Optimal Taxation of Income and Human Capital and Skill-Biased Technical Change*

Bas Jacobs**
Erasmus University Rotterdam, Tinbergen Institute and CESifo

Uwe Thuemmel***
University of Zurich

August 18, 2018

Abstract

How should redistributive governments change tax and education policy in response to skill-biased technical change? To answer this question, this paper merges the canonical model of skill-biased technical change due to Katz and Murphy (1992) with the continuous-type Mirrlees (1971) model. Workers of different ability face an extensive education choice to be come high-skilled. Wages are endogenous. Optimal marginal income tax rates follow the same formula as in Mirrlees (1971). The intercept of the optimal tax function differs for low-skilled and high-skilled workers, while marginal tax rates are the same for high-skilled and low-skilled workers at the cut-off ability where workers are indifferent between being high-skilled or not. We show that education should optimally be taxed on a net basis. Moreover, optimal tax and education policies do not exploit general-equilibrium effects on the wage distribution to reduce pre-tax earnings differentials. SBTC has ambigous effects on optimal marginal tax rates depending only on how social welfare weights change. SBTC has ambiguous effects on income net taxes on education, since distributional benefits and distortions simultaneously increase. Numerical simulations demonstrate that SBTC leads to higher optimal marginal income taxes for middle incomes, while lowering marginal income taxes towards the top. The tax system becomes more progressive in response to SBTC. Education subsidies increase in response to SBTC.

Keywords: Human capital; General equilibrium; Optimal taxation; Education subsidies; Technological Change.
JEL-Codes: H2; H5; I2; J2; O3.

1 Introduction

US President Obama called rising inequality the defining challenge of our time (Obama, 2013; see for empirical evidence also Alvaredo et al., 2017). One of the major contributors to the rise in

*The authors like to thank Bjoern Bruegemann, Albert Jan Hummel, Aart Gerritsen, Dominik Sachs, Kevin Spiritus and seminar participants at Erasmus School of Economics and the CESifo Public Sector Area Conference 2018 for their comments and suggestions.

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

***University of Zurich, Department of Economics, Schoenberggasse 1, 8001 Zurich, Switzerland. E-mail: uwe.thuemmel@uzh.ch. Homepage: http://uwethuemmel.com/.

inequality is the steady increase in skill premia since the 1980s, see e.g., Katz and Murphy (1992), Katz and Autor (1999) Jacobs (2004) Heckman et al. (2006), Violante (2008), and Goldin and Katz (2010).[1] The leading – but certainly not the only – explanation is that technological change is skill-biased, leading to an increase in the relative demand for skilled workers that outpaces the gradual increase in the relative supply of skilled workers.[2] Already in the 1970s, Dutch Nobel laureate Tinbergen (1975) coined this phenomenon the 'race between technological development and education'.[3] In most Western countries the race between education and technology has thus been lost by education.

The main question of this paper is: How should redistributive governments change tax and education policy in response to rising income inequality caused by skill-biased technical change (SBTC)? Despite the obvious relevance of this question, the economic literature has not yet provided convincing answers. For example, Tinbergen (1975) and Goldin and Katz (2010) have argued that investments in (higher) education should expand in response to SBTC so as to compress the pre-tax wage distribution ('predistribution'). By stimulating investment in education, the relative supply of skilled workers keeps up with relative demand for skilled workers, so that rising wage inequality can be prevented. However, to the best of our knowledge no other study has analyzed the consequences of SBTC for optimal education policy. Closest to our analysis is the study by Ales et al. (2015). They have analyzed the setting of optimal nonlinear income taxes in a task-based model of the labor market with job polarization, cf. Acemoglu and Autor (2011). They find that (neutral) technical change lowers optimal marginal tax rates for the middle-income groups.

Our paper analyzes optimal income redistribution and optimal education subsidies when income inequality is rising due to SBTC. We merge the canonical model of SBTC due to Katz and Murphy (1992) and Acemoglu and Autor (2011) with the standard Mirrlees (1971) model of optimal nonlinear income taxation. According to the canonical model of SBTC, output is produced by high- and low-skilled labor, which are imperfectly substitutable factors of production. Both types of labor are inelastically supplied. Over time, the productivity of high-skilled labor rises relative to low-skilled labor, making technology increasingly skill-biased. As a result, demand for high-skilled workers increases relative to demand for low-skilled workers, which drives up the skill-premium. We endogenize high-skilled and low-skilled labor supplies, where individuals supply labor on the intensive margin, as in Mirrlees (1971), and they optimally decide to invest in education and become high-skilled or remain low-skilled. There is a single source of heterogeneity: individuals differ in their ability, which measures their number of efficiency units of labor. Given that high-ability individuals have a comparative advantage in high-skilled work, only sufficiently able individuals invest in education and become high-skilled, while all other individuals remain low-skilled. The wage rates for each occupation, and therefore the skill premium, are endogenously determined by demand and supply for each type of labor.[4]

[1]Recently, Valletta (2019) and Autor (2019) document a flattening of the skill premium over time.

[2]Besides skill-biased technical change, the other main explanation for skill-biased shifts in labor demand in the Western world is international trade, see also Helpman (2016). Moreover, changes in labor-market institutions, such as minimum wages and unions, are considered important to explain rising wage inequality, see also Blau and Kahn (1996), Katz and Autor (1999), Fortin and Lemieux (1997), Lee (1999), and Autor et al. (2008).

[3]Tinbergen's metaphor has been adopted much later by Goldin and Katz (2010).

[4]Low-skilled and high-skilled workers are imperfect substitutes in production. Workers of different abilities

A welfare-maximizing government optimally designs a nonlinear income tax system that can be conditioned on education. This tax system requires that the government needs to observe individual labor earnings and whether individuals are high-skilled or not. These informational assumptions are relatively mild and realistic. Our theoretical contributions are fivefold.

First, we show that optimal marginal tax rates follow the same formula for the optimal nonlinear marginal tax rates as in Mirrlees (1971). Consequently, optimal nonlinear marginal tax rates are continuous in income. What is not surprising is that a Mirrleesian marginal tax schedule is found within each skill group, since the income distributions of low-skilled and high-skilled workers are not overlapping. However, what is surprising is that the marginal tax rates at the top of the low-skilled income distribution and the bottom of the high-skilled income distribution are positive and are exactly the same. Intuitively, a positive marginal tax at the threshold Θ contributes to income redistribution, since the marginally high-skilled pays more tax than the marginally low-skilled. An obvious and realistic policy implementation features a single nonlinear tax schedule in labor earnings, where high-skilled workers face an additional tax (or subsidy), which is independent of their labor earnings.

Second, we show that SBTC raises marginal tax rates especially in the middle of the income distribution. We find that the only mechanism whereby SBTC affects optimal marginal tax rates is via changes in the social welfare weights, which capture the government's demand for income redistribution. The direction in which the social welfare weights change is theoretically ambiguous. By raising the skill premium, SBTC increases income inequality between and within skill-groups, which raises the marginal benefits of income redistribution and thus optimal marginal tax rates. However, as SBTC raises incomes for all workers, it also reduces the social welfare weights for all workers, and more so for low-income and low-skilled workers. The latter reduces the demand for redistribution, and thus leads to lower marginal tax rates. We show numerically that optimal marginal tax rates increase for low-skilled workers and slightly decrease for high-skilled workers. Tax rates rise most around the cut-off of the marginally high-skilled and marginally low-skilled workers. This implies that the government wants to redistribute more from high-skilled to low-skilled workers.

Third, we show that education policies are optimally employed for three reasons. First, education should be subsidized to offset distortions from income taxation. The net tax on education is zero if the subsidy exactly compensates for the distortions from income taxation, see also Bovenberg and Jacobs (2005). Second, the government likes to redistribute income from high-skilled to low-skilled workers. Therefore, it is optimal to tax the infra-marginal rents from ability of the high-skilled. Because the government wants to tax education on a net basis in the policy optimum, optimal education subsidies do not fully offset all tax distortions in skill formation (see also Findeisen and Sachs, 2017). Third, optimal education policy exploits complementarities between labor supply and education. Since the marginal high-skilled worker supplies more labor than the marginal low-skilled worker – while facing the same marginal tax rate – raising investment in education raises labor supply. This does not generate inequality, since the marginal low-skilled worker has the same utility as the marginal high-skilled worker. Lowering the net tax on education thus reduces the tax distortions in labor supply, see also

are perfect substitutes within each skill group.

Jacobs and Bovenberg (2011).

Fourth, SBTC may either decrease or increase optimal net taxes on education. Larger inequality between skilled and unskilled workers raises the social value of income redistribution from high-skilled to low-skilled workers. However, SBTC also increases the distortions from net taxes on education. Therefore, it is not clear whether redistributive governments increase education subsidies in response to SBTC, in contrast to impressions in the literature.

Fifth, we show that neither optimal income taxes nor optimal education subsidies exploit the imperfect substitutability between low-skilled and high-skilled workers to compress before-tax wage differentials.[5] This result demonstrates that Tinbergen's intuition is not incorrect if the government can use condition income tax rates on education type. Intuitively, any redistribution via wage compression can be achieved as well with the nonlinear tax schedule. Consequently, direct income redistribution via education-dependent taxes/tranfers is preferred over indirect redistribution by compressing before-tax income differentials, since the latter generates larger distortions in skill formation.

Finally, we numerically simulate optimal tax and education policies using an empirically plausible calibration of the model to US data. Our simulations confirm our theoretical predictions. SBTC substantially raises inequality and results in more income redistribution from high-skilled to low-skilled workers. SBTC makes the income tax system more progressive. Moreover, SBTC tends to raise marginal tax rates in the interior of the earnings distribution, while lowering marginal tax rates towards the top. Optimal net taxes on education decrease when SBTC becomes more important. This can be interpreted as evidence that optimal education subsidies should increase in response to SBTC. Consequently, this paper does in the end lend support to the policy recommendations of Tinbergen (1975) and Goldin and Katz (2010) to increase tax progression and to promote investment in education in response to SBTC. However, the mechanism is not that these policies should be implemented to win the race with technology and to compress the wage distribution.

The rest of this paper is organized as follows. Section 2 explains the relation of our paper to earlier literature. Section 3 introduces the model. Section 4 derives optimal taxes and education subsidies. Section 5 discusses the consequences of skill-biased technical change for optimal tax and education policies. Section 6 presents numerical simulations of the model. Section 7 concludes. An Appendix contains the proof of the main proposition, presents more details of the simulations and some robustness checks.

2 Relation to the literature

Our paper is related to several strands in the literature. First, and foremost, we contribute to the literature on optimal taxation and education subsidies by extending the standard optimal nonlinear income tax model of Mirrlees (1971) with an education choice on the extensive margin and by allowing for general equilibrium effects on the wage structure.[6] In doing so, we build on the literature which studies optimal Mirrleesian taxation with intensive-margin human capital investment and exogenous wage rates as in Bovenberg and Jacobs (2005), Maldonado (2008),

[5] This is also referred to as the 'tax formula result' by Saez (2004).

[6] See also Diamond (1998) and Saez (2001) for extensive discussions of the optimal nonlinear income tax model.

Bohacek and Kapicka (2008), Anderberg (2009), Jacobs and Bovenberg (2011), and Stantcheva (2017). In all these papers, an important role for the education subsidy is to offset distortions on skill-formation caused by the income tax.[7] By alleviating tax distortions on human capital, education subsidies allow the government to redistribute income at lower efficiency costs. We partially confirm the finding of Bovenberg and Jacobs (2005) that higher income tax rates raise optimal education subsidies – ceteris paribus. However, in our model, the government does not off-set *all* tax-induced distortions on education with subsidies, since the education choice is on the extensive rather than the intensive margin. A discrete education decision gives rise to inframarginal rents, which the government likes to tax for income redistribution, see below.

Second, there is a smaller literature that considers optimal taxation with human capital formation or occupational choice on the extensive margin. Findeisen and Sachs (2017) take the income tax as given to explore how education policies should optimally be conditioned on parental income. Like us, they also find that education features infra-marginal rents that are optimally exploited for income redistribution.

Third, this paper is importantly related to Jacobs and Thuemmel (2018), who study optimal tax and education policies using a similar model as in this paper, but with linear taxes and education subsidies. The main difference is that the government can no longer condition income tax rates on education, hence the income tax can no longer neutralize the impact of general-equilibrium effects on the wage structure. Consequently, in Jacobs and Thuemmel (2018) optimal tax and education policies exploit wage-compression effects. Larger income taxes discourage investment in education and thus raise the skill premium. As a result, linear income taxes are optimally lowered to reduce pre-tax wage inequality. Jacobs and Thuemmel (2018) also show theoretically and numerically that education may optimally be subsidized rather than taxed on a net basis to compress the wage distribution. In this paper, we show that this can never happen if income tax rates can be conditioned on education type. Moreover, Jacobs and Thuemmel (2018) numerically simulate the impact of SBTC on optimal income taxes and education subsidies using a very similar calibration as in this paper. They find that optimal income taxes indeed increase and become more progressive, while optimal education subsidies decline. In our simulations, we also find that optimal income taxes become more progressive in response to SBTC. Moreover, we show that marginal tax rates increase especially around the income level where individuals are indifferent between investing in higher education or not. Furthermore, in contrast to Jacobs and Thuemmel (2018), we find that optimal taxes on education are negative on a net basis and decrease in our baseline simulation. Consequently, this paper shows that whether optimal education subsidies are positive on a net basis and whether they should decrease in response to SBTC is critically dependent on whether the government has access to skill-dependent marginal income tax rates.

Fourth, our paper is related to Stiglitz (1982) and Stern (1982), Rothschild and Scheuer (2013), Ales et al. (2015), and Sachs et al. (2017), who show that general-equilibrium effects on wage rates can be exploited for income redistribution by generalizing Stiglitz (1982) to continuous ability types, multiple occupations and/or tasks. All these papers show that optimal

[7] Related is Krueger and Ludwig (2016), who study optimal income taxation and education subsidies in an advanced OLG model with labor supply, human capital investment, saving and financial frictions, which is simulated for the US. However, Krueger and Ludwig (2016) do not study the effect of SBTC on optimal policy.

taxes are lowered so as to compress the wage distribution. Furthermore, Dur and Teulings (2004) analyze a continuous-type assignment model to analyze optimal log-linear tax and education policies – similar to Heathcote et al. (2014) – and demonstrate that optimal tax and education subsidies should take into account general-equilibrium effects on the wage structure. Jacobs (2012) analyzes a two-type optimal tax model with labor supply and human capital investment on the intensive margin, based on Stiglitz (1982), Stern (1982), and Bovenberg and Jacobs (2005). Optimal nonlinear taxes and education policies are found to exploit general-equilibrium effects.[8]

Our analysis complements these papers by allowing for (potentially) education-dependent nonlinear tax schedules. We confirm Scheuer (2014) and Scheuer and Werning (2016) by showing that general-equilibrium effects are no longer exploited for income redistribution if nonlinear tax schedules can be conditioned on the education decision. Moreover, if the ability distribution has an upper bound, the marginal tax rate at the top is zero, in contrast to the findings in Stern (1982), Stiglitz (1982), Jacobs (2012), and Sachs et al. (2017). Furthermore, we show that a simple policy implementation with a continuous nonlinear tax schedule and different intercepts for high-skilled and low-skilled workers suffices, since heterogeneity in ability is one-dimensional so that earnings distributions of low-skilled and high-skilled workers are non-overlapping. This contrasts with Scheuer (2014) who analyzes multidimensional heterogeneity and requires occupation-specific nonlinear taxes in the policy optimum.

