the tax system also pins down the level of government expenditure  $R$ . When we compute optimal policy, we set the revenue requirement to this level of government expenditure.

Finally, we calibrate the parameters of the cost function for education ( $\pi$  and  $\psi$ ) as well as the parameters of the production function ( $\omega$ , and  $A_{2016}$ ). We calibrate  $\psi$  in the cost function for education by targeting an enrollment elasticity of 0.17. We base this elasticity on estimates in Dynarski (2000). Like many other studies, Dynarski (2000) reports semi-elasticities, which are based on the effect of changes in tuition subsidies (in percent) on college enrollment (in percentage points).<sup>30</sup> It is commonly estimated that a \$1000 increase in tuition subsidies increases college enrollment by 3 to 5 percentage points; see Nielsen et al. (2010) for an overview.<sup>31</sup> We also include robustness checks for the enrollment elasticity. The parameters of the production function ( $\omega$  and  $A_{2016}$ ) are calibrated by targeting levels and changes in the skill premium.

We calibrate the model by setting parameters so as to minimize the sum of squared relative errors between the moments generated by our model and the corresponding empirical moments from the data. Since some moments are much easier to match than others, we impose the constraint that the relative deviation of model moments from their empirical counterparts does not exceed 30%. All calibrated parameters are summarized in Table 2. The implied moments are reported in Table 3.

As expected, our model generates a skill premium that is generally too high, since the wage distributions of low and high-skilled workers do not overlap: the

<sup>28</sup> The utilitarian social welfare function is non-redistributive, since the marginal utility of income is constant due to the quasi-linear utility function.

<sup>29</sup> OECD (2018) data on subsidies and spending on higher education are only available since 1995, which is why we rely on Gumport et al. (1997) for the subsidy in 1980.

<sup>30</sup> See Appendix D how we convert the semi-elasticity from Dynarski (2000) to the enrollment elasticity in our model.

<sup>31</sup> There is no solid empirical evidence on the response of college enrollment to tax changes. In our model, the enrollment elasticities with respect to the tax and subsidy rate are mechanically related; hence, we only target the elasticity of enrollment with respect to the subsidy.

least-earning high-skilled worker still earns a higher wage than the best-earning low-skilled worker. At the same time, the relative change in the skill premium is lower than in the data (59% instead of 76%). The share of high-skilled workers in the model is somewhat higher than in the data for the year 1980, but only slightly higher for the year 2016. The subsidy elasticity of enrollment is lower than our empirical target. Overall, the calibration nevertheless strikes a reasonable balance in matching the various moments in the data.

To illustrate how our model responds to SBTC, we first simulate an increase in skill bias while keeping taxes, subsidies, and transfers at their calibrated values, which we refer to as the ‘status-quo’ economy. The outcomes are plotted in Fig. 1. The share of high-skilled workers is slightly concave in skill bias, while the skill premium increases almost linearly with skill bias. As a benchmark, we also simulate an economy without taxes and education subsidies, which we refer to as the ‘laissez-faire’ economy.<sup>32</sup> Comparing the laissez-faire and the status-quo economy shows the effect of policy: under laissez-faire, the share of high-skilled workers is lower, and correspondingly, the skill premium is higher. We attribute this difference primarily to the education subsidy in the status-quo tax system. However, the differences between the two economies are small. Moreover, in both cases the effect of SBTC on the share of high-skilled and the skill premium is very similar.

## 5.2 Optimal policy and SBTC

We compute optimal policy for different levels of skill bias and show the results in Fig. 2. The optimal marginal tax rate  $t$  increases monotonically with skill bias from about 21% to 31% (Panel 2a). The optimal transfer—expressed as share of average earnings  $\bar{z}$ —increases monotonically from about 4% to 20% (Panel 2b). The optimal subsidy rate  $s$  falls monotonically from about 68% to 49% (Panel 2c).<sup>33</sup> Finally, Panel 2d shows the optimal net tax on skill formation  $\Delta$  as a fraction of average earnings  $\bar{z}$ . Since the optimal net tax is negative, education is subsidized on a net basis. This means that education is distorted upwards compared to the efficient level, i.e., there is ‘over-investment.’ From the simulations, we conclude that the general-equilibrium effects of education subsidies are stronger than the direct distributional losses of education subsidies. The net tax as a fraction of average earnings increases monotonically with SBTC from about –10% to –7%. In other words, the net subsidy on education becomes *smaller* with SBTC.

## 5.3 Decomposition into different channels

We now uncover the three channels through which SBTC affects optimal policy. To do so, we start out from the optimum at  $A = 1$  and then increase the level of skill

<sup>32</sup> In the laissez-faire economy, we adjust the transfer  $b$  to satisfy the government budget constraint, which neither affects the share of high-skilled workers nor the skill premium.

<sup>33</sup> We also compute the optimal tax rate, while keeping the subsidy at its status-quo level, and vice versa. The results are shown in Tables 1 and 2 in Online Appendix C. Qualitatively, optimal taxes and subsidies behave in the same way as in the full optimum.

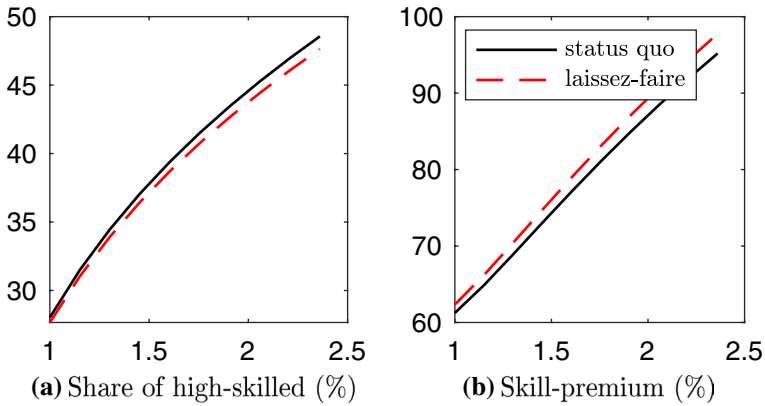
**Table 2** Calibration: parameters

Parameter	Description	Value	Source
$\mu^\theta$	Mean ability	0.22	Calibrated
$\zeta^\theta$	Standard deviation ability	0.01	Calibrated
$\alpha$	Pareto parameter	5.00	Calibrated
$\varepsilon$	Labor-supply elasticity	0.30	Blundell and Macurdy (1999); Meghir and Phillips (2010)
$\sigma$	Elasticity of substitution	2.90	Acemoglu and Autor (2012)
$\pi$	Average cost education	157.9	Calibrated
$\psi$	Elasticity of education	15.8	Calibrated
$\omega$	Share parameter	0.51	Calibrated
$\phi$	Inequality aversion	0.30	Calibrated
$A_{1980}$	Skill bias 1980	1.00	Normalized
$A_{2016}$	Skill bias 2016	2.36	Calibrated
$t_{1980}$	Marginal tax rate in 1980 (%)	34.1	NBER Taxsim
$t_{2016}$	Marginal tax rate in 2016 (%)	27.5	NBER Taxsim
$\bar{t}_{1980}$	Average tax rate in 1980 (%)	18.1	NBER Taxsim
$\bar{t}_{2016}$	Average tax rate in 2016 (%)	15.8	NBER Taxsim
$s_{1980}$	Subsidy rate in 1980 (%)	47.0	Gumport et al. (1997)
$s_{2016}$	Subsidy rate in 2016 (%)	35.0	OECD (2018)
$b_{1980}$	Tax intercept in 1980	0.10	Calibrated
$b_{2016}$	Tax intercept in 2016	0.15	Calibrated
$R_{1980}$	Government revenue in 1980	0.11	Implied
$R_{2016}$	Government revenue in 2016	0.19	Implied

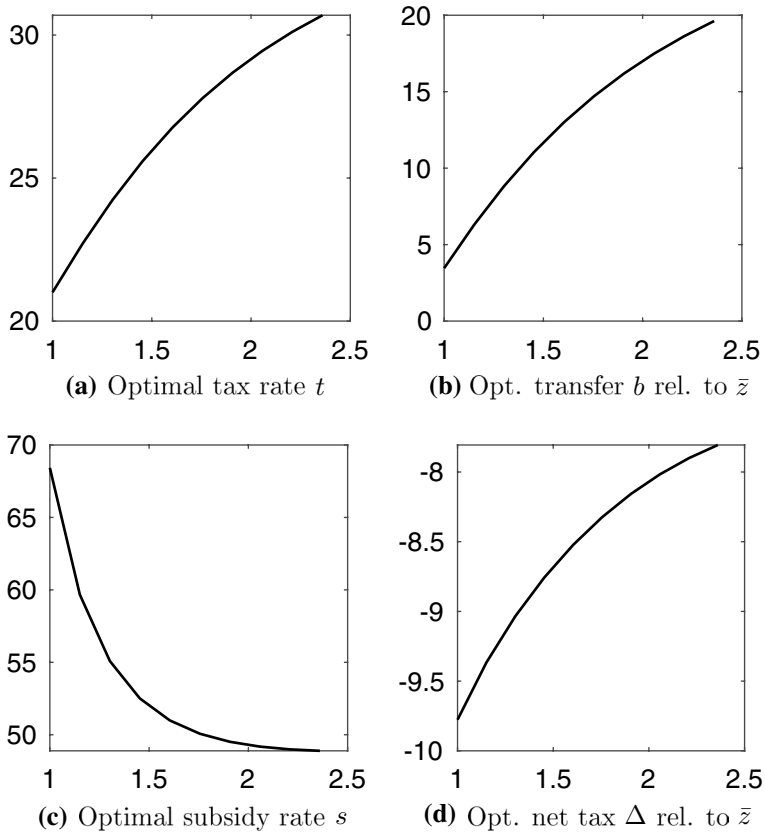
**Table 3** Calibration: model versus data

Moment	Model	Data
Skill premium in 1980 (%)	61.25	47.24
Skill premium in 2016 (%)	97.39	83.24
Share high-skilled in 1980 (%)	28.03	24.06
Share high-skilled in 2016 (%)	47.76	47.18
Enrollment elasticity	0.12	0.17

bias, while holding  $s$  and  $t$  fixed. We then compute how each of the terms in the first-order conditions in Eqs. (22) and (23) is affected by the increase in skill bias. For each term, we report its initial level and its change due to SBTC in Table 4. The terms in the table are grouped according to the three channels by which SBTC affects optimal policy: (1) distributional benefits, (2) distortions in education, and (3) general-equilibrium effects. The effects have already been summarized in Table 1. Table 4 provides the quantification. We now discuss them in detail.



