Differential Taxation and Occupational Choice
Supplementary Material

Renato Gomes* **Jean-Marie Lozachmeur†** **Alessandro Pavan‡**

June 16, 2016

This document contains additional results for the manuscript *"Differential Taxation and Occupational Choice."* All numbered items (i.e., sections, subsections, lemmas, conditions, propositions, and equations) in this document contain the prefix "S". Any numbered reference without the prefix "S" refers to an item in the main text. Please refer to the main text for notation and definitions.

Section S1 contains proofs for the four-type example in Section 2 in the main text. Section S2 shows how the primal-dual approach used in Section 4 in the main text to identify properties of optimal taxation equilibria extends to economies with an arbitrary (finite) number of sectors. Section S3 describes technical details of the calibration exercise of Section 5 not contained in the main text.

S1. Proofs for the four-type example of Section 2

Set up. Consider the simple economy of Section 2 in which $\psi(h) = \frac{1}{2}h^2$ and where the agents' types are drawn from the following discrete distribution with four types

$$f(\underline{n}, \underline{n} - \varepsilon_a) = p_a,\ \ f(\underline{n} - \varepsilon_b, \underline{n}) = p_b,\ \ f(\bar{n}, \bar{n} - \delta_a) = q_a,\ \ f(\bar{n} - \delta_b, \bar{n}) = q_b,$$

with $0 < \underline{n} - \varepsilon_j \leq \underline{n} < \bar{n} - \delta_j \leq \bar{n}$ for $j \in \{a, b\}$ and with $p_j + q_j = \frac{1}{2}$ for $j \in \{a, b\}$.

In this example, the social planner designs a budget-balanced income tax system to maximize the utility of the worst-off agent in the economy (i.e., the social welfare function is Rawlsian).

Optimal tax codes under production efficiency. We consider first the optimal tax system consistent with production efficiency. Under such system, types $(\underline{n}, \underline{n} - \varepsilon_a)$ and $(\bar{n}, \bar{n} - \delta_a)$ work in sector a, whereas types $(\underline{n} - \varepsilon_b, \underline{n})$ and $(\bar{n} - \delta_b, \bar{n})$ work in sector b. In this case, the support for the distribution of productivities is the same across the two sectors. However, the tax and labor supply

*Toulouse School of Economics, University of Toulouse Capitole (CNRS), renato.gomes@tse-fr.eu.
†Toulouse School of Economics, University of Toulouse Capitole (CNRS, IDEI), jean-marie.lozachmeur@tse-fr.eu.
‡Northwestern University and CEPR alepavan@northwestern.edu.

schedules do not necessarily coincide across the two sectors, due to the fact that the skill intensity q_j/p_j may differ across the two sectors.

For each sector $j = a, b$, each $n \in \{\underline{n}, \bar{n}\}$, let

$$u_j(n) \equiv nh_j(n) - T_j\left(nh_j(n)\right) - \psi\left(h_j(n)\right)$$

denote the indirect utility of each agent with sector-j productivity equal to n working in sector j (i.e., the worker's utility evaluated at the optimal choice of labor supply). The combination of the intensive-margin (i.e., labor supply) incentive compatibility constraints with the extensive-margin (i.e., occupational choice) incentive compatibility constraints imply that for $j, k = a, b$, $k \neq j$

$$u_j(\bar{n}) = \max\left\{ \begin{array}{l} \underline{n}h_j(\underline{n}) - T_j\left(\underline{n}h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_j(\underline{n})\right); \\ \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_j}h_k(\underline{n})\right) \end{array} \right\}.$$

Note that the first term in the right-hand side of the previous condition is the utility of a worker of sector-j productivity $\bar{n}$ when he "mimics" a worker of sector-j productivity $\underline{n}$ working in sector j. The second term, instead, is the utility that the same high sector-j productivity worker obtains when he migrates to sector $k \neq j$ and then "mimics" a worker of sector-k productivity $\underline{n}$ (that is, when, after migrating, he produces the same income $y = \underline{n}h_k(\underline{n})$ as any worker of sector-k productivity $\underline{n}$ working in sector k).

To see why this condition must hold under any optimal tax code that is consistent with production efficiency, first suppose that there exists a sector j such that

$$u_j(\bar{n}) > \max\left\{ \begin{array}{l} \underline{n}h_j(\underline{n}) - T_j\left(\underline{n}h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_j(\underline{n})\right); \\ \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_j}h_k(\underline{n})\right); \\ \bar{n}h_k(\bar{n}) - T_k\left(\bar{n}h_k(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_j}h_k(\bar{n})\right) \end{array} \right\}.$$

Then the government could then increase the tax levied to any agent working in sector j with sector-j productivity equal to $\bar{n}$ without violating any of the incentive compatibility constraints. The higher taxes collected from any such agent could then be passed to the worse-off agents in the economy (whose productivity in each sector must necessarily be the lowest) thus contradicting the optimality of the tax code.

Now suppose that

$$\max\left\{ \begin{array}{l} \underline{n}h_j(\underline{n}) - T_j\left(\underline{n}h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_j(\underline{n})\right); \\ \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_j}h_k(\underline{n})\right); \\ \bar{n}h_k(\bar{n}) - T_k\left(\bar{n}h_k(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_j}h_k(\bar{n})\right) \end{array} \right\}$$
$$= \bar{n}h_k(\bar{n}) - T_k\left(\bar{n}h_k(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_j}h_k(\bar{n})\right)$$

and that

$$\bar{n}h_k(\bar{n}) - T_k\left(\bar{n}h_k(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_j}h_k(\bar{n})\right) > \tag{S0}$$
$$\max\left\{\underline{n}h_j(\underline{n}) - T_j\left(\underline{n}h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_j(\underline{n})\right);\ \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_j}h_k(\underline{n})\right)\right\}.$$

Then it must be that

$$u_k(\bar{n}) > \max\left\{\begin{array}{c} \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_k(\underline{n})\right); \\ \underline{n}h_j(\underline{n}) - T_j\left(\underline{n}h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_k}h_j(\underline{n})\right); \\ \bar{n}h_j(\bar{n}) - T_j\left(\bar{n}h_j(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_k}h_j(\bar{n})\right) \end{array}\right\}.$$

To see this, note that, if, instead,

$$u_k(\bar{n}) = \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_k(\underline{n})\right)$$

then, necessarily,

$$\bar{n}h_k(\bar{n}) - T_k\left(\bar{n}h_k(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_j}h_k(\bar{n})\right) < \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_j}h_k(\underline{n})\right).$$

This follows from the fact that incentive compatibility (on the intensive margin) requires that $\underline{n}h_k(\underline{n}) < \bar{n}h_k(\bar{n})$ along with the fact that payoffs satisfy the strict increasing differences property in (y, n) and $\underline{n} < \bar{n} - \delta_j < \bar{n}$. But this would contradicts (S0). Likewise, if

$$u_k(\bar{n}) = \max\left\{\begin{array}{c} \underline{n}h_j(\underline{n}) - T_j\left(\underline{n}h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_k}h_j(\underline{n})\right); \\ \bar{n}h_j(\bar{n}) - T_j\left(\bar{n}h_j(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_k}h_j(\bar{n})\right) \end{array}\right\}$$

then this would be inconsistent with the fact that

$$u_j(\bar{n}) = \bar{n}h_k(\bar{n}) - T_k\left(\bar{n}h_k(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_j}h_k(\bar{n})\right).$$

Hence, if (S0) holds, then necessarily

$$u_k(\bar{n}) > \max\left\{\begin{array}{c} \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_k(\underline{n})\right); \\ \underline{n}h_j(\underline{n}) - T_j\left(\underline{n}h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_k}h_j(\underline{n})\right); \\ \bar{n}h_j(\bar{n}) - T_j\left(\bar{n}h_j(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_k}h_j(\bar{n})\right) \end{array}\right\}.$$

The government could then again increase the tax applied to any agent working in sector k with sector-k productivity equal to $\bar{n}$ without violating any of the incentive compatibility constraints and then increase the utility of the worse-off agents in the economy contradicting once again the optimality of the tax code.

The properties established above thus imply that necessarily (S0) is false, meaning that

$$\bar{n}h_k(\bar{n}) - T_k\left(\bar{n}h_k(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_j}h_k(\bar{n})\right) \leq$$
$$\max\left\{\underline{n}h_j(\underline{n}) - T_j\left(\underline{n}h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_j(\underline{n})\right); \underline{n}h_k(\underline{n}) - T_k\left(\underline{n}h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_j}h_k(\underline{n})\right)\right\}$$

which in turn implies that, at the optimum,

$$u_j(\bar{n}) = \max \left\{ \begin{array}{l} u_j(\underline{n}) + \psi\left(h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_j(\underline{n})\right); \\ u_k(\underline{n}) + \psi\left(h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_j} h_k(\underline{n})\right) \end{array} \right\} \tag{S0-bis}$$

for $j = a, b$ with

$$\max \left\{ \begin{array}{l} u_j(\underline{n}) + \psi\left(h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_j(\underline{n})\right); \\ u_k(\underline{n}) + \psi\left(h_k(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_j} h_k(\underline{n})\right) \end{array} \right\} = u_j(\underline{n}) + \psi\left(h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_j(\underline{n})\right)$$

for at least one sector.

Now let $\Phi_e(q_a, q_b)$ denote welfare at the Rawlsian-optimal allocation respecting production efficiency. Note that the latter maximizes $\min\{u_a(\underline{n}), u_b(\underline{n})\}$ subject to the above constraints (S0-bis) and the budget-balance constraint

$$q_a T_a\left(h_a(\bar{n})\bar{n}\right) + p_a T_a\left(h_a(\underline{n})\underline{n}\right) + q_b T_b\left(h_b(\bar{n})\bar{n}\right) + p_b T_b\left(h_b(\underline{n})\underline{n}\right) = 0.$$

Optimal tax codes favoring sector a. We now turn to Rawlsian-optimal tax codes favoring sector a (those favoring sector b are derived in an analogous way). In particular, we define the latter as the tax codes maximizing the utility of the worse-off agents in the economy among those implementing the occupational choice rule that assigns types $(\underline{n}, \underline{n} - \varepsilon_a)$, $(\underline{n} - \varepsilon_b, \underline{n})$ and $(\bar{n}, \bar{n} - \delta_a)$ to sector a, and types $(\bar{n} - \delta_b, \bar{n})$ to sector b.

Let

$$P_b(q_a, q_b) \equiv \left\{ (\varepsilon_b, \delta_b) : \frac{1}{\frac{1}{2} - q_b} \left(\frac{\underline{n}}{\underline{n} - \varepsilon_b}\right)^2 + \frac{q_b}{\frac{1}{2} - q_a} \left(\frac{\underline{n}}{\bar{n} - \delta_b}\right)^2 < \frac{\frac{1}{2} + q_b}{\frac{1}{2} - q_b} + \frac{\frac{1}{2} + q_b}{\frac{1}{2} - q_a} - \frac{q_a}{\frac{1}{2} - q_a} \left(\frac{\underline{n}}{\bar{n}}\right)^2 \right\}.$$

Note that $(0, 0) \in P_b(q_a, q_b)$, given that $q_a, q_b \in \left(0, \frac{1}{2}\right)$.

Case 1 ***(separation)*** *The parameters of the model are such that* $(\varepsilon_b, \delta_b) \notin P_b(q_a, q_b)$.

We will show below that, whenever $(\varepsilon_b, \delta_b) \notin P_b(q_a, q_b)$, the Rawlsian-optimal tax system favoring sector a entails full separation. That is, in each sector, agents with different productivities in their sector of employment generate different income levels.

In this case, at the optimum, all adjacent downward (intensive-margin) incentive compatibility constraints in sector a must necessarily bind, implying that

$$h_a(\bar{n})\bar{n} - T_a\left(h_a(\bar{n})\bar{n}\right) - \psi\left(h_a(\bar{n})\right) = h_a(\underline{n})\underline{n} - T_a\left(h_a(\underline{n})\underline{n}\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_a(\underline{n})\right),$$

and

$$h_a(\underline{n})\underline{n} - T_a\left(h_a(\underline{n})\underline{n}\right) - \psi\left(h_a(\underline{n})\right) =$$
$$h_a(\underline{n} - \varepsilon_b)\left(\underline{n} - \varepsilon_b\right) - T_a\left(h_a(\underline{n} - \varepsilon_b)\left(\underline{n} - \varepsilon_b\right)\right) - \psi\left(\frac{\underline{n} - \varepsilon_b}{\underline{n}} h_a(\underline{n} - \varepsilon_b)\right),$$

Equivalently,

$$u_a(\bar{n}) = u_a(\underline{n}) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_a(\underline{n})\right).$$

and

$$u_a(\underline{n}) = u_a(\underline{n}-\varepsilon_b) + \psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right). \tag{S1}$$

Furthermore, the (extensive-margin) incentive compatibility constraint for the high-productivity agents working in sector b must also bind, implying that

$$h_b(\bar{n})\bar{n} - T_b\left(h_b(\bar{n})\bar{n}\right) - \psi\left(h_b(\bar{n})\right) = h_a(\underline{n})\underline{n} - T_a\left(h_a(\underline{n})\underline{n}\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b} h_a(\underline{n})\right)$$

or, equivalently,

$$u_b(\bar{n}) = u_a(\underline{n}) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b} h_a(\underline{n})\right)$$

To see why these constraints must bind, first observe that, if

$$u_b(\bar{n}) > \max\left\{\begin{array}{c} u_a(\underline{n}) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b} h_a(\underline{n})\right); \\ u_a(\underline{n}-\varepsilon_b) + \psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\bar{n}-\delta_b} h_a(\underline{n}-\varepsilon_b)\right); \\ u_a(\bar{n}) + \psi\left(h_a(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_b} h_a(\bar{n})\right) \end{array}\right\}$$

then the government could increase the tax levied on sector b's high productivity agents without violating any of the incentive compatibility constraints and then increase the utility of the worse-off agents in the economy contradicting once again the optimality of the tax code. Next observe that if

$$\max\left\{\begin{array}{l} u_a(\underline{n}) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b} h_a(\underline{n})\right); \\ u_a(\underline{n}-\varepsilon_b) + \psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\bar{n}-\delta_b} h_a(\underline{n}-\varepsilon_b)\right); \\ u_a(\bar{n}) + \psi\left(h_a(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_b} h_a(\bar{n})\right) \end{array}\right\}$$
$$= u_a(\bar{n}) + \psi\left(h_a(\bar{n})\right) - \psi\left(\frac{\bar{n}}{\bar{n}-\delta_b} h_a(\bar{n})\right)$$

then this means that none of the incentive-compatibility constraints of the high productivity agents working in sector a binds. The government could then increase the tax levied on sector a's high productivity agents, again increasing welfare. Hence at the optimum, the downstream adjacent incentive compatibility constraint of the highest productivity agent active in sector a must bind, that is, $u_a(\bar{n}) = u_a(\underline{n}) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_a(\underline{n})\right)$. Because $\underline{n} < \bar{n} - \delta_j < \bar{n}$ this in turn means that the incentive compatibility constraint for the high productivity agents in sector b takes the form $u_b(\bar{n}) = u_a(\underline{n}) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b} h_a(\underline{n})\right)$. Finally, that the downstream incentive compatibility constraint of those agents working in sector a with productivity equal to $\underline{n}$ must also bind (meaning that Condition (S1) must hold) follows from the fact that by reducing $u_a(\underline{n})$ the government can reduce the rents of all agents in the economy but those working in sector a with sector-a productivity equal to $\underline{n}-\varepsilon_b$ (the worst-off) and then use the extra taxes collected from all these agents to increase the utility of the worst-off agents, thus boosting welfare.