Fifth, our paper contributes to the few studies that have analyzed the impact of technological change on optimal taxation and education policy. Heckman et al. (1998) develop a dynamic OLG-model with endogenous human capital formation. They find that their model with SBTC is consistent with data on rising wage inequality. Moreover, using the same model, Heckman et al. (1999) demonstrate that general-equilibrium effects on the wage structure largely offset the initial impacts of tax and education policies. Ales et al. (2015) analyze the effect of technological change on nonlinear income taxes in a task-to-talent assignment model of the labor market based on Sattinger (1975) and Acemoglu and Autor (2011). Our paper is complementary in that we analyze optimal taxation and technological change in the canonical model of SBTC. The optimal policy response to SBTC is shown to be the opposite of Ales et al. (2015): we find that marginal tax rates go up for the middle-income groups, whereas they go down in Ales et al. (2015). Heathcote et al. (2014) analyze the impact of skill-biased technological change on the optimal degree of tax progressivity in the US if tax functions are constrained to have constant residual income progression. We show that SBTC should raise tax progressivity, while allowing for general nonlinear tax functions.

Sixth, our paper is also related to the optimal tax literature with both intensive and extensive margins based on Diamond (1980). Saez (2002) analyzes a discrete-type optimal-tax framework where individuals optimally choose their occupation and whether to participate or not.[9] Jacquet et al. (2013) analyze a continuous-type Mirrlees (1971) framework with labor supply on the intensive margin and an extensive participation margin. Our analysis complements both papers,

[8]Although all these contributions show that general-equilibrium effects play a role theoretically, these effects are found to be very modest quantatively in model simulations, see for example Stern (1982), Rothschild and Scheuer (2013), Jacobs (2012), and Sachs et al. (2017).

[9]The occupational choice decision is somewhat confusingly labeled 'intensive margin' by Saez (2002).

since we model labor supply on the intensive margin with an occupational choice on the extensive margin, while not allowing for a participation decision. In our configuration, the optimal tax schedule follows the standard Mirrlees (1971) formula, in contrast to Saez (2002) and Jacquet et al. (2013). Moreover, we find that the optimal net tax on the extensive (education) margin is similar to Diamond (1980), Saez (2002), and Jacquet et al. (2013).

3 Model

3.1 Individuals

There is a continuum of individuals of unit mass. Each worker is endowed with earning ability $\theta \in [\underline{\theta}, \overline{\theta}]$, where the upper bound $\overline{\theta}$ could be infinite. θ is drawn from distribution $F(\theta)$ with corresponding density $f(\theta)$. Superscript $j \in \{L, H\}$ indicates whether an individual is low- or high-skilled, where skill is endogenous.

Individuals derive utility U^j_θ from consumption c^j_θ and disutility from labor supply l^j_θ according to a quasi-linear utility function:

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

where ε is the constant wage-elasticity of labor supply, which is identical for both education groups. With little loss of generality we restrict the analysis to a quasi-linear utility function to avoid unnecessary technical complexity.

The wage rate per efficiency unit of labor is denoted by w^j. Gross earnings are given by $z^j_\theta \equiv w^j \theta l^j_\theta$. Education is a discrete choice to become high-skilled or to remain low-skilled. In order to become high-skilled, workers have to invest a fixed amount of resources p. Consumption is the numéraire commodity and its price is normalized to unity. The government levies eduction-specific nonlinear taxes $T^j(z^j_\theta)$ on labor income z^j_θ earned by workers with education j. By conditioning tax schedules $T^j(z^j_\theta)$ on the education choice, we implicitly allow for net taxes or subsidies on education. Workers of type θ and education j thus face the following budget constraint:[10]

$$c^j_\theta = \begin{cases} z^L_\theta - T^L(z^L_\theta), & \text{if } j = L \\ z^H_\theta - T^H(z^H_\theta) - p, & \text{if } j = H \end{cases}. \tag{2}$$

The informational assumptions of our model are that individual ability θ and labor effort l^j_θ are not verifiable, but individual labor earnings z^j_θ are. Hence, the government can levy nonlinear taxes on labor income. Moreover, the education type j is also verifiable. Hence, the government can differentiate nonlinear income tax schedules $T^j(z^j_\theta)$ by education type j.

Workers maximize utility by choosing consumption, labor supply and education, taking wage rates and government policy as given. Given their education choice, optimal labor supply is obtained by maximizing utility in (1), subject to their budget constraint in (2). First-order

[10] To focus on the redistributive role of education policy, we assume that capital markets are perfect and that there are no externalities of education. Incorporating borrowing constraints or externalities is an interesting avenue for future research.

conditions for utility maximization are given by:[11]

$$l_\theta^j = [w^j\theta(1-T'^j(z_\theta^j))]^\varepsilon, \quad \forall\theta, j. \tag{3}$$

A low-skilled individual chooses to invest in education if and only if she derives higher utility from being high-skilled than low-skilled, i.e., if $U_\theta^H \geq U_\theta^L$. The critical level of ability Θ that separates the high-skilled from the low-skilled follows from:

$$U_\Theta^L = U_\Theta^H. \tag{4}$$

Therefore, the optimal cutoff Θ is implicitly determined by:

$$z_\Theta^H - T^H(z_\Theta^H) - \frac{[w^H\Theta(1-T'^H(z_\Theta^H))]^{1+\varepsilon}}{1+1/\varepsilon} - p = z_\Theta^L - T^L(z_\Theta^L) - \frac{[w^L\Theta(1-T'^L(z_\theta^L))]^{1+\varepsilon}}{1+1/\varepsilon}, \tag{5}$$

where z_Θ^j is the optimal earnings supply of workers with education j at Θ, which follows from (3). Given that wage rates are unequal for low-skilled and high-skilled workers, their earnings are not equal either, i.e., $z_\Theta^L \neq z_\Theta^H$. All individuals with ability $\theta < \Theta$ remain low-skilled, whereas individuals with $\theta \geq \Theta$ become high-skilled. A decrease in Θ corresponds to more individuals becoming high-skilled.

Although the wage rates are endogenous, we assume throughout the paper that the primitives of our model are such that the high-skilled wage is above the low-skilled wage: $w^H > w^L$. If the skill premium w^H/w^L rises, more individuals invest in human capital – ceteris paribus. The same holds true for a decrease in the cost of education p – ceteris paribus. Finally, the income tax potentially distorts the education decision, because taxes on high-skilled workers can be higher than taxes on low-skilled workers. The education choice is also distorted because income taxation reduces labor supply, and thereby lowers the 'utilization rate' of human capital. However, this is the case only if labor supply of the high-skilled is distorted more than that of the low-skilled.

3.2 Firms

A representative firm produces a homogeneous consumption good, using low-skilled labor L and high-skilled labor H as inputs according to a constant-returns-to-scale production technology $Y(\cdot)$:

$$\begin{aligned} Y(L, AH), \quad Y_j(\cdot) &> 0, \quad Y_{jj}(\cdot) < 0, \quad Y_{LH}(\cdot) \geq 0, \\ \lim_{j\to 0} Y_j(\cdot) &= \infty, \quad \lim_{j\to\infty} Y_j(\cdot) = 0, \quad j = L, H. \end{aligned} \tag{6}$$

$A \geq 1$ is the 'skill bias' parameter indicating how much more productive high-skilled workers are relative to low-skilled workers. The subscript on the production function indicates a derivative.[12]

[11] First-order conditions are necessary, but generally not sufficient given the nonlinearities of the tax schedules. We assume that the second-order sufficiency conditions are respected at the optimal second-best allocation. See also the discussion on incentive compatibility below.

[12] With slight abuse of notation, we use a subscript H (rather than AH) to indicate a derivative with respect to the second argument of the production function.

Each labor input has positive, but diminishing marginal products. High- and low-skilled workers are co-operant factors of production ($Y_{LH} \geq 0$), but imperfect substitutes. We impose the Inada conditions on the production technology to ensure that in equilibrium there will be a strictly positive mass of high-skilled individuals, while some individuals remain low-skilled (i.e., $\underline{\theta} < \Theta < \overline{\theta}$) if $w^H > w^L$. The representative firm maximizes profits taking wage rates as given. The first-order conditions are:

$$w^L = Y_L(L, AH), \tag{7}$$

$$w^H = AY_H(L, AH). \tag{8}$$

The marginal product of each labor input should equal its marginal cost.

3.3 General equilibrium

General equilibrium is obtained if labor markets for both education types and the goods market clear:

$$L = \int_{\underline{\theta}}^{\Theta} \theta l_\theta^L \mathrm{d}F(\theta), \tag{9}$$

$$H = \int_{\Theta}^{\overline{\theta}} \theta l_\theta^H \mathrm{d}F(\theta), \tag{10}$$

$$Y = \int_{\underline{\theta}}^{\Theta} c_\theta^L \mathrm{d}F(\theta) + \int_{\Theta}^{\overline{\theta}} (c_\theta^H + p) \mathrm{d}F(\theta). \tag{11}$$

3.4 Government

The government maximizes social welfare, which is the sum of a concave transformation of low- and high-skilled utilities:

$$\int_{\underline{\theta}}^{\Theta} \Psi(U_\theta^L) \mathrm{d}F(\theta) + \int_{\Theta}^{\overline{\theta}} \Psi(U_\theta^H) \mathrm{d}F(\theta), \quad \Psi' > 0, \quad \Psi'' < 0. \tag{12}$$

Given that private marginal utility of income is constant, concavity of the social welfare function is necessary to obtain a social preference for redistribution.[13]

The government budget constraint is given by:

$$\int_{\underline{\theta}}^{\Theta} T^L(z_\theta^L) \mathrm{d}F(\theta) + \int_{\Theta}^{\overline{\theta}} T^H(z_\theta^H) \mathrm{d}F(\theta) = G, \tag{13}$$

which states that total tax revenue from nonlinear taxes on both groups sum to G, which is the exogenous revenue requirement of the government. Without loss of generality, we assume in the theoretical part that the government revenue requirement G is zero.

Since ability is private information, any second-best allocation needs to respect the incentive-compatibility constraints, which state that each individual with ability θ receives higher utility from the bundle $\{c_\theta, z_\theta\}$ of consumption (net income) and earnings (gross income) intended for

[13] With quasi-linear utility, a utilitarian government does not have a preference for redistribution, i.e., if $\Psi' = 0$, $\Psi'' = 0$.

her than any other bundle $\{c_{\theta'}, z_{\theta'}\}$ intended for any other individual with a different ability θ'. Formally, if we write utility as $u(c^j_\theta, z^j_\theta, w^j\theta) \equiv U(c^j_\theta, l^j_\theta)$, then incentive compatibility requires $u(c^j_\theta, z^j_\theta, w^j\theta) \geq u(c^j_{\theta'}, z^j_{\theta'}, w^j\theta)$, $\forall \theta, \theta', j$. This set of incentive-compatibility constraints can be replaced by a first-order differential equation on utility under the first-order approach, which is found by totally differentiating utility $U(c^j_\theta, l^j_\theta)$ with respect to θ, and substituting the first-order condition for labor supply in (3):

$$\frac{\mathrm{d}U^j_\theta}{\mathrm{d}\theta} = \frac{\left(l^j_\theta\right)^{1+1/\varepsilon}}{\theta}, \quad \forall \theta, j. \tag{14}$$

We assume that second-order sufficiency conditions are respected, so that the first-order approach can be applied. In our simulations, we check the second-order conditions and find that they are always respected.[14] Since we assume tax schedules are conditioned on education type j, the incentive compatibility constraints are independent of the endogenous wage rate w^j per efficiency unit of labor. Intuitively, mimicking across education types is not possible, since education is verifiable.

4 Optimal taxation

We find the optimal tax policy by maximizing social welfare in (12) over the entire allocation $\{c^j_\theta, l^j_\theta, U^j_\theta, \Theta\}$ subject to the resource constraint in (11) and incentive compatibility constraints in (33) and (34), while taking into account the definitions of utility in (1), labor market clearing in (9) and (10), and the optimal cutoff Θ in (4). The main theoretical results are given in the following proposition.

Proposition 1. *Define $g_\theta \equiv \frac{\Psi'(\cdot)}{\eta}$ as the social welfare weight of individual θ, where η is the multiplier on the economy's resource constraint. Define $\zeta \equiv \int_\Theta^{\overline{\theta}} (1 - g_\theta) f(\theta) \mathrm{d}\theta > 0$ as the distributional characteristic of education. Then, optimal policy can be characterized as follows.*

- *Optimal marginal tax rates on income follow the standard Mirrlees formula:*

$$\frac{T'(z_\theta)}{1 - T'(z_\theta)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{\int_\theta^{\overline{\theta}} (1 - g_\theta) f(\theta) \mathrm{d}\theta}{\theta f(\theta)}, \quad \forall \theta \in [\underline{\theta}, \overline{\theta}]. \tag{15}$$

- *Optimal marginal tax rates at the bottom and top are zero:*

$$T'(z_{\underline{\theta}}) = T'(z_{\overline{\theta}}) = 0. \tag{16}$$

- *Optimal marginal tax rates at the education cutoff Θ are positive and equal:*

$$\frac{T'(z^L_\Theta)}{1 - T'(z^L_\Theta)} = \frac{T'(z^H_\Theta)}{1 - T'(z^H_\Theta)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{\zeta}{\Theta f(\Theta)} > 0. \tag{17}$$

[14] Second-order conditions are satisfied if the optimal second-best allocation is monotonic in earnings (consumption), since the utility function respects Spence-Mirrlees (or single crossing) conditions, see also Mirrlees (1971), Ebert (1992), and Hellwig (2004).

- *Human capital formation is taxed on a net basis:*

$$T(z_{\Theta}^{H}) - T(z_{\Theta}^{L}) = \frac{\zeta}{\Theta f(\Theta)} \left(\left(l_{\Theta}^{H}\right)^{1+1/\varepsilon} - \left(l_{\Theta}^{L}\right)^{1+1/\varepsilon} \right) > 0. \tag{18}$$

Proof. See Appendix. □

4.1 Optimal income taxation

Our first important finding is that the optimal income tax expression in (15) is identical to the corresponding expressions in Mirrlees (1971), Diamond (1998), and Saez (2001). The reader is referred to these contributions for the interpretation and discussion of the optimal nonlinear tax schedule. Moreover, the end-point results in (16) confirm the standard results that optimal marginal tax rates at the top and the bottom are zero provided that there is a finite top and there is no bunching, for example at zero income. See also Sadka (1976) and Seade (1977) for further discussion and interpretations.

