# SUPPLEMENT TO "NONLINEAR TAX INCIDENCE AND OPTIMAL TAXATION IN GENERAL EQUILIBRIUM"
## (*Econometrica*, Vol. 88, No. 2, March 2020, 469–493)

DOMINIK SACHS
*Department of Economics, University of Munich*

ALEH TSYVINSKI
*Department of Economics, Yale University*

NICOLAS WERQUIN
*Toulouse School of Economics, University of Toulouse Capitole*

## APPENDIX 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 $\theta$. 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, that is, taking a first-order Taylor expansion of the perturbed first-order condition

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

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* $|\varepsilon_w^S(\theta)/\varepsilon_w^D(\theta)| < 1$, *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 toward the fixed point that characterizes the elasticities along the nonlinear budget constraint. Finally, the condition $|\varepsilon_w^S(\theta)/\varepsilon_w^D(\theta)| < 1$ 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 $\theta$ so that wages $w(\theta)$ are strictly increasing in $\theta$ in the initial equilibrium. Next, note that the individual first-order condition (1) implies that the elasticity of income with

Dominik Sachs: dominik.sachs@econ.lmu.de
Aleh Tsyvinski: a.tsyvinski@yale.edu
Nicolas Werquin: nicolas.werquin@tse-fr.eu

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 $\theta$.

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 nonlocal, 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 $\theta$. 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 $y^*$:*

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

*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:*

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

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

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

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

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). *Q.E.D.*

*Formulas for CES Technology.* Wages are $w(\theta) = a(\theta)(L(\theta))^{-\frac{1}{\sigma}}[\int_{\Theta} a(x) \times (L(x))^{\frac{\sigma-1}{\sigma}} dx]^{\frac{1}{\sigma-1}}$, 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

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

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 $\theta$. 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 $\eta$ into output. In equilibrium, a nonlinear price (earnings) schedule $p(\cdot)$ emerges such that an agent who produces $\eta$ units earns income $p(\eta)$, irrespective of the underlying productivity $\theta$. Hence, when an (atomistic) individual $\theta$ provides more effort $l(\theta)$, income moves along the nonlinear schedule $l \mapsto p(\theta \times l)$; for example, in their superstars model with a convex equilibrium earnings schedule, income increases faster than linearly. By contrast, in our framework, different values of $\theta$ 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 $\theta$ provides more effort $l(\theta)$, income increases linearly, as the wage remains constant (since the sector $\theta$ doesn't change). In their framework, Scheuer and Werning show that the general equilibrium effects exactly cancel out at the optimum tax schedule, even though they would of course be nonzero 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]

---

[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.

[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 $\theta$, 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.

### 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 $\theta$ employed in task $\psi$. We assume that high-skill workers have a comparative advantage in tasks with high skill intensity, that is $A(\theta, \psi)$ 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 $\theta$ 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:

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

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

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

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

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

The intermediate good firm of type $\psi$ chooses its demand for labor $L(\theta)$ of each skill $\theta$ 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 $\theta$ as long as their wage is given by

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

Moreover, the wage of any skill $\theta$ that is not employed in task $\psi$ must satisfy

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

*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 $\theta$, 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: $\forall \theta \in \Theta$,

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

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 $\delta$ 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: $\forall \theta \in \Theta$,

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

*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

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

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).

$$\frac{w'(\theta)}{w(\theta)} = \frac{A'_1\big(\theta, M(\theta)\big)}{A\big(\theta, M(\theta)\big)}. \tag{34}$$

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 $\mathcal{L} \equiv \{l(\theta)f(\theta)\}_{\theta \in \Theta}$. 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 $\mathcal{L}$, 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

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

$$= \left\{ \int_{\underline{\theta}}^{\bar{\theta}} a(\theta, M) \left[ l(\theta) f(\theta) \right]^{\frac{\sigma-1}{\sigma}} d\theta \right\}^{\frac{\sigma}{\sigma-1}}, \tag{35}$$

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 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 $\theta'$ (e.g., an immigration inflow) on the wage of type $\theta$, 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 .