**Fig. 1** Effect of SBTC under status-quo tax system and under laissez-faire. *Note* The horizontal axis corresponds to skill bias  $A$ . Status quo refers to the tax system used in the calibration, and summarized in Table 2. Laissez-faire corresponds to  $t = 0$  and  $s = 0$



**Fig. 2** Optimal policy under SBTC (all in %), skill bias  $A$  on the horizontal axis



**Table 4** Decomposition into different channels

Expression	Initial value	Change
Distributional benefits of the income tax and education tax		
$\xi$	806.75	68.31
$\zeta$	0.08	0.03
Tax and subsidy distortions on education		
$\frac{\Delta}{(1-t)\bar{z}}f(\Theta)\Theta\epsilon_{\Theta,t}$	-52.83	-46.46
$\frac{\Delta}{(1-t)\bar{z}}f(\Theta)\Theta\epsilon_{\Theta,s}$	-88.07	-77.44
General-equilibrium effects		
$(\bar{g}^L - \bar{g}^H)\epsilon_{GE}$	61.90	20.27
$\rho(\bar{g}^L - \bar{g}^H)\epsilon_{GE}$	103.19	33.78

All table entries have been multiplied by 1e+04

### 5.3.1 Comparative statics of the optimal tax rate

**5.3.1.1 Distributional benefits of income taxes  $\xi$**  The effect of SBTC on the distributional benefits of income taxes  $\xi$  is determined by changes in the income distribution and in the social welfare weights.

By raising the ratio of wage rates  $w^H/w^L$ , SBTC changes the income distribution: directly by increasing before-tax wage differentials and indirectly by affecting labor-supply and education decisions. Since the increase in labor supply is larger if the wage rate or a worker’s ability are higher, income inequality between and within skill groups increases. Moreover, investment in education rises with SBTC, which also increases income inequality. General-equilibrium effects dampen the labor-supply and education responses by compressing wage differentials, but do not fully offset the direct increase in inequality. For given social welfare weights  $g_\theta$ , SBTC thus increases the distributional benefits of taxing income  $\xi$ .

However, also the social welfare weights change with SBTC. Social welfare weights decline with utility, since the government is inequality averse. High-ability workers experience the largest infra-marginal utility gain due to SBTC. As a consequence, social welfare weights for high-ability workers fall more than for low-ability workers.

The impact of SBTC on  $\xi$  is, therefore, analytically ambiguous: it raises both the utility of the high-ability individuals relatively more and lowers their social welfare weights more. In the numerical comparative statics, we find that SBTC raises the distributional benefits of taxing income (Table 4). The immediate effects on inequality thus dominate changes in social welfare weights. Ceteris paribus, higher distributional benefits of income taxes  $\xi$  thus call for an increase in the optimal tax rate.

**5.3.1.2 Education distortions of income taxes  $\frac{\Delta}{(1-t)\bar{z}}\Theta f(\Theta)\epsilon_{\Theta,t}$**  To disentangle the various effects of SBTC on education distortions, we begin with the first term in the expression for education distortions,  $\frac{\Delta}{(1-t)\bar{z}}$ . On the one hand, the net tax on education  $\Delta \equiv tw^H\Theta l_\Theta^H - tw^L\Theta l_\Theta^L - sp(\Theta)$  increases, because SBTC raises the wage differential

between the marginally high-skilled and the marginally low-skilled worker—*ceteris paribus*. On the other hand, if education is subsidized ( $s > 0$ ), the net tax  $\Delta$  falls, because subsidies increase as SBTC lowers the marginal graduate  $\Theta$ , who has higher costs of education—*ceteris paribus*.<sup>34</sup> Turning to the denominator, SBTC raises average income  $\bar{z}$ . The overall impact of SBTC on  $\frac{\Delta}{(1-t)\bar{z}}$  is that it declines in our simulations.

Next, we turn to the size of the tax base at the marginal graduate,  $\Theta f(\Theta)$ . Analytically, the impact on this expression is ambiguous. SBTC lowers  $\Theta$ , but whether or not  $\Theta f(\Theta)$  increases depends on the location of  $\Theta$  in the skill distribution, i.e., before or after the mode. We find numerically that the tax base  $\Theta f(\Theta)$  increases with SBTC; hence, distortions on education become larger for that reason.

Finally, SBTC changes the elasticity of education with respect to the tax rate  $\varepsilon_{\Theta,t} = \zeta \frac{\sigma+\varepsilon}{\sigma+\varepsilon+\zeta\delta(\beta-\alpha)} > 0$ ; see Appendix A. It raises the income share of the high-skilled workers  $\alpha$  and reduces the measure of the inverse skill premium  $\beta$ . However, the impact of SBTC on  $\delta$ —the mass of labor at the skill cutoff  $\Theta$ —is ambiguous, making its overall impact on  $\varepsilon_{\Theta,t}$  ambiguous as well. In the numerical comparative statics,  $\varepsilon_{\Theta,t}$  slightly increases.

Numerically, we find that education is distorted upwards; the net tax on education is negative ( $\Delta < 0$ ). Moreover, SBTC exacerbates these upward distortions. As upward education distortions become even larger with SBTC, the tax rate should increase, *ceteris paribus* (Table 4).

**5.3.1.3 General-equilibrium effects of income taxes**  $(\tilde{g}^L - \tilde{g}^H)\varepsilon_{GE}$  To understand how SBTC affects general-equilibrium effects of income taxes, we need to know how the difference in income-weighted social welfare weights for the low- and the high-skilled workers,  $\tilde{g}^L - \tilde{g}^H$ , is affected by SBTC. First, SBTC raises income inequality between and within education groups, as argued above. Second, SBTC affects the composition of education groups as more individuals become high-skilled. Since the highest low-skilled worker and the lowest high-skilled worker now have a lower ability, both  $\tilde{g}^L$  and  $\tilde{g}^H$  increase when keeping the schedule of social welfare weights fixed. However, even for a fixed schedule of social welfare weights the net impact on  $\tilde{g}^L - \tilde{g}^H$  is not clear, as it is ambiguous whether  $\tilde{g}^L$  or  $\tilde{g}^H$  increases more. Third, the schedule of social welfare weights changes, as also argued above. Social welfare weights for individuals with higher ability or education decrease relative to social welfare weights of the individuals with lower ability or education, so that  $\tilde{g}^L - \tilde{g}^H$  increases. Taking these effects together, the analytical impact of SBTC on  $\tilde{g}^L - \tilde{g}^H$  is ambiguous, while the numerical impact is negative. Although the average social welfare weight of the low-skilled workers and the high-skilled workers both increase, this increase is found to be smaller for the low-skilled than for the high-skilled workers. Hence, the impact of larger inequality on social welfare weights is offset by the change in the composition of high- and low-skilled workers, and the impact of declining social welfare weights due to larger inequality.

Next, we turn to the impact of SBTC on the general-equilibrium elasticity  $\varepsilon_{GE} = \frac{\alpha(1-\alpha)\zeta\delta}{\sigma+\varepsilon+\zeta\delta(\beta-\alpha)}$ . SBTC raises the income share  $\alpha$  of high-skilled workers and

<sup>34</sup> If in contrast,  $s < 0$ , the net tax  $\Delta$  unambiguously increases with SBTC.

reduces the measure of the inverse skill premium  $\beta$ . However, the analytical impact of SBTC on  $\delta$ , and thus on  $\varepsilon_{GE}$  overall, is ambiguous. Numerically, SBTC increases  $\varepsilon_{GE}$ . Hence, if SBTC becomes more important, the skill premium responds more elastically to changes in policy. Since  $\varepsilon_{GE}$  increases relatively more than  $\tilde{g}^L - \tilde{g}^H$  decreases, we find that general-equilibrium effects of income taxes become more important with SBTC. *Ceteris paribus*, this calls for lower income taxes (Table 4).

**5.3.1.4 All effects combined** Whether the income tax rate rises or falls with SBTC depends on which effects dominate. The increase in distributional benefits as well as larger distortions of net subsidies on education calls for an increase in the income tax, whereas stronger general-equilibrium effects tend to lower income taxes. Numerically, we find that the first two effects dominate (Table 4). As a consequence, SBTC leads to a higher optimal income tax rate.

### 5.3.2 Comparative statics of the optimal subsidy rate

**5.3.2.1 Distributional losses of education subsidies**  $\frac{s\pi}{(1-t)\bar{z}}\zeta$  SBTC affects the distributional characteristic of education  $\zeta$  by changing the social welfare weights  $g_\theta$ , and by lowering the threshold  $\Theta$  as more individuals become high-skilled. Like before, the impact of SBTC on social welfare weights is ambiguous. In contrast to the impact of  $\Theta$  on the distributional characteristic of income  $\xi$ , the decrease in  $\Theta$  leads to a higher distributional characteristic of education  $\zeta$ —*ceteris paribus*. Intuitively, as more individuals with lower social welfare weights become high-skilled, the average welfare weight of low-skilled workers increases, and it becomes more desirable to tax education on a net basis. General-equilibrium effects dampen the labor-supply and education responses, thereby limiting the rise in pre-tax inequality. Numerically, we find that SBTC raises the distributional benefits of taxing education  $\zeta$ . Hence, as the distributional losses of education subsidies increase, the subsidy rate should decrease with SBTC, *ceteris paribus* (Table 4).

**5.3.2.2 Education distortions of education subsidies**  $\frac{\Delta}{(1-t)\bar{z}}\Theta f(\Theta)\varepsilon_{\Theta,s}$  The distortions of taxes and subsidies on education only differ by a factor  $\rho \equiv \frac{s}{(1-s)(1+\varepsilon)} > 0$ , which captures the importance of education subsidies in the total direct costs of education (see also Table 5). This factor is not affected by SBTC. As a consequence, the direction in which SBTC affects distortions on skill formation is the same for taxes and subsidies. As we have argued above, we cannot analytically sign the effect. Numerically, the optimal net tax on education is negative, i.e., there is optimally a net subsidy on education resulting in over-investment. Moreover, we find that SBTC exacerbates these education distortions. Hence, *ceteris paribus*, the optimal subsidy rate should decrease with SBTC (Table 4).

**5.3.2.3 General-equilibrium effects of education subsidies**  $\rho(\tilde{g}^L - \tilde{g}^H)\varepsilon_{GE}$  The general-equilibrium effects of taxes and subsidies also differ only by factor  $\rho$ . It follows from our discussion of the general-equilibrium effects of taxes that we cannot analytically sign the effect, while numerically we find an increase (Table 4). As the general-equilibrium effect of education subsidies becomes more important with SBTC, the optimal subsidy rate should increase, *ceteris paribus*.

**5.3.2.4 Combined effect** While increased distributional losses and larger distortions due to over-investment in education call for a lower subsidy rate, the increased importance of general-equilibrium effects tends to increase optimal subsidies on education. Numerically, we find that the first two effects dominate (Table 4). As a consequence, the optimal subsidy rate falls with SBTC.

## 5.4 Robustness

We find numerically that the optimal tax rate increases with SBTC, while the optimal subsidy rate falls. We now investigate whether these findings are robust to changes in the most important model parameters. If so, even though we cannot sign the impact of SBTC analytically, we can be confident that the impact of SBTC on optimal policy holds more generally. We study the robustness of our findings with regard to (1) the government's inequality aversion, (2) the labor-supply elasticity, and (3) the subsidy elasticity of enrollment into higher education.