Marginal tax rates are implicitly conditioned on education, although the optimal income tax expression seems to condition marginal tax rates only on ability θ. The reason is that the income distributions for low-skilled and high-skilled workers are non-overlapping. In particular, individuals only differ in one dimension (i.e., their ability θ). As a result, there is a one-to-one mapping between education type and ability: all individuals with $\theta \geq \Theta$ are high-skilled and all individuals with $\theta < \Theta$ are low-skilled. Consequently, marginal income tax rates are in fact conditioned on education by setting a different nonlinear tax schedule for $\theta < \Theta$ than for $\theta \geq \Theta$.[15]

4.2 Marginal taxes at the cutoff Θ

The distributional characteristic ζ gives the marginal welfare gain of taxing high-skilled workers. It equals the marginal welfare gain of raising one unit of revenue from all high-skilled workers minus the utility costs – expressed in monetary equivalents – of raising that unit of revenue. Since $\zeta > 0$, we find that marginal tax rates at the education cutoff Θ are positive. Hence, the marginal tax rate 'at the top' for the low-skilled workers and the marginal tax rate 'at the bottom' for the high-skilled workers are non-zero. Surprisingly, marginal tax rates at z_{Θ}^{L} and z_{Θ}^{H} are exactly equal, i.e., $T'(z_{\Theta}^{L}) = T'(z_{\Theta}^{H})$ despite the fact that earnings for low-skilled and high-skilled workers at the cutoff Θ are not, i.e., $z_{\Theta}^{L} \neq z_{\Theta}^{H}$ as can be inferred from (5).

Intuitively, positive marginal tax rates contribute to social welfare only if they increase income redistribution. A positive marginal tax at Θ redistributes more income, since at Θ, high-skilled individuals pay more tax than low-skilled individuals. This intuition is in line with Sadka (1976) and Seade (1977).[16] The optimal marginal tax rate at Θ equates the social

[15] If individuals differ in more than dimension, for example, because their abilities in low-skilled and high-skilled work are different, then there is no longer a one-to-one mapping between multi-dimensional ability and education. In that case, education-dependent nonlinear tax functions are required to implement the second-best optimal allocation, see also Scheuer (2014) and Gomes et al. (2018).

[16] These authors show that optimal tax rates are zero if nobody pays more tax (if the marginal tax rates are raised at the top of the income distribution) or if nobody can benefit from higher transfers (if the marginal tax rates are raised at the bottom of the income distribution). The latter result requires that there is no bunching at the bottom, see Seade (1977).

marginal benefits of a higher marginal tax rate (ζ) and the marginal deadweight losses of doing so ($\frac{T'(z^j_\Theta)}{1-T'(z^j_\Theta)}\left(1+\frac{1}{\varepsilon}\right)^{-1}\Theta f(\Theta)$). Since marginal tax rates distort labor supply of the marginally low- and high-skilled in the same way, and social welfare weights are equalized across these individuals, marginal tax rates are equalized.

Although marginal tax rates are continuous and smooth in ability θ, it contains a hole for incomes z_θ inside the interval (z^L_Θ, z^H_Θ). The reason is that there is no mass of individuals inside this interval. For this reason we will plot optimal nonlinear taxes against ability θ and not against earnings z_θ in our simulations.

4.3 Optimal net taxes on human capital

The third main result is that human capital formation is optimally taxed on a net basis, as shown in (18). Optimal education policy has three roles. First, education policy ensures that education choices are efficient, i.e., $T(z^H_\Theta) = T(z^L_\Theta)$ if there would be no distributional benefits of net taxes on education, i.e., if $\zeta = 0$. This is the same as in Bovenberg and Jacobs (2005): education policy is employed to alleviate the distortions of income taxation on education.

Second, net taxes on education are desirable to redistribute income from high-skilled individuals to low-skilled individuals, since $\zeta > 0$. Intuitively, all infra-marginal high-skilled workers enjoy rents from their ability, which the government likes to redistribute, see also Findeisen and Sachs (2017). Therefore, net tax liabilities of high-skilled workers should be larger than net tax liabilities of low-skilled workers: $T(z^H_\Theta) > T(z^L_\Theta)$, so that education is taxed on a net basis. Hence, in contrast to Bovenberg and Jacobs (2005), it is not optimal to remove all distortions from income taxation on education. The optimal net tax equates marginal distributional gains to marginal distortions. The larger is the distributional gain ζ of taxing human capital on a net basis, the higher is the net tax on human capital. Given that the social welfare weights g_θ are on average equal to one, the average welfare weight of the high-skilled is smaller than one (i.e., $\int_\Theta^{\overline{\theta}} g_\theta f(\theta)\mathrm{d}\theta < 1$), since social welfare weights are declining in income. The larger are distortions on skill formation – as indicated by a larger 'base' of the net tax on education $\Theta f(\Theta)$ – the lower should the optimal net tax on education be.

Third, education policy is used to alleviate the tax distortions on labor supply, as captured by the term $\left(l^H_\Theta\right)^{1+1/\varepsilon} - \left(l^L_\Theta\right)^{1+1/\varepsilon}$. Since wage rates of high-skilled workers are larger (i.e., $w^H > w^L$), marginal tax rates for low-skilled and high-skilled workers equal at Θ (i.e, $T'(z^L_\Theta) = T'(z^H_\Theta)$), and high-skilled workers work more hours, i.e., $l^H_\Theta > l^L_\Theta$. A positive net tax on education reduces the number of high-skilled workers, and thus the total number of hours worked. Given that labor supply is optimally distorted downwards by positive marginal tax rates, the reduction in labor supply of high-skilled workers is larger than the increase in labor supply of low-skilled workers. This gives a first-order welfare loss. Consequently, the more complementary education and hours worked are, the lower should be the net tax on education.[17]

[17]This argument is similar to the Corlett and Hague (1953) argument that optimal commodity taxes should feature lower taxes on goods that are stronger complements to work to alleviate distortions of commodity taxes on labor supply. See also Atkinson and Stiglitz (1976) and Jacobs and Boadway (2014).

4.4 General-equilibrium effects

Our fourth result is that neither the optimal nonlinear income tax nor the optimal net tax on education depend on general-equilibrium effects on the wage structure, in contrast to Stiglitz (1982), Stern (1982), Jacobs (2012), Rothschild and Scheuer (2013), Ales et al. (2015), Sachs et al. (2017), and Jacobs and Thuemmel (2018). The fundamental reason for this result is that the government has access to education-dependent nonlinear taxes. See also Scheuer (2014) and Scheuer and Werning (2016). This also implies that the standard result of zero marginal tax rates at the end-points is preserved, in contrast to the findings in Stern (1982), Stiglitz (1982), Jacobs (2012), and Sachs et al. (2017). Intuitively, compression of the wage structure results in the same distortions on labor supply as a distributionally-equivalent change in nonlinear taxes, but in addition it generates larger distortions in investments in education. Therefore, any redistribution via general-equilibrium effects on the wage structure can also be achieved with the nonlinear tax system, without additionally distorting human capital investments. Hence, it is no longer optimal to exploit general-equilibrium effects for income redistribution if tax rates can be conditioned on the education decision. Hence, the absence of education-dependent income tax rates explains why wage-compression via tax and education policy could be desirable.

4.5 Implementation

Our fifth main finding is that the optimal net tax on human capital only requires a discontinuity in the nonlinear tax function at Θ. Hence, the optimal second-best allocation can also be implemented with a *continuous* nonlinear tax function $\tilde{T}(z_\theta)$ that does not discontinuously jump at Θ, so that $\tilde{T}(z^L_\Theta) = \tilde{T}(z^H_\Theta)$. In this case, the government needs a subsidy (or tax) on education S for the high-skilled workers, which is independent of income. In particular, the tax schedule is then given by:

$$T(z_\theta) \equiv \begin{cases} \tilde{T}(z_\theta), & \theta \in [\underline{\theta}, \Theta) \\ \tilde{T}(z_\theta) - S, & \theta \in [\Theta, \overline{\theta}] \end{cases}, \quad T(z^L_\Theta) = \tilde{T}(z^L_\Theta) = \tilde{T}(z^H_\Theta). \tag{19}$$

This tax implementation would perhaps correspond best to real-world tax and education policies.

4.6 Relation to the optimal tax literature

Our findings relate to the results in Saez (2002) and Jacquet et al. (2013), who also study models of optimal income taxation with an extensive margin. Saez (2002) analyzes optimal taxes on occupational choice and participation, where both choices are on the extensive margin. Jacquet et al. (2013) analyze optimal income taxation with labor supply along the intensive (hours) and the extensive (participation) margin. We analyze optimal taxation of occupational choice, i.e., education, as in Saez (2002), and optimal income taxation with an intensive margin as in Jacquet et al. (2013). We derive an expression for the optimal net tax on education, which very much resembles the expression for the optimal participation tax in Saez (2002) and Jacquet et al. (2013). Moreover, we find the same optimal tax expression as in Mirrlees (1971) in a model with an intensive (hours) margin and an extensive margin in occupational choice. Our formula contrasts with Jacquet et al. (2013), who show that the participation margin affects the optimal nonlinear income tax by lowering the distributional benefits of the marginal tax rate. Intuitively, since marginal tax rates do not create distortions on the extensive margin for all individuals with ability $\theta \neq \Theta$, distortions in education do not determine the optimal income tax. Therefore, a discontinuous jump in the tax schedule at Θ is sufficient to implement optimal net taxes on the extensive margin, and no extensive-margin terms are present in the formula for the optimal nonlinear income tax.

5 Skill-biased technical change

What is the effect of skill-biased technical change on optimal income taxes? From (15) follows that optimal marginal tax rates only change due to a change in the social welfare weights $g_\theta \equiv \frac{\Psi'}{\eta}$, since all other elements in the optimal tax formula are primitives of our model, and thus invariant to skill bias, i.e., θ, $f(\theta)$, and ε. Variations in the social welfare weights are driven by changes in the distribution of indirect utilities. In turn, these changes result from the impact of SBTC on consumption and labor supply.[18] Ultimately, variations in social welfare weights thus capture the distributional impact of SBTC.[19] To derive the change in the social welfare weights, it is necessary to derive the comparative statics of the model in the next Lemma.

Lemma 1. *Denote a relative change in a variable by $\tilde{x} \equiv \mathrm{d}x/x$, except for $\tilde{\Theta} \equiv \mathrm{d}\Theta$. The comparative statics of the model with respect to a marginal change in skill bias $\tilde{A}$ are given by:*

$$\tilde{l}_\theta^j = \varepsilon_\theta \tilde{w}^j, \quad \forall \theta, j, \tag{20}$$

$$\tilde{z}_\theta^j = (1+\varepsilon_\theta)\tilde{w}^j, \quad \forall \theta, j, \tag{21}$$

$$\tilde{\Theta} = \mu^L \tilde{w}^L - \mu^H \tilde{w}^H, \tag{22}$$

$$\tilde{H} = (\bar{\varepsilon}^H + \delta^H \mu^H)\tilde{w}^H - \delta^H \mu^L \tilde{w}^L, \tag{23}$$

$$\tilde{L} = (\bar{\varepsilon}^L + \delta^L \mu^L)\tilde{w}^L - \delta^L \mu^H \tilde{w}^H, \tag{24}$$

$$\tilde{w}^L = \left[\frac{\alpha + \alpha(\bar{\varepsilon}^H + (\delta^L + \delta^H)\mu^H)}{\sigma + (1-\alpha)(\bar{\varepsilon}^H + (\delta^L + \delta^H)\mu^H) + \alpha(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}\right]\tilde{A}, \tag{25}$$

$$\tilde{w}^H = \left[\frac{\sigma - 1 + \alpha + \alpha(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}{\sigma + (1-\alpha)(\bar{\varepsilon}^H + (\delta^L + \delta^H)\mu^H) + \alpha(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}\right]\tilde{A}, \tag{26}$$

where $\sigma \equiv \frac{Y_H Y_L}{Y_{HL} Y}$ is the elasticity of substitution between high-skilled and low-skilled labor, $\alpha \equiv \frac{AHY_H}{Y}$ is the high-skilled income share, $\delta^H \equiv \frac{\Theta f(\Theta) l_\Theta^H}{H}$ and $\delta^L \equiv \frac{\Theta f(\Theta) l_\Theta^L}{L}$ measure the relative mass of high-skilled and low-skilled labor at the education cutoff Θ, $\varepsilon_\theta \equiv \left(\frac{1-T'^j - z_\theta^j T''^j}{1-T'^j + \varepsilon z_\theta^j T''^j}\right)\varepsilon$ is the wage elasticity of labor supply, $\bar{\varepsilon}^H \equiv \int_\Theta^{\overline{\theta}} \frac{\theta l_\theta^H}{H}\varepsilon_\theta \mathrm{d}F(\theta)$ and $\bar{\varepsilon}^L \equiv \int_{\underline{\theta}}^{\Theta} \frac{\theta l_\theta^L}{L}\varepsilon_\theta \mathrm{d}F(\theta)$ are the employment-weighted elasticities of high-skilled and low-skilled labor supply, $\mu \equiv \frac{(1-T'^H)z_\Theta^H}{(1-T'^L)z_\Theta^L}$ is a

[18] We discuss below that due to Roy's identity a *marginal* change in skill-bias only affects indirect utilities by changing income.

[19] Variations in social welfare weights do not capture changes in the government's preferences for redistribution. These preferences are fixed, and given by the social welfare function $\Psi(\cdot)$.

share parameter, $\mu^H \equiv \frac{\mu}{(1-\mu)\varepsilon\Theta}\left[\left(1+\varepsilon\frac{T''^H z^H_\Theta}{(1-T'^H)}\right)(1+\varepsilon_\Theta)-\varepsilon\right] > 0$ *and* $\mu^L \equiv \frac{1}{(1-\mu)\varepsilon\Theta}\left[\left(1+\varepsilon\frac{T''^L z^L_\Theta}{(1-T'^L)}\right)(1+\varepsilon_\Theta)-\varepsilon\right] > 0$ *denote the semi-elasticities of the education cutoff* Θ *with respect high-skilled and low-skilled wages.*

Proof. See Appendix. □

From Lemma 1 follows that if $\sigma > 1$, if the elasticities of labor supply are similar ($\bar{\varepsilon}^H \approx \bar{\varepsilon}^L$), and if human capital responds similarly to a low-skilled wage change and a high-skilled wage change ($\mu^H \approx \mu^L$), then the high-skilled wage increases relatively more than the low-skilled wage in response to skill-biased technical change. This is intuitive. If the relative supply of high-skilled labor to low-skilled labor does not change (much), then $\sigma > 1$ is the condition for skill bias to increase wage inequality.

What happens to individual utility when SBTC appears? First note that – by Roy's identity – utility only changes because income changes, not because consumption and labor change. Hence, we find that the change in utility is a function only of a change in the wage rate: $\mathrm{d}U^j_\theta = w^j\theta(1-T'^j)l^j_\theta\tilde{w}^j = (l^j_\theta)^{1+\varepsilon}\tilde{w}^j$, where the second step follows from the first-order condition in (3). Hence, using Lemma 1, we can derive the change in utility for high-skilled and low-skilled workers as:

$$\mathrm{d}U^H_\theta = \left(l^H_\theta\right)^{1+1/\varepsilon}\left[\frac{\sigma-1+\alpha+\alpha(\bar{\varepsilon}^L+(\delta^L+\delta^H)\mu^L)}{\sigma+(1-\alpha)(\bar{\varepsilon}^H+(\delta^L+\delta^H)\mu^H)+\alpha(\bar{\varepsilon}^L+(\delta^L+\delta^H)\mu^L)}\right]\tilde{A}, \tag{27}$$

$$\mathrm{d}U^L_\theta = \left(l^L_\theta\right)^{1+1/\varepsilon}\left[\frac{\alpha+\alpha(\bar{\varepsilon}^H+(\delta^L+\delta^H)\mu^H)}{\sigma+(1-\alpha)(\bar{\varepsilon}^H+(\delta^L+\delta^H)\mu^H)+\alpha(\bar{\varepsilon}^L+(\delta^L+\delta^H)\mu^L)}\right]\tilde{A}. \tag{28}$$

The terms in brackets are independent of ability type θ (they are dependent on Θ, however). From the monotonicity constraint follows that $\left(l^H_\theta\right)^{1+1/\varepsilon} > \left(l^L_\theta\right)^{1+1/\varepsilon}$, but also $l^j_\theta > l^j_{\theta'}$ for any $\theta > \theta'$. Hence, the impact of skill-biased technical change on the utility of the high-skilled is always larger than the impact of skill-biased technical change on the utility of the low-skilled: i.e., $\mathrm{d}U^H_\theta > \mathrm{d}U^L_\theta$ for all θ. Moreover, the impact of SBTC is always bigger for individuals with a higher ability than with a lower ability: $\mathrm{d}U^j_\theta > \mathrm{d}U^j_{\theta'}$ for any $\theta > \theta'$.