## APPENDIX 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 $\theta$ in response to this perturbation. The labor supply response of type $\theta$ is given by the solution to the perturbed first-order condition

$$0 = v'\big(l(\theta) + \mu \hat{l}(\theta)\big)$$
$$- \big\{ 1 - T'\big[\tilde{w}(\theta) \times \big(l(\theta) + \mu \hat{l}(\theta)\big)\big]$$
$$- \mu \hat{T}'\big[\tilde{w}(\theta) \times \big(l(\theta) + \mu \hat{l}(\theta)\big)\big]\big\} \tilde{w}(\theta), \tag{37}$$

---

[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.

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

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

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

$$
\left\{1 + \frac{1 - T'(y(\theta)) - y(\theta)T''(y(\theta))}{1 - T'(y(\theta)) + \dfrac{v'(l(\theta))}{l(\theta)v''(l(\theta))}y(\theta)T''(y(\theta))} \frac{v'(l(\theta))}{l(\theta)v''(l(\theta))} \frac{1}{\varepsilon^D_w(\theta)}\right\}\frac{\hat{l}(\theta)}{l(\theta)}
$$

$$
= \frac{1 - T'(y(\theta)) - y(\theta)T''(y(\theta))}{1 - T'(y(\theta)) + \dfrac{v'(l(\theta))}{l(\theta)v''(l(\theta))}y(\theta)T''(y(\theta))} \frac{v'(l(\theta))}{l(\theta)v''(l(\theta))} \int_\Theta \gamma(\theta,\theta')\frac{\hat{l}(\theta')}{l(\theta')} \, d\theta'
$$

$$
- \frac{1}{1 - T'(y(\theta)) + \dfrac{v'(l(\theta))}{l(\theta)v''(l(\theta))}y(\theta)T''(y(\theta))} \frac{v'(l(\theta))}{l(\theta)v''(l(\theta))} \hat{T}'(y(\theta)),
$$

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). *Q.E.D.*

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. *Q.E.D.*

PROOF OF EQUATION (12): Suppose that the cross-wage elasticities are multiplicatively separable, that is, 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 $\theta$ to incomes $y(\theta)$ easily leads to (12). Note that this solution is well-defined if $\frac{1}{\sigma\mathbb{E}y}\mathbb{E}[y\varepsilon_w(y)] < 1$. *Q.E.D.*

SUFFICIENT CONDITIONS ENSURING THE CONVERGENCE OF THE RESOLVENT (10): Suppose that the production function is CES with parameter $\sigma$, 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

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

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.                                                                                      *Q.E.D.*

## APPENDIX 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 nondifferentiable 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 $\psi$ 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 $y^*$:

$$\hat{l}(\theta) = -\varepsilon_r(\theta)\frac{\delta_{y^*}(y(\theta))}{1 - T'(y(\theta))} - \varepsilon_w(\theta)\frac{\Gamma(\theta, \theta^*)}{y'(\theta^*)}\varepsilon_r(\theta^*)\frac{1}{1 - T'(y(\theta^*))}$$

$$= -\varepsilon_r(y)\frac{\delta_{y^*}(y)}{1 - T'(y)} - \varepsilon_w(y)\Gamma(y, y^*)\varepsilon_r(y^*)\frac{1}{1 - T'(y^*)}, \tag{39}$$

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 $\theta$'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 $\theta$ to incomes $y(\theta)$)

$$\hat{\mathcal{R}} = \int \hat{T}(y)f_Y(y)\,dy$$

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

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

$$\hat{\mathcal{R}}(y^*) = 1 + \frac{T'(y^*)}{1 - T'(y^*)}\frac{\varepsilon_r^S(y^*)}{\varepsilon_w^S(y^*)}\frac{y^*f_Y(y^*)}{1 - F_Y(y^*)} + \int_{\mathbb{R}_+} T'(y)\left(1 + \frac{1}{\varepsilon_w^S(y)}\right)\cdots$$