To study the effect of different labor-supply elasticities and enrollment elasticities, we recalibrate our model to match the moments in the data under the different elasticities. We present the calibration outcomes in Table 6 in Appendix E. In all calibrations, the change in the skill premium and the share of high-skilled workers is kept at baseline values. All models thus capture SBTC in a comparable way. To study the robustness of our results with respect to inequality aversion, we do not need to recalibrate our model, since the corresponding parameter  $\phi$  does not interact with the other model parameters and can thus be set independently. The range of skill bias  $A$  differs by scenario, as it is calibrated to match SBTC in the data.

### 5.4.1 Inequality aversion

The baseline assumes an elasticity of inequality aversion of  $\phi = 1$ . Figure 3 in Appendix E presents robustness checks for two additional levels of inequality aversion:  $\phi = 0.5$  and  $\phi = 1.5$ . Larger values of inequality aversion correspond to higher optimal tax and subsidy rates. Yet, the qualitative pattern is the same as in the baseline (represented by the solid black line); the optimal tax rate increases with SBTC, while the optimal subsidy rate falls.

### 5.4.2 Labor-supply elasticity

We compute optimal policy for  $\varepsilon = 0.1$  and  $\varepsilon = 0.5$  and present the results, alongside the baseline of  $\varepsilon = 0.3$  in Fig. 4 in Appendix E. As expected, if the labor-supply elasticity is lower, then optimal tax rates are higher. The optimal subsidy rate at  $\varepsilon = 0.5$  is similar to the baseline, while it is higher at  $\varepsilon = 0.1$ . Again, the qualitative pattern of a rising optimal tax and falling optimal subsidy rate with SBTC remains.

### 5.4.3 Subsidy elasticity of enrollment

In the baseline calibration, the subsidy elasticity of enrollment is 0.12. We compute results for two alternative scenarios with enrollment elasticities of 0.10 and 0.14, which we plot in Fig. 5 in Appendix E. If the enrollment elasticity is higher, then optimal tax rates are higher, while the opposite holds for optimal subsidy rates. Once more, we confirm the qualitative pattern of a rising optimal tax and falling optimal subsidy rate with SBTC.

## 5.5 Limitations and future research

We have studied a stylized model of how tax and education policy should respond to SBTC. In doing so, we focused on three first-order issues when thinking about optimal redistributive tax and education policy: direct distributional impacts, distortions in labor supply and education, and general-equilibrium effects of income taxes and education subsidies. Nevertheless, we ignored a number of potentially important real-world features to keep our analysis tractable. In particular, the income distributions of low- and high-skilled workers are not overlapping, and there are no credit constraints, information frictions, or externalities. The latter three features might justify government intervention in education Barr (2004). Allowing for them might therefore change our conclusions, which should, therefore, be taken with caution.

However, for any of these factors to affect our conclusions, they would need to interact with SBTC. It is clearly possible that SBTC—by generating larger income inequality—exacerbates problems with credit constraints. Jacobs and Yang (2016) demonstrate that tighter credit constraints typically raise optimal taxes and education subsidies (lower net taxes on education), thereby strengthening the main findings of this paper. Related is Colas et al. (2021), who analyze the optimal allocation of education subsidies across the income distribution (and other dimensions). They find that education subsidies should optimally be targeted toward students from lower socioeconomic backgrounds, because credit constraints, externalities or information frictions are more severe for them. Their findings also suggest that optimal policies may need to become more progressive in response to SBTC.

The analysis of optimal tax and education policy with SBTC and capital market constraints, externalities or information frictions is therefore an interesting and important avenue for future research.

## 6 Conclusion

This paper studies how optimal linear income tax and education policy should respond to skill-biased technical change (SBTC). To do so, we introduce intensive margin labor supply and a discrete education choice into the canonical model of SBTC based on Katz and Murphy (1992), Violante (2008), and Acemoglu and Autor (2011). We start by deriving expressions for the optimal income tax and education subsidy for a given level of skill bias. The income tax and subsidy trade off direct distributional benefits and general-equilibrium effects of each policy against distortions of each policy on labor supply and

education. Then, we analyze skill-biased technical change (SBTC), which is shown to have theoretically ambiguous impacts on both optimal income taxes and education subsidies, since SBTC simultaneously changes i) distributional benefits, ii) distortions in education, and iii) general-equilibrium effects.

To analyze the importance of each channel, the model is calibrated to the US economy to quantify the impact of SBTC on optimal policy. SBTC is found to make the tax system more progressive, since the distributional benefits of higher income taxes rise more than the tax distortions on education and the general-equilibrium effects of taxes. Moreover, education is subsidized on a net basis and is therefore above its efficient level. Hence, the subsidy indeed exploits general-equilibrium effects for redistribution. However, SBTC lowers optimal education subsidies, since the distributional losses and the distortions of higher education subsidies increase more than the general-equilibrium effects of subsidies.

In line with Tinbergen (1975) and Dur and Teulings (2004), we find that general-equilibrium effects do matter for the optimal design of tax and education policy. Moreover, our findings support the push for more progressive taxation in light of SBTC brought forward by Goldin and Katz (2010). However, Tinbergen and Goldin and Katz also advocate raising education subsidies to win the race against technology and to compress the wage distribution via general-equilibrium effects on wages. Our findings do not support this idea. The reason is that education subsidies not only compress wages, but also entail larger distributional losses and cause more over-investment in education as SBTC becomes more important. The latter are found to be quantitatively stronger than the larger benefits of using education subsidies to exploit general-equilibrium effects.

These results should be taken with caution as they have been derived in a stylized model, which made a number of important simplifications. Future research could fruitfully extend our analysis of optimal tax and education policy under SBTC to allow for overlapping wage distributions, borrowing constraints, information frictions, and externalities.

## Appendix

### A Derivation of elasticities

Table 5 provides the behavioral elasticities of all model variables with respect to the income tax and education subsidy. The derivations are given below.

In order to understand the behavioral elasticities with respect to tax and education policy, it is instructive to first consider the case in which general-equilibrium effects on wages are completely absent, i.e.,  $\sigma \rightarrow \infty$ . In this case, the production function becomes linear, and high- and low-skilled labor are perfect substitutes production. Consequently, all the terms in brackets in the expressions for the elasticities are either zero or one. The first two rows in Table 5 indicate that the wage rates of high-skilled and low-skilled workers are then invariant to taxes and education subsidies ( $\varepsilon_{w^h,t} = \varepsilon_{w^l,s} = 0$ ). Labor supplies only respond to income taxes, but not to education

subsidies ( $\epsilon_{\beta,t} = \epsilon$ ,  $\epsilon_{\beta,s} = 0$ ). An increase in the income tax rate depresses labor supply of both high-skilled and low-skilled workers and more so if the wage elasticity of labor supply  $\epsilon$  is larger. The education subsidy does not affect labor supply of high-skilled and low-skilled workers. With quasi-linear preferences, labor supply only depends on the net after-tax wage, which is unaffected by the education subsidy. Education responds to both taxes and education subsidies ( $\epsilon_{\Theta,t} = \epsilon_{\Theta,s}/\rho = \zeta$ ). A higher income tax rate discourages education, because not all costs of education are deductible. The education response is stronger if the combined elasticity  $\zeta \equiv \frac{1+\epsilon}{1+\epsilon+\psi}$  is larger. Complementarity of education with labor supply makes the education response more elastic. Moreover, the education subsidy boosts education more if the share of direct costs in education  $\rho$  is larger.

The behavioral elasticities change in the presence of general-equilibrium effects on the wage structure (i.e.,  $0 < \sigma < \infty$ ), so that in Table 5 the terms in brackets are no longer equal to 0 or 1. Now, the elasticities of wages with respect to the policy instruments, i.e.,  $\epsilon_{w^H,t}$  and  $\epsilon_{w^L,s}$ , are nonzero. If a policy increases the supply of high-skilled workers relative to the supply of low-skilled workers, the skill premium declines, i.e., the high-skilled wage rate falls relative to the low-skilled wage rate. These general-equilibrium effects change labor supply and education decisions. How strong these general-equilibrium effects on wages are, depends on the combined elasticity  $\zeta$ , the elasticity of substitution in production  $\sigma$ , and the wage elasticity of labor supply  $\epsilon$ . Policy can change relative supplies only via a change in investment in education, and not via changing labor supply (see the discussion below). The lower is  $\sigma$ , the more difficult it is to substitute high- and low-skilled workers in production. The lower is  $\epsilon$ , the less elastic labor supply responds to a change in the wage. The lower is  $\psi$ , the less elastic education responds to changes in wages. Hence, if  $\sigma$ ,  $\epsilon$ , and  $\psi$  are lower, general-equilibrium effects are stronger, i.e.,  $\epsilon_{w^H,t}$  and  $\epsilon_{w^L,s}$  are larger in absolute value.

From the expressions for  $\epsilon_{\beta,t}$  follows that there are two reasons why both high-skilled and low-skilled labor supply decline if the tax rate increases. First, a higher income tax directly distorts individual labor supply downward. Second, an increase in the tax

**Table 5** Elasticities with respect to tax rate  $t$  and subsidy rate  $s$

$\epsilon_{w^H,t} \equiv -\frac{\partial w^H}{\partial t} \frac{1-t}{w^H} = -\zeta \left( \frac{(1-\alpha)\delta}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) < 0$ ,	$\epsilon_{w^H,s} \equiv \frac{\partial w^H}{\partial s} \frac{s}{w^H} = -\zeta \left( \frac{(1-\alpha)\delta}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) \rho < 0$ ,
$\epsilon_{w^L,t} \equiv -\frac{\partial w^L}{\partial t} \frac{1-t}{w^L} = \zeta \left( \frac{\alpha\delta}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) > 0$ ,	$\epsilon_{w^L,s} \equiv \frac{\partial w^L}{\partial s} \frac{s}{w^L} = \zeta \left( \frac{\alpha\delta}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) \rho > 0$ ,
$\epsilon_{\beta,t} \equiv -\frac{\partial \beta}{\partial t} \frac{1-t}{\beta} = \left( \frac{\sigma+\epsilon-\zeta\delta(1-\beta)}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) \epsilon > 0$ ,	$\epsilon_{\beta,s} \equiv \frac{\partial \beta}{\partial s} \frac{s}{\beta} = -\left( \frac{(1-\alpha)\delta\zeta}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) \epsilon \rho < 0$ ,
$\epsilon_{\Theta,t} \equiv -\frac{\partial \Theta}{\partial t} \frac{1-t}{\Theta} = \left( \frac{\sigma+\epsilon+\zeta\delta\beta}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) \epsilon > 0$ ,	$\epsilon_{\Theta,s} \equiv \frac{\partial \Theta}{\partial s} \frac{s}{\Theta} = \left( \frac{\alpha\delta\zeta}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) \epsilon \rho > 0$ ,
$\epsilon_{\Theta,t} \equiv \frac{\partial \Theta}{\partial t} \frac{1-t}{\Theta} = \zeta \left( \frac{\sigma+\epsilon}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) > 0$ ,	$\epsilon_{\Theta,s} \equiv -\frac{\partial \Theta}{\partial s} \frac{s}{\Theta} = \zeta \left( \frac{\sigma+\epsilon}{\sigma+\epsilon+\zeta\delta(\beta-\alpha)} \right) \rho > 0$ .