Lemma 2. *The relative change in the social welfare weight at skill level* θ *for education type* j *is given by:*

$$\tilde{g}_\theta = -\rho_\theta\tilde{U}^j_\theta + \int_{\underline{\theta}}^{\bar{\theta}}\rho_{\theta'}g_{\theta'}\tilde{U}^j_{\theta'}\mathrm{d}F(\theta), \tag{29}$$

where $\rho_\theta \equiv -\frac{\Psi'' U^j_\theta}{\Psi'} > 0$ *is the local elasticity of inequality aversion.*

Proof. See Appendix. □

Lemma 2 proves that SBTC reduces the social welfare weights for all workers because their own utility rises ($-\rho_\theta\tilde{U}^j_\theta < 0$) and the more so for those workers whose utility rises relatively more. The social welfare weights fall more if there is a more inequality averse government (ρ_θ is larger). However, there is a counteracting effect via the change in the multiplier on the government budget constraint η. Since average utility rises in the entire population, the shadow value of public resources declines ($\tilde{\eta} < 0$). Hence, the social welfare weight for every individual increases as a result. The latter effect is captured by $\int_{\underline{\theta}}^{\bar{\theta}} \rho_{\theta'} g_{\theta'} \tilde{U}^j_{\theta'} \mathrm{d}F(\theta)$ and is the same for everyone. Whether an individual's social welfare weight increases or decreases is ambiguous and depends on the relative change in individual utility in comparison to the relative change in total utility. We cannot analytically derive who features the largest change in the social welfare weights. This depends on the initial distribution of utilities. Therefore, we conclude that we cannot – at this stage – say more on how SBTC changes social welfare weights. Moreover, the derivations are valid only for marginal changes in skill-bias. Below, we therefore turn to simulations to study how optimal marginal taxes should respond to (infra-marginal) changes in skill-bias.

How should optimal net taxes on education respond to skill bias? The expression for the optimal net tax on education in (18) reveals that the impact of skill bias on the net tax on education is also fundamentally ambiguous. First, since the wages of high-skilled workers rise more than that of the low-skilled, the distributional benefits of net taxes on education rise, i.e., $\frac{\partial \zeta}{\partial A} > 0$. Second, skill bias lowers the cutoff Θ, i.e., $\frac{\partial \Theta}{\partial A} < 0$, which follows from (4). The latter also changes the density at the cutoff $f(\Theta)$. Depending on where the cutoff Θ is located in the distribution $f(\theta)$, $\frac{\partial f(\Theta)}{\partial A}$ can be either positive or negative. Consequently, the distortions of net taxes on education, as represented by $\Theta f(\Theta)$, may increase or decrease with skill bias. Third, the difference in labor supply at the skill cutoff rises as a result of the rise in the skill premium, i.e., $\frac{\partial[(l^H_\Theta)^{1+1/\varepsilon} - (l^L_\Theta)^{1+1/\varepsilon}]}{\partial A} > 0$. Hence, the government likes to reduce net taxes on education to alleviate the distortions in the labor market. Distributional gains of net taxes on education certainly increase with skill bias, but distortion of net taxes may increase as well. Hence, it is not clear whether skill bias raises net taxes on education.

6 Simulations

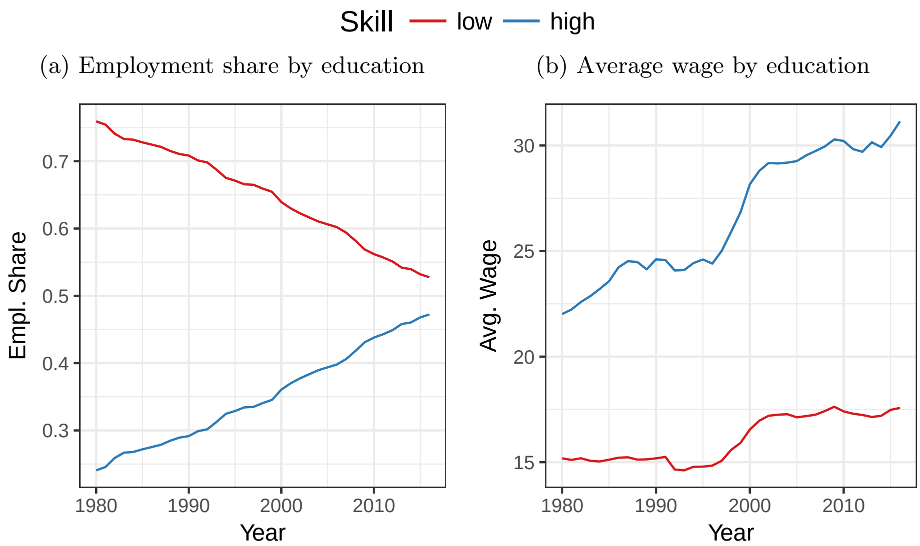
This section conducts simulations for our model calibrated to the US economy to illustrate the impact of SBTC on optimal policy. We take 1980 as the baseline year for the calibration, since SBTC started to take off around that time. We choose 2016 as the final year. For a given tax system, we target the share of college graduates as well as the level and change in the college wage premium. We then use the calibrated model to compute optimal taxes in 1980 and show how optimal taxes should have responded to SBTC in the US labor market.

6.1 Functional forms and calibration

To simulate the model, we need to specify the ability distribution, the production function and the social welfare function. As is common in the optimal tax literature, we assume that ability θ follows a log-normal distribution, which is appended with a Pareto tail. We follow Tuomala (2010) and choose parameters $\mu = 0.4$ and $\varsigma = 0.39$ for the log-normal part of the distribution. The Pareto parameter α in the US has been 2.5 in 1976 and was about 1.5 in 2007 (see e.g. Atkinson et al., 2011). We set it to an intermediate value of $\alpha = 2$.[20]

[20] We append the Pareto tail such that the slopes of the log-normal and Pareto distributions are identical at the cut-off. We proportionately rescale the densities of the resulting distribution to ensure they sum to one.

Figure 1: Employment share and average wage by education



Note: Based on CPS data. High-skilled are individuals with at least a two-year college or associate's degree. All other individuals are classified as low-skilled. The sample includes individuals of working age, working full time. See the Data Appendix for a description of the sample.

We assume a constant-elasticity of substitution (CES) production function as in the canonical model of SBTC (see Katz and Murphy, 1992; Violante, 2008; Acemoglu and Autor, 2011):

$$Y(L,H) = \tilde{A}\left(\omega L^{\frac{\sigma-1}{\sigma}} + (1-\omega)(AH)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}, \quad \tilde{A}, \omega, \sigma > 0, \tag{30}$$

where $\tilde{A}$ is a Hicks-neutral productivity parameter. A represents the skill bias in technology – and SBTC corresponds to an increase in A. ω governs the share of low- and high-skilled labor income in total output. σ is the constant elasticity of substitution between the two types of labor. Since we aim to capture the essence of SBTC, there are four natural calibration targets: levels and changes of the share of high-skilled workers and levels and changes of the skill premium. We plot time series for the share of high- and low-skilled as well as average wages by education based on data from the Current Population Survey (CPS) in Figure 1.

The share of high-skilled workers is 24% in 1980 and 47% in 2016. The average wage per hour worked (converted to 2016 dollar values) of the low-skilled is \$15 in 1980 and \$17.5 in 2016, whereas for the high-skilled it is \$22 in 1980 and \$31 in 2016. The skill premium is defined as the ratio of average wages of high- and low-skilled. It is 1.47 in 1980 and 1.77 in 2016, and thus increased by 21%. In our model, the skill premium corresponds to

$$\text{skill premium} \equiv \frac{w^H}{w^L} \frac{\frac{1}{1-F(\Theta)}\int_{\Theta}^{\overline{\theta}} \theta \mathrm{d}F(\theta)}{\frac{1}{F(\Theta)}\int_{\underline{\theta}}^{\Theta} \theta \mathrm{d}F(\theta)}. \tag{31}$$

When targeting the level of the skill premium, we need to keep in mind that the income distributions of low- and high-skilled do not overlap in our model. Each high-skilled worker earns more than any low-skilled worker. As a result, the model generates a level of the skill premium that is too high compared to the data. Moreover, it is possible that the skill-premium decreases with SBTC due to composition effects. As more individuals sort into college, the average ability of both low- and high-skilled workers decreases. If the average ability of the high-skilled workers falls relatively more than that of the low-skilled workers, the ratio of average abilities $\frac{1}{1-F(\Theta)}\int_{\Theta}^{\overline{\theta}}\theta \mathrm{d}F(\theta)/\frac{1}{F(\Theta)}\int_{\underline{\theta}}^{\Theta}\theta \mathrm{d}F(\theta)$ falls. If this composition effect is strong enough to compensate for the increase in relative wage rates w^H/w^L, the skill premium declines. Since a drop in the skill premium due to SBTC would be counterfactual, we choose our baseline such that w^H/w^L increases strongly. We achieve this by setting $\sigma = 10000$, which implies that low-skilled and high-skilled workers are (close to) perfect substitutes. As a result, SBTC raises the wage rate w^H, whereas w^L remains constant. Setting $\sigma = 10000$ is without much loss of generality, since the optimal policy rules do not exploit general-equilibrium effects due to imperfect labor substitution. We normalize the level of skill bias in 1980 to one: $A_{1980} = 1$. We approximate the income tax system in 1980 by a linear function with intercept $-b$ and marginal tax rate τ.[21] We set $\tau = 35\%$, which is the average marginal income tax in 1980. Moreover, we set the intercept of the tax system such as to match the average tax rate in 1980, which is 18%.[22] Quasi-linear utility implies that individual labor supplies are given by $l_\theta^j = [(1-\tau)w^j]^\varepsilon$, $\forall j$. We set the labor-supply elasticity ε to 0.3, following Blundell and Macurdy (1999) and Meghir and Phillips (2010).

It remains to calibrate parameters $\tilde{A}$, ω, and A_{2016} as well as the cost of college, p, which leaves us with four parameters and four targets. To calibrate the model, we compute the equilibrium of the economy and choose parameters to minimize a weighted distance to our calibration targets. The weights are chosen to achieve a compromise between matching the increase in the skill premium on the one hand, and limiting the level of the skill premium on the other hand. The calibrated parameters are summarized in Table 1. In Table 2, we compare the targeted moments implied by the calibrated model with those from the data. The skill premium is exaggerated by the model. In contrast, the change in the skill premium and employment shares are matched well. We use the calibrated ability distribution and production function, as well as ε, G, and p as inputs in our baseline simulation.

Before we can compute optimal taxes, we need to specify a social welfare function. We assume a standard social welfare function with a constant elasticity of relative inequality aversion ρ:

$$\Psi(U_\theta) = \begin{cases} \frac{U_\theta^{1-\rho}}{1-\rho}, & \rho \neq 1 \\ \ln(U_\theta), & \rho = 1 \end{cases}, \quad \rho > 0. \tag{32}$$

ρ captures the government's desire for redistribution. A value of $\rho = 0$ corresponds to a utilitarian social welfare function, while for $\rho \to \infty$ the welfare function approximates a Rawlsian criterion.

Table 1: Calibrated Parameters

Param.	Description	Value	Source
μ	Ability distribution: mean	0.40	Tuomala (2010)
ς	Ability distribution: st. dev.	0.39	Tuomala (2010)
α	Ability distribution: Pareto parameter	2.00	Atkinson et al. (2011)
ε	Labor-supply elasticity	0.30	Blundell and Macurdy (1999); Meghir and Phillips (2010)
τ	Tax rate	0.35	NBER Taxsim
b	Tax intercept	30.90	calibrated
G	Revenue requirement	32.71	implied
p	Direct cost of college	145.10	calibrated
σ	Substitution elasticity	10000	fixed
$\tilde{A}$	Productivity parameter	42.24	calibrated
ω	Share parameter	0.09	calibrated
A_{1980}	Skill-bias 1980	1.00	normalized
A_{2016}	Skill-bias 2016	1.39	calibrated

Table 2: Calibration: Model vs. Data

Moment	Model	Data
Skill premium in 1980	30.16	1.47
Skill premium in 2016	36.33	1.77
Skill premium: relative change	0.20	0.21
Share of high-skilled in 1980	0.24	0.24
Share of high-skilled in 2016	0.47	0.47

Table 3: Summary statistics based on simulation

Skill bias	Share of high-skilled	Skill premium	Net tax at Θ
No	0.45	26.32	17.10
Yes	0.79	36.84	6.53

Note: Variables from baseline simulation with $\rho = 0.5$ and other parameters given in Table 1. No skill bias corresponds to $A = 1$, skill bias to $A = 1.39$. The skill premium is as defined in (31). The net tax at Θ is defined as $w^H_\Theta \Theta l^H_\Theta - c^H_\Theta - p - (w^L_\Theta \Theta l^L_\Theta - c^L_\Theta)$.

6.2 Results

In order to illustrate the role of SBTC, we simulate the model for two levels of skill bias: $A_{1980} = 1$ (no skill bias) and A_{2016} =1.39 (skill bias).[23] We set $\rho = 0.5$, which generates optimal marginal tax rates in line with those observed empirically. Table 3 shows the impact of SBTC on the share of high-skilled workers, the skill premium, and the net tax on education. The share of high-skilled in both regimes is higher than in the data. As in the calibration, the skill premium is too large compared to the data. Moreover, its relative increase is now 0.40. Finally, optimal net taxes on skill-formation are positive. Education is thus distorted downwards relative to a situation without taxes. The net tax on education is reduced with skill bias. As skill bias increases, the distortions of taxes on education increase, but so do the distributional benefits of higher net taxes on education. In the simulation, the increase in distortions outweighs the rise in distributional benefits, leading to a lower net tax. Hence, SBTC may call for higher subsidies on education.

Figure 2 plots the effect of skill bias on optimal marginal and average tax rates. Marginal tax rates follow the common U-shape, as for example in Saez (2001). We confirm our theoretical result that marginal tax rates at the ability cutoff Θ do not differ between education groups. Skill-bias increases optimal marginal tax rates most around the ability threshold Θ. In contrast, at higher ability levels, skill-bias lowers optimal marginal tax rates. Average tax rates increase, hence the optimal tax system becomes more progressive. Moreover, average tax rates are negative for a large range of ability levels. Individuals with low ability thus receive a transfer, which is funded by taxes paid by individuals of high ability. The largest average tax rates are around 50%. Under skill bias, average tax rates decrease for the low-skilled workers and slightly increase for the high-skilled workers, making the tax system overall more progressive.