$$\times \left[ -\varepsilon_r(y) \frac{\delta(y-y^*)}{1-T'(y)} - \frac{1}{1-T'(y^*)} \varepsilon_w(y) \Gamma(y, y^*) \varepsilon_r(y^*) \right] \frac{y f_Y(y)}{1 - F_Y(y^*)} \, dy$$

$$= \hat{\mathcal{R}}_{\mathrm{ex}}(y^*) + \frac{T'(y^*)}{1-T'(y^*)} \varepsilon_r(y^*) \frac{y^* f_Y(y^*)}{1 - F_Y(y^*)} \left(1 + \varepsilon_w^S(y^*)\right) \frac{1}{\varepsilon_w^D(y^*)}$$

$$- \frac{\varepsilon_r(y^*)}{1-T'(y^*)} \int_{\mathbb{R}_+} T'(y) \left(1 + \varepsilon_w^S(y)\right) \frac{\Gamma(y, y^*)}{1 + \varepsilon_w^S(y)/\varepsilon_w^D(y)} \frac{y f_Y(y)}{1 - F_Y(y^*)} \, dy. \quad (41)$$

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 $\hat{\mathcal{R}}(y^*) = \hat{\mathcal{R}}_{\mathrm{ex}}(y^*)$.
*Q.E.D.*

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

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

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

$$\hat{\mathcal{R}}(y^*) = \hat{\mathcal{R}}_{\mathrm{ex}}(y^*) + \varepsilon_r \left(1 + \varepsilon_w^S\right) \left[ \frac{T'(y^*)}{1-T'(y^*)} \frac{1}{\sigma} \frac{y^* f_Y(y^*)}{1 - F_Y(y^*)} \right.$$

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

Suppose first that $p = 0$, that is, 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 $(1 + \varepsilon_w^S) \frac{\varepsilon_r}{\sigma} = \frac{1+\varepsilon_w^S}{\sigma + \varepsilon_w^S} \varepsilon_r^S$ 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 $\mathrm{Cov}(T'(y); y\Gamma(y, y^*))$ and using $\frac{1}{\mathbb{E}y} \mathrm{Cov}(T'(y); y) = \frac{1}{\mathbb{E}y} \mathbb{E}[yT'(y)] - \mathbb{E}[T'(y)]$ easily leads to (19).
*Q.E.D.*

*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

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

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

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

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

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

$$\hat{\mathcal{W}}(y^*) = \int_{y^*}^\infty \left(1 - g(y)\right) \frac{f_Y(y)}{1 - F_Y(y^*)} \, dy - \varepsilon_r^S(y^*) \frac{T'(y^*)}{1 - T'(y^*)} \frac{y^* f_Y(y^*)}{1 - F_Y(y^*)}$$

$$+ \frac{\dfrac{\varepsilon_r(y^*)}{\varepsilon^D(y^*)}}{1 - T'(y^*)} \psi(y^*) \frac{y^* f_Y(y^*)}{1 - F_Y(y^*)}$$

$$- \frac{\varepsilon_r(y^*)}{1 - T'(y^*)} \int \psi(y) \frac{\Gamma(y, y^*)}{1 + \dfrac{\varepsilon_w^S(y)}{\varepsilon_w^D(y)}} \frac{y f_Y(y)}{1 - F_Y(y^*)} \, dy, \tag{42}$$

where $\psi(y)$ is defined by $\psi(y) = (1 + \varepsilon_w^S(y)) T'(y) + g(y)(1 - T'(y))$. 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 $T'(y)(1 + \varepsilon_w^S(y))$ 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 $g(y)$.

