Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium*

Dominik Sachs LMU Munich

Aleh Tsyvinski Yale University

Nicolas Werquin Toulouse School of Economics†

November 14, 2019

Abstract

We study the incidence of nonlinear labor income taxes in an economy with a continuum of endogenous wages. We derive in closed form the effects of reforming nonlinearly an arbitrary tax system, by showing that this problem can be formalized as an integral equation. Our tax incidence formulas are valid both when the underlying assignment of skills to tasks is fixed or endogenous. We show qualitatively and quantitatively that contrary to conventional wisdom, if the tax system is initially suboptimal and progressive, the general-equilibrium "trickle-down" forces may raise the benefits of increasing the marginal tax rates on high incomes. We finally derive a parsimonious characterization of optimal taxes.

*We thank Laurence Ales, Costas Arkolakis, Andy Atkeson, Alan Auerbach, Felix Bierbrauer, Carlos Da Costa, Cécile Gaubert, François Geerolf, Austan Goolsbee, Piero Gottardi, Nathan Hendren, James Hines, Bas Jacobs, Claus Kreiner, Tim Lee, Arash Nekoei, Michael Peters, Emmanuel Saez, Florian Scheuer, Hakan Selin, Chris Sleet, Stefanie Stantcheva, Philip Ushchev, Gianluca Violante, Ivan Werning, Sevin Yeltekin and Floris Zoutman for comments and discussions. Financial support from the ANR (Programme d'Investissement d'Avenir ANR-17-EURE-0010) is gratefully acknowledged.

†University of Toulouse Capitole, Toulouse, France.

Introduction

This paper connects two classical strands of the public finance literature: the study of tax incidence (Harberger (1962); Kotlikoff and Summers (1987); Fullerton and Metcalf (2002)) and that of optimal nonlinear income taxation in partial and general equilibrium (Mirrlees (1971); Stiglitz (1982); Diamond (1998); Saez (2001); Rothschild and Scheuer (2013)). The objective of the tax incidence analysis is to characterize the first-order effects of locally reforming a given, potentially suboptimal, tax system on the distribution of individual wages, labor supplies, and utilities, as well as on government revenue and social welfare. We derive closed-form analytical formulas for the incidence of any tax reform in a framework with a continuum of endogenous wages and arbitrarily nonlinear taxes. A characterization of optimal taxes in general equilibrium is then obtained by imposing that no tax reform has a positive impact on social welfare.

In our baseline environment, there is a continuum of skills that are imperfectly substitutable in production. Agents choose their labor supply optimally given their wage and the tax schedule. The wage, or marginal productivity, of each worker is decreasing in the aggregate labor of its own skill type, and increasing (resp., decreasing) in the aggregate labor of the skills that are complements (resp., substitutes) in production. We microfound the production function in an environment with a technology over a continuum of tasks to which skills are endogenously assigned, as in Costinot and Vogel (2010); Ales, Kurnaz, and Sleet (2015).

In the model with exogenous wages, the effects of a tax change on the labor effort of a given agent can be easily derived as a function of the elasticity of labor supply (Saez (2001)). The key difficulty in general equilibrium is that this initial response impacts the wage, and thus the labor effort, of every other individual. This further affects the wage distribution, which in turn influences labor supply decisions, and so on. Solving for the fixed point in the labor supply adjustment of each worker is the key step in the tax incidence analysis and the primary technical challenge of our paper. We show that this a priori complex problem can be mathematically formalized as an integral equation. The tools of the theory of integral equations allow us to derive an analytical solution to this problem for a general production function and arbitrary tax reforms. Furthermore, this solution has a clear economic interpretation and is expressed in terms of meaningful, and potentially empirically estimable, labor supply, labor demand, and cross-wage elasticities. It is then straightforward to derive the incidence of tax reforms on individual wages and utilities. Importantly, the elasticities we uncover in general equilibrium are sufficient statistics (see Chetty (2009)): conditional on these parameters, our incidence formulas are valid regardless of whether the underlying structure of the assignment of skills to tasks is fixed or endogenous.

Next, we analyze the aggregate effect of tax reforms on government revenue and social welfare. We derive a general formula that establishes how the deadweight loss of taxes is modified in general equilibrium. We show analytically that the government's revenue gain from reforming the tax schedule in the direction of higher progressivity is larger (the excess burden is smaller) than the conventional formula assuming exogenous wages would predict, if the marginal tax rates being perturbed are initially increasing with income.[1] This result, which is robust to various extensions of our baseline environment, means that accounting for the conventional "trickle-down" forces (Stiglitz (1982); Rothschild and Scheuer (2013)) makes raising top-income marginal tax rates *more*, not less, desirable than in partial equilibrium. Numerical simulations show that in the U.S., assuming exogenous wages implies that 33 percent of the revenue from a tax increase is lost through behavioral responses, while only 17 percent to 29 percent are lost in general equilibrium.

Finally, we derive the implications of our analysis for the optimal tax schedule. In the main body, we focus on deriving a novel characterization which depends on a parsimonious number of parameters that can be easily estimated empirically. To do so, we specialize our production function to have a constant elasticity of substitution (CES) between pairs of types. This leads to particularly sharp and transparent theoretical insights. First, we obtain an optimal taxation formula that generalizes the partial-equilibrium results of Diamond (1998); Saez (2001). We show that marginal tax rates should be lower (resp., higher) for agents whose welfare is valued less (resp., more) than average, because an increase in the marginal tax rate of a given skill type increases their wage at the expense of all the other types. These general equilibrium forces reinforce the U-shaped pattern of optimal taxes. We derive the optimal top tax rate in closed form in terms of the labor supply elasticity, the elasticity of substitution, and the Pareto parameter of the tail of the income distribution.

[1] In this paper, by "progressivity" we mean "increasing marginal tax rates". Another definition would be increasing average, rather than marginal, tax rates. Our result regarding the benefits of raising the progressivity of the tax schedule (if the initial tax code has increasing marginal tax rates) holds under both definitions.

Related literature. This paper is related to the literature on tax incidence: see, e.g., Harberger (1962); Shoven and Whalley (1984) for the seminal papers, Kotlikoff and Summers (1987); Fullerton and Metcalf (2002) for comprehensive surveys, and Hines (2009) for emphasizing the importance of general equilibrium (GE) in taxation. We extend this analysis to the workhorse model of nonlinear income taxation of Mirrlees (1971). The optimal taxation problem in GE with nonlinear tax instruments has originally been studied by Stiglitz (1982) in a model with two types. A series of important contributions by Scheuer (2014); Rothschild and Scheuer (2013, 2014); Scheuer and Werning (2016, 2017); Ales, Kurnaz, and Sleet (2015); Ales and Sleet (2016); Ales, Bellofatto, and Wang (2017) form the modern analysis of optimal nonlinear taxes in GE.[2] Our setting is distinct from those of Scheuer and Werning (2016, 2017), whose modeling of the technology is such that the optimum tax formula of Mirrlees (1971) extends to general production functions; we discuss in detail the difference between our framework and theirs in Appendix A.1.

Most closely related to our paper, Rothschild and Scheuer (2013) generalize Stiglitz (1982) to a Roy setting with several sectors and a continuum of skills in each sector, leading to a multidimensional screening problem,[3] and Ales, Kurnaz, and Sleet (2015) microfound the production function by incorporating an assignment model (as in Sattinger (1975); Teulings (1995); Costinot and Vogel (2010)) into the optimal taxation framework. The key distinction between these papers and ours is that they use mechanism design tools that are only able to characterize the optimum taxes, whereas we study more generally the tax incidence problem by extending the variational, or "tax reform", approach introduced by Piketty (1997); Saez (2001); Golosov et al. (2014) to GE environments. This is important as we show that the (potentially suboptimal) tax system to which the reform is applied is a crucial determinant of the direction and size of the GE effects. Our paper also differs from those mentioned above as it is in the sufficient statistic tradition (Chetty (2009)): conditional on the wage elasticities that we introduce, our baseline tax incidence formulas remain identical for

[2] Rothschild and Scheuer (2016); Piketty, Saez, and Stantcheva (2014) study optimal taxes in the presence of rent-seeking. In this paper we abstract from such considerations and assume that individuals are paid their marginal productivity. Kushnir and Zubrickas (2018) set up a Mirrlees model in which general equilibrium effects occur through product prices rather than wages. Jones (2019) characterizes the optimal top tax rate in an environment where economics growth is driven by endogenous innovation.

[3] Our baseline model is simpler than theirs, as different types earn different wages (there is no overlap in the wage distributions). In the former version of this paper we extended our results to the Roy model.

several underlying primitive environments (namely, whether the assignment of skills to tasks is fixed or endogenous to taxes). Finally, our characterization of optimal income tax rates is also novel: assuming a simple technology leads to parsimonious and transparent formulas generalizing those of Diamond (1998); Saez (2001).

Our paper is also related to the literature that derives simple closed-form expressions for the effects of tax policy in general equilibrium. Heathcote, Storesletten, and Violante (2017) and Antras, De Gortari, and Itskhoki (2017) do so by restricting the production function to be CES and the tax schedule to be CRP. On the one hand, our model is simpler than theirs as we study a static and closed economy with exogenous skills. On the other hand, for most of our theoretical analysis we do not restrict ourselves to particular functional forms for taxes and the production function. Finally, our modeling of the production function is motived by an empirical literature that estimates the impact of immigration on the wage distribution and groups workers according to their position in this distribution (Card (1990); Borjas, Freeman, Katz, DiNardo, and Abowd (1997); Dustmann, Frattini, and Preston (2013)). This empirical literature is a useful benchmark because it studies the impact of labor supply shocks of certain skills on relative wages, which is exactly the channel we want to analyze in our tax setting.

1 Environment

In this section we set up the baseline version of our model. Our main results can be derived most transparently by assuming that individual preferences are quasilinear. Technical details are provided in Appendix A. We extend our analysis to general preferences in Appendix D.

1.1 Initial equilibrium

Individuals. There is a continuum of mass 1 of workers indexed by their skill $\theta \in \Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_+$, distributed according to the pdf $f(\cdot)$ and cdf $F(\cdot)$. Individual preferences over consumption c and labor supply l are represented by the quasilinear utility function $c - v(l)$, where the disutility of labor $v : \mathbb{R}_+ \to \mathbb{R}_+$ is twice continuously differentiable, strictly increasing and strictly convex. Individuals with skill θ earn a wage $w(\theta)$ that they take as given. They choose their labor supply $l(\theta)$ and earn taxable income $y(\theta) = w(\theta) l(\theta)$. Their consumption is equal to $y(\theta) - T(y(\theta))$, where $T : \mathbb{R}_+ \to \mathbb{R}$ is a twice continuously differentiable income tax schedule. Their optimal labor supply choice $l(\theta)$ is the solution to the first-order condition:[4]

$$v'(l(\theta)) = [1 - T'(w(\theta) l(\theta))] w(\theta). \tag{1}$$

We denote by $U(\theta)$ the utility attained by these agents, and by $L(\theta) \equiv l(\theta) f(\theta)$ the total amount of labor supplied by individuals of type θ.

Firms. There is a continuum of mass 1 of identical firms that produce output using the labor of every skill type $\theta \in \Theta$. We posit a constant returns to scale aggregate production function $\mathcal{F}(\mathscr{L})$ over the continuum of labor inputs $\mathscr{L} \equiv \{L(\theta)\}_{s\in\Theta}$.[5] In equilibrium, firms earn no profits and the wage $w(\theta)$ is equal to the marginal productivity of type-θ labor, that is,

$$w(\theta) = \frac{\partial}{\partial L(\theta)} \mathcal{F}(\mathscr{L}). \tag{2}$$

Remark (*Monotonicity*). Without loss of generality we order the skills θ so that the wage function $\theta \mapsto w(\theta)$ is strictly increasing given the tax schedule T.[6] By the Spence-Mirrlees condition, the pre-tax income function $\theta \mapsto y(\theta)$ is then also strictly increasing. Therefore, there are one-to-one relationships between skills θ, wages $w(\theta)$, and pre-tax incomes $y(\theta)$ in the initial equilibrium. We denote by $f_Y(y(\theta)) = (y'(\theta))^{-1} f(\theta)$ the density of incomes and by F_Y the corresponding c.d.f.

Example (CES technology). The production function has a constant elasticity of substitution (CES) if

$$\mathcal{F}(\mathscr{L}) = \left[\int_{\Theta} a(\theta) (L(\theta))^{\frac{\sigma-1}{\sigma}} d\theta\right]^{\frac{\sigma}{\sigma-1}}, \tag{3}$$

[4] The dependence of labor supply on the tax schedule T is left implicit for simplicity. Whenever necessary, we denote the solution to (1) by $l(\theta; T)$.

[5] In Section 1.3 below we provide a microfoundation of this production function. An alternative interpretation of our framework is that different types of workers produce different types of goods that are imperfect substitutes in household consumption.

[6] We can moreover assume w.l.o.g. that the skill set Θ is the interval $[0, 1]$ and that the distribution $f(\theta)$ is uniform, in which case θ indexes the agent's percentile in the wage distribution. Note that this ordering remains unchanged regardless of the tax reform if the production is CES. More generally, our tax incidence analysis does not require that the initial ordering of wages to be unaffected by tax reforms.

for some $\sigma \in [0, \infty)$ and $a(\cdot) \in \mathbb{R}_+$. The wage schedule is then given by $w(\theta) = a(\theta)\left(L(\theta)/\mathcal{F}(\mathscr{L})\right)^{-1/\sigma}$. The cases $\sigma = 0$, $\sigma = 1$, and $\sigma = \infty$ correspond respectively to the Leontieff, Cobb-Douglas, and exogenous-wage technologies.

Government. The government chooses the twice-continuously differentiable tax function $T : \mathbb{R}_+ \to \mathbb{R}$. Tax revenue is given by

$$\mathcal{R} = \int_\Theta T(y(\theta)) f(\theta)\, d\theta.$$

We define the local rate of progressivity of the tax schedule T at income level y as (minus) the elasticity of the retention rate $1 - T'(y)$ with respect to income y,

$$p(y) \equiv -\frac{\partial \ln[1 - T'(y)]}{\partial \ln y} = \frac{yT''(y)}{1 - T'(y)}.$$

Example (CRP taxes). The schedule has a constant rate of progressivity (CRP) if

$$T(y) = y - \frac{1-\tau}{1-p} y^{1-p}, \tag{4}$$

for some $p < 1$.[7] This tax schedule is linear (resp., progressive, regressive), i.e., the marginal tax rates $T'(y)$ and the average tax rates $T(y)/y$ are constant (resp., increasing, decreasing), if $p = 0$ (resp., $p > 0$, $p < 0$).

Equilibrium. An equilibrium given a tax function T is a schedule of labor supplies $\{l(\theta)\}_{\theta\in\Theta}$, labor demands $\{L(\theta)\}_{\theta\in\Theta}$, and wages $\{w(\theta)\}_{\theta\in\Theta}$ such that equations (1) and (2) hold, the labor markets clear: $L(\theta) = l(\theta) f(\theta)$ for all $\theta \in \Theta$, and the goods market clears: $\mathcal{F}(\mathscr{L}) = \int_\Theta w(\theta) L(\theta)\, d\theta$.

1.2 Elasticities

We now define the parameters that determine the economy's adjustment to tax reforms, namely, the elasticities of the labor supply and labor demand curves within each labor market θ, and the cross-price elasticities between labor markets θ, θ'.

[7] See, e.g., Bénabou (2002); Heathcote, Storesletten, and Violante (2017).

Cross-wage elasticities. Consider first two distinct labor markets for skills θ and θ'. We define the elasticity of the wage of type θ', $w(\theta')$, with respect to the aggregate labor of type θ, $L(\theta)$, as

$$\gamma(\theta',\theta) \equiv \frac{\partial \ln w(\theta')}{\partial \ln L(\theta)} = \frac{L(\theta)\mathscr{F}''_{\theta',\theta}(\mathscr{L})}{\mathscr{F}'_{\theta'}(\mathscr{L})}, \quad \forall \theta' \neq \theta \tag{5}$$

where $\mathscr{F}'_{\theta'}$ and $\mathscr{F}''_{\theta',\theta}$ denote the first and second partial derivatives of the production function $\mathscr{F}$ with respect to the labor inputs of types θ' and θ. The cross-wage elasticity between two skills θ,θ' is non-zero if they are imperfect substitutes in production. They are Edgeworth complements if $\gamma(\theta',\theta) > 0$ and substitutes if $\gamma(\theta',\theta) < 0$. In the CES example (3), $\gamma(\theta',\theta) = \frac{1}{\sigma}a(\theta)(L(\theta)/\mathscr{F}(\mathscr{L}))^{\frac{\sigma-1}{\sigma}} > 0$ does not depend on θ', implying that a change in the labor supply of skill θ has the same effect, in percentage terms, on the wage of every skill $\theta' \neq \theta$.

Labor demand elasticities. Next, consider the labor market for a given skill θ. The impact of the aggregate labor effort of skill θ on its own wage, $\frac{\partial \ln w(\theta)}{\partial \ln L(\theta)}$, may be different than its impact on the wage of its close neighbors $\theta' \approx \theta$, $\lim_{\theta'\to\theta} \frac{\partial \ln w(\theta')}{\partial \ln L(\theta)} = \lim_{\theta'\to\theta} \gamma(\theta',\theta)$. That is, the function $\theta' \mapsto \frac{\partial \ln w(\theta')}{\partial \ln L(\theta)}$ may be discontinuous at $\theta' = \theta$. We denote by $\gamma(\theta,\theta) \equiv \lim_{\theta'\to\theta} \frac{\partial \ln w(\theta')}{\partial \ln L(\theta)}$ the complementarity between θ and its neighboring skills, and define the inverse elasticity of labor demand for skill θ, $1/\varepsilon_w^D(\theta)$, as size of the jump between $\frac{\partial \ln w(\theta)}{\partial \ln L(\theta)}$ and $\gamma(\theta,\theta)$. Formally,

$$\frac{\partial \ln w(\theta')}{\partial \ln L(\theta)} \equiv \gamma(\theta',\theta) - \frac{1}{\varepsilon_w^D(\theta)}\delta(\theta'-\theta), \quad \forall (\theta,\theta') \in \Theta^2, \tag{6}$$

where $\delta(\cdot)$ denotes the Dirac delta function. In the CES example (3), this own-wage effect $\varepsilon_w^D(\theta) = \sigma > 0$ captures the fact that the marginal productivity of skill θ is decreasing, whereas θ is Edgeworth complement with every other skill θ'. Note that the tax incidence formulas we derive in this paper are valid whether such a discontinuity indeed occurs (e.g., if the production function is CES) or not (e.g., in the microfoundation of Section 1.3).

Labor supply elasticities. Finally, we define the elasticities of labor supply $l(\theta)$ with respect to the retention rate $r(\theta) \equiv 1 - T'(y(\theta))$ and the wage $w(\theta)$ as

$$\varepsilon_r^S(\theta) \equiv \frac{\partial \ln l(\theta)}{\partial \ln r(\theta)} = \frac{e(\theta)}{1 + p(y(\theta))\, e(\theta)}, \quad \varepsilon_w^S(\theta) \equiv \frac{\partial \ln l(\theta)}{\partial \ln w(\theta)} = (1 - p(y(\theta)))\, \varepsilon_r^S(\theta), \tag{7}$$

where $e(\theta) \equiv \frac{v'(l(\theta))}{l(\theta)v''(l(\theta))}$. The variable $\varepsilon_r^S(\theta)$ is an elasticity along the nonlinear budget constraint:[8] it differs from the standard elasticity parameter $e(\theta)$ as it accounts for the fact that if the tax schedule is nonlinear, a change in individual labor supply $l(\theta)$ causes endogenously a change in the marginal tax rate $T'(w(\theta)\, l(\theta))$ captured by the rate of progressivity $p(y(\theta))$ of the tax schedule, and hence a further labor supply adjustment $e(\theta)$. Solving for the fixed point leads to the correction term $p(y(\theta))\, e(\theta)$ in the denominator of $\varepsilon_r^S(\theta)$.[9]

1.3 Microfoundation and sufficient statistics

The production function we introduced in Section 1.1 can be microfounded as the reduced form of an underlying model of assignment between the worker skills and the tasks involved in production. That is, our analysis encompasses the cases of both fixed and endogenous assignment. To show this, we set up a model that extends Costinot and Vogel (2010) by allowing workers to choose their labor supply endogenously and the government to tax labor income nonlinearly.[10] The technical details are gathered in Appendix A.2.

The final consumption good is produced using a CES technology over a continuum of tasks $\psi \in \Psi$, indexed by their skill intensity (e.g., manual, routine, abstract, etc.). The output of each task is in turn produced linearly using the labor of the skills $\theta \in \Theta$ that are endogenously assigned to this task. Assuming that high-skilled workers have a comparative advantage in tasks with high skill intensities, the market clearing conditions for intermediate goods determine a one-to-one matching function $M : \Theta \to \Psi$ between skills and tasks in equilibrium – there is positive assortative matching.