SBTC has an impact on marginal tax rates only through the changes in the marginal social welfare weights g_θ. To illustrate how the social welfare weights change with SBTC, we plot them in Figure 3. Social welfare weights decline in ability, but interestingly, they exhibit a kink at the education margin Θ.[24] If skill bias gets stronger, social welfare weights change especially around Θ: more weight is put on low-ability individuals and on high-ability individuals around the cutoff Θ. For other ability levels, the social welfare weights change little. These changes in the social welfare weights explain why SBTC affects optimal marginal tax rates mostly around Θ.

6.3 Robustness

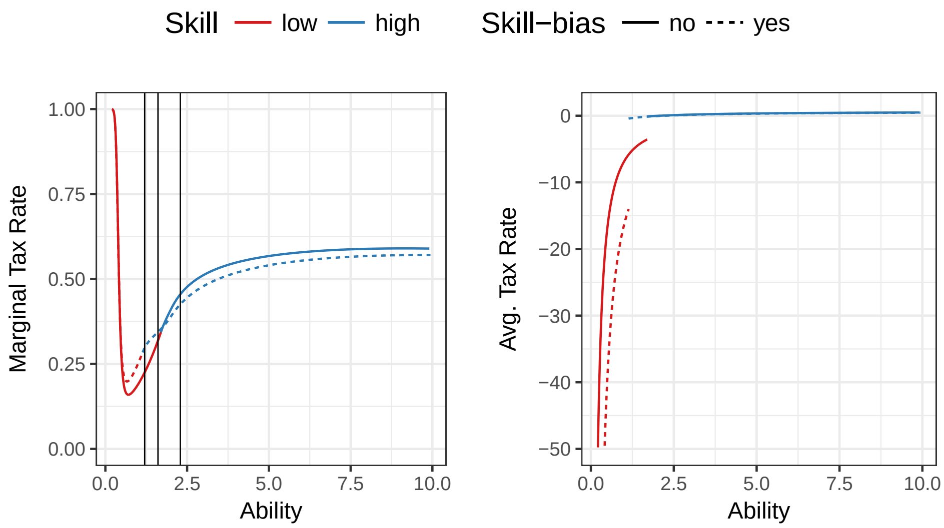
We check the robustness of our results by changing inequality aversion, the labor-supply elasticity, and the elasticity of substitution between low-skilled and high-skilled labor. We do so by varying one parameter at a time, while keeping the other parameters as in the baseline simulation. The labor-supply elasticity ε and the elasticity of substitution σ are parameters that are

[21]The linear tax function is only used in the calibration, but not to obtain optimal taxes.

[22]See http://users.nber.org/˜taxsim/allyup/ally.html.

[23]Simulations are conducted using the optimal-control software GPOPS-II, see Rao et al. (2010). Simulation programs are available upon request from the authors.

[24]The reason is that utilities exhibit a kink at Θ, which is due to a different slope of incomes in θ, generated by $w^H > w^L$.

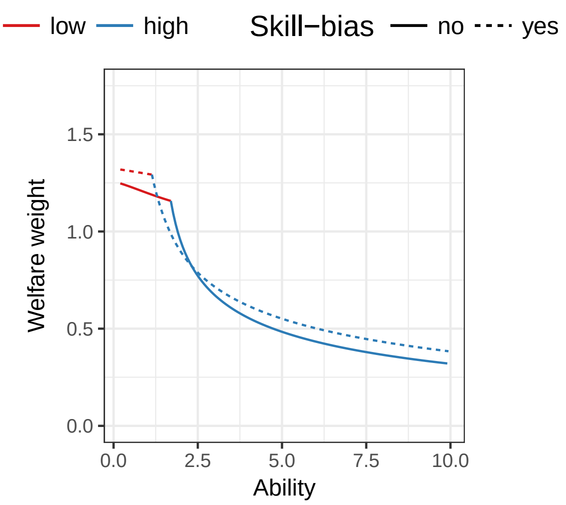


(a) Marginal Tax Rate (b) Average Tax Rate

Figure 2: Effect of skill bias on optimal taxes

Note: No skill bias corresponds to $A_{1980} = 1$, skill bias to A_{2016} =1.39. The vertical lines in the left panel correspond to (from left to right) the 25th, 50th, and 75th percentile of the ability distribution. An ability level of 10 corresponds to the 99th percentile.

Figure 3: Effect of skill bias on social welfare weights



Note: Social welfare weights $U_\theta^{-\rho}/\eta$ are based on social welfare function in (32), where η is the multiplier on the resource constraint.

fixed in the calibration stage. If we change ε and σ, we first need to re-calibrate the model to the data. We then simulate the effect of SBTC based on the re-calibrated model. In contrast, a change in inequality aversion does not require us to re-calibrate the model, since the calibration is independent of the degree of inequality aversion.

The findings are summarized in Table 4 and Figures 4 to 8 in the Appendix. We find that the net tax on education drops with SBTC. While the patterns of marginal and average tax rates differ quantitatively across different parameter settings, the qualitative pattern remains the same: SBTC raises optimal marginal tax rates around the ability cutoff Θ, while lowering them towards the top. Moreover, SBTC makes the tax system more progressive. Quantitatively, marginal tax rates are lower if there is less desire for redistribution and if the labor-supply elasticity is smaller.

7 Conclusions

This paper studies the consequences of skill-biased technical change (SBTC) for optimal nonlinear income taxes and net taxes on education. We merge the canonical model of SBTC due to Katz and Murphy (1992) and the Mirrlees (1971) model of optimal taxation, while allowing for skill-dependent income taxes. For a given level of skill-bias, we find that optimal nonlinear marginal tax rates follow the original nonlinear tax formula of Mirrlees (1971), despite the endogeneity of the education choice and endogenous wage rates. Indeed, optimal taxes do not exploit general-equilibrium effects on the wage structure. Moreover, the optimal marginal tax rates are the same at the cutoff ability level at which individuals are indifferent between becoming high-skilled or remaining low-skilled. The optimal tax function only displays a discontinuity around the cutoff ability level, such that education should optimally be taxed on a net basis. The effect of SBTC on optimal marginal tax rates and the optimal net tax on education is theoretically ambiguous.

We explore the consequences of SBTC on optimal nonlinear taxes and on net taxes on education quantitatively by simulating our model based on US data. We find that SBTC raises optimal marginal tax rates in the middle of the income distribution, while lowering them towards the top. The net tax on education falls with SBTC, whereas the tax system becomes overall more progressive.

Tinbergen (1975) and Goldin and Katz (2010) have advocated higher income tax progressivity and higher education subsidies to win the race with technology. Our quantitative results for optimal policies are in line with these policy suggestions. However, the underlying mechanism is not to win the race with technology. In particular, income taxes and education policies do not exploit general-equilibrium effects to compress the wage distribution. Since the tax schedule can be conditioned on education, any income redistribution via wage compression can be achieved as well with the tax system, without generating additional distortions in education. Education subsidies should increase with SBTC, since larger education distortions dominate the larger distributional losses of higher subsidies. Similarly, income taxes should be made more progressive, because distributional gains increase more than deadweight losses.

Our modeling strategy generated a non-overlapping wage distribution for high-skilled and low-skilled workers. Future research should explore the consequences of wage overlap for the optimal response of income taxes and education subsidies to SBTC. If income distributions overlap across low-and high-skilled workers, skill-dependent nonlinear income taxes will be harder to design, since the government needs to take into account that at each income level, workers may be low-skilled or high-skilled.

Appendix

A Proof Proposition 1

To solve the optimal tax problem, we integrate the incentive constraints for low-skilled and high-skilled workers by parts. In particular, let μ_θ^j be the multiplier on the incentive constraints in the Lagrangian for the optimal tax problem in (35), then we find:

$$\int_{\underline{\theta}}^{\Theta}\left[\mu_\theta^L\frac{\left(l_\theta^L\right)^{1+1/\varepsilon}}{\theta}+U_\theta^L\frac{\mathrm{d}\mu_\theta^L}{\mathrm{d}\theta}\right]\mathrm{d}\theta+\mu_{\underline{\theta}}^L U_{\underline{\theta}}^L-\mu_\Theta^L U_\Theta^L=0, \tag{33}$$

$$\int_{\Theta}^{\overline{\theta}}\left[\mu_\theta^H\frac{\left(l_\theta^H\right)^{1+1/\varepsilon}}{\theta}+U_\theta^H\frac{\mathrm{d}\mu_\theta^H}{\mathrm{d}\theta}\right]\mathrm{d}\theta+\mu_\Theta^H U_\Theta^H-\mu_{\overline{\theta}}^H U_{\overline{\theta}}^H=0. \tag{34}$$

The Lagrangian for maximizing social welfare can be written as follows:

$$\begin{aligned}\max_{\{c_\theta^j,l_\theta^j,U_\theta^j,\Theta,L,H\}}\mathcal{L}\equiv&\int_{\underline{\theta}}^{\Theta}\Psi(U_\theta^L)\mathrm{d}F(\theta)+\int_{\Theta}^{\overline{\theta}}\Psi(U_\theta^H)\mathrm{d}F(\theta)+\pi\left[U_\Theta^H-U_\Theta^L\right]\\&+\eta\left[Y(L,AH)-\int_{\underline{\theta}}^{\Theta}c_\theta^L\mathrm{d}F(\theta)-\int_{\Theta}^{\overline{\theta}}(c_\theta^H+p)\mathrm{d}F(\theta)\right]\\&-\int_{\underline{\theta}}^{\Theta}\left[\mu_\theta^L\frac{\left(l_\theta^L\right)^{1+1/\varepsilon}}{\theta}+U_\theta^L\frac{\mathrm{d}\mu_\theta^L}{\mathrm{d}\theta}\right]\mathrm{d}\theta-\mu_{\underline{\theta}}^L U_{\underline{\theta}}^L+\mu_\Theta^L U_\Theta^L\\&-\int_{\Theta}^{\overline{\theta}}\left[\mu_\theta^H\frac{\left(l_\theta^H\right)^{1+1/\varepsilon}}{\theta}+U_\theta^H\frac{\mathrm{d}\mu_\theta^H}{\mathrm{d}\theta}\right]\mathrm{d}\theta-\mu_\Theta^H U_\Theta^H+\mu_{\overline{\theta}}^H U_{\overline{\theta}}^H\\&+\psi^L\left[\int_{\underline{\theta}}^{\Theta}\theta l_\theta^L\mathrm{d}F(\theta)-L\right]+\psi^H\left[\int_{\Theta}^{\overline{\theta}}\theta l_\theta^H\mathrm{d}F(\theta)-H\right]\\&+\int_{\underline{\theta}}^{\Theta}\lambda_\theta^L\left[c_\theta^L-\frac{(l_\theta^L)^{1+1/\varepsilon}}{1+1/\varepsilon}-U_\theta^L\right]\mathrm{d}F(\theta)+\int_{\Theta}^{\overline{\theta}}\lambda_\theta^H\left[c_\theta^H-\frac{(l_\theta^H)^{1+1/\varepsilon}}{1+1/\varepsilon}-U_\theta^H\right]\mathrm{d}F(\theta),\end{aligned} \tag{35}$$

where π is the Lagrange multiplier on the condition that utilities should be equal for the marginal graduate Θ. If $\pi\neq 0$ skill formation is distorted. η is the Lagrange multiplier on the economy's resource constraint. μ_θ^L (μ_θ^H) is the co-state variable associated with the incentive-compatibility constraint on U_θ^L (U_θ^H). Note that we harmlessly multiplied the ICC's with a minus sign to ensure that the multipliers μ_θ^L and μ_θ^H are positive. ψ^L (ψ^H) is the Lagrange multiplier on the market-clearing condition for low-skilled (high-skilled) labor. λ_θ^L (λ_θ^H) is the Lagrange multiplier on the definition of low-skilled (high-skilled) utility.[25]

A.1 Aggregate production

Take the first-order conditions (FOCs) of the Lagrangian in (35) with respect to L and H, and use the firm's first-order conditions in (7) and (8), to derive the optimality conditions for production:

$$\frac{\partial \mathcal{L}}{\partial L} = \eta w^L - \psi^L = 0 \Leftrightarrow w^L = \psi^L/\eta, \tag{36}$$

$$\frac{\partial \mathcal{L}}{\partial H} = \eta w^H - \psi^H = 0 \Leftrightarrow w^H = \psi^H/\eta. \tag{37}$$

From these equations follows that the shadow value of the labor-market clearing condition – in monetary terms – just equals the wage rate for each education type.

A.2 Expressions for multipliers

We derive results for the Lagrange multipliers in the following Lemmas.

Lemma 3. *The co-state variables associated with the incentive-compatibility constraints are zero at the bounds:* $\mu^L_{\underline{\theta}} = \mu^H_{\overline{\theta}} = 0$.

Proof. This follows from the taking first-order conditions for $U^L_{\underline{\theta}}$ and $U^H_{\overline{\theta}}$ in the Lagrangian (35):

$$\frac{\partial \mathcal{L}}{\partial U^L_{\underline{\theta}}} = -\mu^L_{\underline{\theta}} = 0, \tag{38}$$

$$\frac{\partial \mathcal{L}}{\partial U^H_{\overline{\theta}}} = \mu^H_{\overline{\theta}} = 0. \tag{39}$$

□

Lemma 4. *The multipliers on the definition of utility are equal to the multiplier on the resource constraint:* $\lambda^L_\theta = \lambda^H_\theta = \eta$.

Proof. This follows from the first-order conditions for consumption c^L_θ and c^H_θ in the Lagrangian (35):

$$\frac{\partial \mathcal{L}}{\partial c^L_\theta} = (-\eta + \lambda^L_\theta) f(\theta) = 0, \quad \forall \theta \in [\underline{\theta}, \Theta), \tag{40}$$

$$\frac{\partial \mathcal{L}}{\partial c^H_\theta} = (-\eta + \lambda^H_\theta) f(\theta) = 0, \quad \forall \theta \in [\Theta, \overline{\theta}]. \tag{41}$$

□

[25] We deviate from the standard Mirrlees (1971) model by maximizing over the entire allocation subject to utility constraints. In the Mirrlees model, the utility function is typically inverted to write consumption as a function of the allocation. For notational simplicity we also take the market-clearing conditions for low- and high-skilled labor as separate constraints in the problem, rather than substituting the market-clearing conditions directly in the resource constraint.

Lemma 5.

- *The co-state variables in – monetary terms – for low- and high-skilled are given by*

$$\frac{\mu^L_\theta}{\eta} = \int_\theta^{\overline{\theta}} (1 - g_\vartheta)\, f(\vartheta)\mathrm{d}\vartheta \ \forall\theta \in [\underline{\theta}, \Theta) \tag{42}$$

$$\frac{\mu^H_\theta}{\eta} = \int_\theta^{\overline{\theta}} (1 - g_\vartheta)\, f(\vartheta)\mathrm{d}\vartheta \ \forall\theta \in (\Theta, \overline{\theta}], \tag{43}$$

where the marginal social welfare weight g_θ is defined as

$$g_\theta \equiv \begin{cases} \frac{\Psi'(U^L_\theta)}{\eta}, & \theta < \Theta, \\ \frac{\Psi'(U^H_\theta)}{\eta}, & \theta \geq \Theta. \end{cases} \tag{44}$$

- *The co-state variables – in monetary terms – at the skill cutoff Θ are given by*

$$\frac{\mu^L_\Theta}{\eta} = \frac{\mu^H_\Theta}{\eta} = \frac{\pi}{\eta} = \zeta \equiv \int_\Theta^{\overline{\theta}} (1 - g_\theta)\, f\,(\theta)\,\mathrm{d}\theta, \tag{45}$$

where ζ is the social marginal value of raising resources from the high-skilled.