Combining the above constraints, we have that

$$u_a(\bar{n}) = u_a(\underline{n}-\varepsilon_b) + \psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_a(\underline{n})\right) \tag{S2}$$

and

$$u_b(\bar{n}) = u_a(\underline{n}-\varepsilon_b) + \psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b} h_a(\underline{n})\right). \tag{S3}$$

The budget balance constraint requires that

$$q_a T_a\left(h_a(\bar{n})\bar{n}\right) + p_a T_a\left(h_a(\underline{n})\underline{n}\right) + p_b T_a\left(h_a(\underline{n}-\varepsilon_b)\left(\underline{n}-\varepsilon_b\right)\right) + q_b T_b\left(h_b(\bar{n})\bar{n}\right) = 0.$$

Because of budget balance,

$$\begin{aligned} q_a u_a(\bar{n}) + p_a u_a(\underline{n}) + p_b u_a(\underline{n}-\varepsilon_b) + q_b u_b(\bar{n}) = q_a\left(h_a(\bar{n})\bar{n} - \psi\left(h_a(\bar{n})\right)\right) \\ + p_a\left(h_a(\underline{n})\underline{n} - \psi\left(h_a(\underline{n})\right)\right) + p_b\left(h_a(\underline{n}-\varepsilon_b)(\underline{n}-\varepsilon_b) - \psi\left(h_a(\underline{n}-\varepsilon_b)\right)\right) \\ + q_b\left(h_b(\bar{n})\bar{n} - \psi\left(h_b(\bar{n})\right)\right). \end{aligned}$$

Substituting the expressions in (S1), (S2), and (S3) into the above budget constraint, we have that

$$\begin{aligned} p_b u_a(\underline{n}-\varepsilon_b) = q_a\left(h_a(\bar{n})\bar{n} - \psi\left(h_a(\bar{n})\right)\right) + p_a\left(h_a(\underline{n})\underline{n} - \psi\left(h_a(\underline{n})\right)\right) \\ + p_b\left(h_a(\underline{n}-\varepsilon_b)(\underline{n}-\varepsilon_b) - \psi\left(h_a(\underline{n}-\varepsilon_b)\right)\right) + q_b\left(h_b(\bar{n})\bar{n} - \psi\left(h_b(\bar{n})\right)\right) \\ - q_a\left(u_a(\underline{n}-\varepsilon_b) + \psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_a(\underline{n})\right)\right) \\ - p_a\left(u_a(\underline{n}-\varepsilon_b) + \psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right)\right) \\ - q_b\left(u_a(\underline{n}-\varepsilon_b) + \psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b} h_a(\underline{n})\right)\right), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} u_a(\underline{n}-\varepsilon_b) = q_a\left(h_a(\bar{n})\bar{n} - \psi\left(h_a(\bar{n})\right)\right) + p_a\left(h_a(\underline{n})\underline{n} - \psi\left(h_a(\underline{n})\right)\right) \\ + p_b\left(h_a(\underline{n}-\varepsilon_b)(\underline{n}-\varepsilon_b) - \psi\left(h_a(\underline{n}-\varepsilon_b)\right)\right) + q_b\left(h_b(\bar{n})\bar{n} - \psi\left(h_b(\bar{n})\right)\right) \\ - q_a\left(\psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_a(\underline{n})\right)\right) \\ - p_a\left(\psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right)\right) \\ - q_b\left(\psi\left(h_a(\underline{n}-\varepsilon_b)\right) - \psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}} h_a(\underline{n}-\varepsilon_b)\right) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b} h_a(\underline{n})\right)\right). \end{aligned}$$

The government's problem then consists in choosing $h_a(\bar{n})$, $h_b(\bar{n})$, $h_a(\underline{n})$ and $h_a(\underline{n}-\varepsilon_b)$ to maximize the expression above. Using the fact that $\psi(h)=h^2/2$, we have that the corresponding first-order conditions imply that $h_a(\bar{n})=h_b(\bar{n})=\bar{n}$ and that

$$0=p_a\left(\underline{n}-h_a(\underline{n})\right)-q_a\left(h_a(\underline{n})-\left(\frac{\underline{n}}{\bar{n}}\right)^2 h_a(\underline{n})\right)-q_b\left(h_a(\underline{n})-\left(\frac{\underline{n}}{\bar{n}-\delta_b}\right)^2 h_a(\underline{n})\right),$$

or, equivalently,

$$h_a(\underline{n})=\left(\frac{1-p_b}{p_a}-\frac{q_a}{p_a}\left(\frac{\underline{n}}{\bar{n}}\right)^2-\frac{q_b}{p_a}\left(\frac{\underline{n}}{\bar{n}-\delta_b}\right)^2\right)^{-1}\underline{n}=\underline{n}b(\delta_b)$$

where

$$b(\delta_b)\equiv\left(\frac{1-p_b}{p_a}-\frac{q_a}{p_a}\left(\frac{\underline{n}}{\bar{n}}\right)^2-\frac{q_b}{p_a}\left(\frac{\underline{n}}{\bar{n}-\delta_b}\right)^2\right)^{-1}.$$

Notice that $b(\delta_b)$ is strictly increasing and that $b(\delta_b)\downarrow\left(\frac{1-p_b}{p_a}-\left(\frac{q_a+q_b}{p_a}\right)\left(\frac{\underline{n}}{\bar{n}}\right)^2\right)^{-1}$ as $\delta_b\downarrow 0$.

Moreover,

$$0=p_b\left((\underline{n}-\varepsilon_b)-h_a(\underline{n}-\varepsilon_b)\right)-(q_a+p_a+q_b)\left(h_a(\underline{n}-\varepsilon_b)-\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}}\right)^2 h_a(\underline{n}-\varepsilon_b)\right),$$

or, equivalently,

$$h_a(\underline{n}-\varepsilon_b)=\left(\frac{1}{p_b}-\frac{1-p_b}{p_b}\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}}\right)^2\right)^{-1}(\underline{n}-\varepsilon_b)=(\underline{n}-\varepsilon_b)c(\varepsilon_b)$$

where

$$c(\varepsilon_b)\equiv\left(\frac{1}{p_b}-\frac{1-p_b}{p_b}\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}}\right)^2\right)^{-1}.$$

Notice that $c(\varepsilon_b)$ is strictly decreasing and that $c(\varepsilon_b)\uparrow 1$ as $\varepsilon_b\downarrow 0$.

It follows from the above properties that the solution to the relaxed program (where all incentive compatibility constraints but the one considered above are neglected) violates the remaining incentive compatibility constraints (equivalently, the monotonicity constraints requiring that[1] $h_a(\underline{n}-\varepsilon_b)\leq h_a(\underline{n})$) if and only if

$$\left(\frac{1}{p_b}-\frac{1-p_b}{p_b}\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}}\right)^2\right)^{-1}(\underline{n}-\varepsilon_b)^2>\left(\frac{1-p_b}{p_a}-\frac{q_a}{p_a}\left(\frac{\underline{n}}{\bar{n}}\right)^2-\frac{q_b}{p_a}\left(\frac{\underline{n}}{\bar{n}-\delta_b}\right)^2\right)^{-1}\underline{n}^2,$$

where $p_j=\frac{1}{2}-q_j$ for each $j\in\{a,b\}$. This is equivalent to $(\varepsilon_b,\delta_b)\in P_b(q_a,q_b)$. Hence, when the parameters of the model satisfy the condition in Case 1, the Rawlsian-optimal tax schedule entails separation. In this case, welfare under the optimal tax schedule is given by

$$\hat{\Phi}_a(\varepsilon_b,\delta_b,q_a,q_b)\equiv\max_x\left\{\left(\frac{1}{2}-q_a\right)\left(x\underline{n}-\psi(x)\right)-q_a\left(\psi(x)-\psi\left(\frac{\underline{n}}{\bar{n}}x\right)\right)-q_b\left(\psi(x)-\psi\left(\frac{\underline{n}}{\bar{n}-\delta_b}x\right)\right)\right\}$$
$$+\max_y\left\{\left(\frac{1}{2}-q_b\right)\left(y(\underline{n}-\varepsilon_b)-\psi(y)\right)-\left(\frac{1}{2}+q_b\right)\left(\psi(y)-\psi\left(\frac{\underline{n}-\varepsilon_b}{\underline{n}}y\right)\right)\right\}+(q_a+q_b)\frac{\bar{n}^2}{2}.$$

[1] The constraint that $h_a(\underline{n})\leq h_a(\bar{n})$ is never binding.

By the envelope theorem,

$$\frac{\partial\hat{\Phi}_a}{\partial\varepsilon_b}(\varepsilon_b,\delta_b,q_a,q_b) = -\left(\frac{1}{2}+q_b\right)\frac{\underline{n}-\varepsilon_b}{\underline{n}^2}h_a(\underline{n}-\varepsilon_b)^2 < 0$$

and

$$\frac{\partial\hat{\Phi}_a}{\partial\delta_b}(\varepsilon_b,\delta_b,q_a,q_b) = q_b\frac{\underline{n}^2}{(\bar{n}-\delta_b)^3}h_a(\underline{n})^2 > 0.$$

Moreover,

$$\frac{\partial\hat{\Phi}_a}{\partial q_a}(\varepsilon_b,\delta_b,q_a,q_b) = T_a\left(h_a(\bar{n})\bar{n}\right) - T_a\left(h_a(\underline{n})\underline{n}\right) > 0$$

and

$$\frac{\partial\hat{\Phi}_a}{\partial q_b}(\varepsilon_b,\delta_b,q_a,q_b) = T_b\left(h_b(\bar{n})\bar{n}\right) - T_a\left(h_a(\underline{n}-\varepsilon_b)\left(\underline{n}-\varepsilon_b\right)\right) > 0,$$

where the last two inequalities follow from the incentive compatibility constraints (S1), (S2) and (S3). The incentive constraints also imply that

$$\frac{\partial\hat{\Phi}_a}{\partial q_a}(\varepsilon_b,\delta_b,q_a,q_b) < \frac{\partial\hat{\Phi}_a}{\partial q_b}(\varepsilon_b,\delta_b,q_a,q_b),$$

in which case $\hat{\Phi}_a(\varepsilon_b,\delta_b,q_a,q_b) < \hat{\Phi}_a(\varepsilon_b,\delta_b,q_b,q_a)$ if and only if $q_a > q_b$.

Next, consider the following case:

Case 2 ***(pooling)*** *The parameters of the model are such that* $(\varepsilon_b,\delta_b) \in P_b(q_a,q_b)$.

As established above, in this case, the monotonicity constraints $h_a(\underline{n}-\varepsilon_b) \leq h_a(\underline{n})$ bind implying that

$$(\underline{n}-\varepsilon_b)h_a(\underline{n}-\varepsilon_b) = \underline{n}h_a(\underline{n})$$

or, equivalently,

$$h_a(\underline{n}-\varepsilon_b) = \frac{\underline{n}}{\underline{n}-\varepsilon_b}h_a(\underline{n}).$$

The government's problem then consists in choosing $h_a(\bar{n})$, $h_b(\bar{n})$ and $h_a(\underline{n})$ so as to maximize

$$\begin{aligned} u_a(\underline{n}-\varepsilon_b) &= q_a\left(h_a(\bar{n})\bar{n} - \psi\left(h_a(\bar{n})\right)\right) \\ &+ p_a h_a(\underline{n})\underline{n} + p_b\left(\underline{n}h_a(\underline{n}) - \psi\left(\frac{\underline{n}}{\underline{n}-\varepsilon_b}h_a(\underline{n})\right)\right) + q_b\left(h_b(\bar{n})\bar{n} - \psi\left(h_b(\bar{n})\right)\right) \\ &- q_a\left(\psi\left(\frac{\underline{n}}{\underline{n}-\varepsilon_b}h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}}h_a(\underline{n})\right)\right) - p_a\psi\left(\frac{\underline{n}}{\underline{n}-\varepsilon_b}h_a(\underline{n})\right) \\ &- q_b\left(\psi\left(\frac{\underline{n}}{\underline{n}-\varepsilon_b}h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}-\delta_b}h_a(\underline{n})\right)\right). \end{aligned}$$

The first-order conditions imply that $h_a(\bar{n}) = h_b(\bar{n}) = \bar{n}$ and that

$$\begin{aligned}0 = &(p_a + p_b)\,\underline{n} - p_b \left(\frac{\underline{n}}{\underline{n} - \varepsilon_b}\right)^2 h_a(\underline{n}) \\ &- q_a \left(\left(\frac{\underline{n}}{\underline{n} - \varepsilon_b}\right)^2 h_a(\underline{n}) - \left(\frac{\underline{n}}{\bar{n}}\right)^2 h_a(\underline{n})\right) - p_a \left(\frac{\underline{n}}{\underline{n} - \varepsilon_b}\right)^2 h_a(\underline{n}) \\ &- q_b \left(\left(\frac{\underline{n}}{\underline{n} - \varepsilon_b}\right)^2 h_a(\underline{n}) - \left(\frac{\underline{n}}{\bar{n} - \delta_b}\right)^2 h_a(\underline{n})\right),\end{aligned}$$

or, equivalently,

$$h_a(\underline{n}) = \left(\frac{1}{p_a + p_b}\left(\frac{\underline{n}}{\underline{n} - \varepsilon_b}\right)^2 - \frac{q_a}{p_a + p_b}\left(\frac{\underline{n}}{\bar{n}}\right)^2 - \frac{q_b}{p_a + p_b}\left(\frac{\underline{n}}{\bar{n} - \delta_b}\right)^2\right)^{-1} \underline{n}.$$

Note that $h_a(\underline{n})$ is strictly decreasing in ε_b and strictly increasing in δ_b. Moreover,

$$h_a(\underline{n}) \to \left[\frac{1 - (q_a + q_b)\left(\frac{\underline{n}}{\bar{n}}\right)^2}{p_a + p_b}\right]^{-1} \underline{n} \text{ as } \varepsilon_b, \delta_b \downarrow 0.$$

When the parameters of the model satisfy the conditions in case 2, welfare under the Rawlsian-optimal tax system favoring sector a is then given by:

$$\begin{gathered}\tilde{\Phi}_a(\varepsilon_b, \delta_b, q_a, q_b) \equiv \\ (q_a + q_b)\frac{\bar{n}^2}{2} + \max_x \left\{(1 - q_a - q_b)\,x\underline{n} - \psi\left(\frac{\underline{n}}{\underline{n} - \varepsilon_b}x\right) + q_a\psi\left(\frac{\underline{n}}{\bar{n}}x\right) + q_b\psi\left(\frac{\underline{n}}{\bar{n} - \delta_b}x\right)\right\}.\end{gathered}$$

Note that $\tilde{\Phi}_a(0, 0, q_a, q_b) = \Phi_e(q_a, q_b)$. This follows from the fact that, when $(\varepsilon_b, \delta_b) = (0, 0)$, the Rawlsian-optimal tax system respecting production efficiency satisfies the following properties: $u_b(\underline{n}) = u_a(\underline{n})$ and $h_b(\underline{n}) = h_a(\underline{n})$, which implies that

$$\begin{gathered}\Phi_e(0, 0) = \\ (q_a + q_b)\frac{\bar{n}^2}{2} + \max_x \left\{(1 - q_a - q_b)\,x\underline{n} - \psi\left(\frac{\underline{n}}{\underline{n}}x\right) + (q_a + q_b)\,\psi\left(\frac{\underline{n}}{\bar{n}}x\right)\right\} \\ = \tilde{\Phi}_a(0, 0, q_a, q_b).\end{gathered}$$

Furthermore, by the envelope theorem,

$$\frac{\partial \tilde{\Phi}_a}{\partial \varepsilon_b}(\varepsilon_b, \delta_b, q_a, q_b) = -h_a(\underline{n})^2 \frac{\underline{n}^2}{(\underline{n} - \varepsilon_b)^3} < 0$$

and

$$\frac{\partial \tilde{\Phi}_a}{\partial \delta_b}(\varepsilon_b, \delta_b, q_a, q_b) = q_b h_a(\underline{n})^2 \frac{\underline{n}^2}{(\bar{n} - \delta_b)^3} > 0.$$

Moreover,

$$\frac{\partial \tilde{\Phi}_a}{\partial q_a}(\varepsilon_b, \delta_b, q_a, q_b) = T_a\left(h_a(\bar{n})\bar{n}\right) - T_a\left(h_a(\underline{n})\underline{n}\right) > 0,$$

and

$$\frac{\partial \tilde{\Phi}_a}{\partial q_b}(\varepsilon_b, \delta_b, q_a, q_b) = T_b\left(h_b(\bar{n})\bar{n}\right) - T_a\left(h_a(\underline{n})\underline{n}\right) > 0,$$

where the last two inequalities follow again from the incentive compatibility constraints (S1), (S2) and (S3). The incentive constraints also imply that

$$\frac{\partial \hat{\Phi}_a}{\partial q_a}(\varepsilon_b, \delta_b, q_a, q_b) < \frac{\partial \hat{\Phi}_a}{\partial q_b}(\varepsilon_b, \delta_b, q_a, q_b),$$

in which case $\hat{\Phi}_a(\varepsilon_b, \delta_b, q_a, q_b) < \hat{\Phi}_a(\varepsilon_b, \delta_b, q_b, q_a)$ if and only if $q_a > q_b$.