[8] See also Jacquet and Lehmann (2017).

[9] Since there is a one-to-one map between types θ and incomes $y(\theta)$, one can equivalently index these elasticities by income: $\varepsilon_r^S(y(\theta)) \equiv \varepsilon_r^S(\theta)$. We use these two notations interchangeably in the sequel, and analogously for the labor demand elasticities $\varepsilon_w^D(\theta)$ defined above. On the other hand, the natural change of variables between types θ and incomes $y(\theta)$ for the cross-wage elasticities is $\gamma(y(\theta_1), y(\theta_2)) = (y'(\theta_2))^{-1}\gamma(\theta_1, \theta_2)$, and analogously for the resolvent cross-wage elasticities $\Gamma(\theta_1, \theta_2)$ defined below.

[10] Ales, Kurnaz, and Sleet (2015) characterize optimal taxes in such a model.

It is straightforward to show that this model admits a reduced-form representation where the production of the final good is performed by a technology over skills. This reduced-form technology inherits the CES structure (3) of the original production function over intermediate tasks, except that the technology coefficients $a(\cdot)$ now depend on the matching function M, and are thus endogenous to taxes.

Crucially, we show that tax reforms affect the equilibrium assignment M *only* through their effect on individual labor supply choices $\{L(\theta)\}_{\theta\in\Theta}$. Mathematically, this is a consequence of the fact that, fixing labor supplies, none of the equations that determine the equilibrium depend explicitly on the tax schedule T. Intuitively, this is because individuals always choose the task that maximizes their net wage. Since a tax reform does not alter directly the ranking of net wages, as long as marginal tax rates are below 100 percent, taxes affect the equilibrium sorting of skills only indirectly through the labor supply responses that they induce. Hence the technological coefficients $a(\cdot; M)$ of the reduced-form technology described above can be written without loss of generality as $a(\cdot; \{L(\theta)\}_{\theta\in\Theta})$. Substituting these parameters into (3) yields a production function with the general form postulated in Section 1.1, $\mathcal{F}(\{L(\theta)\}_{\theta\in\Theta})$.

The implied cross-wage elasticities $\gamma(\theta', \theta) = \frac{L(\theta)}{\mathcal{F}'_{\theta'}} \frac{\partial^2 \mathcal{F}}{\partial L(\theta) \partial L(\theta')}$, as defined in equation (5), already account for the potential reassignment of workers into new tasks.[11] That is, they represent the impact of an increase in the labor supply of skill θ on the marginal productivity of skill θ', leaving everyone else's labor supply unchanged and, if assignment is endogenous, letting workers be reassigned into different tasks – i.e., taking into account the adjustment of the technological coefficients $a(\cdot, \{L(\theta)\}_{\theta\in\Theta})$. It follows from this discussion that these cross-wage elasticities are *sufficient statistics*: once expressed as a function of these parameters, the tax incidence formulas that we derive in Sections 2 and 3 are valid both when the underlying structure of assignment is fixed and when it is endogenous to tax reforms.

Graphical representation. We now represent graphically the cross-wage elasticities that arise in the model we just described. As we detail in Section 4 below, we

[11] Note moreover that, while in a setting with exogenous assignment the inverse labor demand elasticities $1/\varepsilon_w^D$ are generally non-zero (i.e., there is a discontinuity in the schedule of elasticities $\frac{\partial \ln w(\theta')}{\partial \ln L(\theta)}$ as $\theta' \approx \theta$), instead with costless reassignment such a discontinuity would cause workers to migrate to neighboring tasks, leading to perfectly elastic labor demand curves (i.e., $1/\varepsilon_w^D = 0$). Our tax incidence formulas are naturally valid in both cases.

follow the calibration of Ales, Kurnaz, and Sleet (2015) who assume a Cobb-Douglas technology over tasks. We compare these elasticities with those obtained in our baseline model of Section 1.1, assuming a CES production function over skills with two calibrations of the elasticity of substitution. The first consists of simply shutting down the endogenous reassignment channel in the former model while keeping all of the other parameters unchanged, hence assuming a Cobb-Douglas production function over skills ($\sigma = 1$). The second, more relevant, calibration consists of directly estimating a CES production function over labor supplies: we then use the value $\sigma = 3.1$ estimated by Heathcote, Storesletten, and Violante (2017).

The left panel of Figure 1 plots the resulting cross-wage elasticities $\gamma(y, y^*)$ in the model of endogenous assignment, in response to changes in the labor supplies of agents who earn $y^* \in \{\$25,000\,;\,\$80,000\,;\,\$200,000\}$. They are V-shaped and increasing in the distance $|y - y^*|$. A higher labor effort of agents y^* lowers wages on a non-degenerate interval of incomes around y^* and raises those of much higher or much lower incomes. Note that the shape of the cross-wage elasticities in Figure 1 is similar to those of Teulings (2005). The right panel compares these elasticities with those obtained with a CES production function (3) and fixed assignment, for $y^* = \$80,000$ and $\sigma \in \{1\,;\,3.1\}$. In this case, as shown above, the wage of agent y^* decreases, while everyone else's wages increase by the same amount in percentage terms. The discontinuity at y^* is represented by the Dirac arrows in the figure. Letting workers be reassigned to different tasks in response to an exogenous increase in the labor supply at y^* thus spreads out the cross-wage effects around $y \approx y^*$ and removes the discontinuity that arises when matching is kept fixed.

2 Incidence of tax reforms

Consider a given initial, potentially suboptimal, tax schedule T, e.g., the U.S. tax code. In this section we derive closed-form formulas for the first-order effects of arbitrary local perturbations of this tax schedule ("tax reforms") on individual labor supplies, wages and utilities. The proofs are gathered in Appendix B.

2.1 Effects on labor supply

As in the case of exogenous wages (Saez (2001)), analyzing the incidence of tax reforms relies crucially on solving for each individual's change in labor supply in terms

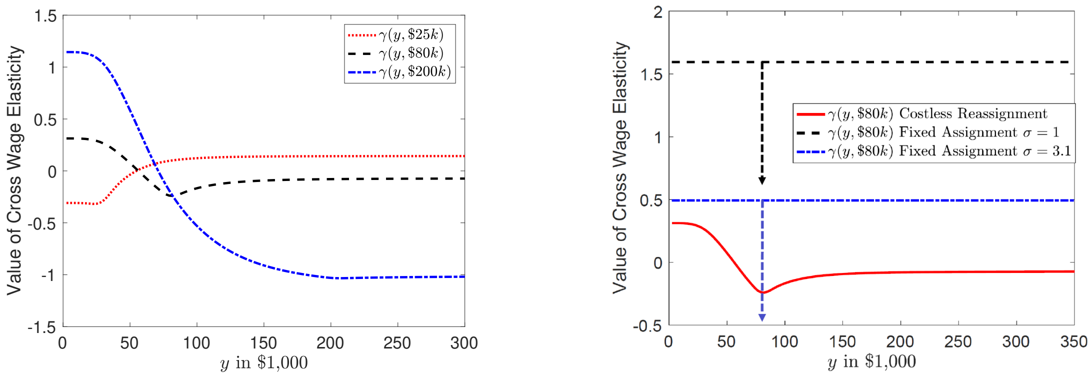


Figure 1: Left panel: Cross wage elasticities $y \mapsto \gamma(y, y^*)$ in the model with endogenous costless reassignment of skills to tasks with y^* equal to \$25,000 (dotted curve), \$80,000 (dashed curve), \$200,000 (dashed-dotted curve). Right panel: Comparison of the cross-wage elasticities $y \mapsto \gamma(y, y^*)$ with $y^* = \$80,000$ in the models with endogenous assignment (solid curve) and with exogenous assignment (CES production) for $\sigma = 1$ (dashed line) and $\sigma = 3.1$ (dashed-dotted line).

of behavioral elasticities. This problem is much more involved in general equilibrium. If wages are exogenous, a change in the tax rate of a given individual, say θ, induces only a change in the labor effort of that agent (measured by the elasticity (7)). In the general equilibrium setting, instead, this labor supply response of type θ affects the wage, and hence the labor supply, of every other skill $\theta' \in \Theta$. This in turn feeds back into the wage distribution, which further impacts labor supplies, and so on. Representing the total effect of this infinite sequence caused by arbitrarily non-linear tax reforms is thus a complex task.[12] The key step towards the general characterization of the economic incidence of taxes, and our first main theoretical contribution, consists of showing that this problem can be mathematically formulated as an integral equation (Lemma 1).[13] Thus, we can apply the tools and results of the theory of integral equations to solve for the labor supply adjustments in closed-form (Proposition 1). The incidence on wages and utilities is then straightforward to obtain (Corollary 2).

[12] We could define, for each specific tax reform one might consider implementing, a "policy elasticity" (as in, e.g., Hendren (2015), Piketty and Saez (2013)), equal to each individual's total labor supply response to the corresponding reform. The key challenge of the incidence problem consists of expressing this total labor supply response in terms of the structural elasticity parameters introduced in Section 1.2. In other words, Proposition 1 below derives the policy elasticity in terms of these structural parameters.

[13] The general theory of linear integral equations is exposed in, e.g., Tricomi (1985) and Zemyan (2012). Moreover, closed-form solutions can be derived in many special cases (see Polyanin and Manzhirov (2008)) and numerical techniques are widely available (see Section 4).

Tax reforms and Gateaux derivatives. Consider an arbitrary non-linear reform of the initial tax schedule $T(\cdot)$. Formally, this tax reform can be represented by a continuously differentiable function $\hat{T}(\cdot)$ on $\mathbb{R}_+$, so that the perturbed tax schedule is $T(\cdot) + \mu \hat{T}(\cdot)$, where $\mu \in \mathbb{R}$ parametrizes the size of the reform.[14] Our aim is to compute the first-order effect of this perturbation on individual labor supply (i.e., the solution to the first-order condition (1)), when the magnitude of the tax change is small, i.e., as $\mu \to 0$. This is formally given by the Gateaux derivative of the labor supply functional $T \mapsto l(\theta; T)$ in the direction $\hat{T}$, that is,[15]

$$\hat{l}(\theta) \equiv \lim_{\mu \to 0} \frac{1}{\mu} \left[l(\theta; T + \mu \hat{T}) - l(\theta; T) \right].$$

The variable $\hat{l}(\theta)$ gives the change in the labor supply of type θ in response to the tax reform $\hat{T}$, taking into account all the general equilibrium effects induced by the endogeneity of wages. We define analogously the changes in individual wages $\hat{w}(\theta)$, utilities $\hat{U}(\theta)$ and government revenue $\hat{\mathcal{R}}$.

Integral equation (IE). The following lemma provides an implicit characterization of the incidence of an arbitrary tax reform $\hat{T}$ on labor supplies.

Lemma 1. *The effect of a tax reform $\hat{T}$ of the initial tax schedule T on individual labor supplies, $\hat{l}(\cdot)$, is the solution to the functional equation: for all $\theta \in \Theta$,*

$$\frac{\hat{l}(\theta)}{l(\theta)} = -\varepsilon_r(\theta) \frac{\hat{T}'(y(\theta))}{1 - T'(y(\theta))} + \varepsilon_w(\theta) \int_{\Theta} \gamma(\theta, \theta') \frac{\hat{l}(\theta')}{l(\theta')} d\theta', \tag{8}$$

where $\varepsilon_r(\theta)$ and $\varepsilon_w(\theta)$ are the elasticities of equilibrium labor of skill θ with respect to the retention rate and the wage, defined respectively by

$$\frac{1}{\varepsilon_r(\theta)} \equiv \frac{1}{\varepsilon_r^S(\theta)} + \frac{1}{\varepsilon_w^D(\theta)} \quad \text{and} \quad \frac{1}{\varepsilon_w(\theta)} \equiv \frac{1}{\varepsilon_w^S(\theta)} + \frac{1}{\varepsilon_w^D(\theta)}.$$

Formula (8) is a linear Fredholm integral equation of the second kind with kernel $\varepsilon_w(\theta) \gamma(\theta, \theta')$. Its unknown, which appears under the integral sign, is the function

[14] An example of this general definition consists of increasing the marginal tax rate on a small income interval, and hence the total tax payment by a constant lump-sum amount above that interval (Piketty (1997); Saez (2001)). We formalize and analyze this important class of perturbations in Section 3.1 below.

[15] The notation $\hat{l}(\theta)$ ignores for simplicity the dependence of this derivatives on the initial tax schedule T and on the tax reform $\hat{T}$.

$\theta \mapsto \hat{l}(\theta)$. Before deriving its solution, we start by providing the economic meaning of this equation.

Due to the reform, the retention rate $r(\theta) = 1 - T'(y(\theta))$ of individual θ changes, in percentage terms, by $\frac{\hat{r}(\theta)}{r(\theta)} = -\frac{\hat{T}'(y(\theta))}{1-T'(y(\theta))}$. This tax reform induces a direct percentage change in labor effort $l(\theta)$ equal to $\varepsilon_r(\theta)\frac{\hat{r}(\theta)}{r(\theta)}$, where $\varepsilon_r(\theta)$ is the elasticity of equilibrium labor on the market for skill θ. This is the partial-equilibrium adjustment, obtained by considering the labor market θ in isolation and ignoring the cross-price effects between markets. It resembles the expression $\varepsilon_r^S(\theta)\frac{\hat{r}(\theta)}{r(\theta)}$ one would get assuming exogenous wages, with one difference: if the marginal product of labor is decreasing, i.e., the labor demand curve is downward sloping, then the initial labor supply adjustment (say, decrease) due to the tax reform causes an own-wage increase determined by $1/\varepsilon_w^D(\theta)$, which in turn raises labor supply and dampens the initial response – hence the relevant elasticity satisfies $\varepsilon_r(\theta) \leq \varepsilon_r^S(\theta)$.

Now, in general equilibrium, the labor supply of type θ is also impacted indirectly by the change in all other individuals' labor supplies, due to the skill complementarities in production. Specifically, the percentage change in labor supply of each type θ', $\frac{\hat{l}(\theta')}{l(\theta')}$, triggers a change in the wage of type θ equal to $\gamma(\theta,\theta')\frac{\hat{l}(\theta')}{l(\theta')}$, and thus a further adjustment in labor supply equal to $\varepsilon_w(\theta)\gamma(\theta,\theta')\frac{\hat{l}(\theta')}{l(\theta')}$. Summing these effects over skills $\theta' \in \Theta$ leads to formula (8).

Solution to the IE and resolvent. We now characterize the solution to the integral equation (8).

Proposition 1. *Assume that the condition $\int_{\Theta^2} |\varepsilon_w(\theta)\gamma(\theta,\theta')|^2 d\theta d\theta' < 1$ holds.*[16] *The unique solution to the integral equation (8) is given by:*

$$\frac{\hat{l}(\theta)}{l(\theta)} = -\varepsilon_r(\theta)\frac{\hat{T}'(y(\theta))}{1-T'(y(\theta))} - \varepsilon_w(\theta)\int_\Theta \Gamma(\theta,\theta')\varepsilon_r(\theta')\frac{\hat{T}'(y(\theta'))}{1-T'(y(\theta'))}d\theta', \quad (9)$$

where for all $(\theta,\theta') \in \Theta^2$, the resolvent $\Gamma(\theta,\theta')$ is defined by

$$\Gamma(\theta,\theta') \equiv \sum_{n=1}^{\infty}\Gamma_n(\theta,\theta'), \quad (10)$$

[16]This technical condition ensures that the infinite series (10) converges. We provide sufficient conditions on primitives such that this condition holds. In more general cases it can be easily verified numerically. Finally, when it is not satisfied, we can more generally express the solution to (8) with a representation similar to (9) but with a more complex resolvent (see Section 2.4 in Zemyan (2012)).

with $\Gamma_1(\theta,\theta') = \gamma(\theta,\theta')$ *and for all* $n \geq 2$,

$$\Gamma_n(\theta,\theta') = \int_\Theta \Gamma_{n-1}(\theta,\theta'')\,\varepsilon_w(\theta'')\,\gamma(\theta'',\theta')\,d\theta''.$$

Sufficient conditions on primitives ensuring the convergence of the resolvent (10) are that the production function is CES, the initial tax schedule is CRP, and the disutility of labor is isoelastic.

The mathematical representation (9) of the solution to the integral equation (8) has the following economic meaning. The first term on the right hand side of (9), $-\varepsilon_r(\theta)\frac{\hat{T}'(y(\theta))}{1-T'(y(\theta))}$, is the partial-equilibrium effect of the reform on labor supply $l(\theta)$, already described in equation (8). The second (integral) term accounts for the cross-wage effects in general equilibrium. Note that this integral term has the same structure as the corresponding term in formula (8), except that: (i) the unknown endogenous labor supply change $\frac{\hat{l}(\theta')}{l(\theta')}$ is now replaced by the (known) partial-equilibrium impact $-\varepsilon_r(\theta')\frac{\hat{T}'(y(\theta'))}{1-T'(y(\theta'))}$; and (ii) the structural cross-wage elasticity $\gamma(\theta,\theta')$ is replaced by the *resolvent* cross-wage elasticity $\Gamma(\theta,\theta')$.[17]

The resolvent elasticity $\Gamma(\theta,\theta')$, defined by the series (10), expresses the total effect of the labor supply of type θ' on the wage of type θ. That is, it accounts for the infinite sequence of general-equilibrium adjustments induced by the complementarities in production. The first iterated kernel ($n = 1$) in the series (10) is simply $\Gamma_1(\theta,\theta') = \gamma(\theta,\theta')$. It accounts for the impact of the labor supply of type θ' on the wage of type θ through direct cross-wage effects. The second iterated kernel ($n = 2$) in (10) accounts for the impact of the labor supply of θ' on the wage of θ, indirectly through the behavior of third parties θ''. This term reads

$$\Gamma_2(\theta,\theta') = \int_\Theta \gamma(\theta,\theta'')\,\varepsilon_w(\theta'')\,\gamma(\theta'',\theta')\,d\theta''. \tag{11}$$

For any θ', a percentage change in the labor supply of θ' induces a percentage change

[17] For applied purposes, one can use both the structural parameters $\gamma(\theta,\theta')$ or the resolvent parameters $\Gamma(\theta,\theta')$ as primitive cross-wage elasticity variables – our tax incidence formulas can be expressed in terms of either of them. Some empirical studies may evaluate the structural parameters $\gamma(\theta,\theta')$ of the production function directly, while others may estimate the full general-equilibrium impact $\Gamma(\theta,\theta')$, including the spillovers generated by the initial shock. In the latter case, it may be useful to recover the structural elasticities $\gamma(\theta,\theta')$ as a function of the higher-order elasticities $\Gamma(\theta,\theta')$, e.g., for counterfactual analysis. It is straightforward to show that $\gamma(\theta,\theta')$ can be expressed as the solution to an integral equation with a kernel determined by $\Gamma(\theta,\theta')$.

in the wage of any other type θ'' by $\gamma(\theta'', \theta')$, and hence a percentage change in the labor supply of θ'' given by $\varepsilon_w(\theta'')\gamma(\theta'', \theta')$. This in turn affects the wage of type θ by the amount $\gamma(\theta, \theta'')\varepsilon_w(\theta'')\gamma(\theta'', \theta')$. Summing over all intermediate types θ'' leads to expression (11). An inductive reasoning shows similarly that the terms $n \geq 3$ in the resolvent series (10) account for the impact of the labor supply of θ' on the wage of θ through n successive stages of cross-wage effects, e.g., for $n = 3$, $\theta' \to \theta'' \to \theta''' \to \theta$.

The case of separable cross-wage elasticities. A particularly tractable special case of Proposition 1 is obtained when the cross-wage elasticities are multiplicatively separable between skills. This occurs in particular when the production function is CES (in which case $\gamma(\theta, \theta')$ depends only on θ') or, more generally, homothetic with a single aggregator (HSA, see Matsuyama and Ushchev (2017)). The following corollary shows that in this case, each round of general equilibrium effects, i.e., each term in the series (10), is a fraction of the first round – so that the resolvent cross-wage elasticity $\Gamma(\theta, \theta')$ is directly proportional to the structural elasticity $\gamma(\theta, \theta')$.

Corollary 1. *Suppose that there exist functions γ_1 and γ_2 such that for all (θ, θ'), $\gamma(\theta, \theta') = \gamma_1(\theta)\gamma_2(\theta')$. The resolvent cross-wage elasticities are then given by*

$$\Gamma(\theta, \theta') = \frac{\gamma(\theta, \theta')}{1 - \int_{\Theta} \varepsilon_w(s)\gamma(s, s)\, ds}. \tag{12}$$

In particular, if the production function is CES, the integral in the denominator of (12) is equal to $\frac{1}{\sigma \mathbb{E} y}\mathbb{E}[y\varepsilon_w(y)]$.