- *The average welfare weight equals one*

$$\int_{\underline{\theta}}^{\overline{\theta}} g_\vartheta f(\vartheta)\mathrm{d}\vartheta = 1. \tag{46}$$

Proof. To derive an expression for $\frac{\mu^H_\theta}{\eta}$ take the first-order condition of the Lagrangian (35) with respect to U^H_θ:

$$\frac{\partial\mathcal{L}}{\partial U^H_\theta} = (\Psi'(U^H_\theta) - \lambda^H_\theta)f(\theta) - \frac{\mathrm{d}\mu^H_\theta}{\mathrm{d}\theta} = 0, \quad \forall\theta \in [\Theta, \overline{\theta}) \tag{47}$$

Substitute $\eta = \lambda^H_\theta$ from Lemma 4 and integrate over $[\theta, \overline{\theta}]$, and use $\mu^H_{\overline{\theta}} = 0$ from Lemma 3 to find:

$$\begin{aligned} \int_\theta^{\overline{\theta}} \frac{d\mu^H_\vartheta}{\mathrm{d}\vartheta}\mathrm{d}\vartheta =& \mu^H_{\overline{\theta}} - \mu^H_\theta = \int_\theta^{\overline{\theta}} (\Psi'(U^H_\vartheta) - \eta)f(\vartheta)\mathrm{d}\vartheta \\ \Leftrightarrow \frac{\mu^H_\theta}{\eta} =& \int_\theta^{\overline{\theta}} (1 - g_\vartheta)f(\vartheta)\mathrm{d}\vartheta. \end{aligned} \tag{48}$$

As a corollary, we derive

$$\frac{\mu^H_\Theta}{\eta} = \zeta \equiv \int_\Theta^{\overline{\theta}} (1 - g_\vartheta)f(\vartheta)\mathrm{d}\vartheta. \tag{49}$$

To derive an expression for $\frac{\mu^L_\theta}{\eta}$ take the first-order condition of the Lagrangian (35) with respect to U^L_θ:

$$\frac{\partial\mathcal{L}}{\partial U^L_\theta} = (\Psi'(U^L_\theta) - \lambda^L_\theta)f(\theta) - \frac{\mathrm{d}\mu^L_\theta}{\mathrm{d}\theta} = 0, \quad \forall\theta \in (\underline{\theta}, \Theta). \tag{50}$$

Substitute $\eta = \lambda_\theta^L$ from Lemma 4, integrate over $[\underline{\theta}, \theta)$, and use $\mu_{\underline{\theta}}^L = 0$ from Lemma 3 to find:

$$\begin{aligned} \int_{\underline{\theta}}^{\theta} \frac{d\mu_\vartheta^L}{\mathrm{d}\vartheta}\mathrm{d}\vartheta =& \mu_\theta^L - \mu_{\underline{\theta}}^L = \int_{\underline{\theta}}^{\theta} (\Psi'(U_\vartheta^L) - \eta) f(\vartheta)\mathrm{d}\vartheta \\ \Leftrightarrow \frac{\mu_\theta^L}{\eta} =& \int_{\underline{\theta}}^{\theta} (g_\vartheta - 1) f(\vartheta)\mathrm{d}\vartheta, \end{aligned} \tag{51}$$

As a corollary, we derive:

$$\frac{\mu_\Theta^L}{\eta} = \int_{\underline{\theta}}^{\Theta} (g_\vartheta - 1) f(\vartheta)\mathrm{d}\vartheta. \tag{52}$$

To derive expressions for the multipliers at the cut-offs μ_Θ^L and μ_Θ^H, take the first-order conditions of the Lagrangian (35) with respect to U_Θ^L and U_Θ^H to find:

$$\frac{\partial \mathcal{L}}{\partial U_\Theta^L} = -\pi + \mu_\Theta^L = 0, \tag{53}$$

$$\frac{\partial \mathcal{L}}{\partial U_\Theta^H} = \pi - \mu_\Theta^H = 0. \tag{54}$$

Rewriting these first-order conditions gives:

$$\frac{\mu_\Theta^L}{\eta} = \frac{\mu_\Theta^H}{\eta} = \frac{\pi}{\eta} = \zeta. \tag{55}$$

Finally, we prove that the social welfare weights sum to one. From (49), (52), and (55) it follows that

$$\begin{aligned} \frac{\mu_\Theta^L}{\eta} - \frac{\mu_\Theta^H}{\eta} =& \int_{\underline{\theta}}^{\Theta} (g_\vartheta - 1) f(\vartheta)\mathrm{d}\vartheta - \int_{\Theta}^{\overline{\theta}} (1 - g_\vartheta) f(\vartheta)\mathrm{d}\vartheta = 0, \\ \Leftrightarrow & \int_{\underline{\theta}}^{\overline{\theta}} g_\vartheta f(\vartheta)\mathrm{d}\vartheta = 1. \end{aligned} \tag{56}$$

The last result implies that the marginal cost of public funds is one at the optimal tax system, see also Jacobs (2018).

The last result allows us to rewrite (51) as:

$$\frac{\mu_\theta^L}{\eta} = \int_{\underline{\theta}}^{\theta} (g_\vartheta - 1) f(\vartheta)\mathrm{d}\vartheta = \int_{\theta}^{\overline{\theta}} (1 - g_\vartheta) f(\vartheta)\mathrm{d}\vartheta. \tag{57}$$

□

A.3 Nonlinear tax low-skilled workers

The first-order conditions of the Lagrangian in (35) with respect to labor of the low-skilled workers l_θ^L are given by

$$\frac{\partial \mathcal{L}}{\partial l_\theta^L} = -\mu_\theta^L (1 + 1/\varepsilon) \frac{(l_\theta^L)^{1/\varepsilon}}{\theta} + \psi^L \theta f(\theta) - \lambda_\theta^L (l_\theta^L)^{1/\varepsilon} f(\theta) = 0, \quad \forall \theta \in [\underline{\theta}, \Theta), \tag{58}$$

We rewrite the first-order conditions for l_θ^L in (58) to find the nonlinear taxes for the low-skilled workers. Substitute the first-order condition for c_θ^L in (40), the shadow wage in (36), and the first-order condition for labor supply in (3) to derive:

$$\frac{T'^L(z_\theta^L)}{1-T'^L(z_\theta^L)} = (1+1/\varepsilon)\frac{\mu_\theta^L/\eta}{\theta f(\theta)}, \quad \forall\theta \in [\underline{\theta}, \Theta). \tag{59}$$

Substitute μ_θ^L/η from Lemma 5 to find:

$$\frac{T'^L(z_\theta^L)}{1-T'^L(z_\theta^L)} = (1+1/\varepsilon)\frac{\int_\theta^{\overline{\theta}}(1-g_\theta)f(\theta)\mathrm{d}\theta}{\theta f(\theta)}, \quad \forall\theta \in [\underline{\theta}, \Theta). \tag{60}$$

The right-hand side is independent of education type, hence we can write $T'^L(z_\theta^L) = T'(z_\theta)$. Finally, we use μ_Θ^L/η from Lemma 5 to find the marginal tax rate at the education cutoff Θ:

$$\frac{T'(z_\Theta^L)}{1-T'(z_\Theta^L)} = (1+1/\varepsilon)\frac{\zeta}{\Theta f(\Theta)}. \tag{61}$$

A.4 nonlinear tax high-skilled workers

The first-order conditions of the Lagrangian in (35) with respect to labor of the high-skilled workers l_θ^H are given by:

$$\frac{\partial\mathcal{L}}{\partial l_\theta^H} = -\mu_\theta^H(1+1/\varepsilon)\frac{(l_\theta^H)^{1/\varepsilon}}{\theta} + \psi^H\theta f(\theta) - \lambda_\theta^H(l_\theta^H)^{1/\varepsilon}f(\theta) = 0, \quad \forall\theta \in [\Theta, \overline{\theta}], \tag{62}$$

We rewrite the first-order conditions for l_θ^H in (62) to find the nonlinear taxes for the low-skilled workers. Substitute the first-order condition for c_θ^H in (41), the shadow wage in (37), and the first-order condition for labor supply in (3) to derive:

$$\frac{T'^H(z_\theta^H)}{1-T'^H(z_\theta^H)} = (1+1/\varepsilon)\frac{\mu_\theta^H/\eta}{\theta f(\theta)}, \quad \forall\theta \in [\Theta, \overline{\theta}]. \tag{63}$$

Substitute μ_θ^H/η from Lemma (5) to find

$$\frac{T'^H(z_\theta^H)}{1-T'^H(z_\theta^H)} = (1+1/\varepsilon)\frac{\int_\theta^{\overline{\theta}}(1-g_\theta)f(\theta)\mathrm{d}\theta}{\theta f(\theta)}, \quad \forall\theta \in [\Theta, \overline{\theta}]. \tag{64}$$

The right-hand side is independent of education type, hence we can write $T'^H(z_\theta^H) = T'(z_\theta)$. Finally, we use μ_Θ^H/η from Lemma (5) to find the marginal tax rate at the education cutoff Θ:

$$\frac{T'(z_\Theta^H)}{1-T'(z_\Theta^H)} = (1+1/\varepsilon)\frac{\zeta}{\Theta f(\Theta)}. \tag{65}$$

A.5 Optimal tax schedule

If we combine (60) and (64), we find the standard Mirrlees expression for the entire domain of θ. Thus the optimal nonlinear tax on labor income is given by:

$$\frac{T'(z_\theta)}{1-T'(z_\theta)} = (1+1/\varepsilon)\frac{\int_\theta^{\overline{\theta}}(1-g_\theta)f(\theta)\mathrm{d}\theta}{\theta f(\theta)}, \quad \forall\theta\in[\underline{\theta},\overline{\theta}]. \tag{66}$$

A.6 Optimal education distortions

Finally, we derive the optimal subsidy on education. The first-order condition of the Lagrangian in (35) with respect to Θ is given by

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\Theta} =&\Psi(U_\Theta^L)f(\Theta)-\Psi(U_\Theta^H)f(\Theta)+\eta\left[-c_\Theta^L f(\Theta)+\left(c_\Theta^H+p\right)f(\Theta)\right]\\ &-\left[\mu_\Theta^L\frac{\left(l_\Theta^L\right)^{1+1/\varepsilon}}{\Theta}+U_\Theta^L\frac{\mathrm{d}\mu_\Theta^L}{\mathrm{d}\theta}\right]+\left[\mu_\Theta^H\frac{\left(l_\Theta^H\right)^{1+1/\varepsilon}}{\Theta}+U_\Theta^H\frac{\mathrm{d}\mu_\Theta^H}{\mathrm{d}\theta}\right]\\ &+\lambda_\Theta^L\left[c_\Theta^L-\frac{(l_\Theta^L)^{1+1/\varepsilon}}{1+1/\varepsilon}-U_\Theta^L\right]f(\Theta)-\lambda_\Theta^H\left[c_\Theta^H-\frac{(l_\Theta^H)^{1+1/\varepsilon}}{1+1/\varepsilon}-U_\Theta^H\right]f(\Theta)\\ &+\psi^L\Theta l_\Theta^L f(\Theta)-\psi^H\Theta l_\Theta^H f(\Theta)=0.\end{aligned} \tag{67}$$

We rewrite this expression by using the property that utilities are equalized ($U_\Theta^L=U_\Theta^H$) at the threshold Θ, which implies that the first term in the first line cancels. We employ the definition of utility to cancel the entire third line. Finally, all terms involving $\frac{d\mu_\Theta^L}{\mathrm{d}\theta}=\frac{d\mu_\Theta^H}{\mathrm{d}\theta}=0$ can be canceled. Divide by η so as to arrive at:

$$\left[(c_\Theta^H+p)-c_\Theta^L\right]f(\Theta)-\frac{\mu_\Theta^L}{\eta}\frac{(l_\Theta^L)^{1+1/\varepsilon}}{\Theta}+\frac{\mu_\Theta^H}{\eta}\frac{(l_\Theta^H)^{1+1/\varepsilon}}{\Theta}+\left[w^L\Theta l_\Theta^L-w^H\Theta l_\Theta^H\right]f(\Theta)=0. \tag{68}$$

where we used (36) and (37). Rearrange by using Lemma (5) $\frac{\mu_\Theta^L}{\eta}=\frac{\mu_\Theta^H}{\eta}=\zeta$:

$$w^H\Theta l_\Theta^H-\left(c_\Theta^H+p\right)-w^L\Theta l_\Theta^L+c_\Theta^L=\frac{\zeta}{f(\Theta)\Theta}\left(\left(l_\Theta^H\right)^{1+1/\varepsilon}-\left(l_\Theta^L\right)^{1+1/\varepsilon}\right). \tag{69}$$

From the household budget constraint in (2) follows that $z_\Theta^L-c_\Theta^L=T^L(z_\Theta^L)$ and $z_\Theta^H-(c_\Theta^H+p)=T^H(z_\Theta^H)$. The optimal net tax on education can thus be written as

$$T^H(z_\Theta^H)-T^L(z_\Theta^L)=\frac{\zeta}{\Theta f \Theta)}\left((l_\Theta^H)^{1+1/\varepsilon}-(l_\Theta^L)^{1+1/\varepsilon}\right). \tag{70}$$

B Proof Lemma 1

To derive the comparative statics of the model with respect to skill-biased technical change, we linearize the following system:

$$w^H = AY_H(L,AH), \tag{71}$$

$$w^L = Y_L(L,AH), \tag{72}$$

$$H = \int_{\Theta}^{\overline{\theta}} \theta l_\theta^H \mathrm{d}F(\theta), \tag{73}$$

$$L = \int_{\underline{\theta}}^{\Theta} \theta l_\theta^L \mathrm{d}F(\theta), \tag{74}$$

$$l_\theta^j = [w^j \theta(1 - T'^j(z_\theta^j))]^\varepsilon, \quad \forall \theta, j, \tag{75}$$

$$z_\theta^j \equiv w^j \theta l_\theta^j, \quad \forall \theta, j, \tag{76}$$

$$z_\Theta^H - T^H(z_\Theta^H) - \frac{[w^H \Theta(1 - T'^H(z_\Theta^H))]^{1+\varepsilon}}{1 + 1/\varepsilon} - p = z_\Theta^L - T^L(z_\Theta^L) - \frac{[w^L \Theta(1 - T'^L(z_\theta^L))]^{1+\varepsilon}}{1 + 1/\varepsilon}. \tag{77}$$

In what follows we will use a tilde to denote a relative change, i.e., $\tilde{x} \equiv \mathrm{d}x/x$ for any variable x, except for $\tilde{\Theta} \equiv \mathrm{d}\Theta$. First, we can linearize the first-order conditions for labor demand in (71) and (72) to find:

$$\tilde{w}^H = \frac{\sigma - (1 - \alpha)}{\sigma} \tilde{A} - \frac{1 - \alpha}{\sigma} (\tilde{H} - \tilde{L}), \tag{78}$$

$$\tilde{w}^L = \frac{\alpha}{\sigma} (\tilde{A} + \tilde{H} - \tilde{L}), \tag{79}$$

where $\sigma \equiv \frac{Y_H Y_L}{Y_{HL} Y}$ is the elasticity of substitution between high-skilled and low-skilled labor, $\alpha \equiv \frac{AHY_H}{Y}$ is the high-skilled income share. Moreover, in deriving both expressions we used $Y_{LL}L = -Y_{LH}AH$ and $Y_{HH}AH = -Y_{LH}L$, since the labor demand functions are homogeneous of degree zero. Hence, the change in relative wages can be written as:

$$\tilde{w}^H - \tilde{w}^L = \frac{\sigma - 1}{\sigma} \tilde{A} - \frac{1}{\sigma} (\tilde{H} - \tilde{L}). \tag{80}$$

Second, we linearize the labor market-clearing conditions in (73) and (74) to find:

$$\tilde{H} = \int_{\Theta}^{\overline{\theta}} \frac{\theta l_\theta^H}{H} \tilde{l}_\theta^H \mathrm{d}F(\theta) - \delta^H \tilde{\Theta}, \tag{81}$$

$$\tilde{L} = \int_{\underline{\theta}}^{\Theta} \frac{\theta l_\theta^L}{L} \tilde{l}_\theta^L \mathrm{d}F(\theta) + \delta^L \tilde{\Theta}, \tag{82}$$

where $\delta^H \equiv \frac{\Theta f(\Theta) l_\Theta^H}{H}$ and $\delta^L \equiv \frac{\Theta f(\Theta) l_\Theta^L}{L}$ measure the relative mass of labor at the education cutoff Θ.