Let us now summarize the conclusions that the analysis above delivers. Welfare under the Rawlsian-optimal tax system favoring sector a is given by

$$\Phi_a(\varepsilon_b, \delta_b, q_a, q_b) = \begin{cases} \tilde{\Phi}_a(\varepsilon_b, \delta_b, q_a, q_b) & \text{if} \quad (\varepsilon_b, \delta_b) \in P_b(q_a, q_b) \\ \hat{\Phi}_a(\varepsilon_b, \delta_b, q_a, q_b) & \text{if} \quad (\varepsilon_b, \delta_b) \notin P_b(q_a, q_b). \end{cases}$$

Note that Φ_a is a continuous function. This function is differentiable everywhere except at $\partial P_b(q_a, q_b)$.[2] Moreover, Φ_a is strictly decreasing in ε_b and strictly increasing in δ_b, q_a and q_b (as Φ_a is continuous and both functions $\tilde{\Phi}_a$ and $\hat{\Phi}_a$ satisfy this property). Finally, Φ_a is such that $\Phi_a(\varepsilon_b, \delta_b, q_a, q_b) < \Phi_a(\varepsilon_b, \delta_b, q_b, q_a)$ if and only if $q_a > q_b$ (as Φ_a is continuous and both functions $\tilde{\Phi}_a$ and $\hat{\Phi}_a$ satisfy this property).

We are now ready to prove Results 1 and 2 in the main text (without loss of generality, the results are stated for the case in which $j = a$ and $k = b$).

Result 1 (Production Inefficiency) *Production efficiency fails at the optimum whenever the degree of skill transferability among low-productivity agents in sector b is sufficiently small. Formally, for any $\delta_b > 0$, there exists $\hat{\varepsilon}$ such that $\Phi_a > \Phi_e$ if and only if $\varepsilon_b < \hat{\varepsilon}$.*

Proof of Result 1. Fix $\delta_b > 0$. Note that $(0, \delta_b) \in P_b(q_a, q_b)$ and recall that $\tilde{\Phi}_a(0, 0, q_a, q_b) = \Phi_e(q_a, q_b)$. Because Φ_a is strictly increasing in δ_b, we know that $\Phi_a(0, \delta_b, q_a, q_b) > \Phi_e(q_a, q_b)$. The result then follows from the above property along with the fact that $\Phi_a(\varepsilon_b, \delta_b, q_a, q_b))$ is continuous and strictly decreasing in ε_b. Q.E.D.

Result 2 (Sectorial Bias) *Social welfare is higher by favoring sector b rather than sector a, that is, $\Phi_b > \Phi_a$, whenever one of the following mutually exclusive conditions hold:*

(1) Sector a enjoys the lowest degree of skill transferability among low-productivity agents and the highest among high-productivity agents (that is, $\varepsilon_a > \varepsilon_b$ and $\delta_a < \delta_b$), and both sectors are equally skill-intensive (that is, $\frac{q_a}{p_a} = \frac{q_b}{p_b}$).

(2) Sector a is less skill-intensive than sector b (that is, $\frac{q_a}{p_a} < \frac{q_b}{p_b}$), and both sectors enjoy the same degree of skill transferability among low and high productivity agents (that is, $\varepsilon_a = \varepsilon_b$ and $\delta_a = \delta_b$).

[2] ∂X denotes the border of the set X. Namely, $\partial X \equiv X \backslash \text{int}\,(X)$, where $\text{int}\,(X)$ is the interior of the set X.

Proof of Result 2. We prove claim 1 first. Let $q_a = q_b \equiv q$ (in which case $p_a = p_b = 1/2 - q$). Then

$$\Phi_b(\varepsilon_a, \delta_a, q, q) < \Phi_b(\varepsilon_b, \delta_a, q, q) < \Phi_b(\varepsilon_b, \delta_b, q, q) = \Phi_a(\varepsilon_b, \delta_b, q, q),$$

where the first inequality follows from the fact that Φ_a is strictly decreasing in ε_b and the second inequality from the fact that $\Phi_b(\varepsilon_b, \delta_a, q, q)$ is strictly increasing in δ_a. The last equality follows directly from the fact that the only difference between the two functions is the label used to define the sectors.

Next, consider claim 2. Let $\varepsilon_a = \varepsilon_b \equiv \varepsilon$, $\delta_a = \delta_b \equiv \delta$. The properties established above imply that

$$\Phi_a(\varepsilon, \delta, q_a, q_b) > \Phi_a(\varepsilon, \delta, q_b, q_a) = \Phi_b(\varepsilon, \delta, q_b, q_a),$$

where the inequality follows from the fact that $\Phi_a(\varepsilon_b, \delta_b, q_a, q_b) < \Phi_a(\varepsilon_b, \delta_b, q_b, q_a)$ if and only if $q_a > q_b$, and the equality follows again from the fact that the only difference between the two functions is the label used to define the sectors. Q.E.D.

S2. K-Sector Economies

Set-up. Consider the following extension of the baseline model of Section 3 in the main text to an economy with $k \in \mathbb{N}$ sectors, indexed by $j \in K = \{1, 2, ..., k\}$. The environment is the same as in Section 3 in the main text, except for the following adjustments. An agent's type is now given by the vector $\mathbf{n} \equiv (n_1, n_2, ..., n_k)$ specifying the agent's productivity in each of the sectors, with each $n_j \in N \equiv (\underline{n}, \overline{n})$, $j \in K$, with $\underline{n} > 0$, $\overline{n} \in \mathbb{R}_{++} \cup \{+\infty\}$ and $\overline{n} > \underline{n}$. Each agent's type is an independent draw from a distribution F with support $\mathbf{N} \equiv N^k$. The distribution F is absolutely continuous with respect to the Lebesgue measure. Let F_j denote the marginal of F with respect to the j-dimension (with bounded density f_j). The conditional distribution of $n_{-j} \equiv (n_1, ..., n_{j-1}, n_{j+1}, ..., n_k)$ given n_j is denoted by $F_{-j|j}(n_{-j}|n_j)$ (with bounded density $f_{-j|j}$). For any set $N_j \subset N$, $\bar{N}_j$ denotes the closure of the set, with $\bar{N} = [\underline{n}, \overline{n}]$.

An agent with sector-j productivity n_j, supplying $h_j \in \mathbb{R}_+$ hours in sector j, produces $n_j h_j$ units of effective labor. The income generated by this agent is $y_j = w_j n_j h_j$, where $w_j \in \mathbb{R}_+$ is the sector-j wage per unit of effective labor, with $\mathbf{w} = (w_1, ..., w_k) \in \mathbb{R}^k$.

Let $T_j(\cdot)$ denote the income tax schedule in sector j. The utility that an agent of type $\mathbf{n}$ obtains by supplying h_j hours in sector j is equal to

$$w_j h_j n_j - T_j(w_j h_j n_j) - \psi\left(h_j\right),$$

where, as in the main text, $\psi\left(h\right) = h^{\frac{1}{\xi}}$, with $\xi \in (0, 1)$.

The production side in each sector $j \in K$ continues to be described by a representative neoclassical firm with linear technology

$$X_j = \mathcal{F}_j(L_j) = L_j$$

where X_j is the amount of good-j produced, and where L_j is the total amount of effective labor hired by the firm. Firm j's profits are then equal to

$$\pi_j = (1 - w_j - \tau_j)L_j,$$

where $\tau_j \in \mathbb{R}$ is the sales tax rate on good j.

Taxation Equilibria. Hereafter, we confine attention to economies in which the government has the ability to use sector-specific income taxes. As argued in the main text, in this case, sales taxes are redundant. We thus drop them from now onwards and set all wages equal to one.

As in the baseline model of Section 3 in the main text, the occupational choice of each agent is described by a rule $\mathcal{C} : \mathbf{N} \to K$ specifying for each type $\mathbf{n} = (n_1, ..., n_k) \in \mathbf{N}$ the sector in which the agent works. In turn, the collection $h \equiv (h_j)_{j=1}^k$ of labor supply schedules $h_j : N_j \to \mathbb{R}_+$ specifies the amount of labor supplied by each individual in each sector as a function of the agent's productivity, with the domain

$$N_j = \{n_j \in N : \exists n_{-j} \in N^{k-1} \text{ such that } \mathcal{C}(n_j, n_{-j}) = j\}$$

of each function h_j corresponding to the set of productivity levels of those agents working in sector j. As in baseline model, the labor supplied by each agent in each sector is independent of the agent's productivity in any of the other sectors.

For any sector $j \in K$ any productivity level $n \in N$ let

$$\tilde{u}_j(n) \equiv \max_h \{hn - T_j(hn) - \psi(h)\}$$

denote the utility that an agent with sector-j productivity n can obtain by choosing optimally his labor supply in sector j. Similarly, for any $j \in K$ any $n \in N_j$, let

$$u_j(n) \equiv h_j(n)n - T_j(h_j(n)n) - \psi(h_j(n))$$

denote the equilibrium payoff of any agent with sector-j productivity n joining sector j.

Letting $\mathcal{T} = (T_j)_{j=1}^k$ denote a tax system, we then have that the definition of a taxation equilibrium $\mathcal{E} \equiv (\mathcal{C}, h, \mathcal{T})$ parallels the one in the baseline model, modulo the adjustment in the government's budget constraint which is now given by:

$$\sum_{j \in K} \int_{\{\mathbf{n}:\mathcal{C}(\mathbf{n})=j\}} T_j(h_j(n_j)n_j) dF(\mathbf{n}) \geq B,$$

where B continues to denote the exogenous government budget requirement.

Let

$$U(\mathbf{n}; \mathcal{E}) \equiv \max_{j \in K} \{\tilde{u}_j(n_j)\}$$

denote the indirect utility of an agent with type $\mathbf{n}$ under the taxation equilibrium $\mathcal{E} \equiv (\mathcal{C}, h, \mathcal{T})$. Hereafter, we restrict attention to a government with a concave utilitarian welfare objective (the case of a government with a Rawlsian objective is similar but omitted to ease the exposition). The government's problem thus consists in choosing a taxation equilibrium $\mathcal{E}$ so as to maximize

$$\Phi\left[U(\cdot;\mathcal{E})\right] \equiv \int_{\mathbf{n}\in\mathbf{N}} \phi\left(U(\mathbf{n};\mathcal{E})\right) dF(\mathbf{n}),$$

where the strictly increasing and weakly concave function ϕ captures the government's preferences for redistribution, as in the baseline model.

Implementability. Lemma (S1) below is the counterpart to Lemma 1 in the main text.

Lemma S1. *The allocation $(\mathcal{C},h)$ is implemented by the tax system $\mathcal{T}$ only if the following conditions jointly hold:*

1. *For every sector $j\in K$, the income schedule $y_j(n)\equiv h_j(n)n$ is nondecreasing over N_j. Moreover, the indirect utility schedule $u_j(n)$ is Lipschitz continuous over N_j with derivative equal to*
$$u_j'(n)=\psi'\left(h_j(n)\right)\frac{h_j(n)}{n} \text{ for almost every } n\in N_j. \tag{S4}$$
2. *There is a collection of absolutely continuous and nondecreasing threshold functions $c\equiv(c_{jl})_{j,l\in K,l\neq j}$ with each $c_{jl}:N\to\bar{N}$ strictly increasing at any point $n_l\in N$ at which $c_{jl}(n_j)\in N$, all $j,l\in K$, $l\neq j$, such that, for all $j\in K$, all $\mathbf{n}\in\mathbf{N}$, (a) $\mathcal{C}(\mathbf{n})=j$ if $n_l<c_{jl}(n_j)$ for all $l\in K$, $l\neq j$ and (b) $\mathcal{C}(\mathbf{n})\neq j$ if there exits $l\in K$, $l\neq j$ such that $n_l>c_{jl}(n_j)$. Furthermore, these threshold functions are such that, for all $j,l\in K$, $l\neq j$, all $\mathbf{n}\in\mathbf{N}$, $n_l<c_{jl}(n_j)$ if and only if $n_j>c_{lj}(n_l)$. Finally, for any $j,l\in K$, $l\neq j$ any $n_j\in N$, (i) $c_{jl}(n_j)\in(\underline{n},\bar{n})$ only if $\tilde{u}_j(n_j)=\tilde{u}_l(c_{jl}(n_j))$, in which case $c_{lj}(c_{jl}(n_j))=c_{jl}^{-1}(c_{jl}(n_j))=n_j$, (ii) $c_{jl}(n_j)=\underline{n}$ if $\tilde{u}_j(n_j)<\tilde{u}_l(n_l)$ for all $n_l\in N$, and (iii) $c_{jl}(n_j)=\bar{n}$ if $\tilde{u}_j(n_j)>\tilde{u}_l(n_l)$ for all $n_l\in N$.*

Suppose the allocation $(\mathcal{C},h)$, along with the tax system $\mathcal{T}$ satisfies the properties in parts 1-2 above. Then, there exists a tax system $\mathcal{T}'$ such that the allocation $(\mathcal{C},h)$ is implemented by the tax system $\mathcal{T}'$.

The proof of this result parallels the one in the main text and is thus omitted. The interesting part is Part 2. It establishes that any occupational choice rule can be described by a collection of nondecreasing threshold functions c_{jl}, one for each pair of sectors $j,l\in K$, $l\neq j$, such that an agent with type $\mathbf{n}\in\mathbf{N}$ is indifferent between working in sector j or in sector l if, and only, if $n_l=c_{jl}(n_j)$ (in which case, $n_j=c_{lj}(n_l)$). Because the payoff from working in a given sector strictly increases with the agent's productivity in that sector, each threshold function $c_{jl}(\cdot)$ is strictly increasing at any point where $c_{jl}(n)\in N$.

Distribution of Productivities. As in the baseline model, it is convenient to describe the endogenous distribution of productivities in each sector in terms of the collection of threshold functions $c=(c_{jl})_{j,l\in K,l\neq j}$ associated with the occupational choice rule $\mathcal{C}$. In order to do so, for any pair of sectors $j, l \in K$, $l \neq j$, let

$$n'_{jl} \equiv \begin{cases} \sup \{n \in N : c_{jl}(n) = \underline{n}\} \text{ if } \{n \in N : c_{jl}(n) = \underline{n}\} \neq \varnothing \\ \underline{n} \text{ if } \{n \in N : c_{jl}(n) = \underline{n}\} = \varnothing \end{cases}$$

and

$$n''_{jl} \equiv \begin{cases} \inf \{n \in N : c_{jl}(n) = \bar{n}\} \text{ if } \{n \in N : c_{jl}(n) = \bar{n}\} \neq \varnothing \\ \bar{n} \text{ if } \{n \in N : c_{jl}(n) = \bar{n}\} = \varnothing. \end{cases}$$

In words, if $c_{jl}(n) = \underline{n}$ for some $n \geq \underline{n}$, then no agent with sector-j productivity below n'_{jl} works in sector j. Furthermore no agent whose sector-j productivity is above n''_{jl} works in sector l. It is also convenient to define the thresholds

$$\underline{n}_j \equiv \sup \{n \in N : c_{jl}(n) = \underline{n} \text{ for some } l \in K,\ l \neq k\}$$
$$\bar{n}_j \equiv \inf \{n \in N : c_{jl}(n) = \bar{n} \text{ for all } l \in K,\ l \neq k\}.$$

Then note that (a) all agents whose sector-j productivity is below $\underline{n}_j$ work in some sector other than j and (b) for any $n_j > \underline{n}_j$ there exists a type $\mathbf{n} = (n_j, n_{-j})$ that works in sector j. This means that $\bar{N}_j = [\underline{n}_j, \bar{n}]$.[3] Finally note that, for any $n > \underline{n}_j$, any $l \in K$, $l \neq j$, $c_{jl}(n) \in (\underline{n}, \bar{n})$ implies that $c_{jl}(n) \in \bar{N}_l$. To see this, recall that $\tilde{u}_j(n) = \tilde{u}_l(c_{jl}(n))$. If $c_{jl}(n) \notin \bar{N}_l$, then there must exist a sector $m \in K$ with $m \neq l$ such that $\tilde{u}_m(n_m) > \tilde{u}_l(c_{jl}(n))$ for all $n_m \in N$. But this means that $\tilde{u}_m(n_m) > \tilde{u}_j(n)$ for all $n_m \in N$, which contradicts the fact that $n > \underline{n}_j$ (equivalently, that $n \in \bar{N}_j$).