*Note:* The term  $\beta \equiv 1/(1-(w^L/w^H)^{1+\epsilon})$  is a measure of the inverse skill premium,  $\delta \equiv (\Theta^L f(\Theta)/L + \Theta^H f(\Theta)/H)\Theta$  measures the importance of the marginal individual with ability  $\Theta$  in aggregate effective labor supply, and  $\rho \equiv \frac{s}{(1-s)(1+\epsilon)} > 0$  captures the importance of education subsidies in the total direct costs of education. Finally,  $\zeta \equiv \frac{1+\epsilon}{1+\epsilon+\psi}$  is a measure of the total education elasticity, which takes into account the feedback with labor supply

reduces investment in education, which in turn reduces relative supply of skilled labor, and wages of high-skilled labor increase relative to low-skilled labor. Hence, the direct effect of a tax increase on high-skilled labor supply  $l_\theta^H$  is dampened by the relative increase in  $w^H$ , whereas the drop in low-skilled labor supply  $l_\theta^L$  is exacerbated by the relative decline in  $w^L$ . As a result, the labor-supply elasticity of low-skilled labor is higher than that of high-skilled labor ( $\epsilon_{l_\theta^L} > \epsilon_{l_\theta^H}$ ).<sup>35</sup> Similarly, by boosting enrollment in education, the subsidy on higher education increases the supply of high-skilled workers relative to the supply of low-skilled workers. This generates general-equilibrium effects on the wage structure: high-skilled wages fall and low-skilled wages rise. Consequently, the education response to education subsidies is muted by general-equilibrium effects on high-skilled and low-skilled wages. High-skilled labor supply falls and low-skilled labor supply increases if the education subsidy rises due to the changes in wage rates.

### A.1 Derivation elasticities

We define  $\tilde{x} \equiv dx/x$  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 in Eq. (4) to obtain

$$\tilde{l}_\theta^H = \epsilon(\tilde{w}^H - \tilde{t}), \tag{28}$$

$$\tilde{l}_\theta^L = \epsilon(\tilde{w}^L - \tilde{t}). \tag{29}$$

Next, we linearize the cutoff ability  $\Theta$  in Eq. (6) to find

$$\tilde{\Theta} = \frac{1}{1 + \epsilon + \psi} \left[ (1 + \epsilon)\tilde{t} - \frac{s}{1 - s}\tilde{s} - (1 + \epsilon)\beta\tilde{w}_H - (1 + \epsilon)(1 - \beta)\tilde{w}_L \right], \tag{30}$$

where we define

$$\beta \equiv \frac{w_H^{1+\epsilon}}{w_H^{1+\epsilon} - w_L^{1+\epsilon}}. \tag{31}$$

Collecting terms, we obtain

$$\tilde{\Theta} = \frac{1 + \epsilon}{1 + \epsilon + \psi} \left[ \tilde{t} - \frac{s}{(1 + \epsilon)(1 - s)}\tilde{s} - \beta\tilde{w}_H - (1 - \beta)\tilde{w}_L \right]. \tag{32}$$

Define  $\zeta \equiv \frac{1+\epsilon}{1+\epsilon+\psi}$  and  $\rho \equiv \frac{s}{(1+\epsilon)(1-s)}$  to write

$$\tilde{\Theta} = \zeta\tilde{t} - \zeta\rho\tilde{s} - \zeta\beta\tilde{w}_H - \zeta(1 - \beta)\tilde{w}_L. \tag{33}$$

<sup>35</sup> Relative wage rates  $w^H/w^L$  change only due to the effect of taxes on the education margin, not due to direct changes in labor supply. This is because the direct effect of a tax increase on individual labor supplies does not lead to a change in relative supply  $H/L$ , since all individual labor supplies fall by the same relative amount due to the constant wage elasticity of labor supply.



Next, we log-linearize the labor market clearing conditions in Eqs. (13) and (14):

$$\tilde{H} = \varepsilon(\tilde{w}^H - \tilde{t}) - \delta_H \tilde{\Theta}, \quad \delta_H \equiv \frac{\Theta^2 l_{\theta}^H f(\Theta)}{H}, \tag{34}$$

$$\tilde{L} = \varepsilon(\tilde{w}^L - \tilde{t}) + \delta_L \tilde{\Theta}, \quad \delta_L \equiv \frac{\Theta^2 l_{\theta}^L f(\Theta)}{L}. \tag{35}$$

Finally, we log-linearize the wage equations in Eqs. (8) and (9) 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

$$\tilde{w}^H = \frac{(1 - \alpha)}{\sigma}(\tilde{L} - \tilde{H}), \tag{36}$$

$$\tilde{w}^L = \frac{\alpha}{\sigma}(\tilde{H} - \tilde{L}), \tag{37}$$

$$\alpha \equiv \frac{HY_H(\cdot)}{Y(\cdot)}, \quad \frac{1}{\sigma} \equiv \frac{Y_{LH}(\cdot)Y(\cdot)}{Y_L(\cdot)Y_H(\cdot)}, \tag{38}$$

where  $\alpha$  denotes the income share of skilled labor in total output and  $\sigma$  is the elasticity of substitution between low-skilled and high-skilled labor in production. We now can solve a system of seven linear equations (Eqs. (28), (29), (33), (34), (35), (36), and (37)) in seven unknowns ( $\tilde{l}_{\theta}^H, \tilde{l}_{\theta}^L, \tilde{\Theta}, \tilde{H}, \tilde{L}, \tilde{w}^H$ , and  $\tilde{w}^L$ ).

First, rewrite Eqs. (34) and (35) by subtracting them from each other

$$\tilde{H} - \tilde{L} = \varepsilon(\tilde{w}_H - \tilde{t}) - \delta_H \tilde{\Theta} - \varepsilon(\tilde{w}_L - \tilde{t}) - \delta_L \tilde{\Theta} = \varepsilon(\tilde{w}_H - \tilde{w}_L) - (\delta_H + \delta_L)\tilde{\Theta}.$$

Define  $\delta \equiv \delta_H + \delta_L$  and substitute Eq. (33) to find

$$\begin{aligned} \tilde{H} - \tilde{L} &= \varepsilon(\tilde{w}_H - \tilde{w}_L) - \delta(\zeta\tilde{t} - \zeta\rho\tilde{s} - \zeta\beta\tilde{w}_H - \zeta(1 - \beta)\tilde{w}_L) \\ &= (\varepsilon + \delta\zeta\beta)\tilde{w}_H + (-\varepsilon + \delta\zeta(1 - \beta))\tilde{w}_L - \delta\zeta\tilde{t} + \delta\zeta\rho\tilde{s}. \end{aligned} \tag{39}$$

Next, substitute  $\tilde{w}_H$  and  $\tilde{w}_L$  from Eqs. (36) and (37) to obtain

$$\begin{aligned} \tilde{H} - \tilde{L} &= -\left(\frac{\delta\zeta\sigma}{\sigma + \varepsilon + \delta\zeta(\beta - \alpha)}\right)\tilde{t} + \left(\frac{\delta\zeta\rho\sigma}{\sigma + \varepsilon + \delta\zeta(\beta - \alpha)}\right)\tilde{s} \\ &= \frac{\delta\zeta\sigma}{\sigma + \varepsilon + \delta\zeta(\beta - \alpha)}(-\tilde{t} + \rho\tilde{s}) \end{aligned} \tag{40}$$

Since  $\beta > 1$  and  $\alpha < 1$  and all other terms in  $\frac{\delta\zeta\sigma}{\sigma + \varepsilon + \delta\zeta(\beta - \alpha)}$  are positive, an increase in the tax rate reduces high-skilled labor relative to low-skilled labor, whereas an increase in the subsidy rate has the opposite effect. Substituting for  $\tilde{H} - \tilde{L}$  from Eqs. (36) and (37) yields

$$\tilde{w}^H = \frac{(1-\alpha)\delta\zeta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)}(\tilde{t} - \rho\tilde{s}), \quad (41)$$

and

$$\tilde{w}^L = \frac{\alpha\delta\zeta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)}(-\tilde{t} + \rho\tilde{s}). \quad (42)$$

Substituting these results into Eqs. (33), (28) and (29) and rearranging yields

$$\tilde{\Theta} = \zeta \left( \frac{\sigma + \varepsilon}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) \tilde{t} - \zeta \left( \frac{\sigma + \varepsilon}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) \rho\tilde{s}, \quad (43)$$

$$\tilde{l}_\theta^H = \varepsilon \left( \frac{\delta\zeta(1-\beta) - (\sigma + \varepsilon)}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \tilde{t} - \frac{(1-\alpha)\delta\zeta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \rho\tilde{s} \right), \quad (44)$$

$$\tilde{l}_\theta^L = \varepsilon \left( -\frac{\sigma + \varepsilon + \zeta\delta\beta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \tilde{t} + \frac{\alpha\delta\zeta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \rho\tilde{s} \right). \quad (45)$$

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

$$\varepsilon_{\Theta,t} \equiv \frac{\partial \Theta}{\partial t} \frac{1-t}{\Theta} = \frac{\tilde{\Theta}}{\tilde{t}} = \zeta \left( \frac{\sigma + \varepsilon}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) > 0, \quad (46)$$

$$\varepsilon_{w^L,t} \equiv -\frac{\partial w^L}{\partial t} \frac{1-t}{w^L} = -\frac{\tilde{w}^L}{\tilde{t}} = \zeta \left( \frac{\alpha\delta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) > 0, \quad (47)$$

$$\varepsilon_{w^H,t} \equiv -\frac{\partial w^H}{\partial t} \frac{1-t}{w^H} = -\frac{\tilde{w}^H}{\tilde{t}} = -\zeta \left( \frac{(1-\alpha)\delta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) < 0. \quad (48)$$

$$\varepsilon_{l_\theta,t} \equiv -\frac{\partial l_\theta^L}{\partial t} \frac{1-t}{l_\theta^L} = -\frac{\tilde{l}_\theta^L}{\tilde{t}} = \left( \frac{\sigma + \varepsilon + \zeta\delta\beta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) \varepsilon > 0, \quad (49)$$

$$\varepsilon_{l_\theta^H,t} \equiv -\frac{\partial l_\theta^H}{\partial t} \frac{1-t}{l_\theta^H} = -\frac{\tilde{l}_\theta^H}{\tilde{t}} = \left( \frac{\sigma + \varepsilon - \delta\zeta(1-\beta)}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) \varepsilon > 0. \quad (50)$$