Third, we linearize the labor supply equations in (75), while taking into account that the marginal tax rate might change due to the nonlinearity of the tax schedule:

$$\tilde{l}_\theta^j = \left(\frac{1 - T'^j - z_\theta^j T''^j}{1 - T'^j + \varepsilon z_\theta^j T''^j} \right) \varepsilon \tilde{w}^j \equiv \varepsilon_\theta \tilde{w}^j. \tag{83}$$

where $\varepsilon_\theta \equiv \left(\frac{1 - T'^j - z_\theta^j T''^j}{1 - T'^j + \varepsilon z_\theta^j T''^j} \right) \varepsilon$ is the wage elasticity of labor supply. Note that if the tax schedule is linear, so that $T''^j = 0$, the wage elasticity of labor supply corresponds to ε. Linearizing earnings supply in (76), and substituting (83), we can derive the relative change in earnings supply as a function of the change in the wage rate:

$$\tilde{z}^j_\theta = \tilde{w}^j + \tilde{l}^j_\theta = (1+\varepsilon_\theta)\tilde{w}^j \tag{84}$$

Fourth, linearize (77) to find an expression for the relative change in the education cutoff Θ:

$$\begin{aligned}&(1-T'^H)z^H_\Theta \tilde{z}^H_\Theta - \varepsilon[w^H\Theta(1-T'^H)]^{1+\varepsilon}\left(\tilde{w}^H + \Theta\tilde{\Theta} - \frac{T''^H z^H_\Theta}{(1-T'^H)}\tilde{z}^H_\Theta\right)\\ &= (1-T'^L)z^L_\Theta \tilde{z}^L_\Theta - \varepsilon[w^L\Theta(1-T'^L)]^{1+\varepsilon}\left(\tilde{w}^L + \Theta\tilde{\Theta} - \frac{T''^L z^L_\Theta}{(1-T'^L)}\tilde{z}^L_\Theta\right).\end{aligned} \tag{85}$$

Use $z^j_\theta \equiv w^j\theta l^j_\theta$ and $l^j_\theta = [w^j\theta(1-T'^j(z^j_\theta))]^\varepsilon$ and rearrange to find:

$$\mu\left[\left(1+\varepsilon\frac{T''^H z^H_\Theta}{(1-T'^H)}\right)(1+\varepsilon_\Theta) - \varepsilon\right]\tilde{w}^H = \left[\left(1+\varepsilon\frac{T''^L z^L_\Theta}{(1-T'^L)}\right)(1+\varepsilon_\Theta) - \varepsilon\right]\tilde{w}^L - (1-\mu)\varepsilon\Theta\tilde{\Theta}. \tag{86}$$

where $\mu \equiv \frac{(1-T'^H)z^H_\Theta}{(1-T'^L)z^L_\Theta}$. Finally, define $\mu^H \equiv \frac{\mu}{(1-\mu)\varepsilon\Theta}\left[\left(1+\varepsilon\frac{T''^H z^H_\Theta}{(1-T'^H)}\right)(1+\varepsilon_\Theta)-\varepsilon\right] > 0$ and $\mu^L \equiv \frac{1}{(1-\mu)\varepsilon\Theta}\left[\left(1+\varepsilon\frac{T''^L z^L_\Theta}{(1-T'^L)}\right)(1+\varepsilon_\Theta)-\varepsilon\right] > 0$, which we signed under the assumption that the second-derivative of the tax function is always of second-order importance. Substitute the definitions for μ^L and μ^H to find:

$$\tilde{\Theta} = \mu^L\tilde{w}^L - \mu^H\tilde{w}^H. \tag{87}$$

Next we solve the system of linearized equations in (78), (79), (81), (82), (83), (84) and (87) for the changes in the wage rates with respect to skill-biased technical change.

Substitute (87) in (81) to obtain

$$\tilde{H} = \int_\Theta^{\bar{\theta}} \frac{\theta l^H_\theta}{H}\varepsilon_\theta \mathrm{d}F(\theta)\tilde{w}^H - \delta^H(\mu^L\tilde{w}^L - \mu^H\tilde{w}^H) = (\bar{\varepsilon}^H + \delta^H\mu^H)\tilde{w}^H - \delta^H\mu^L\tilde{w}^L, \tag{88}$$

where $\bar{\varepsilon}^H \equiv \int_\Theta^{\bar{\theta}} \frac{\theta l^H_\theta}{H}\varepsilon_\theta \mathrm{d}F(\theta)$ is the employment-weighted elasticity of high-skilled labor supply. Similarly, substitute (87) in (82) to obtain:

$$\tilde{L} = \int_{\underline{\theta}}^{\Theta} \frac{\theta l^L_\theta}{L}\varepsilon_\theta \mathrm{d}F(\theta)\tilde{w}^L + \delta^L(\mu^L\tilde{w}^L - \mu^H\tilde{w}^H) = (\bar{\varepsilon}^L + \delta^L\mu^L)\tilde{w}^L - \delta^L\mu^H\tilde{w}^H, \tag{89}$$

where $\bar{\varepsilon}^L \equiv \int_{\underline{\theta}}^{\Theta} \frac{\theta l^L_\theta}{L}\varepsilon_\theta \mathrm{d}F(\theta)$ is the employment-weighted elasticity of low-skilled labor supply.

Subtracting (89) from (88) gives:

$$\tilde{H} - \tilde{L} = (\bar{\varepsilon}^H + (\delta^L+\delta^H)\mu^H)\tilde{w}^H - (\bar{\varepsilon}^L + (\delta^L+\delta^H)\mu^L)\tilde{w}^L. \tag{90}$$

Substitute (90) in (78) and (79) to find:

$$\left[1 + \frac{1-\alpha}{\sigma}(\bar{\varepsilon}^H + (\delta^L+\delta^H)\mu^H)\right]\tilde{w}^H = \frac{\sigma-1+\alpha}{\sigma}\tilde{A} + \frac{1-\alpha}{\sigma}(\bar{\varepsilon}^L + (\delta^L+\delta^H)\mu^L)\tilde{w}^L, \tag{91}$$

$$\left[1 + \frac{\alpha}{\sigma}(\bar{\varepsilon}^L + (\delta^L+\delta^H)\mu^L)\right]\tilde{w}^L = \frac{\alpha}{\sigma}\tilde{A} + \frac{\alpha}{\sigma}(\bar{\varepsilon}^H + (\delta^L+\delta^H)\mu^H)\tilde{w}^H. \tag{92}$$

Rearranging (91) yields:

$$\tilde{w}^L = \frac{\left[\frac{\sigma}{1-\alpha} + (\bar{\varepsilon}^H + (\delta^L + \delta^H)\mu^H)\right]}{(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}\tilde{w}^H - \frac{\frac{\sigma}{1-\alpha} - 1}{(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}\tilde{A}. \tag{93}$$

Substituting (93) in (92) and rearranging gives the change in the high-skilled wage as a function of skill-biased technical change:

$$\tilde{w}^H = \left[\frac{\sigma - (1-\alpha) + \alpha(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}{\sigma + (1-\alpha)(\bar{\varepsilon}^H + (\delta^L + \delta^H)\mu^H) + \alpha(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}\right]\tilde{A} > 0, \tag{94}$$

where we assumed that $\sigma > 1$ to sign this equation.

Finally, substitute (94) in (78), to find:

$$\tilde{A} + \tilde{H} - \tilde{L} = \frac{\sigma}{1-\alpha}\left[1 - \left[\frac{\sigma - (1-\alpha) + \alpha(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}{\sigma + (1-\alpha)(\bar{\varepsilon}^H + (\delta^L + \delta^H)\mu^H) + \alpha(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}\right]\right]\tilde{A}. \tag{95}$$

Next, substitute (95) in (79), and rearrange to find the change in the low-skilled wage as a function of skill-biased technical change:

$$\tilde{w}^L = \left[\frac{\alpha + \alpha(\bar{\varepsilon}^H + (\delta^L + \delta^H)\mu^H)}{\sigma + (1-\alpha)(\bar{\varepsilon}^H + (\delta^L + \delta^H)\mu^H) + \alpha(\bar{\varepsilon}^L + (\delta^L + \delta^H)\mu^L)}\right]\tilde{A} > 0. \tag{96}$$

Previously, we derived that $\tilde{H}$, $\tilde{L}$, $\tilde{l}^j_\theta$, $\tilde{z}^j_\theta$, and $\tilde{\Theta}$, are all functions of the relative changes in the wages, cf. (88), (89), (83), (84) and (87). Hence, we have fully derived the comparative statics of the model with respect to skill-biased technical change.

C Proof Lemma 2

The change in the social welfare weight is given by linearizing its definition $g_\theta \equiv \frac{\Psi'(U^j_\theta)}{\eta}$:

$$\tilde{g}_\theta = -\tilde{\eta} - \rho_\theta \tilde{U}^j_\theta, \quad \rho_\theta \equiv -\frac{\Psi'' U^j_\theta}{\Psi'} > 0. \tag{97}$$

If $\Psi(U^j_\theta)$ has a constant elasticity of inequality aversion, then ρ is the elasticity of inequality aversion. By Lemma 5, the social welfare weights integrate to 1 in the tax optimum:

$$\int_{\underline{\theta}}^{\bar{\theta}} g_{\theta'} \mathrm{d}F(\theta') = 1. \tag{98}$$

Linearizing yields:

$$\int_{\underline{\theta}}^{\bar{\theta}} g_{\theta'}\tilde{g}_{\theta'} \mathrm{d}F(\theta') = 0. \tag{99}$$

Substituting the relative change in the social welfare weights gives:

$$\int_{\underline{\theta}}^{\bar{\theta}} \rho_{\theta'} g_{\theta'} \tilde{U}^j_{\theta'} \mathrm{d}F(\theta) = -\tilde{\eta}. \tag{100}$$

Consequently we find:

$$\tilde{g}_\theta = -\rho_\theta \tilde{U}^j_\theta + \int_{\underline{\theta}}^{\bar{\theta}} \rho_{\theta'} g_{\theta'} \tilde{U}^j_{\theta'} \mathrm{d}F(\theta) \tag{101}$$

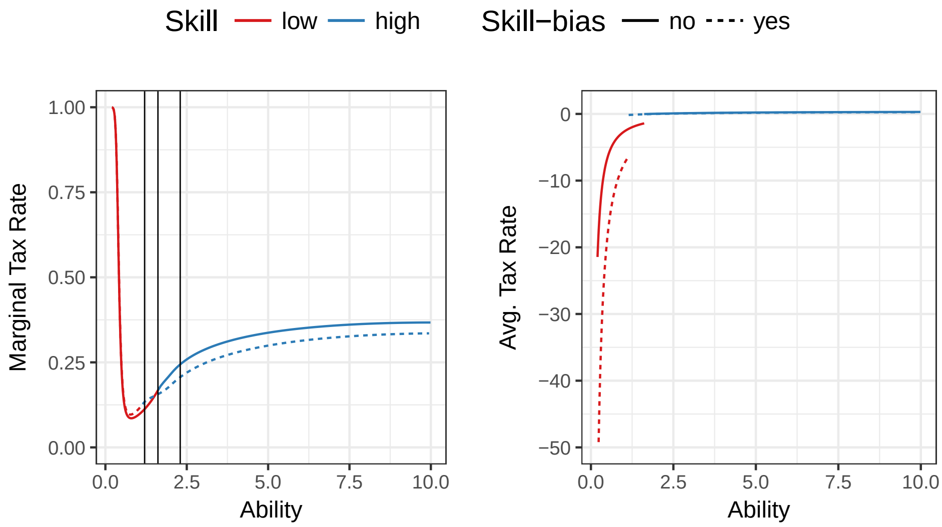
D Robustness

Table 4: Robustness: Summary statistics based on simulations

Skill bias	Share of high-skilled	Skill premium	Net tax at Θ
	$\rho = 0.1$		
No	0.50	26.00	10.08
Yes	0.79	36.77	9.95
	$\rho = 1$		
No	0.43	26.58	20.88
Yes	0.78	36.74	5.64
	$\varepsilon = 0.1$		
No	0.51	30.20	4.36
Yes	0.80	42.86	2.67
	$\varepsilon = 0.5$		
No	0.45	23.13	10.92
Yes	0.79	32.43	4.10
	$\sigma = 10$		
No	0.45	27.91	17.17
Yes	0.79	38.36	6.54

Note: The net tax at Θ is defined as $w^H_\Theta \Theta l^H_\Theta - c^H_\Theta - p - (w^L_\Theta \Theta l^L_\Theta - c^L_\Theta)$.

D.1 Elasticity of inequality aversion

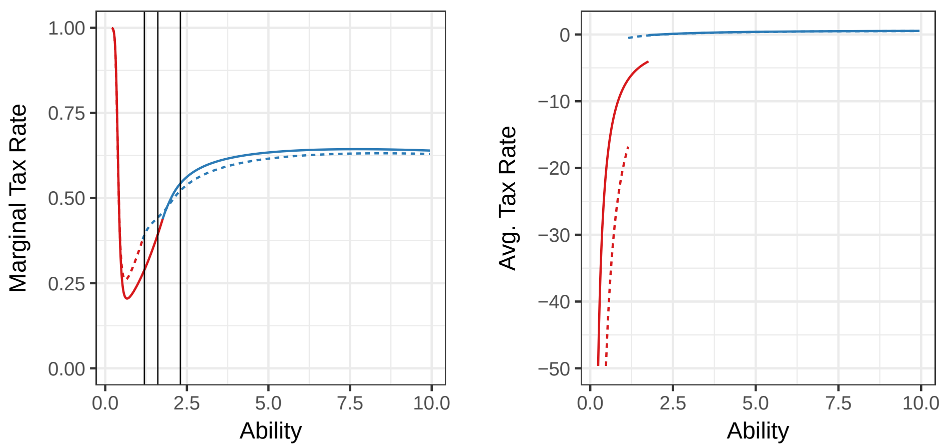


(a) Marginal Tax Rate

(b) Average Tax Rate

Figure 4: Effect of skill bias on optimal taxes: $\rho = 0.1$

Note: Robustness check with parameters $\rho = 0_1$, $\sigma = 10000$, $\varepsilon = 0.3$.