Hereafter, we will say that the occupational choice rule $\mathcal{C}$ is *admissible* if its associated threshold functions c are such that, for all $j, l \in K$, $l \neq k$, (a) $c_{jl}(n)$ are absolutely continuous and strictly increasing over each interval (n'_{jl}, n''_{jl}), (b) equal to $\underline{n}$ for all $n_j < n'_{jl}$, and (c) equal to $\bar{n}$ for all $n_j > n''_{jl}$.

For any admissible rule, we denote by

$$G_j(n_j|c) \equiv \int_{\underline{n}}^{n_j} f_j(n) F_{-j|j}((c_{jl}(n))_{l \in K, l \neq j}|n) dn$$

the measure of agents working in sector j whose sector-j productivity does not exceed n_j. We then denote the density of $G_j(n_j|c)$ by $g_j(n_j|c) \equiv f_j(n_j) F_{-j|j}((c_{jl}(n_j))_{l \in K, l \neq j}|n_j)$.

Characterization procedure: Primal-dual approach. The characterization procedure we follow to identify properties of the optimal taxation equilibria parallels the one in the main text and is based on a primal/dual approach. In the primal, we fix an arbitrary admissible occupational choice rule $\mathcal{C}$ and find a taxation equilibrium that maximizes the government's objective among those that implement $\mathcal{C}$. We denote the *primal problem* by

$$\mathcal{P}_1(\mathcal{C}): \qquad \max_{(h, \mathcal{T})} \ \Phi\left[U(\cdot; (\mathcal{C}, h, \mathcal{T}))\right] \text{ s.t. } (\mathcal{C}, h, \mathcal{T}) \text{ is a taxation equilibrium.}$$

[3] Recall that $\bar{N}_j$ denotes the closure of the set N_j.

In each of the dual problems, instead, we fix the labor supply schedule in all sectors other than sector $s \in K$, and find the taxation equilibrium that maximizes the government's objective among those that implement $h_{-s} = (h_j)_{j \neq s}$:

$$\mathcal{P}_2^s(h_{-s}): \quad \max_{(\mathcal{C}, h_s, \mathcal{T})} \Phi\left[U(\cdot; (\mathcal{C}, (h_s, h_{-s}), \mathcal{T}))\right] \text{ s.t. } (\mathcal{C}, (h_{-s}, h_s), \mathcal{T}) \text{ is a taxation equilibrium.}$$

Clearly, the optimal taxation equilibrium $\mathcal{E}^* = (\mathcal{C}^*, h^*, \mathcal{T}^*)$ must be such that $(h^*, \mathcal{T}^*)$ solves $\mathcal{P}_1(\mathcal{C}^*)$ and $(\mathcal{C}^*, h_s^*, \mathcal{T}^*)$ solves the dual problem $\mathcal{P}_2^s(h_{-s}^*)$, for any sector $s \in K$.

Solution to the primal. As in the main text (and the rest of the literature), we abstract from bunching and corner solutions; that is, we will restrict attention to economies in which the optimality conditions described below identify income schedules $y_j(\cdot)$ that are nondecreasing and such that $y_j(n) > 0$ for all $n \in N_j$ (equivalently, $h_j(n) > 0$ for all $n \in N_j$).

As in the main text, let λ denote the multiplier associated with the government's budget constraint. For any sector $j \in K$, any productivity level $n \in N_j$, then let $m_j(n) \equiv \phi'(u_j(n))/\lambda$ denote the ratio of social marginal utility of all individuals with productivity n working in sector j to the marginal value of public funds for the government. The next proposition parallels Proposition 1 in the main text: It derives properties of optimal marginal tax rates for the solution to the primal problem corresponding to an arbitrary (but admissible) occupational choice rule $\mathcal{C}$.

Let $\mathbb{I}_A$ denote the indicator function, taking value one when property A holds and zero otherwise.

Proposition S1 (Generalized Mirrlees Formula) *Let $c = (c_{jl})_{j,l \in K, l \neq j}$ be the collection of threshold functions corresponding to the admissible occupational choice rule $\mathcal{C}$. The optimal tax system implementing the choice rule $\mathcal{C}$ satisfies the following generalized Mirrlees formulas for all sectors $j \in K$, almost all productivity levels $n \in N_j$:*

$$\underbrace{\xi \frac{T_j'(y_j(n))}{1 - T_j'(y_j(n))} n g_j(n|c)}_{E_j(n)} + \sum\nolimits_{l \neq j} \mathbb{I}_{n < n_{jl}''} \underbrace{\xi \frac{T_l'(y_l(c_{jl}(n)))}{1 - T_l'(y_l(c_{jl}(n)))} c_{jl}(n) g_l(c_{jl}(n)|c)}_{E_{jl}(c_{jl}(n))}$$

$$= \underbrace{\int_n^{\bar{n}} \left[1 - m_j(\tilde{n})\right] g_j(\tilde{n}|c) d\tilde{n},}_{D_j(n)} \tag{S5}$$

together with the occupational choice constraint

$$c_{jl}'(n) = \frac{h_j(n)\left[1 - T_j'(y_j(n))\right]}{h_l(c_{jl}(n))\left[1 - T_l'(y_l(c_{jl}(n)))\right]} \tag{S6}$$

all $l \in K$, $l \neq j$ for which $\mathbb{I}_{n < n_{jl}''}$ (equivalently, for which $c_{jl}(n) < \bar{n}$).

Proposition S1 generalizes the result in Proposition 1 in the main text to an economy with an arbitrary number of sectors. As in the baseline model, the result can be understood heuristically by considering perturbations of the tax code by which the government raises by one dollar the marginal tax rate in sector j at income level $y_j(n)$ and by the same amount the marginal tax rate in each sector $l \neq j$ for which $c_{jl}(n) < \bar{n}$ at income level $y_l(c_{jl}(n))$. The term $D_j(n)$ on the right-hand side of (S5) represents the direct benefit of the additional tax revenue collected from all agents working in sector j with incomes above $y_j(n)$ and of those agents working in sectors $l \in K$, $l \neq j$ with income levels above $y_l(c_{jl}(n))$. The gains of raising this extra money from such agents is naturally discounted by the reduction in these agents' utility, as captured by the terms $m_j(n) = \phi'(u_j(n))/\lambda$. In turn, the terms $E_j(n)$ and $E_{jl}(c_{jl}(n))$ on the left-hand-side of (S5) are the "elasticity effects," capturing the distortions in labor supply stemming from the higher marginal tax rates. Finally, (S6) describes how marginal tax rates relate to the slope of the threshold functions c_{jl}, for each pair of sectors $j, l \in K$, $l \neq j$, and each productivity level $n \in N_j$ for which $c_{jl}(n) < \bar{n}$. For an agent with sector-j productivity equal to n and sector-l productivity equal to $c_{jl}(n)$ to be indifferent as to whether to work in sector j or in sector l, the ratio of marginal net incomes (with respect to productivity) across the two sectors has to equal $c'_{jl}(n)$. Equivalently, note that this condition is simply an envelope condition requiring that, for types who are indifferent as to which of these two sectors to join it must be that $u'_j(n) = u'_l(c_{jl}(n))c'_{jl}(n)$.

As explained in the main text, the conditions in the proposition describe the marginal tax rates (and hence the labor supply choices) of those agents who are indifferent between at least two sectors. Because the labor supply (and utility) of any agent who strictly prefers one sector to all other sectors coincides with that of *some* agent who is indifferent, the result in the above proposition provides a complete characterization of the solution to the primal. Lastly, note that for any sector $j \in K$, any productivity level $n > \bar{n}_j$, the formula in (S5) coincides with the familiar formula from Mirrlees (1971). This is because any agent whose sector-j productivity is above $\bar{n}_j$ prefers working in sector j than in any other sector, irrespective of his productivity in any of the other sectors. For such agents, the only dependence of their marginal tax rate on the tax schedules in the other sectors is through the shadow cost of government funds λ.

We now proceed and prove the result in the above proposition.

Proof of Proposition S1. The government's problem consists in choosing a collection of labor supply and tax schedules $h_j : N_j \to \mathbb{R}_+$, $T_j : \mathbb{R}_+ \to \mathbb{R}$, for each sector $j \in K$, so as to maximize

$$\sum_{j=1}^{k} \int_{N_j} \phi(u_j(n)) dG_j(n|c)$$

where

$$u_j(n) = h_j(n)n - \psi(h_j(n)) - T_j(nh_j(n)) \text{ for every } n \in N_j, \text{ any } j \in K$$

subject to (a) the budget constraint

$$\sum_{j=1}^{k} \int_{N_j} T_j(nh_j(n)) dG_j(n|c) \geq B,$$

(b) the incentive-compatibility constraints on the intensive (labor supply) margin

$$u'_j(n) = \psi'(h_j(n)) \frac{h_j(n)}{n} \text{ for almost every } n \in N_j, \text{ any } j \in K \tag{S7}$$

along with the monotonicity requirement that $y_j(n) = h_j(n)n$ be nondecreasing over N_j, all $j \in K$, and (c) the incentive-compatibility constraints on the extensive (occupational choice) margin

$$u_j(n) = u_l(c_{jl}(n)), \text{ for all } n \in (\underline{n}_j, n''_{jl}) \text{ all } j, l \in K, \ l \neq j. \tag{S8}$$

To see that the constraints in (S8), along with the constraints on the intensive margin, imply that each agent prefers to select the sector specified by the occupational choice rule $\mathcal{C}$ to any other sector, first observe that any agent whose sector-j productivity is less than $\underline{n}_j$ prefers to join a sector other than j, for any $j \in K$. To see this, it suffices to recall that, by definition of the threshold $\underline{n}_j$, there exists some sector $l \neq j$ such that $c_{jl}(n) = \underline{n}$ for all $n \leq \underline{n}_j$. If $\underline{n}_j = \bar{n}$, the choice rule prescribes that no agent should work in sector j. This can always be induced by selecting a tax schedule T_j for sector j that specifies a sufficiently high tax withdraw for all income levels in sector j. Thus consider the more interesting case in which $\underline{n}_j < \bar{n}$. Let sector l be the one for which $c_{jl}(n) = \underline{n}$ for all $n \leq \underline{n}_j$ and $c_{jl}(n) > \underline{n}$ for all $n > \underline{n}_j$. The condition in (S8) implies that $u_j(\underline{n}_j) = u_l(\underline{n})$. Because $u_j(\cdot)$ and $u_l(\cdot)$ are strictly increasing, this means that any agent whose sector-j productivity is less than $\underline{n}_j$ is strictly better off joining sector l rather than sector j, irrespective of his sector-l productivity, which implies that no agent whose sector-j productivity is less than $\underline{n}_j$ joins sector j, as required by the choice rule $\mathcal{C}$.

Next, consider an agent whose sector-j productivity n_j is higher than $\underline{n}_j$. By definition of the threshold $\underline{n}_j$, $c_{jl}(n_j) > \underline{n}$. Suppose that $n_j < n''_{jl}$. Recall that this means that $c_{jl}(n_j) \in (\underline{n}, \bar{n})$, meaning that the agent should prefer to work in sector j than in sector l if and only if his sector-l productivity is less than $c_{jl}(n_j)$. The condition in (S8) along with the monotonicity of the $u_l(\cdot)$ function, guarantees that this is indeed the case. Next, suppose that $n_j > n''_{jl}$, which means that $c_{jl}(n_j) = \bar{n}$, that is, according to $\mathcal{C}$, the agent should prefer working in sector j than in sector l, irrespective of his sector-l productivity. Recall that this also means that $c_{lj}(\bar{n}) < n_j$. Now there are two possibilities. Either $c_{lj}(\bar{n}) > \underline{n}$, or $c_{lj}(\bar{n}) = \underline{n}$. In the former case, that an agent whose sector-j productivity is n_j prefers to work in sector j rather than in sector l, irrespective of his sector-l productivity follows again from (S8) along with the monotonicity of the $u_j(\cdot)$ and $u_l(\cdot)$ function (in fact, $u_j(n_j) > u_j(c_{lj}(\bar{n})) = u_l(\bar{n}) \geq u_l(n_l)$ for all $n_l \in N$). In the latter case, no agent should work in sector l. Again, this can be easily guaranteed by setting taxes sufficiently high in sector l.

Because the properties above apply to all sectors, we conclude that the constraints in (S8), along with the constraints on the intensive margin, imply that each agent prefers to select the sector specified by the occupational choice rule $\mathcal{C}$ to any other sector.

We now proceed by solving the above optimization problem. As anticipated above, we do so by abstracting from the constraints imposing that each income schedule $y_j(n) = h_j(n)n$ be nondecreasing over N_j, which is consistent with the practice commonly followed in the literature.

Pick an arbitrary sector $j \in K$. Combining (S7) with (S8), we have that, for any $l \in K$, $l \neq j$, almost any $n \in (\underline{n}_j, n''_{jl})$, the function $c_{jl}(\cdot)$ satisfies the following differential equation:

$$\psi'\left(h_j(n)\right)\frac{h_j(n)}{n} = c'_{jl}(n)\psi'\left(h_l(c_{jl}(n))\right)\frac{h_l(c_{jl}(n))}{c_{jl}(n)}. \tag{S9}$$

Equivalently, using the isoelastic specification of the disutility of labor, the labor supply schedules in the two sectors are linked by the following relationship, for almost any $n \in (\underline{n}_j, n''_{jl})$,

$$h_l(c_{jl}(n)) = J_{c_{jl}}[n]h_j(n), \tag{S10}$$

where the operator $J_{c_{jl}}$ is defined by

$$J_{c_{jl}}[n] \equiv \begin{cases} \left(\frac{c_{jl}(n)}{nc'_{jl}(n)}\right)^{\xi} & \text{if } c'_{jl}(n) > 0 \\ 1 & \text{otherwise.} \end{cases} \tag{S11}$$

Next, observe that, for any $l \in K$, $l \neq j$, $\bar{N}_l = [c_{jl}(\underline{n}_j), \bar{n}]$. To see this, consider first the case in which $\underline{n}_j = \underline{n}$. If $c_{jl}(\underline{n}) = \underline{n}$, because $c_{jl}(\cdot)$ is strictly increasing and continuous in a right neighborhood of $\underline{n}_j = \underline{n}$, it must be that $\tilde{u}_l(\underline{n}) = u_j(\underline{n})$. Because $\underline{n} \in \bar{N}_j$, then $\underline{n} \in \bar{N}_l$, which establishes the result. If, instead, $c_{jl}(\underline{n}) > \underline{n}$, then there are two possibilities. Either $c_{jl}(\underline{n}) = \bar{n}$, in which case no agent works in sector l, which is formally equivalent to saying that $\bar{N}_l = [c_{jl}(\underline{n}_j), \bar{n}] = \bar{n}$. Or, $c_{jl}(\underline{n}) \in (\underline{n}, \bar{n})$. Again, because $\underline{n} \in \bar{N}_j$ and because $\tilde{u}_l(\cdot)$ is continuous and strictly increasing, then from (S8), it must be that $\bar{N}_l = [c_{jl}(\underline{n}_j), \bar{n}]$. Next, suppose that $\underline{n}_j > \underline{n}$. If $c_{jl}(\underline{n}_j) = \underline{n}$, the result follows again from the fact that (a) $c_{jl}(\cdot)$ is strictly increasing and continuous in a right neighborhood of $\underline{n}_j$ along with (b) the fact that the functions u_j and $\tilde{u}_l$ are continuous and (c) the fact that $\underline{n}_j \in \bar{N}_j$. If, instead, $c_{jl}(\underline{n}_j) > \underline{n}$, the result follows from the following observations. Suppose $c_{jl}(\underline{n}_j) \in (\underline{n}, \bar{n})$. Then again (S8) along with the fact that $\underline{n}_j \in \bar{N}_j$ implies $\bar{N}_l \supset [c_{jl}(\underline{n}_j), \bar{n}]$. To see that $\bar{N}_l = [c_{jl}(\underline{n}_j), \bar{n}]$ then recall that, by the definition of the threshold $\underline{n}_j$, there exists some sector $m \in K$, $m \neq j$ such that $\underline{n}_j = \sup\{n \in N : c_{jm}(n) = \underline{n}\}$. Then observe that for any $n_l < c_{jl}(\underline{n}_j)$,

$$\tilde{u}_l(n_l) < \tilde{u}_l(c_{jl}(\underline{n}_j)) = u_l(c_{jl}(\underline{n}_j)) = u_j(\underline{n}_j) = u_m(\underline{n}).$$

This means that any agent with sector-l productivity $n_l < c_{jl}(\underline{n}_j)$ is strictly better off by joining sector m, irrespective of his sector-m productivity, thus implying that $\bar{N}_l$ does not include any $n_l < c_{jl}(\underline{n}_j)$.