2.2 Effects on wages and utility

We can now easily obtain the incidence of an arbitrary tax reform $\hat{T}$ on individual wages and utilities.

Corollary 2. *The incidence of a tax reform $\hat{T}$ of the initial tax schedule T on individual wages, $\hat{w}(\cdot)$, is given by*

$$\frac{\hat{w}(\theta)}{w(\theta)} = \frac{1}{\varepsilon_w^S(\theta)}\left[\varepsilon_r^S(\theta)\frac{\hat{T}'(y(\theta))}{1 - T'(y(\theta))} + \frac{\hat{l}(\theta)}{l(\theta)}\right], \tag{13}$$

for all $\theta \in \Theta$, where the labor supply response $\hat{l}(\theta)$ is given by (9). The incidence on *individual utilities, $\hat{U}(\cdot)$, is given by*

$$\hat{U}(\theta) = -\hat{T}(y(\theta)) + (1 - T'(y(\theta)))\, y(\theta)\, \frac{\hat{w}(\theta)}{w(\theta)}. \tag{14}$$

Equation (13) gives the changes in individual wages due to the tax reform $\hat{T}$, as a function of the labor supply changes characterized by Proposition 1. Its interpretation is straightforward: multiplying both sides of (13) by $\varepsilon_w^S(\theta)$ gives the percentage adjustment of type-θ labor supply, $\frac{\hat{l}(\theta)}{l(\theta)}$, as the sum of its response in the case of exogenous wages, $-\varepsilon_r^S \frac{\hat{T}'}{1-T'}$, and the effect induced by the percentage wage change, $\varepsilon_w^S \times \frac{\hat{w}}{w}$.

Equation (14) gives the impact of the reform on individual welfare. The first term in the right hand side, $-\hat{T}(y(\theta))$, is due to the fact that a higher tax payment makes the individual poorer and hence reduces utility. The second term accounts for the change in net income due to the wage adjustment $\hat{w}(\theta)$, given by equation (13). If wages were exogenous, so that $\hat{w}(\theta) = 0$ in (14), the utility of agent θ would respond one-for-one to changes in the total tax payment $\hat{T}(y(\theta))$; in particular, by the envelope theorem, it would not be affected by changes in the marginal tax rate $\hat{T}'(y(\theta))$. In general equilibrium, however, this is no longer true because marginal tax rates cause labor supply adjustments which in turn affect wages (second term in (14)) and hence utilities. We can easily show that if all pairs of types are Edgeworth complements and the assignment of workers to tasks is exogenous, then a higher marginal tax rate at income $y(\theta)$ raises the utility of agents with skill θ and lowers that of all other agents.

3 Effects of tax reforms on government revenue

The impact of a tax reform $\hat{T}$ on government revenue follows directly from the changes in equilibrium labor and wages:

$$\hat{\mathcal{R}}(\hat{T}) = \int_\Theta \hat{T}(y(\theta))\, f(\theta)\, d\theta + \int_\Theta T'(y(\theta)) \left[\frac{\hat{l}(\theta)}{l(\theta)} + \frac{\hat{w}(\theta)}{w(\theta)}\right] y(\theta)\, f(\theta)\, d\theta. \tag{15}$$

The first term on the right hand side of (15) is the statutory effect of the tax reform $\hat{T}(\cdot)$, i.e., the mechanical change in government revenue assuming that the individual's labor supply and wage remain constant. The second term is the behavioral effect of the reform. The labor supply and wage adjustments $\hat{l}(\theta)$ and $\hat{w}(\theta)$ both induce a change in government revenue proportional to the marginal tax rate $T'(y(\theta))$. Summing these effects over all individuals using the density $f(\cdot)$ leads to equation (15). The remainder of this section is devoted to deriving the economic implications of this formula. The proofs and technical details are gathered in Appendix C.

3.1 Preliminaries

Elementary tax reforms. From now on, we focus without loss of generality on a specific class of "elementary" tax reforms, represented by the step function $\hat{T}(y) = (1 - F_Y(y^*))^{-1} \mathbb{I}_{\{y \geq y^*\}}$ for a given income level y^*.[18] That is, the total tax liability increases by the constant amount $(1 - F_Y(y^*))^{-1}$ for any income y above y^*, and the marginal tax rates are perturbed by the Dirac delta function at income y^*, i.e. $\hat{T}'(y) = (1 - F_Y(y^*))^{-1} \delta(y - y^*)$. Intuitively, this reform consists of raising the marginal tax rate at only one income level $y^* \in \mathbb{R}_+$, which implies a uniform increase in the total tax payment of agents with income $y > y^*$.[19] The normalization by $(1 - F_Y(y^*))^{-1}$ implies that the statutory increase in government revenue due to the reform (i.e., the first term on the r.h.s. of (15)) is equal to \$1. We denote by $\hat{\mathcal{R}}(y^*)$ the total effect (15) of this elementary tax reform on government revenue.[20]

Exogenous wage benchmark. In the case of exogenous wages, the incidence on government revenue is given by expression (15) with $\hat{w}(\theta) = 0$ and $\hat{l}(\theta) = -\varepsilon_r^S(\theta) \frac{\hat{T}'(y(\theta))}{1-T'(y(\theta))}$. Applying this formula to the elementary tax reform at income

[18] Note that the function $\mathbb{I}_{\{y \geq y^*\}}$ is not differentiable. We can nevertheless use our theory to analyze this reform by applying (9) to a sequence of smooth perturbations $\{\hat{T}_n'(y)\}_{n \geq 1}$ that converges to the Dirac delta function $\delta(y - y^*)$. This notation simplifies the exposition and is made only for convenience. All our formulas can be easily written for any smooth tax reform $\hat{T}$ rather than step functions.

[19] Heuristically, consider a perturbation that raises the marginal tax rate by $\mathrm{d}T'$ on a small income interval $[y^* - \mathrm{d}y, y^*]$, so that the total tax payment above income y^* raises by the amount $\mathrm{d}T' \times \mathrm{d}y$ equal to, say, \$1. This class of tax reforms has been introduced by Piketty (1997); Saez (2001). Then shrink the size of the income interval on which the tax rate is increased, i.e. $\mathrm{d}y \to 0$, while keeping the increase in the tax payment above y^* fixed at \$1. The limit of the marginal tax rate increase $\mathrm{d}T'$ is the Dirac measure at y^*, and the change in the total tax bill converges to its c.d.f., the step function $\mathbb{I}_{\{y \geq y^*\}}$.

[20] Any tax reform $\hat{T}$ can be expressed as a linear combination of such income-specific elementary perturbations: the incidence on tax revenue is given by $\hat{\mathcal{R}}(\hat{T}) = \int \hat{\mathcal{R}}(y^*)(1 - F_Y(y^*)) \hat{T}'(y^*) \, dy^*$. See Golosov, Tsyvinski, and Werquin (2014) for details.

y^* easily leads to (see Saez (2001)):

$$\hat{\mathcal{R}}_{\text{ex}}(y^*) = 1 - T'(y^*) \frac{\varepsilon_r^S(y^*)}{1 - T'(y^*)} \frac{y^* f_Y(y^*)}{1 - F_Y(y^*)}. \quad (16)$$

Equation (16) expresses the impact of an increase in the marginal tax rate at income y^* as the sum of the statutory increase in government revenue, which is normalized to \$1 by construction, and the behavioral revenue loss equal to the product of: (i) the endogenous reduction in the labor income of agent y^*, $\frac{y^*}{1-T'(y^*)}\varepsilon_r^S(y^*)$; (ii) the share $T'(y^*)$ of this income change that accrues to the government; and (iii) the hazard rate of the income distribution, $\frac{f_Y(y^*)}{1-F_Y(y^*)}$. The hazard rate is a cost-benefit ratio that measures the fraction $f_Y(y^*)$ of agents whose labor supply is distorted by the reform, relative to the fraction $1 - F_Y(y^*)$ of agents whose tax bill increases lump-sum. Note that the second term in the right hand side of (16), $\varepsilon_r^S \frac{T'}{1-T'} \frac{y^* f_Y}{1-F_Y}$, is the marginal excess burden of a tax reform: it captures how much revenue, per unit of mechanical increase in taxes, is lost through adjustments in behavior.

3.2 Effects on government revenue

We now derive and analyze the incidence of tax reforms on government revenue in general equilibrium and compare it to the expression (16) obtained assuming exogenous wages.

Proposition 2. *The incidence of the elementary tax reform at income y^* on government revenue is given by*

$$\begin{aligned} \hat{\mathcal{R}}(y^*) = {} & \hat{\mathcal{R}}_{\text{ex}}(y^*) + \frac{\varepsilon_r(y^*)}{1 - T'(y^*)} \\ & \times \int \left[T'(y^*)\left(1 + \varepsilon_w^S(y^*)\right) - T'(y)\left(1 + \varepsilon_w^S(y)\right)\right] \bar{\Gamma}(y, y^*) \frac{y f_Y(y)}{1 - F_Y(y^*)} dy \end{aligned} \quad (17)$$

where $\bar{\Gamma}(y, y^*) \equiv \left(1 + \frac{\varepsilon_w^S(y)}{\varepsilon_w^D(y)}\right)^{-1} \Gamma(y, y^*)$ are normalized resolvent cross-wage elasticities.

To understand formula (17), it is useful to first sketch its proof. The direct effect of the marginal tax rate increase at income y^* is to lower the labor supply of these agents proportionally to $\varepsilon_r(y^*)$. This induces two additional effects in general equilibrium. First, complementarities in production imply that the wage of any agent with income $y \neq y^*$ changes (say, decreases), in percentage terms, by $\Gamma(y, y^*)\,\varepsilon_r(y^*)$, so that their income decreases by $\left(1+\varepsilon_w^S(y)\right) y\,\Gamma(y,y^*)\,\varepsilon_r(y^*)$. A share $T'(y)$ of this income loss accrues to the government, leading to the second term in the square brackets of (17). Second, the non-constant marginal product of labor implies that the wage of agents with income y^* changes (say, increases), in percentage terms, by $\frac{1}{\varepsilon_w^D(y^*)}\varepsilon_r(y^*)$. Thus their income increases by $\left(1+\varepsilon_w^S(y^*)\right) y^* \frac{1}{\varepsilon_w^D(y^*)}\varepsilon_r(y^*)$, a share $T'(y^*)$ of which accrues to the government. The key step is then to sum over the whole population and apply Euler's homogeneous function theorem. Constant returns to scale imply that the own-wage gains of agents with income y^* are exactly compensated by the aggregate cross-wage losses of the other incomes $y \neq y^*$.[21] This gives an expression for the own-wage elasticity $\frac{1}{\varepsilon_w^D(y^*)}$ as an integral of the cross-wage elasticities $\Gamma(y,y^*)$ and leads to the first term in the square brackets of (17).

We now derive the economic implications of Proposition 2. To do so, assume that the labor supply elasticities $\varepsilon_w^S(\cdot)$ are constant (independent of y), which occurs if the disutility of labor is isoelastic and the initial tax schedule is CRP. Since the wage changes of all agents cancel in the aggregate by Euler's theorem, this assumption implies that the income changes of all agents also cancel once we account for the labor supply adjustments. That is, the reshuffling of wages due to the tax reform has distributional effects but keeps the economy's aggregate income constant. This observation turns out to be crucial, as we now discuss.

Linear tax schedule. Suppose first that the initial tax schedule is linear. Since the elasticities $\varepsilon_w^S(\cdot)$ are constant, they can be taken out of the integral in formula (17) and we immediately obtain that the square bracket is equal to zero. Indeed, by Euler's theorem and the fact that every agent faces the same marginal tax rate, the government's tax revenue gain coming from the higher income of agents y^* is exactly compensated by the tax revenue gains or losses coming from the rest of the population. Therefore, tax reforms have the same effect on tax revenue as in the environment with exogenous wages.

Corollary 3. *Suppose that the disutility of labor is isoelastic and the initial tax schedule is linear. Then the incidence of an arbitrary tax reform on government revenue*

[21] Euler's homogeneous function theorem in its most standard form is written in terms of the structural cross-wage elasticities $\gamma(y,y^*)$. This first round of wage changes then induces labor supply changes, which in turn lead to further rounds of own- and cross-wage effects in general equilibrium. Because Euler's theorem applies at every stage, the aggregate effect of all these wage adjustments is again equal to zero, so that the relationship can be expressed in terms of the resolvent cross-wage elasticities $\Gamma(y,y^*)$.

is identical to that obtained assuming exogenous wages: $\hat{\mathcal{R}}(y^*) = \hat{\mathcal{R}}_{\text{ex}}(y^*)$ *for all* y^*.

Non-linear tax schedule. Suppose now, more generally, that the initial tax schedule is non-linear. As above, aggregate income remains unchanged in response to a tax reform. However, the distributional implications of the tax reform now lead to non-trivial effects on government revenue – i.e., the square bracket in formula (17) is non-zero. Indeed, a zero-sum transfer of income from one agent to another is no longer neutral since these workers pay different tax rates to the government on their respective income gains and losses. To further characterize the general-equilibrium contribution to government revenue when the tax schedule is non-linear, assume that the elasticities of labor demand $\varepsilon_w^D(\cdot)$ are also constant, which occurs either when the production function is CES, or in the microfoundation of Section 1.3. The general formula of Proposition 2 can then be simplified as follows.

Corollary 4. *Suppose that the disutility of labor is isoelastic, the initial tax schedule is CRP, and the labor demand elasticities are constant. We then have*

$$\begin{aligned}\hat{\mathcal{R}}(y^*) &= \hat{\mathcal{R}}_{\text{ex}}(y^*) + \frac{\varepsilon_r}{1-T'(y^*)}\frac{y^* f_Y(y^*)}{1-F_Y(y^*)}(1+\varepsilon_w^S) \\ &\quad \times \left\{ \frac{1}{\varepsilon_w^D}\left(T'(y^*) - \mathbb{E}\left[T'(y)\right]\right) - \frac{1}{y^* f_Y(y^*)}\text{Cov}\left(T'(y)\;;\, y\,\bar{\Gamma}(y,y^*)\right)\right\}. \end{aligned} \tag{18}$$

(i) *If the production function is CES, then the covariance term on the right hand side of (18) is constant.*[22] *Letting* $\phi = \frac{1+\varepsilon_w^S}{\sigma+\varepsilon_w^S}$ *and* $\bar{T}' = \mathbb{E}\left[yT'(y)\right]/\mathbb{E}y$, *we then obtain*

$$\hat{\mathcal{R}}(y^*) = \hat{\mathcal{R}}_{\text{ex}}(y^*) + \phi\,\varepsilon_r^S\,\frac{T'(y^*) - \bar{T}'}{1-T'(y^*)}\frac{y^* f_Y(y^*)}{1-F_Y(y^*)}, \tag{19}$$

(ii) *If the production function is microfounded as in the assignment model of Section 1.3, then* $1/\varepsilon_w^D(y) = 0$ *for all* y, *so that the first term in the curly brackets of (18) is equal to zero.*

Corollary 4 delivers novel and important insights. We first discuss both special cases of formula (18) in turn and then conclude on the economic consequences of this result.

[22] This is because we then have $\bar{\Gamma}(y,y^*) = \gamma(y,y^*) = \frac{1}{\sigma\mathbb{E}[y]}y^* f_Y(y^*)$.

CES production. Consider first the case where the production function is CES. Suppose that the marginal tax rates are increasing in the initial economy, i.e., the rate of progressivity is $p > 0$. Consider a reform that raises the marginal tax rate at income y^*. Thus the labor supply of agents with income y^* decreases, which in turn raises their own wage and lowers everyone else's wage. As explained above, by Euler's homogeneous function theorem and the fact that the labor supply elasticities are constant, the resulting income gain of agents with income y^* is exactly compensated in the aggregate by the income losses of the other agents $y \neq y^*$. Now suppose that agents with income y^* are high income earners, so that their marginal tax rate $T'(y^*)$ is larger than the (income-weighted) average marginal tax rate $\bar{T}'$ in the population. Then the government's revenue gain coming from the higher income of agents y^*, which is proportional to $T'(y^*)$, more than compensates the tax revenue loss coming from the rest of the population, which is proportional to $\bar{T}'$. We therefore obtain that $\hat{\mathcal{R}}(y^*) > \hat{\mathcal{R}}_{\text{ex}}(y^*)$. Therefore, formula (19) implies that the general-equilibrium contribution of the tax reform on government revenue is positive (resp., negative) if the marginal tax rate at y^* is larger (resp., smaller) than the income-weighted average marginal tax rate in the economy. Moreover, the larger the income y^* at which the marginal tax rate is increased, the larger the gain in government revenue relative to the exogenous-wage setting. That is, "trickle-down" forces imply higher benefits of *raising*, not lowering, the marginal tax rates on high incomes.[23]

Endogenous assignment. Consider next the case where the production function is microfounded as in Section 1.3, with endogenous and costless sorting of skills into tasks. In this case, the inverse labor demand elasticities $1/\varepsilon_w^D$ are equal to zero and equation (18) implies that the general-equilibrium contribution to the excess burden of the elementary tax reform is determined by the covariance between the initial marginal tax rates $T'(\cdot)$ and the production complementarities $\bar{\Gamma}(\cdot, y^*)$ with agent y^*. If this covariance is positive (resp., negative) at a given income y^*, then the general-equilibrium forces raise (resp., lower) the cost of increasing the marginal tax rate at income y^*, compared to the exogenous-wage benchmark (16). Moreover, if this covariance is increasing with y^* (resp., decreasing), then the general-equilibrium forces raise (resp., lower) the cost of increasing the progressivity of the tax code. Section 4 evaluates this formula numerically for the calibrated values of the cross-

[23] We would obtain the opposite result if the initial tax rate were regressive (i.e., $p < 0$). The benefits of raising the top marginal tax rates would then be smaller than with exogenous wages.

wage elasticities, but we can already anticipate the qualitative results. The left panel of Figure 1 above clearly shows that the covariance between incomes and the cross-wage elasticities is positive for low values of y^* (dotted curve) and negative for large values of y^* (dashed-dotted curve). Therefore, if the marginal tax rates are initially increasing with income, the covariance term $\text{Cov}(T'(y); y\,\bar{\Gamma}(y, y^*))$ decreases with y^*. Consequently, the same qualitative insight as in the CES model holds: the general-equilibrium contribution to government revenue of a tax increase at income y^* increases with y^*. In other words, both terms in the curly brackets of formula (18) push in the same direction.

Conclusion: progressivity and trickle-down. The previous discussion implies that, starting from a progressive tax schedule, the standard partial-equilibrium formula (16) underestimates the tax revenue (or, equivalently, Rawlsian social welfare) gains from raising the marginal tax rates at the top and lowering them at the bottom. In other words, the standard model underestimates the benefits of raising the progressivity of the tax code.[24] Conversely, starting from a regressive tax schedule, the partial-equilibrium formula overestimates the gains (or underestimates the losses) from increasing marginal tax rates at the top. Thus, contrary to conventional wisdom that is based on optimal tax theory (see, e.g., Stiglitz (1982); Rothschild and Scheuer (2013) and Section 5 below), the trickle-down forces caused by the endogeneity of wages may either raise or lower the benefits of raising high-income tax rates, depending on the shape of marginal tax rates in the initial tax system. In particular, since the tax code in the U.S. is progressive (Heathcote, Storesletten, and Violante (2017)), the benefits of raising further its progressivity (i.e., of increasing the marginal tax rates on high incomes) are larger than a model with fixed wages would predict. We therefore conclude that one should be cautious, in practice, when applying the insights of the theory of optimal taxation in general equilibrium to partial reforms of a suboptimal tax code.

[24] In Appendix D, we extend Corollary 4 to the case where agents' utility functions have income effects. This adds a term that dampens the result, but does not overturn it quantitatively for empirically reasonable values of the parameters.

4 Numerical simulations

We calibrate our model to the U.S. economy and evaluate quantitatively the effects of elementary tax reforms on government revenue using formula (18). First, in Section 4.1, we assume that the production function is CES and show that the general-equilibrium effects are sizeable. Second, in Section 4.2, we combine our calibration with that of Ales, Kurnaz, and Sleet (2015) to evaluate our formulas in the environment with endogenous skill-to-task assignment. Details and robustness checks are provided in Appendix E.

4.1 Main specification

We assume that the disutility of labor $v(l)$ is isoelastic with parameter $e = 0.33$ (Chetty (2012)) and that the U.S. tax schedule is CRP with parameters $p = 0.151$ and $\tau = -3$ (Heathcote, Storesletten, and Violante (2017)). To match the U.S. yearly earnings distribution, we assume that $f_Y(\cdot)$ is log-normal with mean 10 and variance 0.95 up to income $y = \$150,000$, above which we append a Pareto distribution with a tail parameter that decreases from $\Pi \approx 2.5$ at $y = \$150,000$ to $\Pi = 1.5$ for $y \geq \$350,000$ (Diamond and Saez (2011)). We follow the approach of Saez (2001) to infer the distribution of wages from the observed earnings distribution and the individual first-order conditions (1). We extend this method to calibrate the production function: assuming a CES technology, choosing an elasticity of substitution σ is enough to pin down all the remaining parameters.