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

$$\varepsilon_{\Theta,s} \equiv -\frac{\partial \Theta}{\partial s} \frac{s}{\Theta} = -\frac{\tilde{\Theta}}{\tilde{s}} = \zeta \left( \frac{\sigma + \varepsilon}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) \rho > 0, \quad (51)$$

$$\varepsilon_{w^L,s} \equiv \frac{\partial w^L}{\partial s} \frac{s}{w^L} = \frac{\tilde{w}^L}{\tilde{s}} = \zeta \left( \frac{\alpha\delta}{\sigma + \varepsilon + \zeta\delta(\beta - \alpha)} \right) \rho > 0, \quad (52)$$

$$\varepsilon_{w^H,s} \equiv \frac{\partial w^H}{\partial s} \frac{s}{w^H} = \frac{\tilde{w}^H}{\tilde{s}} = -\zeta \left( \frac{(1-\alpha)\delta}{\sigma + \varepsilon + \zeta\delta(\beta-\alpha)} \right) \rho < 0, \tag{53}$$

$$\varepsilon_{l^L,s} \equiv \frac{\partial l^L_\theta}{\partial s} \frac{s}{l^L_\theta} = \frac{\tilde{l}^L_\theta}{\tilde{s}} = \left( \frac{\alpha\delta\zeta}{\sigma + \varepsilon + \zeta\delta(\beta-\alpha)} \right) \varepsilon \rho > 0, \tag{54}$$

$$\varepsilon_{l^H,s} \equiv \frac{\partial l^H_\theta}{\partial s} \frac{s}{l^H_\theta} = \frac{\tilde{l}^H_\theta}{\tilde{s}} = -\left( \frac{(1-\alpha)\delta\zeta}{\sigma + \varepsilon + \zeta\delta(\beta-\alpha)} \right) \varepsilon \rho < 0. \tag{55}$$

**A.2 Elasticities with fixed  $\Theta$**

Suppose  $\Theta$  is fixed, and thus  $\tilde{\Theta} = 0$ . Then, Eqs. (34) and (35) simplify to

$$\tilde{H} = \varepsilon(\tilde{w}^H - \tilde{t}), \tag{56}$$

$$\tilde{L} = \varepsilon(\tilde{w}^L - \tilde{t}). \tag{57}$$

Substituting these results in Eqs. (36) and (37) gives

$$\tilde{w}^H - \tilde{w}^L = \frac{(1-\alpha)}{\sigma}(\tilde{L} - \tilde{H}) + \frac{\alpha}{\sigma}(\tilde{L} - \tilde{H}) = (\tilde{L} - \tilde{H})\frac{1}{\sigma} = \varepsilon(\tilde{w}^L - \tilde{w}^H), \tag{58}$$

which holds only if  $\tilde{w}^L - \tilde{w}^H = 0$ . This implies  $\tilde{w}^L = \tilde{w}^H$ , and thus from Eqs. (56) and (57),  $\tilde{L} = \tilde{H}$ . Hence, if  $\Theta$  is fixed, policy does not affect wages. A change  $\tilde{t}$  still affects labor supplies, but it does so symmetrically across skill groups. Hence, both  $s$  and  $t$  affect wages only via changing  $\Theta$ .

**B Optimal policy**

Introducing  $\eta$  as the Lagrange multiplier on the government budget constraint, we can formulate the Lagrangian for maximizing social welfare as

$$\begin{aligned} \max_{b,t,s} \mathcal{L} \equiv & \int_{\underline{\theta}}^{\Theta} \Psi(V^L_\theta) dF(\theta) + \int_{\Theta}^{\bar{\theta}} \Psi(V^H_\theta) dF(\theta) \\ & + \eta \left[ \int_{\underline{\theta}}^{\Theta} tw^L \theta l^L_\theta dF(\theta) + \int_{\Theta}^{\bar{\theta}} (tw^H \theta l^H_\theta - s\pi\theta^{-\psi}) dF(\theta) - b - R \right], \end{aligned} \tag{59}$$

Define marginal social utility as

$$\Psi'_\theta \equiv \begin{cases} \Psi'(V^L_\theta) & \text{if } \theta < \Theta, \\ \Psi'(V^H_\theta) & \text{if } \theta \geq \Theta. \end{cases} \tag{60}$$

The necessary first-order conditions for an optimum are given by

$$\frac{\partial \mathcal{L}}{\partial b} = \int_{\underline{\theta}}^{\Theta} \Psi'_{\theta} \frac{\partial V_{\theta}^L}{\partial b} dF(\theta) + \int_{\Theta}^{\bar{\theta}} \Psi'_{\theta} \frac{\partial V_{\theta}^H}{\partial b} dF(\theta) - \eta = 0, \quad (61)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} &= \int_{\underline{\theta}}^{\Theta} \Psi' \frac{\partial V_{\theta}^L}{\partial t} dF(\theta) + \int_{\Theta}^{\bar{\theta}} \Psi' \frac{\partial V_{\theta}^H}{\partial t} dF(\theta) \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} w^L \theta l_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} w^H \theta l_{\theta}^H dF(\theta) \right] \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} t w^L \theta \frac{\partial l_{\theta}^L}{\partial t} dF(\theta) + \int_{\Theta}^{\bar{\theta}} t w^H \theta \frac{\partial l_{\theta}^H}{\partial t} dF(\theta) \right] \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} t \frac{\partial w^L}{\partial t} \theta l_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} t \frac{\partial w^H}{\partial t} \theta l_{\theta}^H dF(\theta) \right] \\ &+ \underbrace{[\Psi(V_{\Theta}^L) - \Psi(V_{\Theta}^H)]}_{=0} f(\Theta) \frac{\partial \Theta}{\partial t} - \eta [t w^H \Theta l_{\Theta}^H - t w^L \Theta l_{\Theta}^L - s \pi \Theta^{-\psi}] f(\Theta) \frac{\partial \Theta}{\partial t} = 0, \end{aligned} \quad (62)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} &= \int_{\underline{\theta}}^{\Theta} \Psi' \frac{\partial V_{\theta}^L}{\partial s} dF(\theta) + \int_{\Theta}^{\bar{\theta}} \Psi' \frac{\partial V_{\theta}^H}{\partial s} dF(\theta) - \eta \pi \left[ \int_{\Theta}^{\bar{\theta}} \theta^{-\psi} dF(\theta) \right] \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} t w^L \theta \frac{\partial l_{\theta}^L}{\partial s} dF(\theta) + \int_{\Theta}^{\bar{\theta}} t w^H \theta \frac{\partial l_{\theta}^H}{\partial s} dF(\theta) \right] \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} t \frac{\partial w^L}{\partial s} \theta l_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} t \frac{\partial w^H}{\partial s} \theta l_{\theta}^H dF(\theta) \right] \\ &+ \underbrace{[\Psi(V_{\Theta}^L) - \Psi(V_{\Theta}^H)]}_{=0} f(\Theta) \frac{\partial \Theta}{\partial s} - \eta [t w^H \Theta l_{\Theta}^H - t w^L \Theta l_{\Theta}^L - s \pi \Theta^{-\psi}] f(\Theta) \frac{\partial \Theta}{\partial s} = 0. \end{aligned} \quad (63)$$

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

Next, use Roy's identity to derive that

$$\frac{\partial V_{\theta}^i}{\partial b} = 1, \quad (64)$$

$$\frac{\partial V_{\theta}^H}{\partial t} = -\theta w^H l_{\theta}^H + (1-t) \theta l_{\theta}^H \frac{\partial w^H}{\partial t}, \quad (65)$$

$$\frac{\partial V_{\theta}^L}{\partial t} = -\theta w^L l_{\theta}^L + (1-t) \theta l_{\theta}^L \frac{\partial w^L}{\partial t}, \quad (66)$$

$$\frac{\partial V_{\theta}^H}{\partial s} = \pi\theta^{-\psi} + (1-t)\theta l_{\theta}^H \frac{\partial w^H}{\partial s}, \tag{67}$$

$$\frac{\partial V_{\theta}^L}{\partial s} = (1-t)\theta l_{\theta}^L \frac{\partial w^L}{\partial s}. \tag{68}$$

Recall that the net tax wedge on skill formation is defined as  $\Delta \equiv tw^H\Theta l_{\theta}^H - tw^L\Theta l_{\theta}^L - s\pi\Theta^{-\psi}$ . We define  $g_{\theta} \equiv \Psi'/\eta$  as the social welfare weight of individual  $\theta$ , where  $g_{\theta}$  gives the monetized value of providing this individual with an additional euro. Therefore, we can simplify the first-order conditions as

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial b} = 0 &: \int_{\underline{\theta}}^{\Theta} \frac{\Psi'}{\eta} dF(\theta) + \int_{\Theta}^{\bar{\theta}} \frac{\Psi'}{\eta} dF(\theta) \\ &= \int_{\underline{\theta}}^{\Theta} g_{\theta} dF(\theta) + \int_{\Theta}^{\bar{\theta}} g_{\theta} dF(\theta) = 1. \end{aligned} \tag{69}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} &= \int_{\underline{\theta}}^{\Theta} \Psi' \left( -\theta w^L l_{\theta}^L + (1-t)\theta l_{\theta}^L \frac{\partial w^L}{\partial t} \right) dF(\theta) \\ &+ \int_{\Theta}^{\bar{\theta}} \Psi' \left( -\theta w^H l_{\theta}^H + (1-t)\theta l_{\theta}^H \frac{\partial w^H}{\partial t} \right) dF(\theta) \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} w^L \theta l_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} w^H \theta l_{\theta}^H dF(\theta) \right] \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} tw^L \theta \frac{\partial l_{\theta}^L}{\partial t} dF(\theta) + \int_{\Theta}^{\bar{\theta}} tw^H \theta \frac{\partial l_{\theta}^H}{\partial t} dF(\theta) \right] \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} t \frac{\partial w^L}{\partial t} \theta l_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} t \frac{\partial w^H}{\partial t} \theta l_{\theta}^H dF(\theta) \right] - \eta \frac{\Delta}{1-t} \Theta f(\Theta) \frac{\partial \Theta}{\partial t} \frac{1-t}{\Theta} = 0, \end{aligned} \tag{70}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} &= \int_{\underline{\theta}}^{\Theta} \Psi' \left( (1-t)\theta l_{\theta}^L \frac{\partial w^L}{\partial s} \right) dF(\theta) + \int_{\Theta}^{\bar{\theta}} \Psi' \left( \pi\theta^{-\psi} + (1-t)\theta l_{\theta}^H \frac{\partial w^H}{\partial s} \right) dF(\theta) \\ &- \eta \left[ \pi \int_{\Theta}^{\bar{\theta}} \theta^{-\psi} dF(\theta) \right] + \eta \left[ \int_{\underline{\theta}}^{\Theta} tw^L \theta \frac{\partial l_{\theta}^L}{\partial s} dF(\theta) + \int_{\Theta}^{\bar{\theta}} tw^H \theta \frac{\partial l_{\theta}^H}{\partial s} dF(\theta) \right] \\ &+ \eta \left[ \int_{\underline{\theta}}^{\Theta} t \frac{\partial w^L}{\partial s} \theta l_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} t \frac{\partial w^H}{\partial s} \theta l_{\theta}^H dF(\theta) \right] - \eta \frac{\Delta}{s} \Theta f(\Theta) \frac{\partial \Theta}{\partial s} \frac{s}{\Theta} = 0. \end{aligned} \tag{71}$$