## APPENDIX 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, $e_r^S(\theta) \equiv \frac{r(\theta)}{l(\theta)} \cdot \frac{\partial l(\theta)}{\partial r(\theta)}|_{u\text{cst}}$, and the income effect, $e_R(\theta) \equiv r(\theta)w(\theta)\frac{\partial l(\theta)}{\partial R}$, as follows:

$$e_r^c(\theta) = \frac{U_l(\theta)/l(\theta)}{U_{ll}(\theta) + \left(\dfrac{U_l(\theta)}{U_c(\theta)}\right)^2 U_{cc}(\theta) - 2\left(\dfrac{U_l(\theta)}{U_c(\theta)}\right)U_{cl}(\theta)},$$

$$e_R(\theta) = \frac{-\left(\dfrac{U_l(\theta)}{U_c(\theta)}\right)^2 U_{cc}(\theta) + \left(\dfrac{U_l(\theta)}{U_c(\theta)}\right)U_{cl}(\theta)}{U_{ll}(\theta) + \left(\dfrac{U_l(\theta)}{U_c(\theta)}\right)^2 U_{cc}(\theta) - 2\left(\dfrac{U_l(\theta)}{U_c(\theta)}\right)U_{cl}(\theta)}. \tag{43}$$

The labor supply elasticity with respect to the wage is given by $e_w^S(\theta) = (1 - p(y(\theta))) \times e_r^c(\theta) + e_R(\theta)$. 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 $\hat{l}_{pe}(\theta)$ 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 $\theta$. Note that the partial-equilibrium formula for $\hat{l}_{pe}(\theta)$ 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):

$$0 = \left[1 - T'\big((w_\theta + \mu\hat{w}_\theta)(l_\theta + \mu\hat{l}_\theta)\big) - \mu\hat{T}'(w_\theta l_\theta)\right](w_\theta + \mu\hat{w}_\theta) \dots$$

$$\times U_c\big[(w_\theta + \mu\hat{w}_\theta)(l_\theta + \mu\hat{l}_\theta) - T\big((w_\theta + \mu\hat{w}_\theta)(l_\theta + \mu\hat{l}_\theta)\big) - \mu\hat{T}(w_\theta l_\theta), l_\theta + \mu\hat{l}_\theta\big]$$

$$+ U_l\big[(w_\theta + \mu\hat{w}_\theta)(l_\theta + \mu\hat{l}_\theta) - T\big((w_\theta + \mu\hat{w}_\theta)(l_\theta + \mu\hat{l}_\theta)\big) - \mu\hat{T}(w_\theta l_\theta), l_\theta + \mu\hat{l}_\theta\big].$$

A first-order Taylor expansion implies

$$\frac{\hat{l}_\theta}{l_\theta} = \frac{e_R(\theta) + \big(1 - p(y_\theta)\big)e_r^c(\theta)}{1 + p(y_\theta)e_r^c(\theta)}\frac{\hat{w}_\theta}{w_\theta}$$

$$- \frac{e_r^c(\theta)}{1 + p(y_\theta)e_r^c(\theta)}\frac{\hat{T}'(y_\theta)}{1 - T'(y_\theta)} - \frac{e_R(\theta)}{1 + p(y_\theta)e_r^c(\theta)}\frac{\hat{T}(y_\theta)}{\big(1 - T'(y_\theta)\big)y_\theta},$$

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).                                     *Q.E.D.*

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*

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

$$- \phi\varepsilon_r^S(1 - p)\eta\mathbb{E}\left[\frac{T'(y) - \bar{T}'}{1 - T'(y)}\Big|y > y^*\right], \qquad (45)$$

*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 $\Pi$, we have $\hat{\mathcal{R}}(y^*) > \hat{\mathcal{R}}_{\mathrm{ex}}(y^*)$ 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

$$\hat{l}(\theta) = -\frac{\varepsilon_r(\theta)}{1 - T'(y(\theta^*))}\frac{\delta\big(y(\theta) - y\big(\theta^*\big)\big)}{1 - F\big(\theta^*\big)} + \frac{\varepsilon_R(\theta)}{(1 - T'(y(\theta)))y(\theta)}\frac{\mathbb{I}_{\{\theta > \theta^*\}}}{1 - F\big(\theta^*\big)}$$

$$+ \frac{\dfrac{1}{1 - F\big(\theta^*\big)}\varepsilon_w(\theta)}{1 - \displaystyle\int_\Theta \varepsilon_w(\theta')\gamma(\theta, \theta')\,d\theta'}\left[-\gamma\big(\theta, \theta^*\big)\frac{\varepsilon_r\big(\theta^*\big)}{1 - T'(y(\theta^*))}\right.$$