We choose an elasticity of substitution $\sigma \in \{0.6\,;\,3.1\}$. The value $\sigma = 0.6$ is taken from Dustmann, Frattini, and Preston (2013) who study the impact of immigration along the U.K. wage distribution and, as in our framework, group workers according to their position in the wage distribution.[25] The value $\sigma = 3.1$ is taken from Heathcote, Storesletten, and Violante (2017), who structurally estimate this CES parameter for the U.S. by targeting cross-sectional moments of the joint equilibrium distribution of wages, hours, and consumption.[26]

[25]This literature is a useful benchmark because it studies the impact on relative wages of exogenous labor supply shocks of certain skills, which is exactly the channel we want to analyze in our tax setting (except that for us the labor supply shocks are caused by tax reforms rather than immigration inflows).

[26]There is no clear consensus in the empirical literature on how responsive relative wages are to changes in labor supply, and therefore on the appropriate value of σ, see, e.g., the debate on the impact of immigration on wages (Peri and Yasenov (2019); Borjas (2017)). Our two values are on

Our results for the CES specification are illustrated in Figure 2. We plot the impact on government revenue of elementary tax reforms at each income level in the model with exogenous wages (solid curve, equation (16)) and in general equilibrium (dashed curve, equation (19)), as a function of the income $y(\theta)$ at which the marginal tax rate is perturbed. A value of 0.7, say, at a given income $y(\theta)$, means that for each additional dollar of tax revenue mechanically levied by the tax reform at $y(\theta)$, the government effectively gains 70 cents, while 30 cents are lost through the behavioral responses of individuals – that is, the marginal excess burden of this tax reform is 30 percent.

Consider first the solid curve: it has a U-shaped pattern which reflects the shape of the hazard ratio $\frac{y^* f_Y(y^*)}{1-F_Y(y^*)}$ in (16). This is a well-known finding in the literature (Diamond (1998); Saez (2001)). The difference between the dashed and solid curves captures the additional revenue effect due to the endogeneity of wages. In line with our analytical result of formula (19), we observe that this difference is positive for intermediate and high incomes (starting from about \$77,000, where the marginal tax rate equals its income-weighted average). Raising the marginal tax rates for these income levels is more desirable, in terms of government revenue, when the general equilibrium effects are taken into account, while the opposite holds for low income levels. The magnitude of the difference is substantial: the marginal excess burden from increasing the marginal tax rate at \$200,000 is equal to 0.22 cents per dollar if $\sigma = 0.6$ and 0.30 cents per dollar if $\sigma = 3.1$, instead of 0.34 if $\sigma = \infty$. That is, the excess burden is reduced by 35 percent if $\sigma = 0.6$ and 12 percent if $\sigma = 3.1$ due to the general equilibrium effects. Hence, the standard model with exogenous wages significantly underestimates the revenue gains from increasing the progressivity of the tax code.

4.2 Endogenous assignment

We now investigate the effects of tax reforms on government revenue in the endogenous assignment economy described in Section 1.3. We assume a Cobb-Douglas production function over tasks and set the values of its technological parameters using the estimates of Ales, Kurnaz, and Sleet (2015). As we describe in the Appendix, our calibration extends theirs to allow for an unbounded distribution of incomes with a Pareto tail, which is crucial to obtain U-shaped effects of tax reforms on government

the lower and higher sides of the typical empirical estimates.

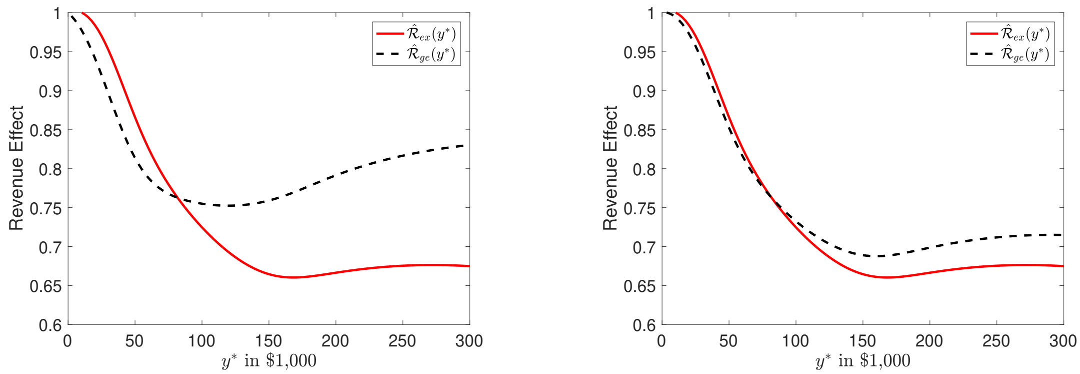


Figure 2: Revenue effects of elementary tax reforms at each income y^* (formula (17)). Solid curves: exogenous wages. Dashed curves: CES technology with $\sigma = 0.6$ (left panel) and $\sigma = 3.1$ (right panel).

revenue. This calibration allows us to compute the cross-wage elasticities, already described in Section 1.3, that enter our tax incidence formula (18). We then compare the effects of tax reforms on government revenue in this environment with those obtained in Section 4.1 assuming a CES technology with fixed assignment.

Effects of tax reforms on government revenue. Figure 3 shows the government revenue impact of elementary tax reforms at each income level. In both panels of Figure 3, the solid curve gives the revenue effects (16) in the model with exogenous wages. The dashed curve is for the model with endogenous and costless reassignment and the dashed-dotted curve is for the model with fixed assignment as in Section 4.1. We consider two calibrations for the latter model. First, in the left panel we assume a Cobb-Douglas production function ($\sigma = 1$), i.e., we shut down the reassignment channel in the calibration of Ales, Kurnaz, and Sleet (2015). Second, more relevant for our purposes, in the right panel we assume a CES production function with $\sigma = 3.1$, following the direct estimation of a technology over labor supplies of different skills by Heathcote, Storesletten, and Violante (2017).

Qualitatively, as shown analytically in Section 3.2, the fixed and endogenous assignment models deliver similar policy implications: the government revenue gains are higher (resp., lower) due to the endogeneity of wages if the marginal tax rates are raised on high (resp., low) incomes. Quantitatively, if we assume a Cobb-Douglas production function in the model with fixed assignment ($\sigma = 1$), we find that the endogenous reassignment of workers into new tasks mitigates the magnitude of the

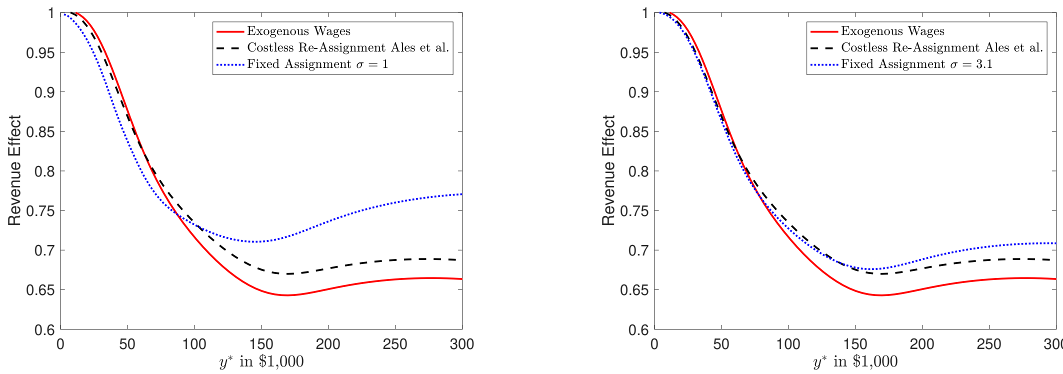


Figure 3: Revenue effects of elementary tax reforms at each income y^* (formula (17)) using the calibration of Ales et al. (2015). Solid curves: exogenous wages. Dashed curves: Cobb-Douglas technology over tasks with endogenous costless reassignment of skills to tasks. Dotted curve: CES technology with fixed assignment and $\sigma = 1$ (left panel) or $\sigma = 3.1$ (right panel).

general-equilibrium effects on revenue:[27] while still significant, they are around 30 percent of those obtained with fixed assignment if the elementary tax reform is conducted at \$200,000. However, if we use a value of σ that is directly estimated for a CES production function over skills ($\sigma = 3.1$), we obtain that the implications of tax reforms for government revenue are quantitatively closer: the effect is now 70% of that obtained with fixed assignment.

5 Optimal taxation

In this section we show that our tax incidence analysis delivers a characterization of the optimal (i.e., social welfare-maximizing) tax schedule as a by-product. We first formally introduce the social welfare criterion. We then present simple extensions of two seminal formulas to the general equilibrium environment: the optimal marginal tax rate formula of Diamond (1998) and the optimal top tax rate formula of Saez (2001).[28] The proofs and technical details are relegated to Appendix F.

[27] Rothschild and Scheuer (2013) derive this result analytically in their model.

[28] Importantly, contrary to Saez (2001), this formula indeed characterizes the optimal top tax rate only if the whole tax schedule is set optimally. Our analysis of Section 3.2 shows that it is no longer valid if the initial tax schedule is suboptimal. We return to this point in Section 5.4 below.

5.1 Welfare function and welfare weights

The government evaluates social welfare by means of a concave function $G : \mathbb{R} \to \mathbb{R}$. Letting λ denote the marginal value of public funds, we define social welfare in monetary units by

$$\mathcal{G} \;=\; \frac{1}{\lambda}\int_{\Theta} G\left(U\left(\theta\right)\right) f\left(\theta\right) d\theta.$$

The optimal tax schedule maximizes social welfare $\mathcal{G}$ subject to the constraint that government revenue $\mathcal{R}$ is non-negative.

We denote by $g\left(\theta\right)$, or equivalently $g\left(y\left(\theta\right)\right)$, the marginal social welfare weight associated with individuals of type θ:

$$g\left(\theta\right) \;=\; \frac{1}{\lambda} G'\left(U\left(\theta\right)\right). \tag{20}$$

The weight $g\left(\theta\right)$ is the social value of giving one additional unit of consumption to individuals with type θ, relative to distributing it uniformly to the whole population.

5.2 Optimal tax schedule

The effects of arbitrary tax reforms $\hat{T}$ on social welfare are easily obtained by adding the effects on government revenue (Section 3.2) to those on individual utilities (Section 2.2), weighing the latter by the marginal social welfare weights $g\left(\theta\right)$. A characterization of the optimum tax schedule can then be obtained by imposing that the welfare effects of any tax reform $\hat{T}$ of the initial tax schedule T are equal to zero. In this section, we focus on the special case of a CES production function. This implies a parsimonious generalization of the result of Stiglitz (1982) derived in a two-skill environment and the formula of Diamond (1998) derived for exogenous wages.

Proposition 3. *Assume that the production function is CES with elasticity of substitution $\sigma > 0$. Then the optimal marginal tax rate at income y^* satisfies*

$$\frac{T'\left(y^*\right)}{1-T'\left(y^*\right)} \;=\; \left(\frac{1}{\varepsilon_r^S\left(y^*\right)}+\frac{1}{\varepsilon_r^D\left(y^*\right)}\right)\left(1-\bar{g}\left(y^*\right)\right)\frac{1-F_Y\left(y^*\right)}{y^* f_Y\left(y^*\right)}+\frac{g\left(y^*\right)-1}{\sigma}, \tag{21}$$

where $\varepsilon_r^D\left(y^*\right)=\sigma$ and $\bar{g}\left(y^*\right) \equiv \mathbb{E}[g\left(y\right)|y \geq y^*]$ is the average marginal social welfare weight above income y^*.

The first term on the right hand side of (21) shows that, analogous to the optimal tax formula obtained in the model with exogenous wages (Diamond (1998), Saez (2001)), the marginal tax rate at income y^* is decreasing in the average social marginal welfare weight $\bar{g}(y^*)$, and increasing in the hazard rate of the income distribution $\frac{1-F_Y(y^*)}{y^* f_Y(y^*)}$. However, the standard inverse elasticity rule is modified: the relevant parameter is now the sum of the inverse elasticity of labor supply and the inverse elasticity of labor demand. Since $\varepsilon_w^D(y^*) = \sigma > 0$, this novel force tends to raise optimal marginal tax rates. Intuitively, increasing the marginal tax rate at y^* leads these agents to lower their labor supply, which raises their own wage and thus mitigates their behavioral response.

The second term, $(g(y^*) - 1)/\sigma$ captures the fact that the wage and welfare of agents θ^* increase in response to a higher marginal tax rate $T'(y^*)$, at the expense of all the other individuals whose wages and welfare decrease (see Section 2.2). Suppose that the government values the welfare of individuals θ^* less than average, i.e., $g(y^*) < 1$.[29] This negative externality induced by the behavior of θ^* implies that the cost of raising the marginal tax rate at y^* is higher than in partial equilibrium, and tends to lower the optimal tax rate. Conversely, the government gains by raising the optimal tax rates of individuals y^* whose welfare is valued more than average, i.e., $g(y^*) > 1$. This induces these agents to work less and earn a higher wage, which makes them strictly better off, at the expense of the other individuals in the economy, whose wage decreases. Therefore, this term creates a force for higher marginal tax rates at the bottom and lower marginal tax rates at the top if the government has a strictly concave social objective.[30]

5.3 Optimal top tax rate

We now derive the implications for the asymptotic optimal marginal tax rate. Let $\Pi > 1$ denote the Pareto coefficient of the tail of the income distribution, that is, $1 - F_Y(y) \sim c\, y^{-\Pi}$ as $y \to \infty$ for some constant c. We show that if the production function is CES and the top marginal tax rate that applies to these incomes is constant, then the tail of the income distribution has the same Pareto coefficient at the optimum as in the current data, even though the wage distribution is endogenous.

[29] Note that the average social marginal welfare weight in the economy is equal to 1.

[30] This result, as well as that of Corollary 5 below, echo those of Rothschild and Scheuer (2013) in the Roy model.

In other words, shifting up or down the top tax rate modifies wages, but the tail parameter Π of the income distribution stays constant. This leads to the following corollary.

Corollary 5. *Assume that the production function is CES with parameter $\sigma > 0$, that the disutility of labor is isoelastic with parameter e, and that incomes are Pareto distributed at the tail with coefficient $\Pi > 1$. Assume moreover that the social marginal welfare weights at the top converge to a constant $\bar{g}$. Then the top rate of the optimal tax schedule is given by*

$$\tau^* = \frac{1-\bar{g}}{1-\bar{g}+\Pi\,\varepsilon_r\,\zeta}, \quad \text{with} \quad \varepsilon_r = \frac{e}{1+\frac{e}{\sigma}} \quad \text{and} \quad \zeta = \frac{1}{1-\Pi\frac{\varepsilon_r}{\sigma}}. \tag{22}$$

In particular, τ^* is strictly smaller than the optimal top tax rate in the model with exogenous wages ($\sigma = \infty$).

Formula (22) generalizes the familiar top tax rate result of Saez (2001) (in which $\varepsilon_r = \varepsilon_r^S$ and $\zeta = 1$) to a CES production function. There is one new parameter, the elasticity of substitution between skills in production σ, that is no longer restricted to being infinite. This proposition implies a strictly lower top marginal tax rate than if wages were exogenous. Immediate calculations of the optimal top tax rate illustrate this formula.[31] Suppose that $\bar{g} = 0$, $\Pi = 2$, $e = 0.5$, and $\sigma = 1.5$.[32] We immediately obtain that the optimal tax rate on top incomes is equal to $\tau^*_{\text{ex}} = 50$ percent in the model with exogenous wages, and falls to $\tau^* = 40$ percent once the general equilibrium effects are taken into account. Suppose instead that $\Pi = 1.5$ and $e = 0.33$, then we get $\tau^*_{\text{ex}} = 66$ percent and $\tau^* = 64$ percent. In this case the trickle-down forces barely affect the optimum tax rate quantitatively.

5.4 Numerical simulations

We now provide a quantitative exploration of the optimum tax schedule (21). The left panel of Figure 4 plots the optimal marginal tax rates for a Rawlsian planner for two different values of the elasticity of substitution. It compares them to the marginal tax rates that a planner would set by applying the standard formula of Diamond (1998),

[31] Again, it is important to keep in mind that equation (22) holds only if the whole tax schedule is set optimally.

[32] These values are meant to be only illustrative but they are in the range of those estimated in the empirical literature. See the calibration in Section 4.

using the same data to calibrate the model and making the same assumptions about the utility function, but assuming that the wage distribution is exogenous. The scale on the horizontal axis is measured in income; e.g., the value of the optimal marginal tax rate at the \$100,000 mark is that of a type θ who earns $y(\theta) = \$100,000$ in the calibration to the U.S data – the income that this agent would earn in the optimal allocation would be different. The exogenous-wage optimum is U-shaped, reflecting the shape of the hazard rate of the wage distribution. When general equilibrium effects are taken into account, the optimal top tax rate is reduced (Corollary 5) and the U-shape is more pronounced.

To understand these results, the right panel of Figure 4 plots the shape of the general-equilibrium correction to optimal tax rates. We do so by applying our incidence formula (17) using the exogenous-wage optimum as our initial tax schedule (i.e., the dotted curve in the left panel of Figure 4). The red line gives the government revenue impact of elementary tax reforms under the (erroneous) assumption that wages are exogenous. In this scenario, the effect would be uniformly equal to zero by construction. The dashed curve gives the correct effect, taking into account the endogenous adjustment of wages. In this case, the gains from raising the marginal tax rates are themselves U-shaped and, except at the very bottom of the income distribution, negative. Therefore, when the low-income marginal tax rates are high (as in the exogenous-wage optimum) rather than low (as in the CRP tax code), the general equilibrium forces call for higher marginal tax rates on low incomes, and lower tax rates on intermediate and high incomes.

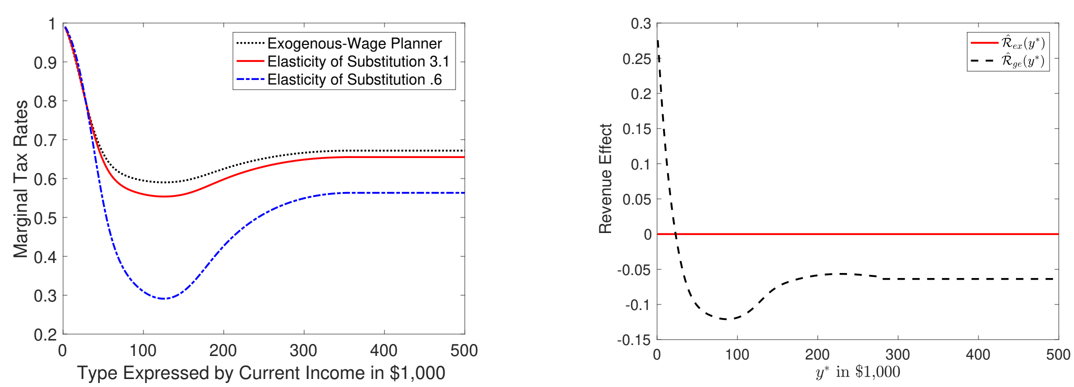


Figure 4: Left panel: Optimal marginal tax rates. Dotted curve: exogenous wages; Bold: CES technology with $\sigma = 3.1$; Dashed-dotted: $\sigma = 0.6$. Right panel: Revenue effects of elementary reforms (formula (17)) around the exogenous-wage optimum. Bold line: exogenous wages; Dashed: CES technology with $\sigma = 3.1$.

Discussion. These observations allow us to reconcile the insights of Section 3.2 (according to which the endogeneity of wages raises the benefits of increasing the marginal tax rates on high incomes) and those of Section 5.3 (according to which the optimal top tax rate is lower than in partial equilibrium). The reason for this discrepancy is that the optimum tax code is U-shaped and therefore has a form of regressivity – relatively high marginal tax rates at the bottom and relatively low marginal tax rates at the top. Instead, in Section 3.2, we analyzed partial reforms of a suboptimal tax code, with low marginal tax rates at the bottom and high rates at the top. In the latter environment, even though the overall gains of reforming the tax code always point towards the optimal U-shaped schedule, the general-equilibrium contribution to these overall gains tends to mitigate the partial-equilibrium benefits if the tax system being reformed is progressive. The converse is true if the tax schedule being reformed is regressive. Intuitively, since the fraction of the endogenous wage changes that accrues to the government is proportional to the marginal tax rate, the general-equilibrium effects of tax reforms on government revenue inherit the shape of the initial tax schedule. Therefore, the key take-away is that insights about the optimum tax schedule may actually be reversed when considering partial reforms of the current, suboptimal tax code.