Having established that $\bar{N}_j = [\underline{n}_j, \bar{n}]$ and $\bar{N}_l = [c_{jl}(\underline{n}_j), \bar{n}]$ for all $l \in K$, $l \neq j$, we then observe that there always exist a sector j for which $c_{jl}(\bar{n}) = \bar{n}$ all $l \in K$, $l \neq j$. To see this, suppose that $c_{jl}(\bar{n}) < \bar{n}$ for some $l \neq j$. Then observe that $c_{lj}(\bar{n}) = \bar{n}$. If for any $m \in K$, $m \neq l$ $c_{lm}(\bar{n}) = \bar{n}$ it then suffices to let sector l be the one for which the property holds. If, instead, there exists $m \in K$, $m \neq l, j$ such that $c_{lm}(\bar{n}) < \bar{n}$, then again note that $c_{ml}(\bar{n}) = \bar{n}$. Furthermore, $c_{mj}(\bar{n}) = \bar{n}$. The latter property follows from the combination of $c_{lj}(\bar{n}) = \bar{n}$ with $c_{ml}(\bar{n}) = \bar{n}$. In fact, condition $c_{lj}(\bar{n}) = \bar{n}$ means that an agent whose sector-l productivity is equal to $\bar{n}$ prefers working in sector l than in sector j, irrespective of his sector-j productivity. In turn, condition $c_{ml}(\bar{n}) = \bar{n}$ means that an agent whose sector-m productivity is equal to $\bar{n}$ prefers working in sector m than in sector l, irrespective of his sector-l productivity. Together, these properties imply that $c_{mj}(\bar{n}) = \bar{n}$. This transitivity property rules out cycles, thus establishing the result.

From now on, let j be the sector for which $c_{jl}(\bar{n}) = \bar{n}$ all $l \in K$, $l \neq j$. Next, observe that, for any $n \in (\underline{n}_j, \bar{n})$,

$$T_j(nh_j(n))) = h_j(n)n - \psi\left(h_j(n)\right) - u_j(n),$$

whereas for any $l \in K$, $l \neq j$, any $n \in (c_{jl}(\underline{n}_j), \bar{n})$,[4]

$$\begin{aligned} T_l(nh_l(n)) &= nh_l(n) - \psi\left(h_l(n)\right) - u_l(n) \\ &= nJ_{c_{jl}}[c_{jl}^{-1}(n)]h_j(c_{jl}^{-1}(n)) - \psi\left(J_c[c_{jl}^{-1}(n)]h_j(c_{jl}^{-1}(n)\right) - u_j(c_{jl}^{-1}(n)). \end{aligned}$$

Lastly, for any $n \in N_j$, any $l \in K$, $l \neq k$, let

$$c_{-l}^{j}(n;c) \equiv (c_{l1}(c_{jl}(n)), ..., c_{ll-1}(c_{jl}(n)), c_{ll+1}(c_{jl}(n)), ..., c_{lk}(c_{jl}(n)))$$

and similarly let

$$c_{-j}(n;c) \equiv (c_{j1}(n), ..., c_{jj-1}(n), c_{jj+1}(n), ..., c_{jk}(n)).$$

These functions conveniently permit us to describe the probability distribution F over the entire type space as a function of the productivity level in sector j alone. This simplification in turn will permit us to turn the government problem into a canonical optimal-control problem. To see this, note that the results and definitions above imply that, for any sector-j productivity level $n_j \in [\underline{n}_j, \bar{n}]$,

$$\begin{aligned} &F(c_{j1}(n_j), ..., c_{jj-1}(n_j), n_j, c_{jj+1}(n_j), ..., c_{jk}(n_j)) \\ &= \int_{\underline{n}_j}^{n_j} f_j(n)F_{-j|j}(c_{-j}(n;c)|n)dn + \sum\nolimits_{l\neq j} \int_{c_{jl}(\underline{n}_j)}^{c_{jl}(n_j)} f_l(n)F_{-l|l}(c_{-l}(n;c)|n)dn \\ &= \int_{\underline{n}_j}^{n_j} g_j(n|c)dn + \sum\nolimits_{l\neq j} \int_{c_{jl}(\underline{n}_j)}^{c_{jl}(n_j)} g_l(n|c)dn \\ &= \int_{\underline{n}_j}^{n_j} \left\{ f_j(n)F_{-j|j}(c_{-j}(n;c)|n) + \sum\nolimits_{l\neq j} c'_{jl}(n)\, f_l(c_{jl}(n))F_{-l|l}(c_{-l}^{j}(n;c)|c_{jl}(n)) \right\} dn \\ &= \int_{\underline{n}_j}^{n_j} \bar{g}_j(n|c)dn \end{aligned}$$

where the third equality follows from changing variables using $n_l = c_{jl}(n)$ along with the first equality and where the last equality follows from letting

$$\bar{g}_j(n|c) \equiv f_j(n)F_{-j|j}(c_{-j}(n;c)|n) + \sum\nolimits_{l\neq j} c'_{jl}(n)\, f_l(c_{jl}(n))F_{-l|l}(c_{-l}^{j}(n;c)|c_{jl}(n)).$$

[4]Note that here we are using the fact that $c_{jl}(\bar{n}) = \bar{n}$, which implies that for any $n \in (c_{jl}(\underline{n}_j), \bar{n})$, $c_{jl}^{-1}(n)$ is well defined and $c_{jl}^{-1}(n) \in [\underline{n}_j, \bar{n}]$.

Equipped with this notation, we can then conveniently rewrite the government's problem as consisting in choosing a pair of functions $u_j : N_j \to \mathbb{R}$ and $h_j : N_j \to \mathbb{R}_+$ so as to maximize

$$\int_{N_j} \phi(u_j(n))\bar{g}_j(n|c)dn \tag{S12}$$

subject to the budget constraint

$$\begin{aligned} &\int_{N_j} \left\{ [h_j(n)n - \psi(h_j(n)) - u_j(n)] f_j(n) F_{-j|j}(c_{-j}(n;c)|n) \right\} dn \\ &+ \sum\nolimits_{l \neq j} \int_{N_j} \left\{ \begin{array}{c} \left\{ J_{c_{jl}}[n] h_j(n) c_{jl}(n) - \psi\left(J_{c_{jl}}[n] h_j(n)\right) - u_j(n) \right\} \cdot \\ \cdot c'_{jl}(n) f_l(c_{jl}(n)) F_{-l|l}(c^j_{-l}(n;c)|c_{jl}(n)) \end{array} \right\} dn, \\ &\geq B \end{aligned} \tag{S13}$$

and the IC constraints

$$u'_j(n) = \psi'(h_j(n)) \frac{h_j(n)}{n} \text{ for almost every } n \in N_j.$$

As anticipated above, this is the key step in the proof, where we used the property that the type distribution, labor supply, and utility schedule in each sector other than sector j can be expressed in terms of the sector-j labor supply and utility schedule.[5] This simplification in turn permits us to reduce the government's problem to a standard optimal control problem with control variable h_j and state variable u_j. The Hamiltonian associated to this problem is:

$$\begin{aligned} H = &\phi(u_j(n))\bar{g}_j(n|c) + \lambda [h_j(n)n - \psi(h_j(n)) - u_j(n)] f_j(n) F_{-j|j}(c_{-j}(n;c)|n) \\ &+ \sum\nolimits_{l \neq j} \lambda \left\{ J_{c_{jl}}[n] h_j(n) c_{jl}(n) - \psi\left(J_{c_{jl}}[n] h_j(n)\right) - u_j(n) \right\} c'_{jl}(n) f_l(c_{jl}(n)) F_{-l|l}(c^j_{-l}(n;c)|c_{jl}(n)) \\ &+ \mu(n) \cdot \psi'(h_j(n)) \frac{h_j(n)}{n} - \lambda B, \end{aligned}$$

where λ is the Lagrange multiplier associated to the government budget constraint (S13) and where $\mu(n)$ is the co-state variable associated with the law of motion of the state variable u_j. The transversality conditions are:

$$\mu(\underline{n}_j) = \mu(\bar{n}) = 0. \tag{S14}$$

From Pontryagin Maximum Principle,

$$\mu'(n) = -\frac{\partial H}{\partial u_j} = \left[\lambda - \phi'(u_j(n))\right] \bar{g}_j(n|c). \tag{S15}$$

Integrating the right-hand side of (S15) and using the transversality condition (S14), we have that

$$\mu(n) = -\lambda \int_n^{\overline{n}} [1 - m_j(\tilde{n})] \bar{g}_j(\tilde{n}|c) d\tilde{n}, \tag{S16}$$

[5]Note that, for the simplification to be possible, it is essential to select a sector j for which $c_{jl}(\bar{n}) = \bar{n}$ all $l \in K$, $l \neq j$. As explained above, such a sector always exists, under any admissible occupational choice rule.

where we used the definition

$$m_j(n) \equiv \frac{\phi'\left(u_j(n)\right)}{\lambda}.$$

Furthermore, for any $n \in N_j$ such that $h_j(n) > 0$, the following first order condition must hold:

$$\begin{aligned}
&\frac{\partial H}{\partial h_j(n)} = \lambda\left[n - \psi'\left(h_j(n)\right)\right] f_j(n) F_{-j|j}(c_{-j}(n;c)|n) + \\
&\lambda \sum\nolimits_{l\neq j} \left\{J_{c_{jl}}[n]c_{jl}(n) - \psi'\left(J_{c_{jl}}[n]h_j(n)\right) J_{c_{jl}}[n]\right\} c'_{jl}(n)\, f_l(c_{jl}(n)) F_{-l|l}(c^j_{-l}(n;c)|c_{jl}(n)) + \\
&\mu(n)\frac{\psi'\left(h_j(n)\right) + \psi''\left(h_j(n)\right) h_j(n)}{n} = 0. \qquad \text{(S17)}
\end{aligned}$$

Combining (S16) with (S17) and using the definitions of the densities

$g_j(n|c) = f_j(n)F_{-j|j}(c_{-j}(n;c)|n)$ and, for all $l \in K,\ l \neq j,\ g_l(c_{jl}(n)|c) = f_l(c_{jl}(n))F_{-l|l}(c^j_{-l}(n;c)|c_{jl}(n))$

we obtain that

$$\begin{aligned}
&\left[n - \psi'\left(h_j(n)\right)\right] g_j(n_j|c) + \sum\nolimits_{l\neq j} \left\{J_{c_{jl}}[n]c_{jl}(n) - \psi'\left(J_{c_{jl}}[n]h_j(n)\right) J_{c_{jl}}[n]\right\} c'_{jl}(n)\, g_l(c_{jl}(n)|c) \qquad \text{(S18)} \\
&= \left\{\frac{\psi'\left(h_j(n)\right) + \psi''\left(h_j(n)\right) h_j(n)}{n}\right\} \int_n^{\overline{n}} \left[1 - m_j(\tilde{n})\right] \bar{g}_j(\tilde{n}|c) d\tilde{n}.
\end{aligned}$$

Next, note that, any interior point where the schedule T_j is differentiable, the following first-order condition for the optimal choice of labor supply $h_j(n)$ must hold:

$$n - \psi'\left(h_j(n)\right) = nT'_j(y_j(n)) \qquad \text{(S19)}$$

where $y_j(n) = nh_j(n)$. Using (S19), we have that

$$\begin{aligned}
\frac{\psi'\left(h_j(n)\right) + \psi''\left(h_j(n)\right) h_j(n)}{n} &= \frac{\psi'\left(h_j(n)\right)}{n}\left\{1 + \frac{\psi''\left(h_j(n)\right) h_j(n)}{\psi'\left(h_j(n)\right)}\right\} \\
&= \left[1 - T'_j(y_j(n))\right] \xi^{-1}.
\end{aligned}$$

Hence, using (S19), for any $n > \bar{n}_j$, we can rewrite the optimality condition (S18) as the usual Mirrlees condition

$$\xi \frac{T'_j(y_j(n))}{1 - T'_j(y_j(n))} n f_j(n) = \int_n^{\overline{n}} \left[1 - m_j(\tilde{n})\right] f_j(\tilde{n}) d\tilde{n}, \qquad \text{(S20)}$$

where we used the fact that, for $n > \bar{n}_j$, $c_{jl}(n) = \bar{n}$ all $l \in K$, $l \neq j$, and $g_j(n|c) = f_j(n)$.

Next, consider any $n \in (\underline{n}_j, \bar{n}_j)$. Using (S19) along with the fact that, for any sector $l \in K$, $l \neq j$, for which $n \leq n''_{jl}$ (i.e., for which $c_{jl}(n) < \bar{n}$), $h_l(c_{jl}(n)) = J_{c_{jl}}[n]h_j(n)$, we can rewrite Condition (S18) as follows:

$$\begin{aligned}
&\frac{T'_j(y_j(n))}{1 - T'_j(y_j(n))} n g_j(n|c) + \sum\nolimits_{l\neq j} \mathbb{I}_{n<n''_{jl}} \frac{T'_l(y_l(c_{jl}(n)))}{1 - T'_j(y_j(n))} J_{c_{jl}}[n] c'_{jl}(n) c_{jl}(n) g_l(c_{jl}(n)|c) \qquad \text{(S21)} \\
&= \xi^{-1} \int_n^{\overline{n}} \left[1 - m_j(\tilde{n})\right] \bar{g}_j(\tilde{n}|c) d\tilde{n}.
\end{aligned}$$

Furthermore, from Condition (S9), we have that, for any $n \in (\underline{n}_j, \bar{n}_j)$, any $l \in K$, $l \neq j$ for which $n \leq n''_{jl}$

$$c'_{jl}(n) = \frac{\psi'\left(h_j(n)\right)\frac{h_j(n)}{n}}{\psi'\left(h_l(c_{jl}(n))\right)\frac{h_l(c_{jl}(n))}{c_{jl}(n)}},$$

which, using (S19), we can rewrite as

$$c'_{jl}(n) = \frac{\left[1 - T'_j(y_j(n))\right]h_j(n)}{\left[1 - T'_l(y_l(c_{jl}(n)))\right]h_l(c_{jl}(n))}. \tag{S22}$$

Replacing (S22) into (S21) and using again the fact that, for any $l \in K$, $l \neq j$ for which $n \leq n''_{jl}$, $J_{c_{jl}}[n]h_j(n) = h_l(c_{jl}(n))$, we then have that, for any $n \in (\underline{n}_j, \bar{n}_j)$,

$$\begin{aligned} &\xi\frac{T'_j(y_j(n))}{1 - T'_j(y_j(n))}ng_j(n|c) + \xi\sum\nolimits_{l\neq j}\mathbb{I}_{n<n''_{jl}}\frac{T'_l(y_l(c_{jl}(n)))}{1 - T'_l(y_l(c_{jl}(n)))}c_{jl}(n)g_l(c_{jl}(n)|c) \\ &= \int_n^{\overline{n}}\left[1 - m_j(\tilde{n})\right]\bar{g}_j(\tilde{n}|c)d\tilde{n}. \end{aligned} \tag{S23}$$

Combining the above results completes the proof for the claim in the proposition. Q.E.D.

Solution to the dual. We now turn to the dual problem $\mathcal{P}_2^s(h_{-s})$, where the labor supply schedules $h_{-s} = (h_j)_{j\neq s}$ in all sectors other than sector $s \in K$ are held fixed, and where the sector-s labor supply schedule h_s (or equivalently, the occupational choice rule $\mathcal{C}$) is chosen so as to maximize the government's objective. The result in the next proposition is for a sector s for which there exists another sector $j \neq s$ such that $c_{jl}(\bar{n}) = \bar{n}$ all $l \in K$, $l \neq j$ (such a sector j always exists, under any occupational choice rule, as established in the proof of the preceding proposition). For simplicity, we also assume that, if $\underline{n}_j > \underline{n}$, then $c_{js}(\underline{n}_j) = \underline{n}$ (which implies that $\bar{N}_s = [\underline{n}, \bar{n}]$). This last assumption only matters for a certain transversality condition discussed in the proof of the proposition below. The Euler equation for the case in which $c_{js}(\underline{n}_j) > \underline{n}$ is similar but omitted.

Let $\varepsilon_{y_s}(n) \equiv y'_s(n)n/y_s(n)$ denote the elasticity of income with respect to productivity, in sector s.