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  $z_\theta^L \equiv w^L \theta l_\theta^L$  and  $z_\theta^H \equiv w^H \theta l_\theta^H$  to find

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} = & - \left[ \int_{\underline{\theta}}^{\Theta} \Psi' z_\theta^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} \Psi' z_\theta^H dF(\theta) \right] + \eta \left[ \int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} z_\theta^H dF(\theta) \right] \\ & + \frac{t}{1-t} \eta \left[ \int_{\underline{\theta}}^{\Theta} z_\theta^L \frac{\partial l_\theta^L}{\partial t} \frac{1-t}{l_\theta^L} dF(\theta) + \int_{\Theta}^{\bar{\theta}} z_\theta^H \frac{\partial l_\theta^H}{\partial t} \frac{1-t}{l_\theta^H} dF(\theta) \right] \\ & + \int_{\underline{\theta}}^{\Theta} \left[ \Psi' + \eta \frac{t}{1-t} \right] z_\theta^L \frac{\partial w^L}{\partial t} \frac{1-t}{w^L} dF(\theta) \\ & + \int_{\Theta}^{\bar{\theta}} \left[ \Psi' + \eta \frac{t}{1-t} \right] z_\theta^H \frac{\partial w^H}{\partial t} \frac{1-t}{w^H} dF(\theta) - \eta \frac{\Delta}{1-t} \Theta f(\Theta) \frac{\partial \Theta}{\partial t} \frac{1-t}{\Theta} = 0. \end{aligned} \quad (72)$$

And, simplify the first-order condition for  $t$  using the definitions of the elasticities from Table 5:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial t} = & - \left[ \int_{\underline{\theta}}^{\Theta} \Psi' z_\theta^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} \Psi' z_\theta^H dF(\theta) \right] + \eta \left[ \int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} z_\theta^H dF(\theta) \right] \\ & - \frac{t}{1-t} \eta \left[ \int_{\underline{\theta}}^{\Theta} z_\theta^L \varepsilon_{l,t} dF(\theta) + \int_{\Theta}^{\bar{\theta}} z_\theta^H \varepsilon_{l,t} dF(\theta) \right] \\ & - \int_{\underline{\theta}}^{\Theta} \left[ \Psi' + \eta \frac{t}{1-t} \right] z_\theta^L \varepsilon_{w,t} dF(\theta) \\ & - \int_{\Theta}^{\bar{\theta}} \left[ \Psi' + \eta \frac{t}{1-t} \right] z_\theta^H \varepsilon_{w,t} dF(\theta) - \eta \frac{\Delta}{1-t} \Theta f(\Theta) \varepsilon_{\Theta,t} = 0. \end{aligned} \quad (73)$$

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

$$\bar{z}^L \equiv \int_{\underline{\theta}}^{\Theta} z_\theta^L dF(\theta), \quad \bar{z}^H \equiv \int_{\Theta}^{\bar{\theta}} z_\theta^H dF(\theta). \quad (74)$$

By dividing Eq. (73) by  $\eta$  and substituting for the definitions, we obtain

$$\begin{aligned}
 & - \left[ \int_{\underline{\theta}}^{\Theta} g_{\theta} z_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} g_{\theta} z_{\theta}^H dF(\theta) \right] + \bar{z}^L + \bar{z}^H - \frac{t}{1-t} [\varepsilon_{\mu,t} \bar{z}^L + \varepsilon_{\mu,t} \bar{z}^H] \\
 & - \varepsilon_{w^L,t} \int_{\underline{\theta}}^{\Theta} \left[ g_{\theta} + \frac{t}{1-t} \right] z_{\theta}^L dF(\theta) - \varepsilon_{w^H,t} \int_{\Theta}^{\bar{\theta}} \left[ g_{\theta} + \frac{t}{1-t} \right] z_{\theta}^H dF(\theta) \\
 & - \frac{\Delta}{1-t} \Theta f(\Theta) \varepsilon_{\Theta,t} = 0.
 \end{aligned} \tag{75}$$

Next, define the distributional characteristic of labor income as

$$\xi \equiv 1 - \frac{\int_{\underline{\theta}}^{\Theta} g_{\theta} z_{\theta}^L dF(\theta) + \int_{\Theta}^{\bar{\theta}} g_{\theta} z_{\theta}^H dF(\theta)}{[\bar{z}^L + \bar{z}^H] \int_{\underline{\theta}}^{\bar{\theta}} g_{\theta} dF(\theta)}. \tag{76}$$

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:

$$\alpha = \frac{\bar{z}^H}{\bar{z}^L + \bar{z}^H}, \quad 1 - \alpha = \frac{\bar{z}^L}{\bar{z}^L + \bar{z}^H}. \tag{77}$$

Hence, the optimal income tax expression can be written as

$$\begin{aligned}
 \xi &= \frac{t}{1-t} [(1 - \alpha)(\varepsilon_{\mu,t} + \varepsilon_{w^L,t}) + \alpha(\varepsilon_{\mu,t} + \varepsilon_{w^H,t})] + \frac{\Delta}{1-t} \frac{\Theta f(\Theta)}{\bar{z}} \varepsilon_{\Theta,t} \\
 &+ \varepsilon_{w^L,t} \frac{\int_{\underline{\theta}}^{\Theta} g_{\theta} z_{\theta}^L dF(\theta)}{[\bar{z}^L + \bar{z}^H]} + \varepsilon_{w^H,t} \frac{\int_{\Theta}^{\bar{\theta}} g_{\theta} z_{\theta}^H dF(\theta)}{[\bar{z}^L + \bar{z}^H]}.
 \end{aligned} \tag{78}$$

Substitute the income-weighted social welfare weights of each skill group:  $\tilde{g}^L \equiv \int_{\underline{\theta}}^{\Theta} g_{\theta} z_{\theta}^L dF(\theta) / \bar{z}^L$  and  $\tilde{g}^H \equiv \int_{\Theta}^{\bar{\theta}} g_{\theta} z_{\theta}^H dF(\theta) / \bar{z}^H$  to find the optimal tax in the proposition:

$$\begin{aligned}
 & \frac{t}{1-t} [(1 - \alpha)(\varepsilon_{\mu,t} + \varepsilon_{w^L,t}) + \alpha(\varepsilon_{\mu,t} + \varepsilon_{w^H,t})] + \frac{\Delta}{(1-t)} \frac{\Theta f(\Theta)}{\bar{z}} \varepsilon_{\Theta,t} \\
 &= \xi - \varepsilon_{w^H,t} \alpha \tilde{g}^H - \varepsilon_{w^L,t} (1 - \alpha) \tilde{g}^L.
 \end{aligned} \tag{79}$$

Finally, substitute for the elasticities from Table 5 to find

$$\begin{aligned}
 & \frac{t}{(1-t)} \varepsilon + \frac{\Delta}{(1-t)} \frac{\Theta f(\Theta)}{\bar{z}} \left( \frac{\sigma + \varepsilon}{\sigma + \varepsilon + \delta(\beta - \alpha)} \right) \\
 &= \xi - \frac{(1 - \alpha) \alpha \delta}{(\sigma + \varepsilon + \delta(\beta - \alpha))} (\tilde{g}^L - \tilde{g}^H).
 \end{aligned} \tag{80}$$

## B.2 Optimal education subsidy

Using similar steps as a above, we rewrite the optimal education subsidy using the definitions for  $z_\theta^L \equiv w^L \theta l_\theta^L$  and  $z_\theta^H \equiv w^H \theta l_\theta^H$  to find

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} = & \int_{\underline{\theta}}^{\bar{\theta}} \Psi' \left( \frac{(1-t)}{s} z_\theta^L \frac{\partial w^L}{\partial s} \frac{s}{w^L} \right) dF(\theta) \\ & + \int_{\underline{\theta}}^{\bar{\theta}} \Psi' \left( \pi \theta^{-\psi} + \frac{(1-t)}{s} z_\theta^H \frac{\partial w^H}{\partial s} \frac{s}{w^H} \right) dF(\theta) \\ & - \eta \left[ \pi \int_{\underline{\theta}}^{\bar{\theta}} \theta^{-\psi} dF(\theta) \right] + \eta \left[ \int_{\underline{\theta}}^{\bar{\theta}} \frac{t}{s} z_\theta^L \frac{\partial l_\theta^L}{\partial s} \frac{s}{l_\theta^L} dF(\theta) + \int_{\underline{\theta}}^{\bar{\theta}} \frac{t}{s} z_\theta^H \frac{\partial l_\theta^H}{\partial s} \frac{s}{l_\theta^H} dF(\theta) \right] \\ & + \eta \left[ \int_{\underline{\theta}}^{\bar{\theta}} \frac{t}{s} \frac{\partial w^L}{\partial s} \frac{s}{w^L} z_\theta^L dF(\theta) + \int_{\underline{\theta}}^{\bar{\theta}} \frac{t}{s} \frac{\partial w^H}{\partial s} \frac{s}{w^H} z_\theta^H dF(\theta) \right] - \eta \frac{\Delta}{s} \Theta f(\theta) \frac{\partial \Theta}{\partial s} \frac{s}{\Theta} = 0. \end{aligned} \quad (81)$$

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

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} = & \int_{\underline{\theta}}^{\bar{\theta}} \Psi' \left( \frac{(1-t)}{s} z_\theta^L \varepsilon_{w^L,s} \right) dF(\theta) + \int_{\underline{\theta}}^{\bar{\theta}} \Psi' \left( \pi \theta^{-\psi} + \frac{(1-t)}{s} z_\theta^H \varepsilon_{w^H,s} \right) dF(\theta) \\ & - \eta \pi \int_{\underline{\theta}}^{\bar{\theta}} \theta^{-\psi} dF(\theta) + \eta \left[ \frac{t}{s} (\varepsilon_{l^H,s} + \varepsilon_{w^L,s}) \bar{z}^L + \frac{t}{s} (\varepsilon_{l^H,s} + \varepsilon_{w^H,s}) \bar{z}^H \right] \\ & + \eta \frac{\Delta}{s} \Theta f(\theta) \varepsilon_{\Theta,s} = 0. \end{aligned} \quad (82)$$