6 Conclusion

We developed a variational approach for the study of nonlinear tax reforms in general equilibrium. Our methodology consisted of using the tools of the theory of integral equations to characterize: (i) the incidence of reforming a given tax schedule, e.g. the current U.S. tax code, and (ii) the optimal tax schedule. The formulas we derived are expressed in terms of sufficient statistics. The direct empirical estimation of these cross-wage elasticities is an important avenue for future research.

References

Ales, L., A. A. Bellofatto, and J. J. Wang (2017): "Taxing Atlas: Executive compensation, firm size, and their impact on optimal top income tax rates," *Review of Economic Dynamics*, 26, 62–90.

Ales, L., M. Kurnaz, and C. Sleet (2015): "Technical change, wage inequality, and taxes," *The American Economic Review*, 105, 3061–3101.

Ales, L. and C. Sleet (2016): "Taxing top CEO incomes," *American Economic Review*, 106, 3331–66.

Antras, P., A. De Gortari, and O. Itskhoki (2017): "Globalization, inequality and welfare," *Journal of International Economics*, 108, 387–412.

Bénabou, R. (2002): "Tax and Education Policy in a Heterogenous-Agent Economy: What of Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica*, 70, 481–517.

Borjas, G. J. (2017): "The wage impact of the Marielitos: A reappraisal," *ILR Review*, 70, 1077–1110.

Borjas, G. J., R. B. Freeman, L. F. Katz, J. DiNardo, and J. M. Abowd (1997): "How much do immigration and trade affect labor market outcomes?" *Brookings papers on economic activity*, 1997, 1–90.

Card, D. (1990): "The impact of the Mariel boatlift on the Miami labor market," *ILR Review*, 43, 245–257.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Östling (2017): "The effect of wealth on individual and household labor supply: evidence from Swedish lotteries," *American Economic Review*, 107, 3917–46.

Chetty, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–488.

——— (2012): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80, 969–1018.

Costinot, A. and J. Vogel (2010): "Matching and inequality in the world economy," *Journal of Political Economy*, 118, 747–786.

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

Diamond, P. A. and E. Saez (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165–90.

Dustmann, C., T. Frattini, and I. P. Preston (2013): "The effect of immigration along the distribution of wages," *The Review of Economic Studies*, 80, 145–173.

Fullerton, D. and G. E. Metcalf (2002): "Tax incidence," *Handbook of public economics*, 4, 1787–1872.

Golosov, M., A. Tsyvinski, and N. Werquin (2014): "A Variational Approach to the Analysis of Tax Systems," *Working Paper*.

Guner, N., R. Kaygusuz, and G. Ventura (2014): "Income taxation of US households: Facts and parametric estimates," *Review of Economic Dynamics*, 17, 559–581.

Guner, N., C. Rauh, and G. Ventura (2017): "Means-Tested Transfers in the US: Facts and Parametric Estimates," *Working Paper*.

Harberger, A. C. (1962): "The incidence of the corporation income tax," *The Journal of Political Economy*, 215–240.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): "Optimal tax progressivity: An analytical framework," *The Quarterly Journal of Economics*, 132, 1693–1754.

Hendren, N. (2015): "The Policy Elasticity," in *Tax Policy and the Economy, Volume 30*, University of Chicago Press.

Hines, J. R. (2009): "Peter Mieszkowski and the General Equilibrium Revolution in Public Finance," in *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, vol. 102, 213–216.

Imbens, G. W., D. B. Rubin, and B. I. Sacerdote (2001): "Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players," *American Economic Review*, 778–794.

Jacquet, L. and E. Lehmann (2017): "Optimal Income Taxation with Composition Effects," *IZA Discussion Paper No. 11019.*

Jones, C. I. (2019): "Taxing Top Incomes in a World of Ideas," *Working Paper.*

Kotlikoff, L. J. and L. H. Summers (1987): "Tax incidence," *Handbook of public economics*, 2, 1043–1092.

Kushnir, A. and R. Zubrickas (2018): "Optimal Income Taxation with Endogenous Prices," *Working Paper.*

Matsuyama, K. and P. Ushchev (2017): "Beyond CES: Three Alternative Cases of Flexible Homothetic Demand Systems," *Buffett Institute Global Poverty Research Lab Working Paper.*

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

Peri, G. and V. Yasenov (2019): "The Labor Market Effects of a Refugee Wave Synthetic Control Method Meets the Mariel Boatlift," *Journal of Human Resources*, 54, 267–309.

Piketty, T. (1997): "La redistribution fiscale face au chomage," *Revue française d'économie*, 12, 157–201.

Piketty, T. and E. Saez (2013): "A Theory of Optimal Inheritance Taxation," *Econometrica*, 81, 1851–1886.

Piketty, T., E. Saez, and S. Stantcheva (2014): "Optimal taxation of top labor incomes: A tale of three elasticities," *American Economic Journal: Economic Policy*, 6, 230–271.

Polyanin, A. D. and A. V. Manzhirov (2008): *Handbook of integral equations*, CRC press.

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

——— (2014): "A theory of income taxation under multidimensional skill heterogeneity," *NBER Working Paper No. 19822.*

——— (2016): "Optimal taxation with rent-seeking," *The Review of Economic Studies*, 83, 1225–1262.

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

Sattinger, M. (1975): "Comparative advantage and the distributions of earnings and abilities," *Econometrica*, 455–468.

Scheuer, F. (2014): "Entrepreneurial Taxation with Endogenous Entry," *American Economic Journal: Economic Policy*, 6, 126–163.

Scheuer, F. and I. Werning (2016): "Mirrlees meets Diamond-Mirrlees," *NBER Working Paper No. 22076*.

——— (2017): "The taxation of superstars," *The Quarterly Journal of Economics*, 132, 211–270.

Shoven, J. B. and J. Whalley (1984): "Applied general-equilibrium models of taxation and international trade: an introduction and survey," *Journal of Economic Literature*, 22, 1007–1051.

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

Teulings, C. N. (1995): "The wage distribution in a model of the assignment of skills to jobs," *Journal of Political Economy*, 280–315.

——— (2005): "Comparative advantage, relative wages, and the accumulation of human capital," *Journal of Political Economy*, 113, 425–461.

Tricomi, F. G. (1985): *Integral equations*, vol. 5, Courier Corporation.

Zemyan, S. M. (2012): *The classical theory of integral equations: a concise treatment*, Springer Science & Business Media.

Online Appendix for "Nonlinear Tax Incidence and Optimal Taxation in General Equilibrium"

Dominik Sachs, Aleh Tsyvinski, and Nicolas Werquin

A Proofs of Section 1

A.1 Reduced-form production function

Labor supply elasticities. The first-order condition (1) can be rewritten as $v'(l(\theta)) = r(\theta)w(\theta)$, where $r(\theta) = 1 - T'(w(\theta)l(\theta))$ is the retention rate of agent θ. Ignoring the endogeneity of $r(\theta)$ and applying the implicit function theorem (IFT) to this equation gives the labor supply elasticity along the linear budget constraint, $e(\theta) = \frac{r(\theta)}{l(\theta)}\frac{\partial l(\theta)}{\partial r(\theta)} = \frac{v'(l(\theta))}{l(\theta)v''(l(\theta))}$. Applying the IFT again but accounting for the endogeneity of $T'(w(\theta)l(\theta))$ to labor supply – i.e., taking a first-order Taylor expansion of the perturbed first-order condition

$$v'(l(\theta) + \delta l(\theta)) = [1 - T'(w(\theta)(l(\theta) + \delta l(\theta))) - \delta r(\theta)]\, w(\theta)$$

and solving for $\delta l(\theta)$ – leads to the expression (7) for the labor supply elasticity along the nonlinear budget constraint $\varepsilon_r^S(\theta)$. The elasticity with respect to the wage, $\varepsilon_w^S(\theta)$, can be derived analogously. Throughout the paper we make the following assumption.

Assumption 1. *The first-order condition (1) has a unique solution $l(\theta)$. For all $\theta \in \Theta$, we have $|p(y(\theta))e(\theta)| < 1$ and $\left|\varepsilon_w^S(\theta)/\varepsilon_w^D(\theta)\right| < 1$, where the labor supply and demand elasticities $e(\theta), \varepsilon_w^S(\theta), \varepsilon_w^D(\theta)$ are defined in Section 1.2.*

As in the partial-equilibrium environment with exogenous wages, the uniqueness of the solution to the individual first-order condition allows us to apply the IFT. The condition $|p(y(\theta))e(\theta)| < 1$ ensures that the elasticities $\varepsilon_r^S(\theta), \varepsilon_w^S(\theta)$ are well-defined. Specifically, the condition $p(y(\theta))e(\theta) > -1$ ensures that the second-order condition of the individual problem is satisfied. The condition $p(y(\theta))e(\theta) < 1$ ensures the convergence of the labor supply responses towards the fixed point that characterizes the elasticities along the nonlinear budget constraint. Finally, the condition $\left|\varepsilon_w^S(\theta)/\varepsilon_w^D(\theta)\right| < 1$ ensures that the equilibrium labor elasticities $\varepsilon_w(\theta)$ introduced in Lemma 1 are well defined.

□

One-to-one map between skills, wages, and incomes. Without loss of generality, we order skills θ so that wages $w(\theta)$ are strictly increasing in θ in the initial equilibrium. Next, note that the individual first-order condition (1) implies that the elasticity of income with respect to the wage is given by $\frac{w(\theta)}{y(w(\theta))} y'(w(\theta)) = 1 + \varepsilon_w^S(\theta)$, so that incomes are strictly increasing in wages if and only if $\varepsilon_w^S(\theta) > -1$, or equivalently $e(\theta) > -1$, which is equivalent to the Spence-Mirrlees condition. Hence, imposing the Spence-Mirrlees condition implies that there is a one-to-one map between incomes $y(\theta)$ and skills θ .

Importantly, note that for our analysis we do not need to impose that this monotone mapping is preserved after the tax reform is implemented because the reforms we consider are marginal. Nevertheless, we now show that when the production function is CES, this ordering remains satisfied after any, possibly non-local, tax reform. This is useful because it implies that the ordering of types does not change between the wage distribution calibrated using current data and the one implied by the optimal tax schedule. Without loss of generality we assume that types are uniformly distributed on the unit interval $\Theta = [0, 1]$, so that $f(\theta) = 1$ for all θ. For a CES production function, we have

$$\frac{w'(\theta)}{w(\theta)} = \frac{a'(\theta)}{a(\theta)} - \frac{1}{\sigma}\frac{l'(\theta)}{l(\theta)} = \frac{a'(\theta)}{a(\theta)} - \frac{\varepsilon_w^S(\theta)}{\sigma}\frac{w'(\theta)}{w(\theta)}.$$

Assumption 1 above implies $1 + \varepsilon_w^S(\theta)/\sigma > 0$, so that the sign of $w'(\theta)$ is the same as that of $a'(\theta)$ independently of the tax system.

□

Lemma 2 (**Euler's homogeneous function theorem**). *The following relationship between the own-wage elasticity and the structural cross-wage elasticities is satisfied for all y^*:*

$$-\frac{1}{\varepsilon_w^D(y^*)} y^* f_Y(y^*) + \int_{\mathbb{R}_+} \gamma(y, y^*) y f_Y(y)\, dy = 0, \tag{23}$$

where we define $\gamma(y(\theta), y(\theta')) \equiv (y'(\theta'))^{-1} \gamma(\theta, \theta')$. Equivalently, this can be expressed in terms of the resolvent cross-wage elasticities:

$$-\frac{1}{\varepsilon_w^D(y^*)} y^* f_Y(y^*) + \int_{\mathbb{R}_+} \frac{\Gamma(y, y^*)}{1 + \varepsilon_w^S(y)/\varepsilon_w^D(y)} y f_Y(y)\, dy = 0. \tag{24}$$

Proof of Lemma 2. Constant returns to scale imply $\frac{1}{\varepsilon_w^D(\theta')} y(\theta') f(\theta') = \int_\Theta \gamma(\theta, \theta') y(\theta)\, dF(\theta)$ for all θ'. Changing variables from types θ to incomes $y(\theta)$ leads to (23). Now this equation implies that

$$\begin{aligned}\int_\Theta \frac{\hat{w}(\theta)}{w(\theta)} y(\theta) f(\theta)\, d\theta &= \int_\Theta \left[-\frac{1}{\varepsilon_w^D(\theta)} \frac{\hat{l}(\theta)}{l(\theta)} + \int_\Theta \gamma(\theta, \theta') \frac{\hat{l}(\theta')}{l(\theta')} d\theta' \right] y(\theta) f(\theta)\, d\theta \\ &= -\int_\Theta \left[\frac{1}{\varepsilon_w^D(\theta)} y(\theta) f(\theta) + \int_\Theta \gamma(\theta', \theta) y(\theta') f(\theta')\, d\theta' \right] \frac{\hat{l}(\theta)}{l(\theta)} d\theta = 0.\end{aligned}$$

We can use equation (13) to substitute for $\frac{\hat{w}(\theta)}{w(\theta)}$ in the previous equality, and then equation (9) to substitute for $\frac{\hat{l}(\theta)}{l(\theta)}$. Applying the formula to the elementary tax reform $\hat{T}'(y) = \delta(y - y^*)$ and changing variables from skills to incomes leads to:

$$0=\int_{\mathbb{R}_+}\frac{1}{\varepsilon_w^S(y)}\left[\varepsilon_r^S(y)\frac{\delta(y-y^*)}{1-T'(y)}-\left(\varepsilon_r(y)\frac{\delta(y-y^*)}{1-T'(y)}+\varepsilon_w(y)\frac{\Gamma(y,y^*)\varepsilon_r(y^*)}{1-T'(y^*)}\right)\right]yf(y)\,dy.$$

This easily leads to formula (24).

□

Formulas for CES technology. Wages are $w(\theta)=a(\theta)(L(\theta))^{-\frac{1}{\sigma}}\left[\int_\Theta a(x)(L(x))^{\frac{\sigma-1}{\sigma}}dx\right]^{\frac{1}{\sigma-1}}$, so that the cross-wage and own-wage elasticities are given by

$$\gamma(\theta,\theta')=\frac{1}{\sigma}\frac{a(\theta')(L(\theta'))^{\frac{\sigma-1}{\sigma}}}{\int_\Theta a(x)(L(x))^{\frac{\sigma-1}{\sigma}}dx}\quad\text{and}\quad\frac{1}{\varepsilon_w^D(\theta)}=\frac{1}{\sigma}. \tag{25}$$

This implies in particular, for all $\theta\in\Theta$, $\int_\Theta\gamma(\theta,\theta')\,d\theta'=\frac{1}{\sigma}$. Applying Euler's homogeneous function theorem to rewrite expression (25) for $\gamma(\theta,\theta')$ and changing variables leads to:

$$\gamma(y,y')=\frac{1}{\sigma}\frac{y'f_Y(y')}{\int_{\mathbb{R}_+}xf_Y(x)\,dx}. \tag{26}$$

Assume in addition that the disutility of labor is isoelastic with parameter e and that the initial tax schedule is CRP with parameter p. The labor supply elasticities (7) and the equilibrium labor elasticities (introduced in Lemma 1) are then all constant and given by $\varepsilon_r^S(y)=\frac{e}{1+pe}$, $\varepsilon_w^S(y)=\frac{(1-p)e}{1+pe}$, $\varepsilon_r(y)=\frac{e}{1+pe+(1-p)\frac{e}{\sigma}}$, $\varepsilon_w(y)=\frac{(1-p)e}{1+pe+(1-p)\frac{e}{\sigma}}$.

□

Relationship with Scheuer and Werning (2016, 2017). These papers analyze a general equilibrium extension of Mirrlees (1971) and prove a neutrality result: in their model, the optimal tax formula is the same as in partial equilibrium, even though they consider a more general production function than Mirrlees (1971).[33] The key modeling difference between our framework and theirs is the following. In theirs, all the agents produce the same input with different productivities θ. Denoting by $\eta(\theta)=\theta l(\theta)$ the agent's production of that input (i.e., the efficiency units of labor), the aggregate production function then maps the distribution of η into output. In equilibrium, a nonlinear price (earnings) schedule $p(\cdot)$ emerges such that an agent who produces η units earns income $p(\eta)$, irrespective of the underlying productivity θ. Hence, when an (atomistic) individual θ provides more effort $l(\theta)$, income moves along the non-linear schedule $l\mapsto p(\theta\times l)$; e.g., in their superstars model with a convex equilibrium earnings schedule, income increases faster than linearly. By contrast, in our framework, different values of θ index different inputs in the aggregate production function; for each of these inputs, there is one specific price (wage) $w(\theta)$, and hence a linear earnings schedule $l\mapsto w(\theta)\times l$. Therefore, when an individual θ provides more effort $l(\theta)$, income increases linearly, as the wage remains constant (since the sector θ doesn't change). In their framework,

[33] The policy implications can nevertheless be different. For instance, in Scheuer and Werning (2017), the relevant earnings elasticity in the formula written in terms of the observed income distribution is higher due to the superstar effects.

Scheuer and Werning show that the general equilibrium effects exactly cancel out at the optimum tax schedule, even though they would of course be non-zero in the characterization of the incidence effects of tax reforms around a suboptimal tax code. In our framework, as in those of Stiglitz (1982); Rothschild and Scheuer (2014); Ales, Kurnaz, and Sleet (2015), these general equilibrium forces are also present at the optimum.[34]

□

A.2 Microfoundation of the production function

Our microfoundation of the production function $Y = \mathcal{F}(\{L(\theta)\}_{\theta\in\Theta})$ extends the Costinot and Vogel (2010) model of endogenous assignment of skills to tasks to incorporate endogenous labor supply choices by agents and nonlinear labor income taxes. There is a continuum of mass one of agents indexed by their skill $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$ and a continuum of tasks (e.g., manual, routine, abstract, etc.) indexed by their skill intensity, $\psi \in \Psi = [\underline{\psi}, \bar{\psi}]$. Let $A(\theta,\psi)$ be the product of a unit of labor of skill θ employed in task ψ. We assume that high-skill workers have a comparative advantage in tasks with high skill intensity, i.e., $A(\theta,\psi)$ is strictly log-supermodular: $A(\theta',\psi')\,A(\theta,\psi) > A(\theta,\psi')\,A(\theta',\psi)$ for all $\theta' > \theta$ and $\psi' > \psi$.

Individuals. Agents with skill θ earn wage $w(\theta)$ which they take as given. Labor supply satisfies (1). We denote by $c(\theta)$ the agent's consumption of the final good.

Final good firm. The final good Y is produced using as inputs the output $Y(\psi)$ of each task $\psi \in \Psi$ with the following CES production function:

$$Y = \left\{\int_{\underline{\psi}}^{\bar{\psi}} B(\psi)\,[Y(\psi)]^{\frac{\sigma-1}{\sigma}}\,d\psi\right\}^{\frac{\sigma}{\sigma-1}}.$$

The final good firm chooses the quantities of inputs $Y(\psi)$ of each type ψ to maximize its profit $Y - \int_\Psi p(\psi)\,Y(\psi)\,d\psi$, where $p(\psi)$ is the price of task ψ which the firm takes as given. The first-order conditions read: $\forall\psi\in\Psi$,

$$Y(\psi) = [p(\psi)]^{-\sigma}[B(\psi)]^{\sigma}Y. \tag{27}$$

Intermediate good firms. The output of task ψ is produced linearly by intermediate firms that hire the labor $L(\theta \mid \psi)$ of skills $\theta \in \Theta$ that they hire, so that

$$Y(\psi) = \int_\Theta A(\theta,\psi)\,L(\theta\mid\psi)\,d\theta.$$

[34] Another perspective to understand the distinction between our two papers is as a difference in the utility function. In Scheuer-Werning, individuals can pick one element within the set of effective labor $H = \mathbb{R}_+^*$. In our setting, each element of H corresponds to one type θ, different types of individuals supply different kinds of effective labor and choose the quantity with which they supply this variety. We are grateful to an anonymous referee for suggesting this interpretation.