---

[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.

$$+ \int_{\theta^*}^{\bar{\theta}} \gamma(\theta, \theta') \frac{\varepsilon_R(\theta')}{(1 - T'(y(\theta')))y(\theta')} \, d\theta' \Bigg].$$

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

$$\hat{\mathcal{R}}(y(\theta^*)) = \hat{\mathcal{R}}_{\text{ex}}(y(\theta^*))$$

$$+ \frac{\frac{1}{1 - F(\theta^*)} \varepsilon_r^S}{1 - T(y(\theta^*))} \left( \frac{1}{\varepsilon_w^D} \frac{1 + \varepsilon_w^S}{1 + \frac{\varepsilon_w^S}{\varepsilon_w^D}} T'(y(\theta^*)) y(\theta^*) f(\theta^*) \right.$$

$$\left. - \int T'(y(\theta)) y(\theta) \frac{\gamma(\theta, \theta^*) \frac{1 + \varepsilon_w^S}{1 + \varepsilon_w^S / \varepsilon_w^D} \frac{1}{1 + \varepsilon_w^S / \varepsilon_w^D}}{1 - \int \varepsilon_w \gamma(x, x) \, dx} \, dF(\theta) \right)$$

$$+ \int_{\theta^*}^{\bar{\theta}} \frac{\frac{1}{1 - F(\theta^*)} \varepsilon_R^S}{(1 - T'(y(\theta'))) y(\theta')} \left( \frac{1}{\varepsilon_w^D} \frac{1 + \varepsilon_w^S}{1 + \frac{\varepsilon_w^S}{\varepsilon_w^D}} T'(y(\theta')) y(\theta') f(\theta') \right.$$

$$\left. - \int T'(y(\theta)) y(\theta) \frac{\gamma(\theta, \theta') \frac{1 + \varepsilon_w^S}{1 + \varepsilon_w^S / \varepsilon_w^D} \frac{1}{1 + \varepsilon_w^S / \varepsilon_w^D}}{1 - \int \varepsilon_w \gamma(x, x) \, dx} \, dF(\theta) \right) d\theta'.$$

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

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

$$\left. - \eta(1 - p) \left( \frac{\Pi}{\Pi - p} \frac{1 - \bar{T}'}{1 - T'(y^*)} - 1 \right) \right]. \tag{46}$$

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

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 approximately 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%, resp.) if the elasticity of substitution is $\sigma = 0.66$ ($\sigma = 3.1$, resp.).

## APPENDIX 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, that is, 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, Figure 2).

*CES Production Function with Exogenous Assignment.* Denote by $\theta_y$ 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 $1 - T'(y) = v'(\frac{y}{w})\frac{1}{w}$ 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 $\sigma$. 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 $y'(\theta_y) \equiv \frac{dy(\theta)}{d\theta}|_{\theta_y}$. We can without loss of generality assume that $\theta$ 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 $\alpha_1 = 0.41$ and $\alpha_2 = 3.01$. The parameter $\alpha_1$ represents the pure returns to skill and $\alpha_2$ 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 $\alpha_2$. But we replace the constant $\alpha_1$ with
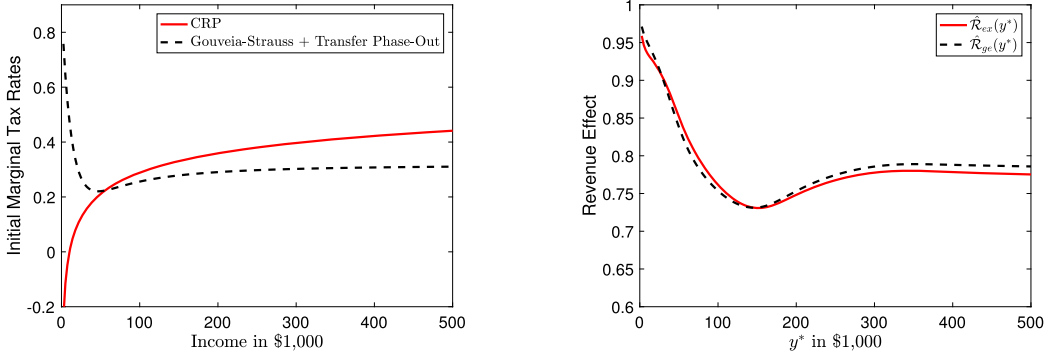
FIGURE 5.—Alternative specification of the baseline tax schedule. Bold curve: CRP tax schedule. Dashed curve: Gouveia–Strauss approximation with additional distortions due to means-tested transfers. Left panel: Right panel: Revenue gains of elementary tax reforms. Bold curve: exogenous wages. Dashed curve: CES production function with $\sigma = 3.1$.