All elasticities are independent from  $\theta$  (they do depend on  $\Theta$ ). Hence, they can be taken out of the integral signs. After dividing by  $\eta$  and multiplication with  $s/(1-t)$ , we obtain

$$\begin{aligned} \varepsilon_{w^L,s} \int_{\underline{\theta}}^{\bar{\theta}} g_\theta z_\theta^L dF(\theta) + \varepsilon_{w^H,s} \int_{\underline{\theta}}^{\bar{\theta}} g_\theta z_\theta^H dF(\theta) - \frac{s}{1-t} \pi \int_{\underline{\theta}}^{\bar{\theta}} \theta^{-\psi} (1-g_\theta) dF(\theta) \\ + \frac{t}{1-t} \varepsilon_{l^L,s} \bar{z}^L + \frac{t}{1-t} \varepsilon_{l^H,s} \bar{z}^H + \frac{t}{1-t} \varepsilon_{w^L,s} \bar{z}^L + \frac{t}{1-t} \varepsilon_{w^H,s} \bar{z}^H + \frac{\Delta}{1-t} \Theta f(\theta) \varepsilon_{\Theta,s} = 0. \end{aligned} \quad (83)$$

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



$$\begin{aligned} &\varepsilon_{w^L,s}(1-\alpha)\tilde{g}^L + \varepsilon_{w^H,s}\alpha\tilde{g}^H - \frac{1}{z}\frac{s}{1-t}\pi \int_{\Theta}^{\bar{\theta}} \theta^{-\psi}(1-g_{\theta})dF(\theta) \\ &+ \frac{t}{1-t}\varepsilon_{\mu,s}(1-\alpha) + \frac{t}{1-t}\varepsilon_{\mu,s}\alpha + \frac{t}{1-t}\varepsilon_{w^L,s}(1-\alpha) \\ &+ \frac{t}{1-t}\varepsilon_{w^H,s}\alpha + \frac{1}{z}\frac{\Delta}{1-t}\Theta f(\theta)\varepsilon_{\Theta,s} = 0. \end{aligned} \tag{84}$$

Collect terms and rewrite to arrive at

$$\begin{aligned} &\varepsilon_{w^L,s}(1-\alpha)\tilde{g}^L + \varepsilon_{w^H,s}\alpha\tilde{g}^H - \frac{1}{z}\frac{s}{1-t}\pi \int_{\Theta}^{\bar{\theta}} \theta^{-\psi}(1-g_{\theta})dF(\theta) \\ &+ \frac{t}{1-t}(1-\alpha)(\varepsilon_{\mu,s} + \varepsilon_{w^L,s}) + \frac{t}{1-t}\alpha(\varepsilon_{\mu,s} + \varepsilon_{w^H,s}) + \frac{1}{z}\frac{\Delta}{1-t}\Theta f(\theta)\varepsilon_{\Theta,s} = 0. \end{aligned} \tag{85}$$

Now, substitute the definitions of the elasticities from Table 5 to derive the following results:

$$\begin{aligned} &\left(\frac{\alpha\delta}{\sigma + \varepsilon + \delta(\beta - \alpha)}\right)\rho(1-\alpha)\tilde{g}^L - \left(\frac{(1-\alpha)\delta}{\sigma + \varepsilon + \delta(\beta - \alpha)}\right)\rho\alpha\tilde{g}^H \\ &= \left(\frac{\alpha(1-\alpha)\delta}{\sigma + \varepsilon + \delta(\beta - \alpha)}\right)\rho(\tilde{g}^L - \tilde{g}^H), \end{aligned} \tag{86}$$

$$(1-\alpha)(\varepsilon_{\mu,s} + \varepsilon_{w^L,s}) = (1-\alpha)(1+\varepsilon)\frac{\alpha\delta}{\sigma + \varepsilon + \delta(\beta - \alpha)}\rho, \tag{87}$$

$$\alpha(\varepsilon_{\mu,s} + \varepsilon_{w^H,s}) = -\alpha(1+\varepsilon)\frac{(1-\alpha)\delta}{\sigma + \varepsilon + \delta(\beta - \alpha)}\rho. \tag{88}$$

Thus, we find

$$\frac{t}{1-t}(1-\alpha)(\varepsilon_{\mu,s} + \varepsilon_{w^L,s}) + \frac{t}{1-t}\alpha(\varepsilon_{\mu,s} + \varepsilon_{w^H,s}) = 0. \tag{89}$$

The condition for the optimal subsidy Eq. (85) then simplifies to

$$\begin{aligned} &\left(\frac{\alpha(1-\alpha)\delta}{\sigma + \varepsilon + \delta(\beta - \alpha)}\right)\rho(\tilde{g}^L - \tilde{g}^H) - \frac{1}{z}\frac{s}{1-t}\pi \int_{\Theta}^{\bar{\theta}} \theta^{-\psi}(1-g_{\theta})dF(\theta) \\ &+ \frac{1}{z}\frac{\Delta}{1-t}\Theta f(\theta)\varepsilon_{\Theta,s} = 0. \end{aligned} \tag{90}$$

Substituting for  $\varepsilon_{\Theta,s}$  from Table 5 then yields

$$\left( \frac{\alpha(1-\alpha)\delta}{\sigma+\varepsilon+\delta(\beta-\alpha)} \right) \rho(\tilde{g}^L - \tilde{g}^H) - \frac{1}{\bar{z}} \frac{s}{1-t} \pi \int_{\Theta}^{\bar{\theta}} \theta^{-\psi} (1-g_{\theta}) dF(\theta) + \frac{\Delta}{1-t} \frac{\Theta f(\theta)}{\bar{z}} \frac{\sigma+\varepsilon}{\sigma+\varepsilon+\delta(\beta-\alpha)} \rho = 0. \quad (91)$$

Substitute  $\varepsilon_{GE} \equiv (1-\alpha)\varepsilon_{w^L,t} = -\alpha\varepsilon_{w^H,t} = \frac{\alpha(1-\alpha)\delta}{(\sigma+\varepsilon+\delta(\beta-\alpha))}$ , and the distributional characteristic of the education subsidy  $\zeta$ , to find the optimal subsidy in the proposition:

$$\frac{\Delta}{1-t} \frac{\Theta f(\theta)}{\bar{z}} \varepsilon_{\Theta,s} = \frac{1}{\bar{z}} \frac{s\pi}{1-t} \zeta - \rho(\tilde{g}^L - \tilde{g}^H) \varepsilon_{GE}. \quad (92)$$

## C Data appendix

Data on wages and educational attainment are taken from the Current Population Survey (CPS) Merged Outgoing Rotation Groups (MORG) as prepared by the National Bureau for Economic Research (2022). The data cover the years from 1979 to 2016, where we focus on the period 1980 to 2016. We use the same sample selection criteria as Acemoglu and Autor (2011). In particular, individuals are of age 16 to 64 and their usual weekly hours worked exceed 35. We obtain hourly wages by dividing weekly earnings by usual hours worked. We convert all wages into 2016 dollar values using the personal consumption expenditures chain-type price index.<sup>36</sup> The highest earnings in the CPS are top-coded.

**Table 6** Calibration for robustness checks

Parameter	Description	Robustness:		Robustness:	
		Labor supply elasticity		Enrollment elasticity	
$\varepsilon$	Labor-supply elasticity	0.1	0.5	0.3	0.3
$\pi$	Average cost education	146.97	210.06	209.91	199.56
$\psi$	Elasticity of education	16.21	16.44	16.63	19.55
$\omega$	Share parameter	0.51	0.50	0.52	0.55
$A_{2016}$	Skill bias (2016)	2.06	2.30	2.47	1.13
$b_{1980}$	Tax intercept (1980)	0.12	0.09	0.10	0.08
$b_{2016}$	Tax intercept (2016)	0.14	0.15	0.16	0.06
$R_{1980}$	Government revenue (1980)	0.13	0.10	0.12	0.09
$R_{2016}$	Government revenue (2016)	0.17	0.18	0.20	0.08
	Skill premium in 1980 (%)	60.43	65.94	57.93	82.96
	Skill premium in 2016 (%)	93.28	103.32	92.66	90.93
	Share high-skilled 1980 (%)	30.63	26.82	26.92	15.89
	Share high-skilled 2016 (%)	47.30	44.04	48.54	17.01
	Enrollment elasticity 1980	0.12	0.14	0.10	0.14

<sup>36</sup> We obtain the price index from Federal Reserve Bank of St. Louis (2022).

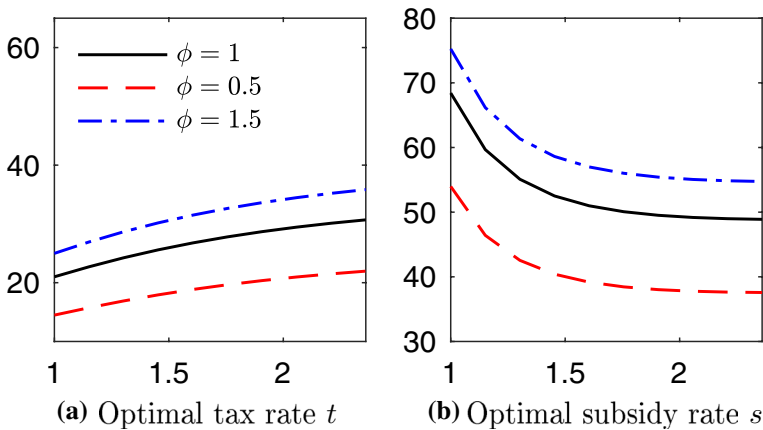
Top-coded earnings are therefore replaced by draws from a Pareto distribution. Like Acemoglu and Autor (2011), we exclude individuals who earn less than 50% of the 1982 minimum wage (\$3.35) converted to 2016-dollars. We also exclude self-employed individuals, as well as individuals whose occupation does not have an `occ1990dd` classification. We weight observations by CPS sample weights. We code education levels based on the highest grade attended (before 1992) and the highest grade completed (after 1992).

### D Enrollment elasticity

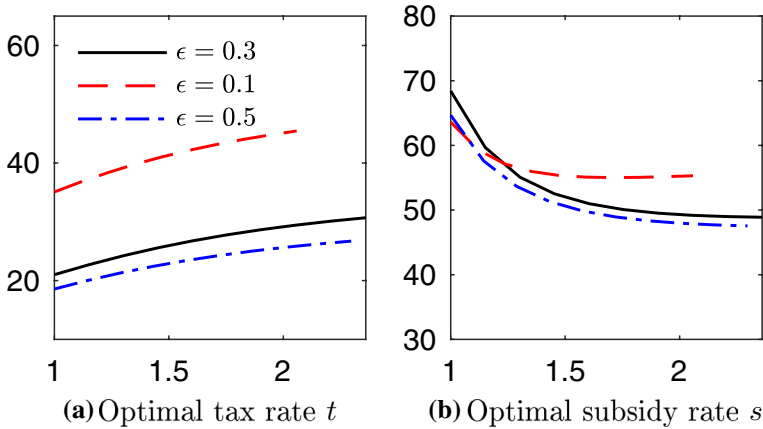
Dynarski (2000) finds that a \$1000 increase in financial aid raises college attendance rates in Georgia between 3.7 and 4.2 percentage points. Before the introduction of the scholarship, average tuition per student was \$1900. Based on data from the US Department of Education, Gumpert et al. (1997) document that in 1992 government funding as a percentage of all funding for higher education in the US was around 40%, which we treat as the initial subsidy rate. We consider the tuition of \$1900 as the private cost of higher education, which equals 60% of the total cost of \$3167. A reduction of \$1000 corresponds to a change in the subsidy rate of 30 percentage points. Using an initial college enrollment rate in Georgia of 0.32, and assuming an increase of 0.04 in the enrollment share due to the HOPE scholarship, we compute the relative change in enrollment as  $0.04/0.32$  and the relative change in the subsidy rate as  $0.3/0.4$ . The resulting enrollment elasticity of the subsidy is then equal to 0.17.