The intermediate good firm of type ψ chooses its demand for labor $L(\theta)$ of each skill θ to maximize its profit $p(\psi) Y(\psi) - \int_\Theta w(\theta) L(\theta \mid \psi) d\theta$ taking the wage $w(\theta)$ as given. The first-order condition implies that this firm is willing to hire any quantity of labor that is supplied by the workers of type θ as long as their wage is given by

$$w(\theta) = p(\psi) A(\theta, \psi), \text{ if } L(\theta \mid \psi) > 0. \tag{28}$$

Moreover, the wage of any skill θ that is not employed in task ψ must satisfy

$$w(\theta) \geq p(\psi) A(\theta, \psi), \text{ if } L(\theta \mid \psi) = 0. \tag{29}$$

Market clearing. We first impose that the market for the final good market clears. This condition reads $Y = \int_\Theta c(\theta) f(\theta) d\theta + \mathcal{R}$, where f the density of skills $\theta \in \Theta$ in the population and $\mathcal{R} \equiv \int_\Theta T(w(\theta) l(\theta)) f(\theta) d\theta$ is the government revenue which is used to buy the final good. Using the agents' and the government budget constraints, this can be rewritten as:

$$Y = \int_\Theta w(\theta) l(\theta) f(\theta) d\theta. \tag{30}$$

Second, we impose that the market for each intermediate good $\psi \in \Psi$ clears. For simplicity, we assume at the outset that there is a one-to-one matching function $M : \Theta \to \Psi$ between skills and tasks – we show below that it is indeed the case in equilibrium. Letting $\psi = M(\theta)$ be the task assigned to skill θ, we must then have $\int_{\underline{\psi}}^{M(\theta)} Y(\psi) d\psi = \int_{\underline{\theta}}^{\theta} A(\theta', M(\theta')) L(\theta' \mid M(\theta')) d\theta'$, or simply $Y(\psi) d\psi = A(\theta, M(\theta)) L(\theta \mid M(\theta)) d\theta$. This implies: $\forall \theta \in \Theta$,

$$Y(M(\theta)) M'(\theta) = A(\theta, M(\theta)) L(\theta \mid M(\theta)). \tag{31}$$

Formally, this condition is obtained by substituting for $L(\theta \mid \psi) = \delta_{\{\psi = M(\theta)\}}$ in the equation $Y(\psi) = \int_\Theta A(\theta, \psi) L(\theta \mid \psi) d\theta$, where δ is the dirac delta function, and changing variables from skills to tasks to compute the integral. Third, we impose that the market for labor of each skill $\theta \in \Theta$ clears: $\forall \theta \in \Theta$,

$$l(\theta) f(\theta) = L(\theta \mid M(\theta)). \tag{32}$$

Competitive equilibrium. Given a tax function $T : \mathbb{R}_+ \to \mathbb{R}$, an equilibrium consists of a schedule of labor supplies $\{l(\theta)\}_{\theta \in \Theta}$, labor demands $\{L(\theta \mid \psi)\}_{\theta \in \Theta, \psi \in \Psi}$, intermediate goods $\{Y(\psi)\}_{\psi \in \Psi}$, final good Y, wages $\{w(\theta)\}_{\theta \in \Theta}$, prices $\{p(\psi)\}_{\psi \in \Psi}$, and a matching function $M : \Theta \to \Psi$ such that equations (1), (27), (28), (29), (30), (31), (32) hold.

Equilibrium assignment. The first part of the analysis consists of proving the existence of the continuous and strictly increasing one-to-one matching function $M : \Theta \to \Psi$ with $M(\underline{\theta}) = \underline{\psi}$ and $M(\bar{\theta}) = \bar{\psi}$. That is, there is positive assortative matching. The proof is identical to that in Costinot and Vogel (2010). The second part of the analysis consists of characterizing the matching function and the wage schedule. We find

$$M'(\theta) = \frac{A(\theta, M(\theta))\, l(\theta)\, f(\theta)}{[p(M(\theta))]^{-\sigma}[B(M(\theta))]^{\sigma} Y} \tag{33}$$

with $M(\underline{\theta}) = \underline{\psi}$ and $M(\bar{\theta}) = \bar{\psi}$, and where Y is given by (30) and $p(M(\theta))$ is given by (28).

$$\frac{w'(\theta)}{w(\theta)} = \frac{A_1'(\theta, M(\theta))}{A(\theta, M(\theta))}. \tag{34}$$

Equation (33), which characterizes the equilibrium matching as the solution to a nonlinear differential equation, is a direct consequence of the market clearing equation (31), in which we use (27) to substitute for $Y(M(\theta))$. Equation (34), which characterizes the equilibrium wage schedule, is a consequence of the firms' profit maximization conditions (28) and follows the same steps as Costinot and Vogel (2010).

Reduced form production function. Equilibrium assignment of skills to tasks is endogenous to taxes. We denote by $M(\cdot \mid T) : \Theta \to \Psi$ the matching function with T as an explicit argument. The main result, for our purposes, is that the tax schedule T affects the equilibrium assignment only through its effect on agents' labor supply choices $\mathscr{L} \equiv \{l(\theta) f(\theta)\}_{\theta \in \Theta}$. Indeed, note that none of the equations (27)-(32), which define the equilibrium for given labor supply levels $\{l(\theta)\}_{\theta\in\Theta}$, depend directly on T. This implies that if two distinct tax schedules lead to the same equilibrium labor supply choices $\mathscr{L}$, they will also lead to the same assignment of skills to tasks M. Therefore, the matching function $M(\cdot \mid T)$ can be rewritten as $M(\cdot \mid \mathscr{L})$. This result implies that the model can be summarized by a reduced-form production function $\mathcal{F}(\mathscr{L})$ over the labor supplies of different skills in the population. To see this, note that the production function (over tasks) of the final good can be written as

$$\begin{aligned} Y &= \left\{\int_{\underline{\psi}}^{\bar{\psi}} B(\psi)[Y(\psi)]^{\frac{\sigma-1}{\sigma}} d\psi\right\}^{\frac{\sigma}{\sigma-1}} = \left\{\int_{\underline{\theta}}^{\bar{\theta}} B(M(\theta))[Y(M(\theta))]^{\frac{\sigma-1}{\sigma}} M'(\theta)\, d\theta\right\}^{\frac{\sigma}{\sigma-1}} \\ &= \left\{\int_{\underline{\theta}}^{\bar{\theta}} a(\theta, M)[l(\theta) f(\theta)]^{\frac{\sigma-1}{\sigma}} d\theta\right\}^{\frac{\sigma}{\sigma-1}}, \end{aligned} \tag{35}$$

where $a(\theta, M) \equiv B(M(\theta))[A(\theta, M(\theta))]^{\frac{\sigma-1}{\sigma}}[M'(\theta)]^{\frac{1}{\sigma}}$.[35] The second equality follows from a change of variables from tasks to skills using the one-to-one map M between the two variables, and the third equality uses the market clearing conditions (31) and (32) to substitute for $Y(M(\theta))$. Equation (35) defines a production function over skills $\theta \in \Theta$. This production function inherits the CES structure of the original production function, except that the technological coefficients $a(\theta, M)$ are now endogenous to taxes since they depend on the matching function M. We can write (35) as a

[35] Note that, of course, this reduced-form production function is consistent with the wage schedule derived above. We find that $w(\theta) = B(M(\theta))\, A(\theta, M(\theta))\, [\frac{Y}{Y(M(\theta))}]^{1/\sigma}$ by combining (28) and (27). Differentiating the reduced-form production function (35) with respect to $l(\theta) f(\theta)$ and using (31) leads to the same expression.

function $\tilde{\mathcal{F}}(\{l(\theta) f(\theta)\}_{\theta\in\Theta}, M) \equiv \tilde{\mathcal{F}}(\mathscr{L}, M)$. Now, using the result proved above that the function $M \equiv M(\cdot \mid \mathscr{L})$ depends on taxes only through the equilibrium labor supplies $\mathscr{L}$, we finally obtain the following reduced form production function:

$$Y = \mathcal{F}(\mathscr{L}). \tag{36}$$

Using the reduced-form production function (36), all of the results we have derived go through. We can still define wages and the cross-wage elasticities as $w(\theta) = \frac{\partial \mathcal{F}(\mathscr{L})}{\partial [l(\theta) f(\theta)]}$ and $\gamma(\theta, \theta') \equiv \frac{\partial \ln w(\theta)}{\partial \ln[l(\theta') f(\theta')]}$. These cross-wage elasticities are defined as the impact of an exogenous shock to the supply of labor of type θ' (e.g., an immigration inflow) on the wage of type θ, keeping everyone's labor supply constant otherwise, but allowing for the endogenous re-assignment of skills to tasks following this exogenous shock. Indeed, the reduced-form production function $\mathcal{F}$ accounts for the dependence of the matching function on agents' labor supplies.

B Proofs of Section 2

Proof of Lemma 1 and Corollary 2. Denote the perturbed tax function by $\tilde{T}(y) = T(y) + \mu \hat{T}(y)$ and by $\hat{l}(\theta)$ the Gateaux derivative of the labor supply of type θ in response to this perturbation. The labor supply response of type θ is given by the solution to the perturbed first-order condition

$$\begin{aligned} 0 =& v'\left(l(\theta) + \mu \hat{l}(\theta)\right) \\ & - \left\{1 - T'\left[\tilde{w}(\theta) \times \left(l(\theta) + \mu \hat{l}(\theta)\right)\right] - \mu \hat{T}'\left[\tilde{w}(\theta) \times \left(l(\theta) + \mu \hat{l}(\theta)\right)\right]\right\} \tilde{w}(\theta), \end{aligned} \tag{37}$$

where $\tilde{w}(\theta)$ is the perturbed wage schedule, which satisfies

$$\begin{aligned} \frac{\tilde{w}(\theta) - w(\theta)}{\mu} &= \frac{1}{\mu}\left\{\mathscr{F}'_\theta(\{(l(\theta') + \mu \hat{l}(\theta')) f(\theta')\}_{\theta'\in\Theta}) - \mathscr{F}'_\theta(\{l(\theta') f(\theta')\}_{\theta'\in\Theta})\right\} \\ \underset{\mu\to 0}{=} \quad \mathscr{F}'_\theta \int_\Theta \frac{L(\theta') \mathscr{F}''_{\theta,\theta'}}{\mathscr{F}'_\theta} \frac{\hat{l}(\theta')}{l(\theta')} d\theta' &= w(\theta)\left[-\frac{1}{\varepsilon_w^D(\theta)} \frac{\hat{l}(\theta)}{l(\theta)} + \int_\Theta \gamma(\theta, \theta') \frac{\hat{l}(\theta')}{l(\theta')} d\theta'\right]. \end{aligned} \tag{38}$$

Taking a first-order Taylor expansion of the perturbed first-order conditions (37) around the baseline allocation, using (38) to substitute for $\tilde{w}(\theta) - w(\theta)$, and solving for $\hat{l}(\theta)$ yields

$$\begin{aligned} &\left\{1 + \frac{1 - T'(y(\theta)) - y(\theta) T''(y(\theta))}{1 - T'(y(\theta)) + \frac{v'(l(\theta))}{l(\theta) v''(l(\theta))} y(\theta) T''(y(\theta))} \frac{v'(l(\theta))}{l(\theta) v''(l(\theta))} \frac{1}{\varepsilon_w^D(\theta)}\right\} \frac{\hat{l}(\theta)}{l(\theta)} \\ =& \frac{1 - T'(y(\theta)) - y(\theta) T''(y(\theta))}{1 - T'(y(\theta)) + \frac{v'(l(\theta))}{l(\theta) v''(l(\theta))} y(\theta) T''(y(\theta))} \frac{v'(l(\theta))}{l(\theta) v''(l(\theta))} \int_\Theta \gamma(\theta, \theta') \frac{\hat{l}(\theta')}{l(\theta')} d\theta' \\ & - \frac{1}{1 - T'(y(\theta)) + \frac{v'(l(\theta))}{l(\theta) v''(l(\theta))} y(\theta) T''(y(\theta))} \frac{v'(l(\theta))}{l(\theta) v''(l(\theta))} \hat{T}'(y(\theta)), \end{aligned}$$

which leads to equation (8). Equation (13) follows easily from (38). Substituting into (8) leads to formula (13). Equation (14) follows by taking the Gateaux derivative of the agent's indirect utility and using the first order condition (1).

□

Proof of Proposition 1. Equation (8) is a Fredholm integral equation of the second kind. Assume that the condition $\int_{\Theta^2} |\varepsilon_w(\theta)\gamma(\theta,\theta')|^2 \, d\theta d\theta' < 1$ holds. Theorem 2.3.1 in Zemyan (2012) gives the unique solution (9) to this equation.

□

Proof of equation (12). Suppose that the cross-wage elasticities are multiplicatively separable, i.e., of the form $\gamma(\theta,\theta') = \gamma_1(\theta)\gamma_2(\theta')$. Theorem 1.3.1 in Zemyan (2012) (or 4.9.1 in Polyanin and Manzhirov (2008)) gives the solution to the integral equation (9). If the production function is CES, we have $\gamma_1(\theta) = 1$ and $\gamma_2(\theta) = \frac{1}{\sigma \mathbb{E} y} y(\theta) f_Y(y(\theta)) y'(\theta)$. A change of variables from skills θ to incomes $y(\theta)$ easily leads to (12). Note that this solution is well defined if $\frac{1}{\sigma \mathbb{E} y}\mathbb{E}[y\varepsilon_w(y)] < 1$.

□

Sufficient conditions ensuring the convergence of the resolvent (10). Suppose that the production function is CES with parameter σ, that the disutility of labor is isoelastic with parameter e, and that the initial tax schedule is CRP with parameter $p < 1$. Corollary 1 implies that the resolvent series converges if

$$\frac{1}{\sigma \mathbb{E} y}\mathbb{E}[y\varepsilon_w(y)] = \frac{(1-p)e}{1+pe+(1-p)\frac{e}{\sigma}} < 1,$$

where we used the expression for $\varepsilon_w(y)$ derived in Section A.1 above. Since $(1-p)e > 0$, this condition is satisfied if $1+pe > 0$. Recall that this condition is the second-order condition of the individual problem, which is satisfied by Assumption 1 above. In particular, in the calibration to the U.S. economy, we have $p = 0.15 > 0 > -\frac{1}{e} \approx -3$ so this clearly holds.

□

C Proofs of Section 3

Elementary tax reforms. Suppose that the tax reform T is the step function $T(y) = \mathbb{I}_{\{y \geq y^*\}}$, so that $T'(y) = \delta(y-y^*)$ is the Dirac delta function – that is, marginal tax rates are perturbed at income y^* only. To apply formula (9) to this non-differentiable perturbation, construct a sequence of smooth funtions $\varphi_{y^*,\epsilon}(y)$ such that $\delta(y-y^*) = \lim_{\epsilon\to 0}\varphi_{y^*,\epsilon}(y)$, in the sense that for all continuous functions ψ with compact support, $\psi(y^*) = \lim_{\epsilon\to 0}\int_{\mathbb{R}}\varphi_{y^*,\epsilon}(y)\psi(y)\,dy = \lim \int_{\Theta}\varphi_{y^*,\epsilon}(y(\theta'))\{\psi(y(\theta'))y'(\theta')\}\,d\theta'$, where the second equality follows from a change of variables in the integral. This can be obtained by defining an absolutely integrable and smooth function $\varphi_{y^*}(y)$ with compact support and $\int_{\mathbb{R}}\varphi_{y^*}(y)\,dy = 1$, and letting $\varphi_{y^*,\epsilon}(y) = \epsilon^{-1}\varphi_{y^*}(\frac{y}{\epsilon})$. Letting

$\Phi_{y^*,\epsilon}$ be such that $\Phi'_{y^*,\epsilon} = \varphi_{y^*,\epsilon}$, we then have, for all $\epsilon > 0$, the following labor supply incidence formula:

$$\hat{l}(\theta, \Phi_{y^*,\epsilon}) = -\varepsilon_r(\theta) \frac{\varphi_{y^*,\epsilon}(y(\theta))}{1 - T'(y(\theta))} - \varepsilon_w(\theta) \int_\Theta \Gamma(\theta, \theta') \varepsilon_r(\theta') \frac{\varphi_{y^*,\epsilon}(y(\theta'))}{1 - T'(y(\theta'))} d\theta'.$$

Letting $\epsilon \to 0$, we obtain the incidence of the elementary tax reform at y^*:

$$\begin{aligned} \hat{l}(\theta) &= -\varepsilon_r(\theta) \frac{\delta_{y^*}(y(\theta))}{1 - T'(y(\theta))} - \varepsilon_w(\theta) \frac{\Gamma(\theta, \theta^*)}{y'(\theta^*)} \varepsilon_r(\theta^*) \frac{1}{1 - T'(y(\theta^*))} \\ &= -\varepsilon_r(y) \frac{\delta_{y^*}(y)}{1 - T'(y)} - \varepsilon_w(y) \Gamma(y, y^*) \varepsilon_r(y^*) \frac{1}{1 - T'(y^*)}, \end{aligned} \tag{39}$$

where in the last equality we let $y = y(\theta)$ and $y^* = y(\theta^*)$, and we use the change of variables $\Gamma(y, y^*) = \frac{\Gamma(\theta,\theta^*)}{y'(\theta^*)}$.

□

Proof of Proposition 2 and Corollary 3. The first-order effects of a tax reform $\hat{T}$ on individual θ's tax payment are given by $\hat{T}(y(\theta)) + [\frac{\hat{w}(\theta)}{w(\theta)} + \frac{\hat{l}(\theta)}{l(\theta)}] y(\theta) T'(y(\theta))$ so that the first-order effects of the tax reform $\hat{T}$ on government revenue are given by (changing variables from types θ to incomes $y(\theta)$)

$$\hat{\mathcal{R}} = \int \hat{T}(y) f_Y(y)\, dy + \int T'(y) \left[\frac{\varepsilon_r^S(y)}{\varepsilon_w^S(y)} \frac{\hat{T}'(y)}{1 - T'(y)} + \left(1 + \frac{1}{\varepsilon_w^S(y)}\right) \frac{\hat{l}(y)}{l(y)} \right] y f_Y(y)\, dy, \tag{40}$$

where $\hat{l}(y)$ is the change in labor supply of agents with income initially equal to y. Using formula (9), this implies that the effect of the elementary tax reform at income y^* is given by

$$\begin{aligned} \hat{\mathcal{R}}(y^*) \;&=\; 1 + \frac{T'(y^*)}{1 - T'(y^*)} \frac{\varepsilon_r^S(y^*)}{\varepsilon_w^S(y^*)} \frac{y^* f_Y(y^*)}{1 - F_Y(y^*)} + \int_{\mathbb{R}_+} T'(y) \left(1 + \frac{1}{\varepsilon_w^S(y)}\right) \cdots \\ &\quad \times \left[-\varepsilon_r(y) \frac{\delta(y - y^*)}{1 - T'(y)} - \frac{1}{1 - T'(y^*)} \varepsilon_w(y) \Gamma(y, y^*) \varepsilon_r(y^*) \right] \frac{y f_Y(y)}{1 - F_Y(y^*)} dy \\ &=\; \hat{\mathcal{R}}_{\text{ex}}(y^*) + \frac{T'(y^*)}{1 - T'(y^*)} \varepsilon_r(y^*) \frac{y^* f_Y(y^*)}{1 - F_Y(y^*)} \left(1 + \varepsilon_w^S(y^*)\right) \frac{1}{\varepsilon_w^D(y^*)} \\ &\quad - \frac{\varepsilon_r(y^*)}{1 - T'(y^*)} \int_{\mathbb{R}_+} T'(y) \left(1 + \varepsilon_w^S(y)\right) \frac{\Gamma(y, y^*)}{1 + \varepsilon_w^S(y) / \varepsilon_w^D(y)} \frac{y f_Y(y)}{1 - F_Y(y^*)} dy. \end{aligned} \tag{41}$$

Using Euler's theorem (24) easily leads to equation (17). If the disutility of labor is isoelastic and the initial tax schedule is linear, then the marginal tax rate $T'(y)$ and the elasticity $\varepsilon_w^S(y)$ are constant. Applying equation (17) immediately implies that $\hat{\mathcal{R}}(y^*) = \hat{\mathcal{R}}_{\text{ex}}(y^*)$.