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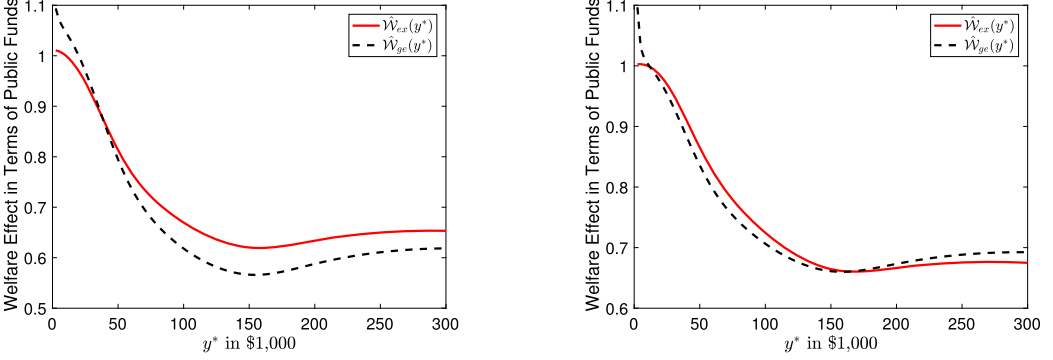
FIGURE E.6.—Welfare effect of elementary tax reforms for the social welfare function $G(u) = \frac{u^{1-\kappa}}{1-\kappa}$. Left panel: $\kappa = 1$. Right panel: $\kappa = 3$.

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.          Q.E.D.

### APPENDIX F: OPTIMAL TAXATION

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

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

COROLLARY 7—Optimal Tax Schedule in General Equilibrium: *The welfare-maximizing tax schedule T satisfies*: *for all $y^* \in \mathbb{R}_+$,*

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

$$\left. \times \int_{\mathbb{R}_+} \left[\psi(y^*) - \psi(y)\right]\frac{\Gamma(y, y^*)}{1 + \dfrac{\varepsilon_r^S(y^*)}{\varepsilon_r^D(y^*)}}\frac{y f_Y(y)}{1 - F_Y(y^*)}\, dy\right\}, \qquad (47)$$

*where $\psi(y) = (1 + \varepsilon_w^S(y))T'(y) + g(y)(1 - T'(y))$. 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

$$\frac{T'(y^*)}{1 - T'(y^*)} = \frac{1}{\varepsilon_r^S(y^*)}(1 - \bar{g}(y))\frac{1 - F_Y(y^*)}{y^* f_Y(y^*)}$$

$$+ \frac{\varepsilon_r(y^*)}{\varepsilon_r^S(y^*)}\frac{1}{1 - T'(y^*)}\int_{\mathbb{R}_+}[\psi(y^*) - \psi(y)]\frac{\Gamma(y, y^*)}{1 + \frac{\varepsilon_w^S(y)}{\varepsilon_w^D(y)}}\frac{yf_Y(y)}{y^* f_Y(y^*)}\,dy.$$