Proposition S2 (Occupational Choice). *Let $s \in K$ be any sector for which there exists another sector $j \neq s$ such that $c_{jl}(\bar{n}) = \bar{n}$ all $l \in K$, $l \neq j$. Pick any productivity level $n \in (\underline{n}_j, n''_{js})$.*[6] The *solution to the dual problem $\mathcal{P}_2^s(h_{-s})$ – equivalently, the optimal tax system implementing the labor supply schedules $h_{-s} = (h_l)_{l\neq s}$ – satisfies the following integral-form Euler equation*

$$W_s(c_{js}(n)) = R_s(c_{js}(n)) + M(n) + \underbrace{E_s(c_{js}(n))(1 - T'_s(y_s(c_{js}(n)))y_s(c_{js}(n))}_{\text{continuity correction: } \Delta_b\left(c(n_a)\right)} \tag{S24}$$

where

$$W_s(c_{js}(n)) \equiv \int_{\underline{n}}^{c_{js}(n)} m_s\left(n_s\right)\left[1 - T'_s(y_s(n_s))\right]y_s(n_s)dG_s(n_s|c) \tag{S25}$$

[6] Recall that $n''_{js} = \inf\{n \in N : c_{js}(n) = \bar{n}\}$.

is the "welfare effect,"

$$R_s(c_{js}(n)) \equiv \int_{\underline{n}}^{c_{js}(n)} \left[1 - T_s'(y_s(n_s))\varepsilon_{y_s}(n_s)\right] y_s(n_s) dG_s(n_s|c) \tag{S26}$$

is the "revenue collection effect,"

$$M(n) \equiv \sum\nolimits_{l\neq s} \int_{\underline{n}_l}^{c_{sl}(c_{js}(n))} \left[T_l(y_l(n_l)) - T_s(y_s(c_{ls}(n_l)))\right] c_{ls}(n_l) f_{-(l,s)}(n_l, c_{ls}(n_l)) dn_l \tag{S27}$$

is the "migration effect," $E_s(c_{js}(n))$ *is the elasticity effect described in (*S5*), and* $f_{-(l,s)}(n_l, c_{ls}(n_l))$ *is the density of agents working in sector* l *with sector-*l *productivity* n_l *and sector-*s *productivity* $c_{ls}(n_l)$*.*

Before we formally prove the proposition, we first provide an explanation based on the same heuristics as in the main text, but adapted to the general multi-sector economy under examination here.

Heuristic derivation of the Euler condition in Proposition S2. Let $(\mathcal{C}, h_s, \mathcal{T})$ be a solution to the dual problem $\mathcal{P}_2^s(h_{-s})$. Pick an arbitrary sector $j \neq s$ along with a sector-j productivity level $n \in (\underline{n}_j, n''_{js})$. Let $c_{js}(n)$ be the sector-s productivity level specified by the occupational choice rule $\mathcal{C}$ such that any agent whose sector-s productivity is equal to $c_{js}(n)$ and whose sector-j productivity is equal to n is indifferent as to whether to work in sector s or in sector j.

Starting from $\mathcal{T}$, say that the tax system $\mathcal{T}^\alpha$ is a *sector-*s*-*α*-payroll-tax reform up to income level* $y_s(c_{js}(n))$ if for all sectors $l \neq s$, all y, $T_l^\alpha(y) = T_l(y)$, whereas, for sector s

$$T_s^\alpha(y) \equiv \begin{cases} \alpha y + T_s\left((1-\alpha)y\right) & \text{if} \quad y < y_s(c_{js}(n)) \\ T_s\left(y\right) & \text{if} \quad y \geq y_s(c_{js}(n)) \end{cases} \tag{S28}$$

Clearly, if $(\mathcal{C}, h_s, \mathcal{T})$ is a solution to the dual problem $\mathcal{P}_2^s(h_{-s})$, then no sector-s-α-payroll-tax reform up to income level $y_s(c_{js}(n))$ must increase the government's objective.

Next, observe that, under the perturbed tax schedule $\mathcal{T}^\alpha$, any agent working in sector s with productivity $n_s < c_{js}(n)$ supplying an amount of labor $h_s < y_s(c_{js}(n))/n_s$ obtains an utility equal to

$$n_s h_s - \psi(h_s) - T_s^\alpha(n_s h_s) = (1-\alpha) n_s h_s - \psi(h_s) - T_s\left((1-\alpha) n_s h_s\right). \tag{S29}$$

From (S29), one can see that such an utility is the same as that of an agent with sector-s productivity equal to $(1-\alpha)n_s$ under the original tax system $\mathcal{T}$. This means that, for α arbitrarily small, the indirect utility u_s^α of each agent working in sector s with sector-s productivity $n_s \geq \underline{n}/(1-\alpha)$ is given by

$$u_s^\alpha(n_s) \equiv \begin{cases} u_s\left((1-\alpha)n_s\right) & \text{if} \quad n_s < c_{js}(n) \\ u_s\left(n_s\right) & \text{if} \quad n_s > c_{js}(n). \end{cases} \tag{S30}$$

where u_s is the indirect utility function under the original tax system $\mathcal{T}$. As a consequence, the occupational choice rule under the perturbed tax system $\mathcal{T}^\alpha$ can be described by a collection of threshold functions c^α such that, for any $n_j < n$

$$c_{js}^\alpha(n) = \frac{1}{1-\alpha} c_{js}(n_j) \tag{S31}$$

of the various threshold functions c_{jm} under the original schedule T_m, for any $n_j < n$.

Next, consider the effects on the government's objective of such a perturbation, when α is infinitesimally small. The first effect, which in analogy to Proposition 2 in the main text, we call a **welfare effect** controls for the effect of the reform on the agents' utility. From (S30), it is easy to see that, at $\alpha = 0$, the marginal effect of such a reform on the indirect utility of any agent whose sector-s productivity is $n_s < c_{js}(n)$ is equal to $-n_s u_s'(n_s)$. Taking into account the social marginal weight that the planner assigns to such an agent is equal to $m_s(n_s)$ we then have that the welfare effect of such reform on any agent whose sector-s productivity is equal to n_s, with $n_s < c_{js}(n)$, is equal to

$$-m_s(n_s)\, u_s'(n_s)\, n_s = -m(n_s)\left[1 - T_s'(y_s(n_s))\right] y_s(n_s),$$

where the equality follows because incentive compatibility requires that

$$u_s'(n_s) = \psi'(h_s(n_s)) \frac{h_s(n_s)}{n_s}$$

and, moreover, at any point of differentiability of the tax schedule T_s,

$$n_s - \psi'(h_s(n_s)) = n_s T_s'(y_s(n_s)).$$

Integrating the expression above for all $n_s < c_{js}(n)$ leads to the term $W_s(c_{js}(n))$ in the Euler equation, as defined in (S25).

The second effect of such reform is a **revenue collection effect.** Using (S28), it is easy to see that, under the new system $\mathcal{T}^\alpha$, the revenue collected from each agent working in sector s with productivity $n_s < c_{js}(n)$ is given by[7]

$$\alpha n_s h_s((1-\alpha)n_s) + T_s\left((1-\alpha)n_s h_s((1-\alpha)n_s)\right). \tag{S32}$$

Differentiating the right-hand-side in (S32) with respect to α, and evaluating the expression at $\alpha = 0$, we obtain that the marginal effect of the reform on the revenue collected from each agent whose sector-s productivity $n_s < c_{js}(n)$ is equal to

$$\left[1 - T_s'(y_s(n_s))\varepsilon_{y_s}(n_s)\right] y_s(n_s).$$

Integrating the expression above for all $n_s < c_{js}(n)$ leads to the revenue collection effect $R(n)$ in the Euler equation, as defined in (S26).

[7] Again, we are assuming that α is small and that $n_s(1-\alpha) \geq \underline{n}$.

The third effect of such reform is the **migration effect**, which accounts for the fact that agents change occupation in response to the tax reform. Differentiating equation (S31) with respect to α, and evaluating the derivative at $\alpha = 0$, we have that the occupational choice rule shifts at a rate $c_{js}(n_j)$, at each productivity level $n_j < n$ in response to an infinitesimal reform. Accordingly, for any $n_j < n$, the mass of agents whose sector-j productivity is n_j and who change occupations is given by $c_{js}(n_j) f_{-(j,s)}(n_j, c_{js}(n_j))$ where $f_{-(j,s)}(n_j, c_{js}(n_j))$ is the density of agents with sector-j productivity equal to n_j, sector-s productivity equal to $c_{js}(n_j)$, and who prefer working in one of these two sectors than in any other sector. As a consequence, the impact on tax revenues from the migration of these agents from sector s to sector j is equal to:

$$[T_j(y_j(n_j)) - T_s(y_s(c_{js}(n_j)))]\, c_{js}(n_j) f_{-(j,s)}(n_j, c_{js}(n_j)).$$

Integrating the above expression for all $n_j < n$ leads to the migration effect

$$\int_{\underline{n}_j}^{n} [T_j(y_j(n_j)) - T_s(y_s(c_{js}(n_j)))]\, c_{js}(n_j) f_{-(j,s)}(n_j, c_{js}(n_j)) dn_j$$

in the Euler equation in (S27) – observe that, when $l = j$, $c_{sl}(c_{js}(n)) = c_{sj}(c_{js}(n)) = n$. Note that the welfare effect and the revenue collection effect pertain to individuals working in sector s both before and after the reform. In contrast, the migration effect pertains to individual leaving sector s to join one of the other sectors. In an economy with more than two sectors, the migration effect naturally accounts for the fact that agents migrate from sector s to each of the other sectors. In particular, the agents migrating from sector s to an arbitrary sector $l \neq s, j$ are those who, before the reform, are just indifferent between working in sector s and working in sector l and prefer working in one of these two sectors to working in any other sector. Integrating over all productivity levels from $\underline{n}_l$ to $c_{sl}(c_{js}(n))$,[8] we then have that the effect of the migration from sector s to sector l is given by:

$$\int_{\underline{n}_l}^{c_{sl}(c_{js}(n))} [T_l(y_l(n_l)) - T_s(y_s(c_{ls}(n_l)))]\, c_{ls}(n_l) f_{-(l,s)}(n_l, c_{ls}(n_l)) dn_l$$

Summing across all sectors $l \neq s$ (including sector j) then yields the expression in $M(n)$ in the Euler equation.

The last effect is a **continuity correction** and corresponds to the term $\Delta_s(c_{js}(n))$ in the right-hand side of the Euler equation (S24). As one can see from equation (S30), the reform introduces a discontinuity in the sector-s indirect utility schedule at $n_s = c_{js}(n)$. Indeed,

$$\lim_{n_s \to c_{js}(n)^-} u_s^\alpha(n_s) = u_s((1-\alpha)c_{js}(n)) < u_s(c_{js}(n)) = \lim_{n_s \to c_{js}(n)^+} u_s^\alpha(n_s),$$

[8] Recall that $\underline{n}_l$ is the lowest productivity level among those of all agents working in sector l and $c_{sl}(c_{js}(n))$ is the sector-l productivity level such that an agent working in sector l with productivity $c_{sl}(c_{js}(n))$ obtains the same utility as working in sector s with productivity $n_s = c_{js}(n)$.

for any $\alpha > 0$. Accordingly, for an α-reform to lead to an implementable allocation, it has to be coupled with transfers to sector-s agents with productivities in a neighborhood of $c_{js}(n)$, so as to restore the continuity of the indirect utility schedule. For incremental reforms (i.e., for $\alpha \approx 0$) only sector-s agents with productivity $c_{js}(n)$ need to receive such transfers. In order to reduce the indirect utility of those agents whose sector-s productivity is equal to $c_{js}(n)$ to its "continuity level" $\lim_{n_s \to c_{js}(n)^-} u_s^\alpha(n_s)$, the government charges a lump-sum tax to all agents working in sector s with productivity $c_{js}(n)$ (producing income $y_s(c_{js}(n))$) equal to the extra taxes that these agents would pay were they also subject to the reform. This lump-sum charge is the term $\Delta_s(c_{js}(n))$ in the right-hand side of the Euler equation (S24). It is equal to the product of (a) the elasticity effect $E_s(c_{js}(n))$ (capturing the foregone tax revenues per unit of marginal-tax increase) and (b) the change in marginal taxes that such agents would face were they also subject to the reform. At $\alpha \approx 0$, the change in marginal taxes faced by such agents is approximated (up to second-order effects) by the variation in their indirect utility, which is equal to $[1 - T_s'(y_s(c_{js}(n)))]\, y_s(c_{js}(n))$, as discussed in the derivation of the welfare effect.

We now turn to the formal proof of Proposition S2.

Proof of Proposition S2. Fix the labor supply schedule h_j for all sectors $j \neq s$ (each with domain $N_j = [\underline{n}_j, \bar{n}]$, $j \in K$, $j \neq s$). The planner's problem is as in the proof of Proposition S1, except that the control policies are now (i) the collection of threshold functions $c_{lr} : N \to \bar{N}$ defining the occupational choice rule, with each function c_{lr} continuous over N, strictly increasing over $(\underline{n}'_{lr}, n''_{lr})$ for some $\underline{n} \leq \underline{n}'_{lr}, n''_{lr} \leq \bar{n}$, and such that $c_{lr}(n) = \underline{n}$ for all $n < \underline{n}'_{lr}$ and $c_{lr}(n) = \bar{n}$ for all $n > n''_{lr}$, $l, r \in K$, $l \neq r$, (ii) sector-s labor supply schedule $h_s : N_s \to \mathbb{R}_+$, for some $N_s = [\underline{n}_s, \bar{n}]$, and the tax schedules $T_l : \mathbb{R}_+ \to \mathbb{R}$, one for each sector $l \in K$.

Note that, because the labor supply schedules for all sectors $j \neq s$ are fixed, so are the threshold functions c_{lr} for all pairs of sectors $l, r \neq s$ (this follows directly from the fact that the equilibrium utility functions $u_l(\cdot)$ in all sectors $l \neq s$ are uniquely pinned down by the labor supply schedules $h_l(\cdot)$ up to a constant that is uniform across all sectors and that corresponds to the utility of the agent with the lowest payoff).

The planner's problem can be conveniently rewritten by introducing the functions

$$R_l(n) \equiv \phi(u_l(n)) + \lambda \{h_l(n)n - \psi(h_l(n)) - u_l(n)\}, \tag{S33}$$

for each $l \in K$. These functions denote the value the planner assigns to the utility of an agent working in sector l, with sector-l productivity n, adjusted for the opportunity cost of raising funds from the agent, where λ is the multiplier associated with the government's budget constraint.

The planner's problem can then be reformulated as consisting in choosing a collection of threshold functions $c = (c_{jl})_{j,l \in K, l \neq j}$ satisfying the properties above, along with a labor supply schedule h_s :

$N_s \to \mathbb{R}_+$, and a collection of utility functions $u_l : N_l \to \mathbb{R}$, one for each $l \in K$, that jointly maximize

$$\sum_{l \in K} \int_{\underline{n}}^{\bar{n}} R_l\left(n\right) g_l(n|c) dn, \tag{S34}$$

subject to (a) the incentive compatibility constraints for the intensive (labor supply) margin

$$u_l'(n) = \psi'\left(h_l(n)\right) \frac{h_l(n_l)}{n_l} \text{ for almost every } n \in N_l, \text{ all } l \in K, \tag{S35}$$

and (b) the incentive compatibility constraints for the extensive (occupational choice) margin

$$h_r(c_{lr}(n)) = J_{c_{l,r}}[n] h_l(n),$$

for all $n \in (n'_{lr}, n''_{lr})$, all $l, r \in K$, $l \neq r$, with the functions $J_{c_{l,r}}$ as defined in the proof of Proposition S1.

Now let sector $j \in K$ be the one for which $c_{jl}(\bar{n}) = \bar{n}$ all $l \in K$, $l \neq j$ (that such a sector exists was established in the proof of the preceding proposition; that such a sector is different from sector s follows from the assumption in the proposition).