□

Proof of Corollary 4. If the disutility of labor is isoelastic, the initial tax schedule is CRP, the elasticities $\varepsilon_w^S, \varepsilon_w^D$ are constant and the integral in equation (41) can be simplified. The resulting expectation $\mathbb{E}[T'(y)\frac{y\Gamma(y,y^*)}{y^* f_Y(y^*)}]$ can be rewritten as

$$\mathrm{Cov}\left(T'(y);\frac{y\Gamma(y,y^*)}{y^* f_Y(y^*)}\right)+\frac{1}{y^* f_Y(y^*)}\mathbb{E}\left[T'(y)\right]\mathbb{E}\left[y\Gamma(y,y^*)\right].$$

But by Euler's theorem (equation (24)), we have $\frac{1}{1+\varepsilon_w^S/\varepsilon_w^D}\mathbb{E}[y\Gamma(y,y^*)]=\frac{1}{\varepsilon_w^D}y^* f_Y(y^*)$. Substituting into the previous expression easily leads to (18). Now suppose in addition that the production function is CES, so that the elasticities $\varepsilon_r,\varepsilon_w$ are constant and $\Gamma(y,y^*)$ is given by formula (12) with $\gamma(y,y^*)=\frac{1}{\sigma\mathbb{E}y}y^* f_Y(y^*)$. Substituting into (41) implies

$$\hat{\mathcal{R}}(y^*)=\hat{\mathcal{R}}_{\text{ex}}(y^*)+\varepsilon_r\left(1+\varepsilon_w^S\right)\left[\frac{T'(y^*)}{1-T'(y^*)}\frac{1}{\sigma}\frac{y^* f_Y(y^*)}{1-F_Y(y^*)}-\int_{\mathbb{R}_+}\frac{T'(y)}{1-T'(y^*)}\gamma(y,y^*)\frac{y f_Y(y)\,dy}{1-F_Y(y^*)}\right].$$

Suppose first that $p=0$, i.e., the initial tax schedule is linear. In this case, we have $T'(y^*)=T'(y)$ for all y, so that the term in the square brackets is equal to 0 by Euler's homogeneous function theorem. More generally, with a nonlinear tax schedule, we can use expression (26) for $\gamma(y,y^*)$ to rewrite the term in square brackets as

$$\frac{1}{1-T'(y^*)}\frac{1}{\sigma}\frac{y^* f_Y(y^*)}{1-F_Y(y^*)}\left[T'(y^*)-\int_{\mathbb{R}_+}T'(y)\frac{y}{\mathbb{E}y}f_Y(y)\,dy\right].$$

Using the fact that $\left(1+\varepsilon_w^S\right)\frac{\varepsilon_r}{\sigma}=\frac{1+\varepsilon_w^S}{\sigma+\varepsilon_w^S}\varepsilon_r^S$ leads to equation (19). Note that we can also derive this result from equation (18): substituting for $\Gamma(y,y^*)=\frac{1}{\sigma\mathbb{E}y}(1+\frac{\varepsilon_w^S}{\sigma})y^* f_Y(y^*)$ into $\mathrm{Cov}\left(T'(y);y\Gamma(y,y^*)\right)$ and using $\frac{1}{\mathbb{E}y}\mathrm{Cov}\left(T'(y);y\right)=\frac{1}{\mathbb{E}y}\mathbb{E}\left[yT'(y)\right]-\mathbb{E}\left[T'(y)\right]$ easily leads to (19).

□

Incidence on social welfare. The first-order effect of a tax reform $\hat{T}$ on the government objective $\mathcal{G}=\frac{1}{\lambda}\int G(U(\theta))f(\theta)\,d\theta$ is given by

$$\hat{\mathcal{G}}=-\int\hat{T}(y)g(y)f_Y(y)\,dy+\int(1-T'(y))\,y\frac{\hat{w}(y)}{w(y)}g(y)f_Y(y)\,dy,$$

where $g(y)=\frac{G'(U(\theta))}{\lambda}$ denotes the marginal social welfare weight at income y, and where $\hat{w}(y)$ is the change in labor supply of agents with income initially equal to y. Therefore, we obtain that the tax reform affects social welfare by

$$\hat{\mathcal{W}}=\hat{\mathcal{R}}+\hat{\mathcal{G}}=\int(1-g(y))\hat{T}(y)f_Y(y)\,dy-\int\frac{T'(y)}{1-T'(y)}\varepsilon_r^S(y)\hat{T}'(y)\,yf_Y(y)\,dy$$
$$+\int\left[\left(1+\varepsilon_w^S(y)\right)T'(y)+g(y)(1-T'(y))\right]\frac{\hat{w}(y)}{w(y)}yf_Y(y)\,dy.$$

Using equations (13) and (9), and applying this formula to the elementary tax reform at y^*, we get:

$$\begin{aligned}\hat{\mathcal{W}}(y^*) =& \int_{y^*}^{\infty} (1-g(y)) \frac{f_Y(y)}{1-F_Y(y^*)} dy - \varepsilon_r^S(y^*) \frac{T'(y^*)}{1-T'(y^*)} \frac{y^* f_Y(y^*)}{1-F_Y(y^*)} \\ &+ \frac{\frac{\varepsilon_r(y^*)}{\varepsilon^D(y^*)}}{1-T'(y^*)} \psi(y^*) \frac{y^* f_Y(y^*)}{1-F_Y(y^*)} - \frac{\varepsilon_r(y^*)}{1-T'(y^*)} \int \psi(y) \frac{\Gamma(y,y^*)}{1+\frac{\varepsilon_w^S(y)}{\varepsilon_w^D(y)}} \frac{y f_Y(y)}{1-F_Y(y^*)} dy,\end{aligned} \tag{42}$$

where $\psi(y)$ is defined by $\psi(y) = \left(1+\varepsilon_w^S(y)\right) T'(y) + g(y)(1-T'(y))$. Assume for simplicity that the production function is CES, the disutility of labor is isoelastic, and the tax schedule is CRP. The labor supply and demand elasticities are then constant, and we have $\Gamma(y,y^*) = \frac{\gamma(y,y^*)}{1-\varepsilon_w/\sigma} = \frac{1}{1-\varepsilon_w/\sigma} \frac{y^* f_Y(y^*)}{\sigma \mathbb{E} y}$. It follows that the second line in the previous expression can be rewritten as

$$\frac{\varepsilon_r/\sigma}{1-T'(y^*)} \left[\psi(y^*) - \int_{\mathbb{R}_+} \psi(y) \frac{y}{\mathbb{E}y} f_Y(y)\, dy\right] \frac{y^* f_Y(y^*)}{1-F_Y(y^*)}.$$

Thus, the variable $T'(y)\left(1+\varepsilon_w^S(y)\right)$ in equation (17), which measures the total impact of a wage adjustment $\hat{w}(y)$ on the government budget, is now replaced by the more general expression $\psi(y)$. Its second term comes from the fact that the share $1-T'(y)$ of the income gain due to the wage adjustment $\hat{w}(y)$ is kept by the individual; this in turn raises social welfare in proportion to the welfare weight $g(y)$.

□

D Generalizations: preferences with income effects

In this section we extend the model of Section 1 to a general utility function over consumption and labor supply $U(c,l)$, where $U_c, U_{cc} > 0$ and $U_l, U_{ll} < 0$. This specification allows for arbitrary substitution and income effects.

Elasticity concepts. The first-order condition of the agent reads $r(\theta) w(\theta) U_c(\theta) + U_l(\theta) = 0$, where $U_c(\theta)$ is a short-hand notation for $U_c(y(\theta) - T(y(\theta)), l(\theta))$ and $r(\theta) = 1 - T'(y(\theta))$ is the agent's retention rate. Differentiating this equation allows us to define the compensated (Hicksian) elasticity of labor supply with respect to the retention rate, $e_r^S(\theta) \equiv \frac{r(\theta)}{l(\theta)} \left.\frac{\partial l(\theta)}{\partial r(\theta)}\right|_{u \text{ cst}}$, and the income effect, $e_R(\theta) \equiv r(\theta) w(\theta) \frac{\partial l(\theta)}{\partial R}$, as follows:

$$\begin{aligned} e_r^c(\theta) =& \frac{U_l(\theta)/l(\theta)}{U_{ll}(\theta) + \left(\frac{U_l(\theta)}{U_c(\theta)}\right)^2 U_{cc}(\theta) - 2\left(\frac{U_l(\theta)}{U_c(\theta)}\right) U_{cl}(\theta)}, \\ e_R(\theta) =& \frac{-\left(\frac{U_l(\theta)}{U_c(\theta)}\right)^2 U_{cc}(\theta) + \left(\frac{U_l(\theta)}{U_c(\theta)}\right) U_{cl}(\theta)}{U_{ll}(\theta) + \left(\frac{U_l(\theta)}{U_c(\theta)}\right)^2 U_{cc}(\theta) - 2\left(\frac{U_l(\theta)}{U_c(\theta)}\right) U_{cl}(\theta)}. \end{aligned} \tag{43}$$

The labor supply elasticity with respect to the wage is given by $e_w^S(\theta) = (1-p(y(\theta)))\, e_r^c(\theta) + e_R(\theta)$.

As in Sections 1.2 and 2.1, we then normalize $e_r^{c,S}(\theta), e_R^S(\theta), e_w^S(\theta)$ by $1+p(y(\theta))e_r^c(\theta)$ to get the corresponding elasticities along the nonlinear budget constraint $\varepsilon_r^{c,S}(\theta), \varepsilon_R^S(\theta), \varepsilon_w^S(\theta)$, and further by $1+\varepsilon_w^S(\theta)/\varepsilon_w^D(\theta)$ to get the elasticities of equilibrium labor $\varepsilon_r^c(\theta), \varepsilon_R(\theta), \varepsilon_w(\theta)$. The cross-wage and own-wage elasticities $\gamma(\theta,\theta'), 1/\varepsilon_w^D(\theta)$ are defined as in (5) and (6). Finally, the resolvent cross-wage elasticity $\Gamma(\theta,\theta')$ is defined as in (10).

□

Proposition 4 (Generalization of Proposition 1). *The incidence of an arbitrary tax reform $\hat{T}$ on individual labor supply is given by the following formula, which generalizes (9):*

$$\hat{l}(\theta) = \hat{l}_{pe}(\theta) + \varepsilon_w(\theta)\int_\Theta \Gamma(\theta,\theta')\,\hat{l}_{pe}(\theta')\,d\theta', \tag{44}$$

where $\varepsilon_w(\theta)$, and $\Gamma(\theta,\theta')$ are given by their generalized definitions above, and where

$$\hat{l}_{pe}(\theta) \equiv -\varepsilon_r(\theta)\frac{\hat{T}'(y(\theta))}{1-T'(y(\theta))} + \varepsilon_R(\theta)\frac{\hat{T}(y(\theta))}{(1-T'(y(\theta)))\,y(\theta)}.$$

The incidence on wages, utilities and government revenue are derived as the corresponding formulas in Section 2.2.

The interpretation of this formula is identical to that of (9), except that the partial-equilibrium impact of the reform $\hat{l}_{\text{pe}}(\theta)$ is modified: in addition to the substitution effect already described in the quasilinear model, labor supply now also rises by an amount proportional to $\varepsilon_R(\theta)$ due to an income effect induced by the higher total tax payment $\hat{T}(y(\theta))$ of agent θ. Note that the partial-equilibrium formula for $\hat{l}_{\text{pe}}(\theta)$ is identical to that derived in models with exogenous wages by Saez (2001) and Golosov, Tsyvinski, and Werquin (2014), except that that now the elasticities $\varepsilon_r(\theta)$ and $\varepsilon_R(\theta)$ take into account the own-wage effects $\varepsilon_w^D(\theta)$.

Proof of Proposition (4). Consider a tax reform $\hat{T}$. The perturbed first order condition reads (letting $w_\theta = w(\theta)$, etc. for conciseness):

$$\begin{aligned}0 = &\left[1 - T'\left((w_\theta+\mu\hat{w}_\theta)\left(l_\theta+\mu\hat{l}_\theta\right)\right) - \mu\hat{T}'(w_\theta l_\theta)\right](w_\theta+\mu\hat{w}_\theta)\ldots\\ &\times U_c\left[(w_\theta+\mu\hat{w}_\theta)\left(l_\theta+\mu\hat{l}_\theta\right) - T\left((w_\theta+\mu\hat{w}_\theta)\left(l_\theta+\mu\hat{l}_\theta\right)\right) - \mu\hat{T}(w_\theta l_\theta)\ ,\ l_\theta+\mu\hat{l}_\theta\right]\\ &+ U_l\left[(w_\theta+\mu\hat{w}_\theta)\left(l_\theta+\mu\hat{l}_\theta\right) - T\left((w_\theta+\mu\hat{w}_\theta)\left(l_\theta+\mu\hat{l}_\theta\right)\right) - \mu\hat{T}(w_\theta l_\theta)\ ,\ l_\theta+\mu\hat{l}_\theta\right].\end{aligned}$$

A first-order Taylor expansion implies:

$$\begin{aligned}\frac{\hat{l}_\theta}{l_\theta} =& \frac{e_R(\theta)+(1-p(y_\theta))e_r^c(\theta)}{1+p(y_\theta)e_r^c(\theta)}\frac{\hat{w}_\theta}{w_\theta}\\ &-\frac{e_r^c(\theta)}{1+p(y_\theta)e_r^c(\theta)}\frac{\hat{T}'(y_\theta)}{1-T'(y_\theta)} - \frac{e_R(\theta)}{1+p(y_\theta)e_r^c(\theta)}\frac{\hat{T}(y_\theta)}{(1-T'(y_\theta))\,y_\theta},\end{aligned}$$

where the first-order change in the wage $w(\theta)$ is given by equation (38). This leads to an integral equation for $\hat{l}_\theta/l_\theta$ which can be solved following the same steps as in Proposition 1 to obtain equation

(44).

□

Corollary 6 **(Generalization of Corollary 4).** *Assume that the production function is CES, the tax schedule is CRP, and the utility function has the form* $U(c,l) = \frac{c^{1-\eta}}{1-\eta} - \frac{l^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$. *The revenue effect of the elementary tax reform at income* y^* *is then given by*

$$\hat{\mathcal{R}}(y^*) = \hat{\mathcal{R}}_{ex}(y^*) + \phi\varepsilon_r^S \frac{T'(y^*) - \bar{T}'}{1 - T'(y^*)} \frac{y^* f_Y(y^*)}{1 - F_Y(y^*)} - \phi\varepsilon_r^S (1-p)\eta\mathbb{E}\left[\frac{T'(y) - \bar{T}'}{1 - T'(y)} | y > y^*\right], \tag{45}$$

where $\bar{T}' = \mathbb{E}[yT'(y)]/\mathbb{E}y$ *is the income-weighted average marginal tax rate in the economy and where* $\phi = \frac{1+\varepsilon_w^S}{\sigma+\varepsilon_w^S}$.[36] *If in addition top incomes are Pareto distributed with parameter* Π, *we have* $\hat{\mathcal{R}}(y^*) > \hat{\mathcal{R}}_{ex}(y^*)$ *as* $y^* \to \infty$ *if an only if* $\Pi > p + \eta - p\eta$. *In this case, the theoretical insights of Section 3.2 remain qualitatively valid with income effects.*

Proof of Corollary 6. Under the assumed functional form assumptions, the labor supply and demand elasticities are constant and we have $\varepsilon_r^{c,S} = \frac{e}{\eta e(1-p)+pe+1}$, $\varepsilon_R^S = -(1-p)\eta\varepsilon_r^{c,S}(\theta)$, and $\varepsilon_w^S = (1-p)(1-\eta)\varepsilon_r^{c,S}$. Since the production function is CES, the integral equation for $\hat{l}(\theta)/l(\theta)$ has a multiplicatively separable kernel and its solution for an elementary tax reform at income $y(\theta^*)$ is given by:

$$\hat{l}(\theta) = -\frac{\varepsilon_r(\theta)}{1 - T'(y(\theta^*))} \frac{\delta(y(\theta) - y(\theta^*))}{1 - F(\theta^*)} + \frac{\varepsilon_R(\theta)}{(1 - T'(y(\theta))y(\theta)} \frac{\mathbb{I}_{\{\theta > \theta^*\}}}{1 - F(\theta^*)} + \frac{\frac{1}{1-F(\theta^*)}\varepsilon_w(\theta)}{1 - \int_\Theta \varepsilon_w(\theta')\gamma(\theta,\theta')d\theta'} \left[-\gamma(\theta,\theta^*)\frac{\varepsilon_r(\theta^*)}{1 - T'(y(\theta^*))} + \int_{\theta^*}^{\bar{\theta}} \gamma(\theta,\theta')\frac{\varepsilon_R(\theta')}{(1 - T'(y(\theta'))y(\theta')}d\theta'\right].$$

Therefore, the effect of the tax reform on government revenue, $\hat{\mathcal{R}} = \int \hat{T}dF + \int T'y(\hat{l} + \hat{w})dF$, is given by:

$$\hat{\mathcal{R}}(y(\theta^*)) = \hat{\mathcal{R}}_{\text{ex}}(y(\theta^*))$$
$$+ \frac{\frac{1}{1-F(\theta^*)}\varepsilon_r^S}{1 - T(y(\theta^*))}\left(\frac{1}{\varepsilon_w^D}\frac{1+\varepsilon_w^S}{1+\frac{\varepsilon_w^S}{\varepsilon_w^D}}T'(y(\theta^*))y(\theta^*)f(\theta^*) - \int T'(y(\theta))y(\theta)\frac{\gamma(\theta,\theta^*)\frac{1+\varepsilon_w^S}{1+\varepsilon_w^S/\varepsilon_w^D}\frac{1}{1+\varepsilon_w^S/\varepsilon_w^D}}{1 - \int \varepsilon_w\gamma(x,x)dx}dF(\theta)\right)$$
$$+ \int_{\theta^*}^{\bar{\theta}} \frac{\frac{1}{1-F(\theta^*)}\varepsilon_R^S}{(1 - T'(y(\theta'))y(\theta')}\left(\frac{1}{\varepsilon_w^D}\frac{1+\varepsilon_w^S}{1+\frac{\varepsilon_w^S}{\varepsilon_w^D}}T'(y(\theta'))y(\theta')f(\theta') - \int T'(y(\theta))y(\theta)\frac{\gamma(\theta,\theta')\frac{1+\varepsilon_w^S}{1+\varepsilon_w^S/\varepsilon_w^D}\frac{1}{1+\varepsilon_w^S/\varepsilon_w^D}}{1 - \int \varepsilon_w\gamma(x,x)dx}dF(\theta)\right)d\theta'.$$

This expression easily leads to (45). Now, since the tax schedule is CRP, we have $\frac{T'(y)-\bar{T}'}{1-T'(y)} = \frac{y^p}{\bar{y}}\mathbb{E}[y^{1-p}] - 1 = \frac{1-\bar{T}'}{1-T'(y)} - 1$. If incomes above $y(\theta^*)$ are Pareto distributed with tail parameter Π,

[36] Note that for $\eta = 0$, this formula reduces to equation (19). If $\eta > 0$ and the baseline tax schedule is progressive, then the first and second general-equilibrium contributions have opposite signs.

we have $\mathbb{E}[y^p|y > y^*] = \frac{\Pi}{\Pi-p} y^{*p}$ and hence

$$\hat{\mathcal{R}}(y^*) = \hat{\mathcal{R}}_{\text{ex}}(y^*) + \phi\varepsilon_r^S \left[\Pi \left(\frac{1-\bar{T}'}{1-T'(y^*)} - 1 \right) - \eta(1-p) \left(\frac{\Pi}{\Pi - p} \frac{1-\bar{T}'}{1-T'(y^*)} - 1 \right) \right]. \tag{46}$$

The term in square brackets is positive for y large enough if and only if $\Pi > \eta(1-p)\frac{\Pi}{\Pi-p}$, i.e., $\Pi > \eta + p(1-\eta)$, because $T'(y) \to 1$ as $y \to \infty$.

□

Equation (46) leads to simple calculations of the additional general equilibrium effect on government revenue. To illustrate this, we consider a parameterization that is based on the empirical literature that estimates the impact of lottery wins on labor supply (Imbens, Rubin, and Sacerdote (2001), Cesarini et al. (2017)). Using these wealth shocks they find that a one dollar increase in wealth leads to a decrease in life-cycle labor income (in net present value) of 10-11 cents. Thus, we calibrate our (static) model such that an increase in unearned income of 1 dollar implies a decrease in earnings of 10-11 cents. Further, we set $\varepsilon_r^{c,S}(\theta) = 0.33$ Chetty (2012). As in our benchmark calibration in the main body, we assume that $p = 0.15$. To target the value of the lottery papers, we set $\varepsilon_R^S(\theta) = -0.08$, which captures approximatey a 10-11 cents decrease in gross income if the marginal tax rate is around 25%. The relationship $\varepsilon_R^S(\theta) = -(1-p)\eta\varepsilon_r^{c,S}$ then yields a value of $\eta \approx 0.29$. Finally, the value for e that is consistent with $\varepsilon_r^{c,S} = 0.33$ is $e = 0.38$. Evaluating the second term on the right hand side of (46) for these numbers reveals that it becomes positive for income levels where the marginal tax rate is above 27.6%, a number that is slightly higher than the income-weighted average marginal tax rate, which is equal to 26%. The income levels that correspond to these tax rates are approximately \$85,000 and \$77,000. A last simple exercise is then to evaluate general equilibrium revenue effect at a higher income level and compare it to the value that is obtained in the absence of income effects. We do this comparison for the income level of \$200,000 and find that the additional revenue effect coming from the endogeneity of wages is reduced by 28% (32% respectively) if the elasticity of substitution is $\sigma = 0.66$ ($\sigma = 3.1$ respectively).

E Numerical simulations

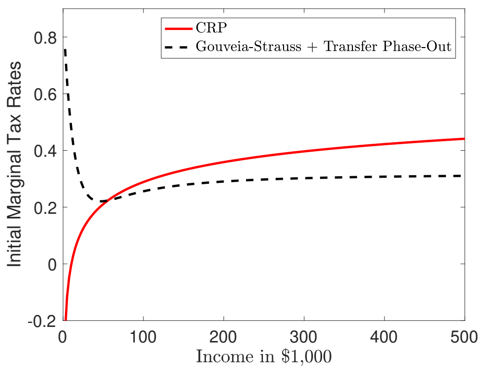
Calibration of the model. We assume that incomes are log-normally distributed apart from the top, where we append a Pareto distribution for incomes above \$150,000. To obtain a smooth hazard ratio $\frac{1-F_y(y)}{yf_y(y)}$, we decrease the thinness parameter of the Pareto distribution linearly between \$150,000 and \$350,000 and let it be constant at 1.5 afterwards (Diamond and Saez, 2011). In the last step we use a standard kernel smoother to ensure differentiability of the hazard ratios at \$150,000 and \$350,000. We set the mean and variance of the lognormal distribution at 10 and 0.95, respectively. The mean parameter is chosen such that the resulting income distribution has a mean of \$64,000, i.e., approximately the average US yearly earnings. The variance parameter was chosen such that the hazard ratio at level \$150,000 is equal to that reported by Diamond and Saez (2011,

Fig.2).