### E Robustness

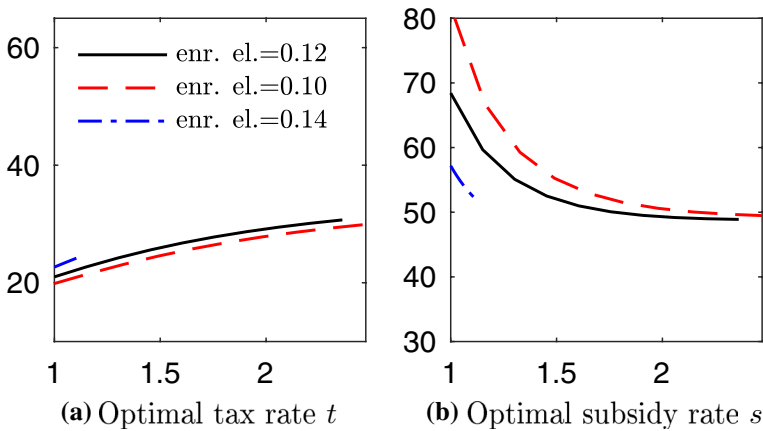
See Figs. 3, 4, 5.



**Fig. 3** Optimal policy with SBTC for different levels of inequality aversion  $\phi$ . Note Skill bias  $A$  on the horizontal axis. The solid black line corresponds to the baseline calibration



**Fig. 4** Optimal policy with SBTC for different levels of the labor-supply elasticity  $\epsilon$ . *Note* Skill bias  $A$  on the horizontal axis. The solid black line corresponds to the baseline calibration



**Fig. 5** Optimal policy with SBTC for different levels of the enrollment elasticity enr. el. *Note* Skill bias  $A$  on the horizontal axis. The solid black line corresponds to the baseline calibration

**Electronic supplementary material** The online version contains supplementary material available at <https://doi.org/10.1007/s10797-022-09756-8>.

**Acknowledgements** The authors like to thank the editor Ron Davies, two anonymous referees, Bjoern Bruegemann, seminar participants of Erasmus University Rotterdam, and participants of the 2014 IIPF Conference in Lugano and the 2015 EEA Conference in Mannheim for useful comments and suggestions.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is

not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In O. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics*. Elsevier, Vol. 4-B, chapter 12, pp. 1043–1171.
- Acemoglu, D., & Autor, D. (2012). What Does Human Capital Do? A Review of Goldin and Katz's The Race between Education and Technology'. *Journal of Economic Literature*, 50, 426–463.
- Ales, L., Kurnaz, M., & Sleet, C. (2015). Technical Change, Wage Inequality and Taxes. *American Economic Review*, 105, 3061–3101.
- Allen, F. (1982). Optimal linear income taxation with general equilibrium effects on wages. *Journal of Public Economics*, 17, 135–143.
- Anderberg, D. (2009). Optimal Policy and the risk properties of human capital reconsidered. *Journal of Public Economics*, 93, 1017–1026.
- Barr, N. (2004). Higher education funding. *Oxford Review of Economic Policy*, 20, 264–283.
- Blundell, R., & MaCurdy, T. (1999). Labor supply: A review of alternative approaches. In O. C. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics*. Elsevier, Vol. 3-A, chapter 27, pp. 1559–1695.
- Bohacek, R., & Kapicka, M. (2008). Optimal Human Capital Policies. *Journal of Monetary Economics*, 55, 1–16.
- Bovenberg, A. L., & Jacobs, B. (2005). Redistribution and Education Subsidies Are Siamese Twins. *Journal of Public Economics*, 89, 2005–2035.
- Colas, M., Findeisen, S., & Sachs, D. (2021). Optimal Need-Based Financial Aid. *Journal of Political Economy*, 129, 492–533.
- Diamond, P. A. (1980). Income Taxation with Fixed Hours of Work. *Journal of Public Economics*, 13, 101–110.
- Dixit, A., & Sandmo, A. (1977). Some Simplified Formulae for Optimal Income Taxation. *Scandinavian Journal of Economics*, 79, 417.
- Dur, R. & Teulings, C. N. (2004). Are Education Subsidies an Efficient Redistributive Device? In J. Agell, M. Keen, & A. J. Weichenrieder (Eds.), *Labor Market Institutions and Public Regulation*. Cambridge, MA: The MIT Press, chapter 4, pp. 123–161.
- Dynarski, S. (2000). Hope for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance. *National Tax Journal*, 53, 629–662.
- Federal Reserve Bank of St. Louis. (2022). *Personal Consumption Expenditures: Chain-type Price Index*. St. Louis: FED of St. Louis. <https://fred.stlouisfed.org/series/DPCERG3A086NBEA>
- Feldstein, M. S. (1972). Distributional Equity and the Optimal Structure of Public Prices. *American Economic Review*, 62, 32–36.
- Findeisen, S., & Sachs, D. (2016). Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans. *Journal of Public Economics*, 138, 1–21.
- Goldin, C., & Katz, L. F. (2010). *The Race between Education and Technology*. Belknap Press of Harvard University Press.
- Gomes, R., Lozachmeur, J.-M., & Pavan, A. (2018). Differential Taxation and Occupational Choice. *Review of Economic Studies*, 85, 511–557.
- Gumport, P. J., Iannozzi, M., Shaman, S., & Zemsky, R. (1997). *"The United States" Country Report: Trends in Higher Education from Massification to Post-Massification*. Hiroshima: Six Nation Educational Research Project, Hiroshima University.
- Heathcote, J., Storesletten, K., & Violante, G. L. (2014). Optimal Tax Progressivity: An Analytical Framework, Working Paper 19899, National Bureau of Economic Research (NBER).
- Heathcote, J., Storesletten, K., & Violante, G. L. (2017). Optimal Tax Progressivity: An Analytical Framework. *Quarterly Journal of Economics*, 132, 1–62.

- Heckman, J. J., Lochner, L., & Taber, C. (1998). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents. *Review of Economic Dynamics*, 1, 1–58.
- Heckman, J. J., Lochner, L., & Taber, C. (1999). Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy. *Fiscal Studies*, 20, 25–40.
- Hellwig, M. F. (1986). The Optimal Linear Income Tax Revisited. *Journal of Public Economics*, 31, 163–179.
- Jacobs, B. (2005). Optimal Income Taxation with Endogenous Human Capital. *Journal of Public Economic Theory*, 7, 295–315.
- Jacobs, B. (2012). Optimal Redistributive Tax and Education Policies in General Equilibrium. *International Tax and Public Finance*, 20, 1–26.
- Jacobs, B. (2018). The Marginal Cost of Public Funds Is One at the Optimal Tax System. *International Tax and Public Finance*, 25, 1–30.
- Jacobs, B., & Bovenberg, A. L. (2011). Optimal Taxation of Human Capital and the Earnings Function. *Journal of Public Economic Theory*, 13, 957–971.
- Jacobs, B., & Thuemmel, U. (2021). *Optimal Taxation of Income and Human Capital and Skill-Biased Technical Change*, mimeo. Rotterdam: Erasmus University Rotterdam.
- Jacobs, B., & Yang, H. (2016). Second-Best Income Taxation and Education Policy with Endogenous Human Capital and Borrowing Constraints. *International Tax and Public Finance*, 23, 234–268.
- Katz, L. F., & Murphy, K. M. (1992). Changes in Relative Wages, 1963–1987: Supply and Demand Factors. *Quarterly Journal of Economics*, 107, 35–78.
- Krueger, D., & Ludwig, A. (2016). On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium. *Journal of Monetary Economics*, 77, 72–98.
- Loebbing, J. (2020). Redistributive Income Taxation with Directed Technical Change, CESifo Working Paper no. 8743, Munich: CESifo.
- Maldonado, D. (2008). Education Policies and Optimal Taxation. *International Tax and Public Finance*, 15, 131–143.
- Meghir, C., & Phillips, D. (2010). Labour Supply and Taxes. In J. A. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, & J. Poterba (Eds.), *Dimensions of Tax Design: The Mirrlees Review*. Oxford University Press, chapter 3, pp. 202–274.
- National Bureau for Economic Research. (2021). *TAXSIM*. Cambridge-MA: NBER. <https://users.nber.org/~taxsim/>
- National Bureau for Economic Research. (2022). *Current Population Survey (CPS) - Merged Outgoing Rotation Group Earnings Data*. Cambridge-MA: NBER. <https://www.nber.org/research/data/current-population-survey-cps-merged-outgoing-rotation-group-earnings-data>
- Nielsen, H. S., Sørensen, T., & Taber, C. (2010). Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform. *American Economic Journal: Economic Policy*, 2, 185–215.
- OECD. (2018). *Spending on Tertiary Education*. Organisation for Economic Co-operation and Development (OECD): Indicator.
- Reed, W. J., & Jorgensen, M. (2004). The Double Pareto-Lognormal Distribution—A New Parametric Model for Size Distributions. *Communications in Statistics - Theory and Methods*, 33, 1733–1753.
- Rothschild, C., & Scheuer, F. (2013). Redistributive Taxation in the Roy Model. *Quarterly Journal of Economics*, 128, 623–668.
- Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers*, 3, 135–146.
- Sachs, D., Tsyvinski, A., & Werquin, N. (2020). Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium. *Econometrica*, 88, 469–493.
- Saez, E. (2002). Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses. *Quarterly Journal of Economics*, 117, 1039–1073.
- Saez, E., Gieritz, S. H., & Slemrod, J. B. (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature*, 50, 3–50.
- Sheshinski, E. (1972). The Optimal Linear Income Tax. *Review of Economic Studies*, 39, 297–302.
- Stantcheva, S. (2017). Optimal Taxation and Human Capital Policies over the Life Cycle. *Journal of Political Economy*, 125, 1931–1990.
- Tinbergen, J. (1975). *Income Distribution: Analysis and Policies*. Amsterdam: North-Holland Publishing Company.

- Van Reenen, J. (2011). Wage Inequality, Technology and Trade: 21st Century Evidence. *Labour Economics*, 18, 730–741.
- Violante, G. L. (2008). Skill-Biased Technical Change. In S. N. Durlauf & L. E. Blume (Eds.), *New Palgrave Dictionary of Economics* (2nd ed.). Basingstoke: Nature Publishing Group, pp. 520–523.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.