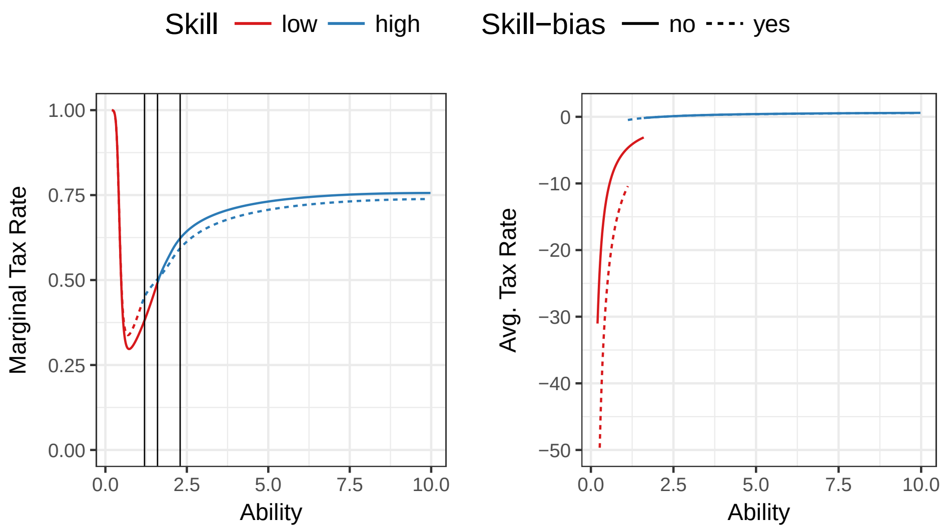
(a) Marginal Tax Rate

(b) Average Tax Rate

Figure 5: Effect of skill bias on optimal taxes: $\rho = 1$

Note: Robustness check with parameters $\rho = 1$, $\sigma = 10000$, $\varepsilon = 0.3$.

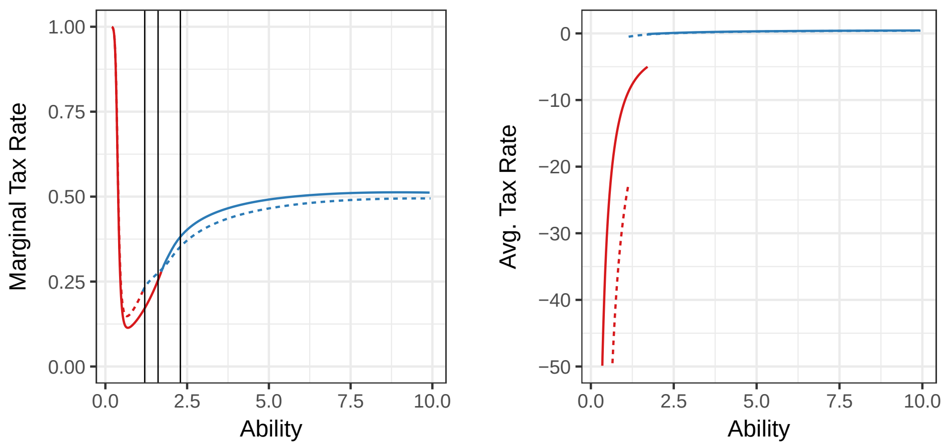
D.2 Labor-supply elasticities



(a) Marginal Tax Rate (b) Average Tax Rate

Figure 6: Effect of skill bias on optimal taxes: $\varepsilon = 0.1$

Note: Robustness check with parameters $\rho = 0.5$, $\sigma = 10000$, $\varepsilon = 0.1$.

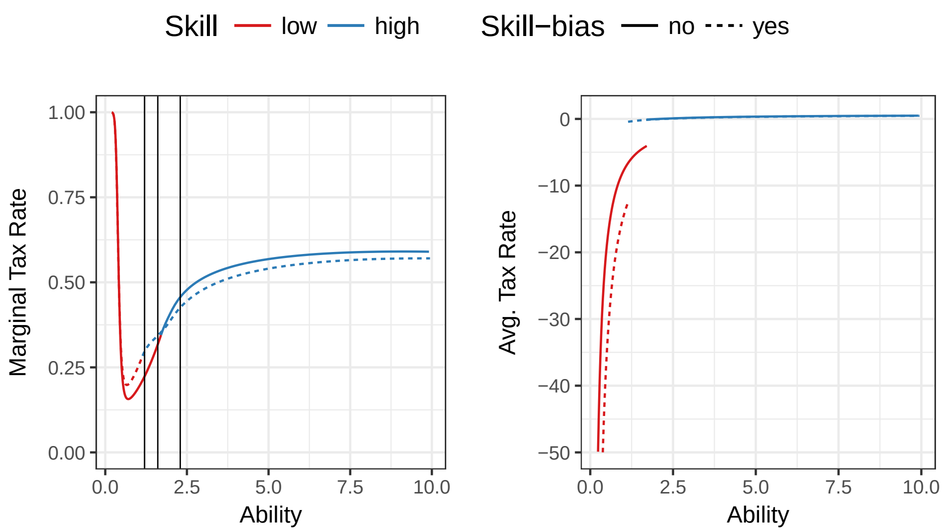


(a) Marginal Tax Rate (b) Average Tax Rate

Figure 7: Effect of skill bias on optimal taxes: $\varepsilon = 0.5$

Note: Robustness check with parameters $\rho = 0.5$, $\sigma = 10000$, $\varepsilon = 0.5$.

D.3 Substitution elasticity between high- and low-skilled labor



(a) Marginal Tax Rate (b) Average Tax Rate

Figure 8: Effect of skill bias on optimal taxes: $\sigma = 10$

Note: Robustness check with parameters $\rho = 0.5$, $\sigma = 10$, $\varepsilon = 0.3$.

E Data Appendix

We obtain data on wages and educational attainment from the Current Population Survey (CPS) Merged Outgoing Rotation Groups (MORG) as prepared by the National Bureau of Economic Research (NBER).[26] The data cover the years from 1979 to 2016. We focus on the period 1980 to 2016.

Selection of the sample follows Acemoglu and Autor (2011). We include individuals aged 16 to 64 whose usual weekly hours worked exceed 35. Hourly wages are obtained by dividing weakly earnings by usual hours worked. All wages are converted into 2016 dollar values using the personal consumption expenditures chain-type price index.[27] The highest earnings in the CPS are top-coded. We therefore windsorize earnings by multiplying top-coded earnings by 1.5. Like Acemoglu and Autor (2011), we exclude those individuals who earn less than 50% of the 1982 minimum wage ($3.35) converted to 2016-dollars. Self-employed individuals are excluded, as are individuals whose occupation does not have an `occ1990dd` classification. Observations are weighted by CPS sample weights. Education levels are coded based on the highest grade attended (before 1992) and the highest grade completed (after 1992).

[26]see `http://www.nber.org/data/morg.html`
[27]I obtain the price index from https://fred.stlouisfed.org/series/DPCERG3A086NBEA

References

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

Ales, L., M. Kurnaz, and C. Sleet (2015): "Technical Change, Wage Inequality and Taxes," *American Economic Review*, 105, 3061–3101.

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2017): *World Inequality Report 2018*, Paris: The World Inequality Lab.

Anderberg, D. (2009): "Optimal Policy and the Risk Properties of Human Capital Reconsidered," *Journal of Public Economics*, 93, 1017–1026.

Atkinson, A. and J. Stiglitz (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6, 55–75.

Atkinson, A. B., T. Piketty, and E. Saez (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49, 3–71.

Autor, D. (2019): "How Long Has this Been Going on? A Discussion of "Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both?"," in *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, ed. by C. R. Hulten and V. A. Ramey, University of Chicago Press, chap. 9.

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

Blau, F. D. and L. M. Kahn (1996): "International Differences in Male Wage Inequality: Institutions versus Market Forces," *Journal of Political Economy*, 104, 791–837.

Blundell, R. and T. Macurdy (1999): "Labor Supply: A Review of Alternative Approaches," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter and D. Card, Elsevier, vol. Volume 3, Part A, chap. 27, 1559–1695.

Bohacek, R. and M. Kapicka (2008): "Optimal Human Capital Policies," *Journal of Monetary Economics*, 55, 1–16.

Bovenberg, A. L. and B. Jacobs (2005): "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 89, 2005–2035.

Corlett, W. J. and D. C. Hague (1953): "Complementarity and the Excess Burden of Taxation," *Review of Economic Studies*, 21–30.

Diamond, P. A. (1980): "Income Taxation with Fixed Hours of Work," *Journal of Public Economics*, 13, 101–110.

——— (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95.

DUR, R. AND C. N. TEULINGS (2004): "Are Education Subsidies an Efficient Redistributive Device?" in *Labor Market Institutions and Public Regulation*, ed. by J. Agell, M. Keen, and A. J. Weichenrieder, Cambridge, MA: The MIT Press, chap. 4, 123–161.

EBERT, U. (1992): "A Reexamination of the Optimal Nonlinear Income Tax," *Journal of Public Economics*, 49, 47–73.

FINDEISEN, S. AND D. SACHS (2017): "Optimal Need-Based Financial Aid," mimeo: University of Mannheim/LMU Munich, Mannheim/Munich.

FORTIN, N. M. AND T. LEMIEUX (1997): "Institutional Changes and Rising Wage Inequality: Is There a Linkage?" *Journal of Economic Perspectives*, 11, 75–96.

GOLDIN, C. AND L. F. KATZ (2010): *The Race between Education and Technology*, Cambridge-MA: Belknap Press of Harvard University Press.

GOMES, R., J. LOZACHMEUR, AND A. PAVAN (2018): "Differential Taxation and Occupational Choice," *Review of Economic Studies*, 85, 511–557.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): "Optimal Tax Progressivity: An Analytical Framework," Working Paper 19899, National Bureau of Economic Research.

HECKMAN, J. J., L. LOCHNER, AND C. TABER (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.

——— (1999): "Human Capital Formation and General Equilibrium Treatment Effects: A Study of Tax and Tuition Policy," *Fiscal Studies*, 20, 25–40.

HECKMAN, J. J., L. J. LOCHNER, AND P. E. TODD (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, ed. by E. Hanushek and F. Welch, Elsevier, vol. 1, chap. 7, 307–458.

HELLWIG, M. F. (2004): "Optimal Income Taxation, Public-Goods Provision and Public-Sector Pricing: A Contribution to the Foundations of Public Economics," Max Planck Institut Technical Report 2014/14, Bonn: Max Planck Institut.

HELPMAN, E. (2016): "Globalization and Wage Inequality," NBER Working Paper No. 22944, Cambridge-MA: National Bureau for Economic Research.

JACOBS, B. (2004): "The Lost Race Between Schooling and Technology," *De Economist*, 152, 47–78.

——— (2012): "Optimal Redistributive Tax and Education Policies in General Equilibrium," *International Tax and Public Finance*, 1–26.

——— (2018): "The Marginal Cost of Public Funds is One at the Optimal Tax System," *International Tax and Public Finance*, 25, 883–912.

Jacobs, B. and R. Boadway (2014): "Optimal Linear Commodity Taxation under Optimal Non-linear Income Taxation," *Journal of Public Economics*, 117, 201–210.

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

Jacobs, B. and U. Thuemmel (2018): "Optimal Linear Income Taxation and Education Subsidies with Skill-Biased Technical Change," mimeo: Erasmus University Rotterdam/University of Zürich, Rotterdam/Zürich.

Jacquet, L., E. Lehmann, and B. Van der Linden (2013): "Optimal Redistributive Taxation with Both Extensive and Intensive Responses," *Journal of Economic Theory*, 148, 1770–1805.

Katz, L. F. and D. H. Autor (1999): "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter and D. Card, Elsevier, vol. 3 of *Handbook of Labor Economics*, chap. 26, 1463 – 1555.

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

Krueger, D. and A. Ludwig (2016): "On the Optimal Provision of Social Insurance: Progressive Taxation Versus Education Subsidies in General Equilibrium," *Journal of Monetary Economics*, 77, 72–98.

Lee, D. S. (1999): "Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage?*," *Quarterly Journal of Economics*, 114, 977–1023.

Maldonado, D. (2008): "Education Policies and Optimal Taxation," *International Tax and Public Finance*, 15, 131–143.

Meghir, C. and D. Phillips (2010): "Labour supply and taxes," in *Dimensions of Tax Design: The Mirrlees Review*, ed. by J. A. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba, Oxford University Press, chap. 3, 202–274.

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

Obama, B. H. (2013): "Remarks by the President on Economic Mobility," The White House: Office of the Press Secretary, Washington, D.C.

Rao, A. V., D. A. Benson, C. Darby, M. A. Patterson, C. Francolin, I. Sanders, and G. T. Huntington (2010): "Algorithm 902: GPOPS, A MATLAB Software for Solving Multiple-phase Optimal Control Problems Using the Gauss Pseudospectral Method," *ACM Transactions on Mathematical Software*, 37, 22:1–22:39.

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

Sachs, D., A. Tsyvinski, and N. Werquin (2017): "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium," NBER Working Paper No.22646, Cambridge-MA: National Bureau for Economic Research.

Sadka, E. (1976): "On Income Distribution, Incentive Effects and Optimal Income Taxation," *Review of Economic Studies*, 43, 261–267.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229.

——— (2002): "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, 117, 1039–1073.

——— (2004): "Direct or Indirect Tax Instruments for Redistribution: Short-run Versus Long-run," *Journal of Public Economics*, 88, 503–518.

Sattinger, M. (1975): "Comparative Advantage and the Distributions of Earnings and Abilities," *Econometrica*, 43, 455–468.

Scheuer, F. (2014): "Entrepreneurial Taxation with Endogenous Entry," *American Economic Journal: Economic Policy*, 6, 126–63.

Scheuer, F. and I. Werning (2016): "Mirrlees meets Diamond-Mirrlees," NBER Working Paper No. 22076, Cambridge-MA: National Bureau of Economic Research.

Seade, J. K. (1977): "On the Shape of Optimal Tax Schedules," *Journal of Public Economics*, 7, 203–235.

Stantcheva, S. (2017): "Optimal Taxation and Human Capital Policies over the Life Cycle," *Journal of Political Economy*, 125, 1931–1990.

Stern, N. (1982): "Optimum Taxation with Errors in Administration," *Journal of Public Economics*, 17, 181–211.

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

Tinbergen, J. (1975): *Income Distribution: Analysis and Policies*, North-Holland: Elsevier.

Tuomala, M. (2010): "On Optimal Non-linear Income Taxation: Numerical Results Revisited," *International Tax and Public Finance*, 17, 259–270.

Valletta, R. G. (2019): "Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both?" in *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, ed. by C. R. Hulten and V. A. Ramey, University of Chicago Press, chap. 9.

Violante, G. L. (2008): "Skill-Biased Technical Change," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Basingstoke: Nature Publishing Group, 520–523, 2 ed.