CES production function with exogenous assignment. Denote by θ_y the type of an agent who earns income y given the current tax system. Our first step is then the same as in Saez (2001): we use the individual's first order condition $1 - T'(y) = v'\left(\frac{y}{w}\right)\frac{1}{w}$ and the observed income and marginal tax rate in the data, to back out the wage. As in Saez (2001), this gives us both the wage $w(\theta_y)$ as well as the labor supply $l(\theta_y) = \frac{y}{w(\theta_y)}$ that correspond to that income level y, given the current tax schedule. Assume that the production function is CES with a given parameter σ. Once we know the wage $w(\theta_y)$, the labor supply $l(\theta_y)$, and the density of incomes $f_Y(y)$, we can back out the primitive parameters $a(\theta_y)$ of the CES production function (3) using the formula $w(\theta_y) = a(\theta_y)[l(\theta_y)f_Y(y)y'(\theta_y)/\mathscr{F}(\mathscr{L})]^{1/\sigma}$, where we know everything but $a(\theta_y)$ and $y'(\theta_y) \equiv \left.\frac{dy(\theta)}{d\theta}\right|_{\theta_y}$. We can without loss of generality assume that θ is uniformly distributed in the unit interval. This pins down $y'(\theta_y)$, since we observe the income percentiles in the data. We can therefore infer the parameter $a(\theta_y)$ for each y.

Microfoundation with endogenous assignment. Now consider the model of Section A.2. Ales, Kurnaz, and Sleet (2015, p.30) calibrate the following relation $\frac{A_1'(\theta,M(\theta))}{A(\theta,M(\theta))} = \alpha_1 + \alpha_2 M(\theta)$ with $\alpha_1 = 0.41$ an $\alpha_2 = 3.01$. The parameter α_1 represents the pure returns to skill and α_2 represents the complementarity with tasks. We extend this functional form as follows: $\frac{A_1'(\theta,M(\theta))}{A(\theta,M(\theta))} = \alpha_1(\theta) + \alpha_2 M(\theta)$. That is, we keep the linearity assumption as well as the value of the complementarity parameter α_2. But we replace the constant α_1 with a function $\alpha_1(\theta)$ that ensures that the empirical wage distribution is exactly matched. Crucially, this allows us to depart from the restriction of a bounded income distribution (which leads to inverse-U-shaped optimal tax rates) and to capture instead the Pareto tail of the distribution. To estimate the relevant parameters $\alpha_1(\theta)$, we start by calibrating the wage distribution using the same method as Saez (2001), as explained in the main body of the paper. We then plug the parameters of the Cobb Douglas function estimated by Ales, Kurnaz, and Sleet (2015, p.27) into equation (33). Solving this equation gives us $M(\theta)$ for the current allocation. We can then find the function $\alpha_1(\theta)$ such that the following equation holds: $\frac{w'(\theta)}{w(\theta)} = \alpha_1(\theta) + \alpha_2 M(\theta)$, where the left hand side is the empirical wage distribution.

Robustness: alternative baseline tax function. We propose several robustness exercises for our tax incidence results. First, we depart from the assumption that the initial tax schedule is CRP and consider an alternative calibration that differs in two ways: (i) we use a Gouveia-Strauss approximation for the income tax, taken from Guner, Kaygusuz, and Ventura (2014); (ii) we also account for the phasing-out of means-tested transfer programs that increase effective marginal tax rates, in par-



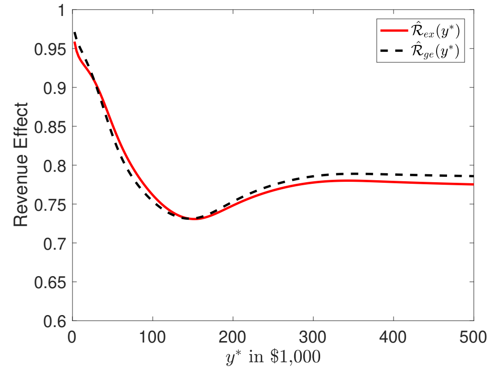


Figure 5: Alternative specification of the baseline tax schedule. Bold curve: CRP tax schedule. Dashed curve: Gouveia-Strauss approximation with additional distortions due to means-tested transfers. Left panel: Right panel: Revenue gains of elementary tax reforms. Bold curve: exogenous wages. Dashed curve: CES production function with $\sigma = 3.1$.

ticular for low incomes. The Gouveia-Strauss specification we use is the third to last column in Table 12 of Guner, Kaygusuz, and Ventura (2014). For the phasing-out of transfers, we use parametric estimates from Guner, Rauh, and Ventura (2017), namely, $T(I) = \exp(-1.816)\exp(-4.29I)I^{-0.006}$ where I is expressed in multiples of average income (we use a CPI deflator and express everything in terms of year 2000 dollars). Figure 5 shows the resulting schedule of marginal tax rates (left panel) and the normalized revenue gains of elementary tax reforms for a CES parameter $\sigma = 3.1$ (right panel). The additional general-equilibrium revenue effects due to the endogeneity of wages are naturally smaller in magnitude than for a CRP initial tax schedule because of the very large bottom marginal tax rates. Nevertheless, the general insight of Figure 2 is unchanged.

Robustness: incidence on social welfare. Second, we depart from our focus on revenue effects (i.e., Rawlsian welfare) and consider alternative concave social welfare functions $G(u) = \frac{u^{1-\kappa}}{1-\kappa}$. The CES parameter in Figure 6 is $\sigma = 3.1$. Welfare gains are expressed in terms of public funds. For a low taste for redistribution ($\kappa = 1$, left panel), the welfare gains of raising tax rates on high incomes are reversed due to general equilibrium. For a stronger taste for redistribution ($\kappa = 3$, right panel), general equilibrium effects imply that raising the top tax rates is more desirable. On the one hand, general equilibrium effects raise tax revenue (as in thee main body of the paper). On the other hand, the implied wage decreases for the working poor make them worse-off. In case of very strong redistributive tastes (i.e., when the social marginal welfare weights decrease sufficiently fast with income, the extreme case being the Rawlsian welfare criterion), the tax revenue gain gets a higher weight (since these gains are used for lump-sum redistribution). If relatively richer workers (for whom the lump-transfer is less important relative to the very poor) still have significant welfare weights, the wage effects dominates.

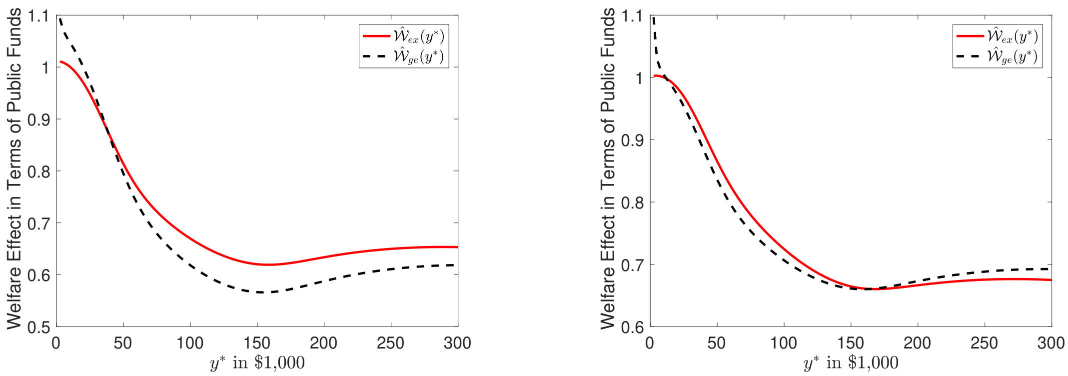


Figure 6: Welfare effect of elementary tax reforms for the social welfare function $G(u) = \frac{u^{1-\kappa}}{1-\kappa}$. Left panel: $\kappa = 1$. Right panel: $\kappa = 3$.

$\square$

F Optimal taxation

In the model with exogenous wages (Diamond, 1998), the optimum schedule $T'_{\text{pe}}(\cdot)$ is characterized by

$$\frac{T'_{\text{pe}}(y^*)}{1-T'_{\text{pe}}(y^*)} = \frac{1}{\varepsilon_r^S(y^*)}\left(1-\bar{g}(y^*)\right)\frac{1-F_Y(y^*)}{y^* f_Y(y^*)}.$$

Corollary 7 (Optimal tax schedule in general equilibrium). *The welfare-maximizing tax schedule T satisfies: for all $y^* \in \mathbb{R}_+$,*

$$\begin{aligned}\frac{T'(y^*)}{1-T'_{pe}(y^*)} = \frac{1}{\varepsilon_r^S(y^*)}\frac{1-F_Y(y^*)}{y^* f_Y(y^*)}\Bigg\{&1-\bar{g}(y^*)+\varepsilon_r(y^*)\ldots \\ &\times\int_{\mathbb{R}_+}\left[\psi(y^*)-\psi(y)\right]\frac{\Gamma(y,y^*)}{1+\frac{\varepsilon_r^S(y^*)}{\varepsilon_r^D(y^*)}}\frac{yf_Y(y)}{1-F_Y(y^*)}dy\Bigg\},\end{aligned} \tag{47}$$

where $\psi(y) = \left(1+\varepsilon_w^S(y)\right)T'(y)+g(y)\left(1-T'(y)\right)$. This optimal tax formula (47) can be straightforwardly transformed into an integral equation in $T'(\cdot)$, which can then be solved using similar techniques as in Section 2.1.

Proof of Corollary 7. The impact of the elementary tax reforms on social welfare is given by (42). Using Euler's theorem (24), imposing $\hat{\mathcal{W}}(y^*) = 0$ for all y^* and rearranging the terms leads to

$$\begin{aligned}\frac{T'(y^*)}{1-T'(y^*)} =&\frac{1}{\varepsilon_r^S(y^*)}\left(1-\bar{g}(y)\right)\frac{1-F_Y(y^*)}{y^* f_Y(y^*)} \\ &+\frac{\varepsilon_r(y^*)}{\varepsilon_r^S(y^*)}\frac{1}{1-T'(y^*)}\int_{\mathbb{R}_+}\left[\psi(y^*)-\psi(y)\right]\frac{\Gamma(y,y^*)}{1+\frac{\varepsilon_w^S(y)}{\varepsilon_w^D(y)}}\frac{yf_Y(y)}{y^* f_Y(y^*)}dy.\end{aligned}$$

Multiplying this equation by $1-T'(y^*)$ and solving for $T'(y^*)$ easily leads to (47).

□

Proof of Proposition 3. If the production function is CES, we have $\varepsilon_w^D(y)=\sigma$ and $\Gamma(y,y^*)=\frac{y^* f_Y(y^*)}{\sigma\mathbb{E}[(1+\frac{1}{\sigma}\varepsilon_w^S(x))^{-1}x]}$. Using these expressions, formula (47) can then be rewritten as

$$\left[1+\frac{1}{\sigma}\left(g\left(y^*\right)-1\right)\right]T'\left(y^*\right)=\frac{1-T'\left(y^*\right)}{\varepsilon_r\left(y^*\right)}\left(1-\bar{g}\left(y^*\right)\right)\frac{1-F_Y\left(y^*\right)}{y^* f_Y\left(y^*\right)}+\frac{1}{\sigma}g\left(y^*\right)-\frac{A}{\sigma},$$

where A is a constant (independent of y^*) equal to

$$A\equiv\frac{1}{\mathbb{E}[\frac{y}{1+\frac{1}{\sigma}\varepsilon_w^S(y)}]}\int\frac{g\left(y\right)+\left[\left(1-g\left(y\right)\right)+\varepsilon_w^S\left(y\right)\right]T'\left(y\right)}{1+\frac{1}{\sigma}\varepsilon_w^S\left(y\right)}yf_Y\left(y\right)dy. \tag{48}$$

The previous equation can then be rewritten as

$$T'\left(y^*\right)=\frac{\frac{1}{\varepsilon_r(y^*)}\left(1-\bar{g}\left(y^*\right)\right)\frac{1-F_Y(y^*)}{y^* f_Y(y^*)}+\frac{1}{\sigma}\left(g\left(y^*\right)-A\right)}{1+\frac{1}{\varepsilon_r(y^*)}\left(1-\bar{g}\left(y^*\right)\right)\frac{1-F_Y(y^*)}{y^* f_Y(y^*)}+\frac{1}{\sigma}\left(g\left(y^*\right)-1\right)} \tag{49}$$

We now show that $A=1$, which easily leads to expression (21). Consider the following tax reform:

$$\hat{T}_2\left(y\right)=-\frac{\varepsilon_r\left(y^*\right)}{1-T'\left(y^*\right)}\gamma\left(y,y^*\right)\left(1-T'\left(y\right)\right)y,$$
$$\hat{T}_2'\left(y\right)=-\frac{\varepsilon_r\left(y^*\right)}{1-T'\left(y^*\right)}\gamma\left(y,y^*\right)\left(1-T'\left(y\right)-yT''\left(y\right)\right),$$

where $\gamma\left(y,y^*\right)=\frac{1}{\sigma}\frac{y^* f_Y(y^*)}{\int xf_Y(x)dx}$ is independent of y since the production function is CES. (It is easy to show that this is the tax reform that cancels out the general equilibrium effects on individual labor supply of the elementary reform at y^*.) Tedious but straightforward algebra shows that the incidence of this counteracting tax reform $\hat{T}_2$ on social welfare is given by

$$\begin{aligned}\hat{\mathcal{W}}(\hat{T}_2)&=\int\hat{\mathcal{W}}\left(y^*\right)\hat{T}_2'\left(y^*\right)\left(1-F_Y\left(y^*\right)\right)dy^*\\&=-\frac{1}{\sigma}\frac{\varepsilon_r\left(y^*\right)}{1-T'\left(y^*\right)}\frac{y^* f_Y\left(y^*\right)}{\int xf_Y\left(x\right)dx}\left\{\int\left(1-g\left(y\right)\right)\left(1-T'\left(y\right)\right)yf_Y\left(y\right)dy\ldots\right.\\&\quad-\int\varepsilon_w\left(y\right)\left(\left[1+\frac{1}{\sigma}\left(g\left(y\right)-1\right)\right]T'\left(y\right)-\frac{1}{\sigma}g\left(y\right)\right)yf_Y\left(y\right)dy\\&\quad\left.-\frac{1}{\sigma}\frac{\int\varepsilon_w\left(y\right)ydF_Y\left(y\right)}{\mathbb{E}\left[\frac{x}{1+\frac{1}{\sigma}\varepsilon_w^S(x)}\right]}\int\frac{1}{1+\frac{1}{\sigma}\varepsilon_w^S\left(x\right)}\left[\left(1-g\left(x\right)+\varepsilon_w^S\left(x\right)\right)T'\left(x\right)+g\left(x\right)\right]xdF_Y\left(x\right)\right\}.\end{aligned}$$

Using expression (48) for A and imposing that $\hat{\mathcal{W}}(\hat{T}_2)=0$ leads to

$$\int\frac{\left(1-g\left(y\right)\right)+\varepsilon_w^S\left(y\right)}{1+\frac{1}{\sigma}\varepsilon_w^S\left(y\right)}T'\left(y\right)yf_Y\left(y\right)dy=\int\frac{\left(1-g\left(y\right)\right)+\frac{1-A}{\sigma}\varepsilon_w^S\left(y\right)}{1+\frac{1}{\sigma}\varepsilon_w^S\left(y\right)}yf_Y\left(y\right)dy. \tag{50}$$

Now compare expressions (48) and (50). These two equations imply

$$\int \frac{\left[(1-g(y))+\varepsilon_w^S(y)\right]T'(y)}{1+\frac{1}{\sigma}\varepsilon_w^S(y)}yf_Y(y)\,dy = \mathbb{E}\left[\frac{A-g(y)}{1+\frac{1}{\sigma}\varepsilon_w^S(y)}y\right]$$
$$= \int \frac{(1-g(y))+\frac{1-A}{\sigma}\varepsilon_w^S(y)}{1+\frac{1}{\sigma}\varepsilon_w^S(y)}yf_Y(y)\,dy.$$

Solving for A implies $A=1$.

□

Proof of Corollary 5. Suppose that in the data (i.e., given the current tax schedule and constant top tax rate), the income distribution has a Pareto tail, so that the (observed) hazard rate $\frac{1-F_Y(y^*)}{yf_Y(y^*)}$ converges to a constant. We show that under these assumptions, the income distribution at the optimum tax schedule is also Pareto distributed at the tail with the same Pareto coefficient. We have

$$\frac{1-F_Y(y(\theta))}{y(\theta)f_Y(y(\theta))} = \frac{1-F(\theta)}{\frac{y(\theta)}{y'(\theta)}f(\theta)} = \frac{1-F(\theta)}{\theta f(\theta)}\frac{\theta y'(\theta)}{y(\theta)} = \frac{1-F(\theta)}{\theta f(\theta)}\varepsilon_{y,\theta}, \tag{51}$$

where we define the income elasticity $\varepsilon_{y,\theta} \equiv d\ln y(\theta)/d\ln\theta$. To compute this elasticity, use the individual first order condition (1) with isoelastic disutility of labor to get $l(\theta) = r(\theta)^e w(\theta)^e$, where $r(\theta)$ is agent θ's retention rate. Thus we have $\varepsilon_{l,\theta} \equiv \frac{d\ln l(\theta)}{d\ln\theta} = e\frac{d\ln r(\theta)}{d\ln\theta} + e\frac{d\ln w(\theta)}{d\ln\theta}$. But since the production function is CES, we have

$$\frac{d\ln w(\theta)}{d\ln\theta} = \frac{d\ln a(\theta)}{d\ln\theta} - \frac{1}{\sigma}\frac{d\ln l(\theta)}{d\ln\theta} - \frac{1}{\sigma}\frac{d\ln f(\theta)}{d\ln\theta} = \frac{\theta a'(\theta)}{a(\theta)} - \frac{1}{\sigma}\varepsilon_{l,\theta} - \frac{1}{\sigma}\frac{\theta f'(\theta)}{f(\theta)}.$$

Using this expression, we obtain

$$\varepsilon_{l,\theta} = e\left[\frac{\theta a'(\theta)}{a(\theta)} - \frac{1}{\sigma}\varepsilon_{l,\theta} - \frac{1}{\sigma}\frac{\theta f'(\theta)}{f(\theta)} + \frac{\theta r'(\theta)}{r(\theta)}\right].$$

Since we assume that the second derivative of the optimal marginal tax rate, $T''(y)$, converges to zero for high incomes, we have $\lim_{\theta\to\infty} r'(\theta) = 1$. Moreover, the variables $\frac{\theta a'(\theta)}{a(\theta)}$ and $\frac{\theta f'(\theta)}{f(\theta)}$ are primitive parameters that do not depend on the tax rate. Assuming that they converge to constants as $\theta\to\infty$, we obtain that $\lim_{\theta\to\infty}\varepsilon_{l,\theta}$ is constant, and hence $\varepsilon_{y,\theta} = \varepsilon_{l,\theta} + \varepsilon_{w,\theta} = \left(1+\frac{1}{e}\right)\varepsilon_{l,\theta}$ converges to a constant independent of the tax rate. Therefore, the hazard rate of the income distribution at the optimum tax schedule, given by (51), converges to the same constant as the hazard rate of incomes observed in the data. Now let $y^*\to\infty$ in equation (21), to obtain an expression for the optimal top tax rate $\tau^* = \lim_{y^*\to\infty} T'(y^*)$. We have seen that $\lim_{y^*\to\infty}\varepsilon_r(y^*) = \frac{e}{1+e/\sigma}$. Furthermore assume that $\lim_{y^*\to\infty} g(y^*) = \bar{g}$, so that $\lim_{y^*\to\infty}\bar{g}(y^*) = \bar{g}$. Therefore (21) implies

$$\frac{\tau^*}{1-\tau^*} = \frac{1+e/\sigma}{e}(1-\bar{g})\frac{1}{\Pi} + \frac{\bar{g}-1}{\sigma} = \frac{1-\bar{g}}{\Pi e} + \frac{1-\bar{g}}{\Pi\sigma} + \frac{\bar{g}-1}{\sigma},$$

where Π is the Pareto parameter. Solving for τ^* leads to (22).

□