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

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

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

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

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

The previous equation can then be rewritten as

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

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

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

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

where $\gamma(y, y^*) = \frac{1}{\sigma}\frac{y^* f_Y(y^*)}{\int xf_Y(x)\,dx}$ 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

$$\hat{\mathcal{W}}(\hat{T}_2) = \int \hat{\mathcal{W}}(y^*)\hat{T}_2'(y^*)(1 - F_Y(y^*))\,dy^*$$

$$= -\frac{1}{\sigma}\frac{\varepsilon_r(y^*)}{1 - T'(y^*)}\frac{y^*f_Y(y^*)}{\int xf_Y(x)\,dx}\left\{\int (1 - g(y))(1 - T'(y))yf_Y(y)\,dy\ldots\right.$$

$$-\int \varepsilon_w(y)\left(\left[1 + \frac{1}{\sigma}(g(y) - 1)\right]T'(y) - \frac{1}{\sigma}g(y)\right)yf_Y(y)\,dy$$

$$-\frac{1}{\sigma}\frac{\int \varepsilon_w(y)y\,dF_Y(y)}{\mathbb{E}\left[\dfrac{x}{1 + \dfrac{1}{\sigma}\varepsilon_w^S(x)}\right]}$$

$$\times \int \frac{1}{1 + \dfrac{1}{\sigma}\varepsilon_w^S(x)}\left[(1 - g(x) + \varepsilon_w^S(x))T'(x) + g(x)\right]x\,dF_Y(x)\bigg\}.$$

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

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

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

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

$$= \int \frac{(1 - g(y)) + \dfrac{1 - A}{\sigma}\varepsilon_w^S(y)}{1 + \dfrac{1}{\sigma}\varepsilon_w^S(y)}yf_Y(y)\,dy.$$

Solving for $A$ implies $A = 1$.                                                            *Q.E.D.*

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

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

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 $\theta$'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 $\varepsilon_{y,\theta} = \varepsilon_{l,\theta} + \varepsilon_{w,\theta} = (1 + \frac{1}{e})\varepsilon_{l,\theta}$ 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 $\Pi$ is the Pareto parameter. Solving for $\tau^*$ leads to (22).              *Q.E.D.*

## REFERENCES

ALES, L., M. KURNAZ, AND C. SLEET (2015): "Technical Change, Wage Inequality, and Taxes," *The American Economic Review*, 105, 3061–3101. [3,14,15]

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, 3917–3946. [13]

CHETTY, R. (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80, 969–1018. [14]

COSTINOT, A., AND J. VOGEL (2010): "Matching and Inequality in the World Economy," *Journal of Political Economy*, 118, 747–786. [4,5]

DIAMOND, P. A., AND E. SAEZ (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165–190. [14]

GOLOSOV, M., A. TSYVINSKI, AND N. WERQUIN (2014): "A Variational Approach to the Analysis of Tax Systems," Working Paper. [11]

GUNER, N., R. KAYGUSUZ, AND G. VENTURA (2014): "Income Taxation of US Households: Facts and Parametric Estimates," *Review of Economic Dynamics*, 17, 559–581. [15]

GUNER, N., C. RAUH, AND G. VENTURA (2017): "Means-Tested Transfers in the US: Facts and Parametric Estimates," Working Paper. [15]

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. [13]

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38, 175–208. [3]

POLYANIN, A. D., AND A. V. MANZHIROV (2008): *Handbook of Integral Equations*. CRC press. [7]

ROTHSCHILD, C., AND F. SCHEUER (2014): "A Theory of Income Taxation Under Multidimensional Skill Heterogeneity," NBER Working Paper No. 19822. [3]

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229. [11,14,15]

SCHEUER, F., AND I. WERNING (2016): "Mirrlees Meets Diamond-Mirrlees," NBER Working Paper No. 22076. [3]

———— (2017): "The Taxation of Superstars," *The Quarterly Journal of Economics*, 132, 211–270. [3]

STIGLITZ, J. E. (1982): "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 17, 213–240. [3]

ZEMYAN, S. M. (2012): *The Classical Theory of Integral Equations: A Concise Treatment*. Springer Science & Business Media. [7]