Arguments similar to those in the proof of Proposition S1 then permit us to write the planner's objective as[9]

$$\int_{\underline{n}_j}^{\bar{n}} \left[R_j\left(n\right) g_j(n|c) + \sum\nolimits_{l \neq j} \hat{R}_l\left(n|c\right) c'_{jl}\left(n\right) g_l(c_{jl}(n)|c)\right] dn, \tag{S36}$$

where, for any $n \in (\underline{n}_j, n''_{jl})$,

$$\begin{aligned}
\hat{R}_l\left(n|c\right) &= R_l\left(c_{jl}(n)\right) \\
&= \phi\left(u_l(c_{jl}(n))\right) + \lambda\left\{h_l(c_{jl}(n))c_{jl}(n) - \psi\left(h_l(c_{jl}(n))\right) - u_l\left(c_{jl}(n)\right)\right\} \\
&= \phi\left(u_j\left(n\right)\right) + \lambda\left\{J_{c_{jl}}[n]h_j(n)c_{jl}(n) - \psi\left(J_{c_{jl}}[n]h_j(n)\right) - u_j\left(n\right)\right\}
\end{aligned}$$

whereas $\hat{R}_l\left(n|c\right) = 0$ if $n \geq n''_{jl}$. Also recall that for any sector $l \neq j, s$, the threshold function c_{lj} is uniquely pinned down by the labor supply schedules h_j and h_l, as explained above.

Because the labor supply schedule h_l and the occupational choice thresholds c_{lr} are fixed for all sectors $l, r \in K$, $l, r \neq m$, the problem described above de facto reduces to choosing (i) a scalar $\underline{u}$ describing the utility of the lowest-productivity agent active in each sector $l \in K$, i.e., such that $u_l(\underline{n}_l) = \underline{u}$ all $l \in K$, and (ii) an absolutely continuous function $c_{js} : [\underline{n}_j, \bar{n}] \to \bar{N}$, strictly increasing over $[\underline{n}_j, n''_{js}]$ for some $n''_{js} \leq \bar{n}$, and satisfying $c_{js}(n) = \bar{n}$ for all $n \geq n''_{js}$, so as to maximize (S36).

[9] Recall that the transformation uses (a) the change in variables corresponding to the collection of pairwise threshold functions c_{jl} associated with the occupational choice rule $\mathcal{C}$ along with the fact that (a) the collection of sector-j utility levels $[u_j(\underline{n}_j), u_j(\bar{n})]$ spans the entire range of utility levels in the economy, (c) the labor supply schedule in each sector $l \neq j$ can be expressed as a function of the labor supply in sector j and the threshold functions c_{jl}, and (d) the definition of the endogenous densities for the distributions of productivities.

Given $\underline{u}$, because the labor supply schedule $h_l : (\underline{n}_l, \bar{n}) \to \mathbb{R}_+$ is fixed for all $l \neq s$, the utility function $u_l : (\underline{n}_l, \bar{n}) \to \mathbb{R}$ in each sector $l \neq s$ is then uniquely determined by

$$u_l(n) = \int_{\underline{n}_l}^{n} \psi' \left(h_l(\tilde{n})\right) \frac{h_l(\tilde{n})}{\tilde{n}} d\tilde{n} + \underline{u}.$$

Furthermore, once the threshold function $c_{js} : [\underline{n}_j, \bar{n}] \to \bar{N}$ specifying the locus of types who are indifferent between working in sector j and in sector s is in hands, (a) the set of productivity levels for sector s is then given by $\bar{N}_s = [c_{js}(\underline{n}_j), \bar{n}]$, (b) the sector-$s$ labor supply schedule is given by

$$h_s(n) = J_{c_{js}}[c_{js}^{-1}(n)] h_j(c_{js}^{-1}(n)), \tag{S37}$$

for all $n \in [c_{js}(\underline{n}_j), \bar{n}]$ and (c) the sector-s utility schedule is then given by

$$u_s(n) = u_j\left(c_{js}^{-1}(n)\right) \tag{S38}$$

for all $n \in [c_{js}(\underline{n}_j), \bar{n}]$. It is also easy to see that, once the threshold function $c_{js} : [\underline{n}_j, \bar{n}] \to \bar{N}$ for the pair of sectors $j, s \in K$ has been determined, the remaining threshold functions $c_{ls} : [\underline{n}_l, \bar{n}] \to \bar{N}$ involving sector s are uniquely determined by the utility schedules in the corresponding sectors.[10] More precisely, for any $l \in K$, $l \neq s, j$, any $n \in \bar{N}_l$, let

$$c_{ls}(n) = \begin{cases} u_s^{-1}(u_l(n)) \text{ if } u_l(n) \leq u_s(\bar{n}) \\ \bar{n} \text{ otherwise} \end{cases}$$

Finally, note that, once the scalar $\underline{u}$ and the threshold function $c_{js} : [\underline{n}_j, \bar{n}] \to \bar{N}$ have been determined, the tax schedule for each sector $l \in K$ can be set equal to

$$T_l(y) = \begin{cases} y - \psi\left(\frac{y}{y_l^{-1}(y)}\right) - u_l(y_l^{-1}(y)) & \text{if } \exists\, n \in N_l \text{ s.t. } y_l(n) = y \\ \bar{T} & \text{otherwise} \end{cases}$$

with $\bar{T}$ arbitrarily high to discourage agents working in sector l from generating any income $y \notin y_l(N_l)$.

Now, for simplicity, assume that sector s is such that $c_{js}(\underline{n}_j) = \underline{n}$ if $\underline{n}_j > \underline{n}$ (as it will be clear from the analysis below, this assumption only matters for the determination of a certain transversality condition; the case for $c_{js}(\underline{n}_j) > \underline{n}$ is similar, modulo an adjustment in one of the boundary conditions). As a first step towards the characterization of the optimal threshold function $c_{js} : [\underline{n}_j, \bar{n}] \to \bar{N}$, fix $n''_{js} \geq \underline{n}_j$ and look at the optimality conditions for the threshold function c_{js} over the interval $[\underline{n}_j, n''_{js}]$.[11] To ease the exposition, let $\tilde{R} : \mathbb{R}^3 \to \mathbb{R}$ be the function defined by

$$\tilde{R}_s(n, c, J) \equiv \phi\left(u_j(n)\right) + \lambda \left\{J h_j(n) c - \psi\left(J h_j(n)\right) - u_j(n)\right\}, \tag{S39}$$

[10] As explained above, the thresholds functions c_{lr} for all sectors $l, r \neq s$ are already pinned down by the labor supply schedules in these sectors, which are exogenously given as primitive of the dual problem.

[11] Recall that $n''_{js} = \inf\{n \in N : c_{js}(n) = \bar{n}\}$.

and denote by $\partial\tilde{R}_s/\partial n$, $\partial\tilde{R}_s/\partial c$, and $\partial\tilde{R}_s/\partial J$ its partial derivatives. Then, for any $n \in [\underline{n}_j, \bar{n}]$, let $J : \mathbb{R}^3 \to \mathbb{R}$ be the function defined by

$$J(n, c, c') = \left(\frac{c}{nc'}\right)^{\xi}, \tag{S40}$$

and note that, for $n \in (\underline{n}_j, n''_{js})$, $J(n, c_{js}(n), c'_{js}(n)) = J_{c_{js}}[n] = \Big(c_{js}(n)/\Big(n \cdot c'_{js}(n)\Big)\Big)^{\xi}$. Hereafter, we then denote by J_n, J_c, and J_c the partial derivatives of J with respect to n, c and c', respectively. Finally, note that, given the threshold functions $c_{rq} : [\underline{n}_r, \bar{n}] \to \bar{N}$, $r, q \in K$, $q, r \neq s$, for any $n \in [\underline{n}_j, \bar{n}]$, the endogenous densities $g_j(n|c)$ and $g_l(c_{jl}(n)|c)$ in the government's objective in (S36) depend on the threshold function $c_{js} : [\underline{n}_j, \bar{n}] \to \bar{N}$ only through the value that this function takes at n, and are given by;

$$g_j(n|c) \equiv f_j(n)F_{-j|j}(c_{-j}(n; c)|n) \tag{S41}$$

and

$$g_l(c_{jl}(n)|c) = f_l(c_{jl}(n))F_{-l|l}(c^j_{-l}(n; c)|n) \tag{S42}$$

with the thresholds $c_{-j}(n; c)$ and $c^j_{-l}(n; c)$ as defined in the proof of Proposition S1. In other words, given the thresholds functions c_{rq}, $r, q \neq s$, the value of the densities $g_j(n|c)$ and $g_l(c_{jl}(n)|c)$ depend on the function $c_{js} : [\underline{n}_j, \bar{n}] \to \bar{N}$ only through $c_{js}(n)$. This means that the optimality conditions for the threshold function c_{js} can be obtained as a solution to a *calculus of variations* problem with control c_{js} and objective

$$\int_{\underline{n}_j}^{n''_{js}} \left[\begin{array}{c} R_j(n)\, g_j(n|c) + \sum_{l \neq j,s} \hat{R}_l(n|c)\, c'_{jl}(n)\, g_l(c_{jl}(n)|c) \\ +\tilde{R}_s\Big(n, c_{js}(n), J(n, c_{js}(n), c'_{js}(n))\Big) \cdot c'_{js}(n) \cdot g_s(c_{js}(n)|c) \end{array} \right] dn.$$

Dropping the arguments from $\tilde{R}_s$, and J, to facilitate the writing, and holding fixed the thresholds $c_{rq}(n)$ all $r, q \neq s$, we then have that, at any $(\underline{n}_j, n''_{js})$, the point-wise Euler equation of this problem is given by

$$\begin{aligned} & R_j(n)\,\frac{\partial g_j(n|c)}{\partial c_{js}(n)} + \sum\nolimits_{l \neq j,s} \hat{R}_l(n|c)\, c'_{jl}(n)\,\frac{\partial g_l(c_{jl}(n)|c)}{\partial c_{js}(n)} + \frac{\partial \tilde{R}_s}{\partial c} c'_{js}(n) g_s(c_{js}(n)|c) \\ & + \frac{\partial \tilde{R}_s}{\partial J} J_c c'_{js}(n) g_s(c_{js}(n)|c) + \tilde{R}_s c'_{js}(n) \frac{\partial g_s(c_{js}(n)|c)}{\partial c_{js}(n)} \\ & = \frac{d}{dn}\left[\frac{\partial \tilde{R}_s}{\partial J} J_{c'} c'_{js}(n) g_s(c_{js}(n)|c) + \tilde{R}_s g_s(c_{js}(n)|c)\right]. \end{aligned} \tag{S43}$$

Next observe that

$$g_j(n|c) \equiv f_j(n)F_{-j|j}(c_{-j}(n; c)|n) = \int_{\underline{n}}^{c_{js}(n)} f_{-(j,s)}(n, n_s)\,dn_s,$$

and, likewise, for all $l \neq j, s$,

$$g_l(c_{jl}(n)|c) = f_l(c_{jl}(n))F_{-l|l}(c^j_{-l}(n; c)|n) = \int_{\underline{n}}^{c_{js}(n)} f_{-(s,l)}(c_{jl}(n), n_s)\,dn_s$$

while for sector s,

$$g_s(c_{js}(n)|c) = f_s(c_{js}(n))F_{-s|s}(c^j_{-s}(n;c)|n) = \int_{\underline{n}}^{n} f_{-(j,s)}(n_j, c_{js}(n))dn_j$$

where the densities $f_{-(j,s)}(n, n_s)$, $f_{-(s,l)}(c_{jl}(n), n_s)$ are defined as follows

$$f_{-(j,s)}(n, n_s) = \int_{\underline{n}}^{c_{j1}(n)} \cdots \int_{\underline{n}}^{c_{jk}(n)} f(n_1, ..., n_s, n, ...n_k)dn_1 \cdots dn_k$$

$$f_{-(s,l)}(c_{jl}(n), n_s) = \int_{\underline{n}}^{c_{j1}(n)} \cdots \int_{\underline{n}}^{c_{jk}(n)} f(n_1, ..., n_s, c_{jl}(n), ...n_k)dn_1 \cdots dn_k$$

Note that $f_{-(j,s)}(n_j, n_s)$ is the density of agents with sector-j productivity equal to n_j and sector-s productivity equal to n_s and who prefer working in one of these two sectors to working in any other sector. Similarly $f_{-(s,l)}(c_{jl}(n), n_s)$ is the density of agents with sector-s productivity equal to n_s and sector-l productivity equal to $c_{jl}(n)$ who prefer working in one of these two sectors to working in any other sector.

Using these definitions, the fifth term of the left-hand-side of (S43) can be developed as follows:

$$\begin{aligned}
\tilde{R}_s c'_{js}(n)\frac{\partial g_s(c_{js}(n)|c)}{\partial c_{js}(n)} &= \tilde{R}_s c'_{js}(n)\left(\int_{\underline{n}}^{n} \frac{\partial}{\partial n_s} f_{-(j,s)}(n_j, c_{js}(n))dn_j\right) \\
&= \tilde{R}_s \frac{d}{dn}\left[g_s(c_{js}(n)|c)\right] - \tilde{R}_s f_{-(j,s)}(n, c_{js}(n)) \\
&- \tilde{R}_s \sum\nolimits_{l\neq j,s} f_{-(l,s)}(c_{jl}(n), c_{js}(n))c'_{jl}(n) \\
&= \frac{d}{dn}\left[\tilde{R}_s g_s(c_{js}(n)|c)\right] - \frac{d\tilde{R}_s}{dn} g_s(c_{js}(n)|c) - \tilde{R}_s f_{-(j,s)}(n, c_{js}(n)) \\
&- \tilde{R}_s \sum\nolimits_{l\neq j,s} f_{-(l,s)}(c_{jl}(n), c_{js}(n))c'_{jl}(n)
\end{aligned} \tag{S44}$$

Similarly, for any $l \neq s, j$, the second term in the left-hand-side of (S43) can be rewritten as

$$\hat{R}_l(n|c)\, c'_{jl}(n)\, \frac{\partial g_l(c_{jl}(n)|c)}{\partial c_{js}(n)} = \hat{R}_l(n|c)\, c'_{jl}(n)\, f_{-(s,l)}(c_{jl}(n), c_{js}(n)) \tag{S45}$$

Substituting (S44) and (S45) into (S43) and simplifying, we can rewrite the point-wise Euler equation (S43) as follows

$$\begin{aligned}
&\left[R_j(n) - \tilde{R}_s\right] f_{-(j,s)}(n, c_{js}(n)) + \sum\nolimits_{l\neq j,s}\left[\hat{R}_l(n|c) - \tilde{R}_s\right] c'_{jl}(n) f_{-(s,l)}(c_{jl}(n), c_{js}(n)) \\
&+ \frac{\partial\tilde{R}_s}{\partial c} c'_{js}(n) g_s(c_{js}(n)|c) + \frac{\partial\tilde{R}_s}{\partial J} J_c c'_{js}(n) g_s(c_{js}(n)|c) - \frac{d\tilde{R}_s}{dn} g_s(c_{js}(n)|c) \\
&= \frac{d}{dn}\left[\frac{\partial\tilde{R}_s}{\partial J} J_{c'} c'_{js}(n) g_s(c_{js}(n)|c)\right].
\end{aligned} \tag{S46}$$

Multiplying both sides of (S46) by $c_{js}(n)$ and rearranging terms, we obtain that

$$\left[R_j(n) - \tilde{R}_s\right] f_{-(j,s)}(n, c_{js}(n))c_{js}(n) + \sum\nolimits_{l\neq j,s} \left[\hat{R}_l(n|c) - \tilde{R}_s\right] c'_{jl}(n)c_{js}(n)f_{-(s,l)}(c_{jl}(n), c_{js}(n))$$
$$+\frac{\partial \tilde{R}_s}{\partial c} c'_{js}(n)c_{js}(n)g_s(c_{js}(n)|c) - \frac{d\tilde{R}_s}{dn} g_s(c_{js}(n)|c)c_{js}(n)$$
$$= \frac{d}{dn}\left[\frac{\partial \tilde{R}_s}{\partial J} J_{c'} c'_{js}(n)g_s(c_{js}(n)|c)\right] c_{js}(n) - \frac{\partial \tilde{R}_s}{\partial J} J_c g_s(c_{js}(n)|c)c_{js}(n)c'_{js}(n). \tag{S47}$$

Next, note that, for any $n \in (\underline{n}_j, n''_{js})$,

$$J_c = J_c(n, c_{js}(n), c'_{js}(n)) = \xi\left(\frac{c_{js}(n)}{nc'_{js}(n)}\right)^{\xi-1} \frac{1}{nc'_{js}(n)} = \xi J \frac{1}{c_{js}(n)},$$
$$J_{c'} = J_{c'}(n, c_{js}(n), c'_{js}(n)) = -\xi\left(\frac{c_{js}(n)}{nc'_{js}(n)}\right)^{\xi-1} \frac{c_{js}(n)}{nc'_{js}(n)}\frac{1}{c'_{js}(n)} = -\xi J \frac{1}{c'_{js}(n)}.$$

Note that the expressions above are always well-defined, as $c'_{js}(n) > 0$ for all $n \in (\underline{n}_j, n''_{js})$.

Replacing these expressions into the right-hand side of (S47), we obtain that for any $n \in (\underline{n}_j, n''_{js})$, the right-hand side of the Euler equation becomes

$$\frac{d}{dn}\left[\frac{\partial \tilde{R}_s}{\partial J} J_{c'} c'_{js}(n)g_s(c_{js}(n)|c)\right] c_{js}(n) - \frac{\partial \tilde{R}_s}{\partial J} J_c g_s(c_{js}(n)|c)c_{js}(n)c'_{js}(n)$$
$$= -\xi\left\{\frac{d}{dn}\left[J\frac{\partial \tilde{R}_s}{\partial J} g_s(c_{js}(n)|c)\right] c_{js}(n) + J\frac{\partial \tilde{R}_s}{\partial J} g_s(c_{js}(n)|c)c'_{js}(n)\right\}$$
$$= -\xi\frac{d}{dn}\left[J\frac{\partial \tilde{R}_s}{\partial J} g_s(c_{js}(n)|c)c_{js}(n)\right]. \tag{S48}$$

Substituting (S48) into (S47), we then have that, for any $n \in (\underline{n}_j, n''_{js})$, the Euler equation becomes:

$$\left[R_j(n) - \tilde{R}_s\right] f_{-(j,s)}(n, c_{js}(n))c_{js}(n) + \sum\nolimits_{l\neq j,s} \left[\hat{R}_l(n|c) - \tilde{R}_s\right] c'_{jl}(n)c_{js}(n)f_{-(s,l)}(c_{jl}(n), c_{js}(n))$$
$$+ \frac{\partial \tilde{R}_s}{\partial c} c'_{js}(n)c_{js}(n)g_s(c_{js}(n)|c) - \frac{d\tilde{R}_s}{dn} g_s(c_{js}(n)|c)c_{js}(n)$$
$$= -\xi\frac{d}{dn}\left[J\frac{\partial \tilde{R}_s}{\partial J} g_s(c_{js}(n)|c)c_{js}(n)\right] \tag{S49}$$

Integrating (S49) from $\underline{n}_j$ to $n_j \in (\underline{n}_j, n''_{js})$ we then obtain that

$$\int_{\underline{n}_j}^{n_j} \left[R_j(n) - \tilde{R}_s(n)\right] f_{-(j,s)}(n, c_{js}(n)) c_{js}(n) dn \tag{S50}$$
$$+ \int_{\underline{n}_j}^{n_j} \sum\nolimits_{l \neq j,s} \left[\hat{R}_l(n|c) - \tilde{R}_s(n)\right] c'_{jl}(n) c_{js}(n) f_{-(s,l)}(c_{jl}(n), c_{js}(n)) dn$$
$$+ \int_{\underline{n}_j}^{n_j} \frac{\partial \tilde{R}_s}{\partial c} c'_{js}(n) c_{js}(n) g_s(c_{js}(n)|c) dn - \int_{\underline{n}_j}^{n_j} \frac{d\tilde{R}_s}{dn} g_s(c_{js}(n)|c) c_{js}(n) dn$$
$$= -\xi J(n_j) \frac{\partial \tilde{R}_s(n_j)}{\partial J} g_s(c_{js}(n_j)|c) c_{js}(n_j) + \lim_{n \to \underline{n}_j} \xi J(n) \frac{\partial \tilde{R}_s(n)}{\partial J} g_s(c_{js}(n)|c) c_{js}(n)$$

where we have highlighted the dependence of $\tilde{R}_s$ and of J on n to avoid possible confusion.

Consider the second term in the right-hand side of (S50). This term is zero if $\underline{n}_j = \underline{n}$, as in this case,

$$\lim_{n \to \underline{n}} g_s(c_{js}(n)|c) = \lim_{n_j \to \underline{n}} \left\{ \int_{\underline{n}}^{n_j} f_{-(s,j)}(n, c_{js}(n)) dn \right\} = 0,$$

and all remaining terms are bounded. When $\underline{n}_j > \underline{n}$, the optimal choice of $\underline{n}_j$ implies that the following transversality condition holds:

$$\lim_{n \to \underline{n}_j} \xi J(n) \frac{\partial \tilde{R}_s(n)}{\partial J} g_s(c_{js}(n)|c) c_{js}(n) = 0,$$

which is exactly the second term in the right-hand side of (S50).

We now express each term in (S50) as a function of the income tax schedules. By definition of R_j and $\tilde{R}_s$ in (S33) and (S39), the first term in (S50) is simply:

$$\int_{\underline{n}_j}^{n_j} \left[R_j(n) - \tilde{R}_s(n)\right] f_{-(j,s)}(n, c_{js}(n)) c_{js}(n) dn =$$
$$\lambda \int_{\underline{n}_j}^{n_j} \left[T_j(y_j(n)) - T_s(y_s(c_{js}(n)))\right] c_{js}(n) f_{-(j,s)}(n, c_{js}(n)) dn. \tag{S51}$$

Similarly, using the change of variables, $n_l = c_{jl}(n)$, the second term in (S50) can be rewritten as

$$\sum\nolimits_{l \neq j,s} \int_{\underline{n}_l}^{c_{jl}(n_j)} \left[T_l(y_l(n)) - T_s(y_s(c_{ls}(n)))\right] c_{ls}(n) f_{-(l,s)}(n, c_{ls}(n)) dn$$

The third term in (S50) is obtained by differentiating (S39) with respect to c, which yields:

$$\int_{\underline{n}_j}^{n_j} \frac{\partial \tilde{R}_s(n)}{\partial c} c_{js}(n) c'_{js}(n) g_s(c_{js}(n)|c) dn$$
$$= \int_{\underline{n}_j}^{n_j} \lambda J_{c_{js}}(n) h_j(n) c_{js}(n) c'_{js}(n) g_s(c_{js}(n)|c) dn$$
$$= \lambda \int_{\underline{n}_j}^{n_j} y_s(c_{js}(n)) c'_{js}(n) g_s(c_{js}(n)|c) dn = \lambda \int_{c_{js}(\underline{n}_j)}^{c_{js}(n_j)} y_s(n_s) g_s(n_s|c) dn_s \tag{S52}$$

where the last equality follows from changing the variable of integration from n to n_s (using the relation $n_s = c_{js}(n)$).

The fourth term in (S50) is obtained by totally differentiating (S39) with respect to n_j, which gives

$$\begin{aligned}
&\int_{\underline{n}_j}^{n_j} \frac{d\tilde{R}_s(n)}{dn} c_{js}(n) g_s(c_{js}(n)|c) dn \\
&= \int_{\underline{n}_j}^{n_j} \left\{ \frac{d}{dn} \left[\phi\left(u_j\left(n\right)\right) + \lambda T_s(y_s(c_{js}(n)) \right] \right\} c_{js}(n) g_s(c_{js}(n)|c) dn \\
&= \int_{\underline{n}_j}^{n_j} \left\{ \phi'\left(u_j\left(n\right)\right) u_j'\left(n\right) + \lambda T_s'(y_s(c_{js}(n)) \frac{dy_s(c_{js}(n))}{dn_s} c_{js}'(n) \right\} c_{js}(n) g_s(c_{js}(n)|c) dn \\
&= \int_{\underline{n}_j}^{n_j} \left\{ \begin{array}{c} \phi'\left(u_j\left(n\right)\right) \psi'\left(h_s(c_{js}(n))\right) h_s(c_{js}(n)) \\ +\lambda T_s'(y_s(c_{js}(n)) \frac{dy_s(c_{js}(n))}{dn_s} c_{js}(n) \end{array} \right\} c_{js}'(n) g_s(c_{js}(n)|c) dn \\
&= \int_{\underline{n}_j}^{n_j} \left\{ \begin{array}{c} \phi'\left(u_j\left(n\right)\right) \left[1 - T_s'(y_s(c_{js}(n)))\right] y_s(c_{js}(n)) \\ +\lambda T_s'(y_s(c_{js}(n)) \frac{dy_s(c_{js}(n))}{dn_s} c_{js}(n) \end{array} \right\} c_{js}'(n) g_s(c_{js}(n)|c) dn
\end{aligned}$$

where the last two equalities use (S19) and (S9). Changing again the variables of integration using the relation $n_s = c_{js}(n)$, we then obtain that the third term in (S50) is equal to

$$\begin{aligned}
&\int_{\underline{n}_j}^{n_j} \frac{d\tilde{R}_s(n)}{dn} c_{js}(n) g_s(c_{js}(n)|c) dn \\
&= \int_{c_{js}(\underline{n}_j)}^{c_{js}(n_j)} \left\{ \phi'\left(u_s\left(n_s\right)\right) \left[1 - T_s'(y_s(n_s))\right] + \lambda T_s'(y_s(n_s)) \varepsilon_{y_s}(n_s) \right\} y_s(n_s) g_s(n_s|c) dn_s,
\end{aligned}$$

where

$$\varepsilon_{y_s}(n_s) \equiv \frac{dy_s(n_s)}{dn_s} \frac{n_s}{y_s(n_s)}$$

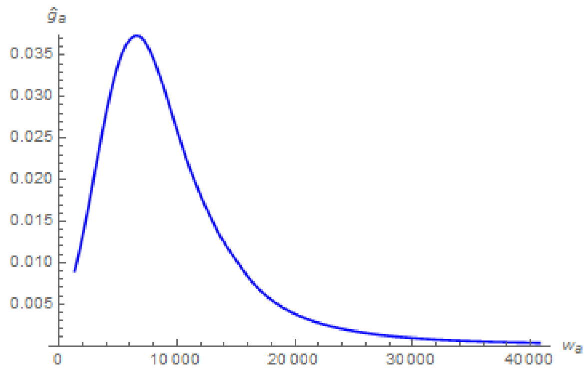
Finally, the right-hand-side of (S50) is obtained by differentiating (S39) with respect to J which yields:

$$\frac{\partial \tilde{R}_s(n)}{\partial J} = \lambda h_j(n) \left[c_{js}(n) - \psi'\left(J_{c_{js}}(n) h_j(n) \right) \right] = \lambda h_j(n_j) c_{js}(n) T_s'(J_{c_{js}}(n) h_j(n) c_{js}(n)).$$

We then have that the right-hand-side of (S50) can be rewritten as:

$$\begin{aligned}
-\xi J_{c_{js}}(n) \frac{\partial \tilde{R}_s(n)}{\partial J} g_s(c_{js}(n)|c) c_{js}(n) &= -\xi \lambda y_s(c_{js}(n)) T_s'(y_s(c_{js}(n))) g_s(c_{js}(n)|c) c_{js}(n) \\
&= \xi \frac{T_s'(y_s\left(c_{js}(n)\right))}{1 - T_s'(y_s(c_{js}(n)))} c_{js}(n) g_s(c_{js}(n)|c) \cdot \left\{ (1 - T_s'(y_s(c_{js}(n)))) y_s(c_{js}(n) \right\}.
\end{aligned} \tag{S53}$$

Substituting (S51)-(S53) into (S50) and rearranging yields (S24). Q.E.D.



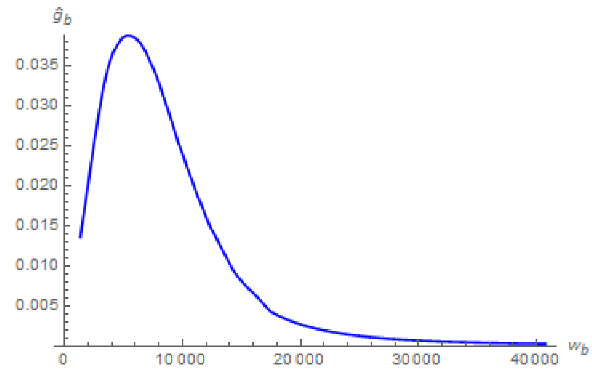


Figure 1: The left-hand side (resp., right-hand side) panel depicts the distribution of skills in sector a (resp., sector b) under production efficiency.

S3. Details for the Numerical Analysis of Section 5

We use data from the Current Population Survey of March 2014 on wage earnings and industry classification. We assign industries to either one of two economic sectors: manufacturing and services. The manufacturing sector (indexed by a) comprises mining, utilities, construction and all manufacturing industries. These correspond to industry census codes 0370 to 3990. The services sector (indexed by b) contains the trade, the transportation, the financial and banking industry as well as other industries. These correspond to industry census codes 4000 to 9290. We select annual labor earnings levels belonging to the interval $[\$5000, \$200000]$ in each sector.

Calibration of effective productivities

We interpret the wage earnings data as generated by a (sub-optimal) uniform tax system where production efficiency prevails. As such, each agent chooses the sector in which his productivity is the highest. We refer to this productivity as the effective productivity of an agent. We follow the literature in assuming a constant wage-elasticity of labor supply equal to 0.25. We then employ standard methods (as in Saez (2001)) to obtain the distribution of effective productivities using as inputs (i) the workers' earnings data, and (ii) the U.S. schedule of marginal tax rates for singles. The real tax schedule of singles without dependent children is well approximated by a linear tax function at rate 29% with an intercept of about $\$4,000$.[12] We use the first-order condition of the intensive labor supply decision to infer the skill level from each observation of earnings.

To smooth the income distribution, we use a quadratic kernel with a bandwidth of $\$2,613$. Although underrepresented in the CPS, Saez (2001) argues that the skill distribution exhibits a fat upper tail in the U.S., which has dramatic consequences for the shape of optimal marginal tax rates. We therefore expand (in a continuously differentiable way) our kernel estimation by taking a Pareto distribution,[13] with an index $\ell = 2.5$ for skill levels between $n_a = \$15,527$, $n_b = \$17,892$, and

[12] See http://www.oecd-ilibrary.org/taxation/taxing-wages-2015/united-states_tax_wages-2015-43-en

[13] An (untruncated) Pareto distribution with Pareto index $\ell > 1$ is such that $\Pr(n > \hat{n}) = C/\hat{n}^{\ell}$ with $\ell, C \in \mathbb{R}_{++}$.

$\bar{n}_a = \bar{n}_b = \$40,748$. This represents the top 6% of our approximation of the skill distribution. The lower bound of the skill distribution is $\underline{n}_a = \underline{n}_b = \$1,340$ and the length of the skill interval is equal to \$390. The effective productivity densities $\hat{g}$ are illustrated in Figures 1.

Given this calibration of the current economy, we find that the government budget constraint is verified only when we set the exogenous revenue requirement to $G = \$9997$ per capita which represents 17.26% of GDP per capita.

Calibration of latent productivities

The productivity grid is such that productivity levels are indexed by $\omega = 1, ..., 100$. As described in the main text, the binomial c.d.f with parameter $p_j(\omega)$ is used to parametrize the conditional distributions of latent productivity. Denote by ω_{med} the index corresponding to the median effective skill level. In scenario 4, we set, for sector b, $p_b(\omega) = (2\omega_{med} - \beta_1\omega)/2\omega_{med}$ for any $\omega < \omega_{med}$ and $p_b(\omega) = (100 - \beta_2\omega)/2(100 - \omega_{med})$ otherwise. We chose β_1 to guarantee that agents with the lowest skill have a degree of skill transferability around 95%, and β_2 such that agents with the highest skill have a degree of skill transferability close to 45%. For sector a, we use $p_a(\omega) = (2\omega_{med} - \alpha_1\omega)/2\omega_{med}$ for any $\omega < \omega_{med}$ and $p_a(\omega) = (100 - \alpha_2\omega)/2(100 - \omega_{med})$ otherwise. We chose α_1 to guarantee that agents with the lowest skill have a degree of skill transferability around 90%, and α_2 such that agents with the highest skill have a degree of skill transferability close to 60%. Our simulations reveal that, as long as sector a is favored relative to b at the optimum, the results do not depend on the values of α_1 and α_2.

The occupational choice rule

In the case of sector-specific nonlinear income taxation, the second-best occupational choice rule is approximated by finding the third-degree polynomial

$$c(n_a) = \upsilon_0 + \upsilon_1 n_a + \upsilon_2 n_a^2 + \upsilon_3 n_a^3$$

that maximizes the Rawlsian social welfare function. In the case of sector-specific sales taxes (and uniform income taxes), the second-best occupational choice rule is the linear function $c(n_a) = \tilde{\upsilon}_1 n_a$ that maximizes the Rawlsian social welfare function.