Working Papers

Differential Taxation and Occupational Choice

Renato Gomes
Jean-Marie Lozachmeur
Alessandro Pavan

CESIFO WORKING PAPER NO. 5054
CATEGORY 1: PUBLIC FINANCE
NOVEMBER 2014

CESifo *Working Paper No. 5054*

Differential Taxation and Occupational Choice

Abstract

We study nonlinear income taxation in a Roy model in which agents' productivity is sectorspecific. We show that when income taxes can be sector-specific, the Diamond-Mirrlees theorem (according to which the second-best displays production efficiency) fails: social welfare (be it Rawlsian or Weighed Utilitarian) can be increased by assigning some agents to their least productive sector. By sacrificing production efficiency, the planner incurs second-order losses in total output, but obtains a first-order reduction in the informational costs of redistribution. The same result obtains when the government is constrained to a uniform income tax schedule, as long as sales taxes can be made sector-specific. In this latter case, our result also implies failure of the Atkinson-Stiglitz theorem (according to which, when preferences over consumption and leisure are separable, as they are in our economy, the second-best can be implemented with zero sales taxes).

JEL-Code: C720, D620.

Keywords: income taxation, occupational choice, sales taxes, sector-specific taxation, production efficiency.

Renato Gomes
Toulouse School of Economics
Toulouse / France
renato.gomes@tse-fr.eu

Jean-Marie Lozachmeur
Toulouse School of Economics
Toulouse / France
jean-marie.lozachmeur@tse-fr.eu

Alessandro Pavan
Northwestern University
Evanston / Illinois / USA
alepavan@northwestern.edu

This version: July 2014
For comments and suggestions, we thank seminar participants at LAGV 2012 (Marseille), SED 2013 (Seoul), PET 2013 (Lisbon), PSE 2013 (Munich), Université de Rennes and TTC 2014 (Köln). We are grateful to Felix Bierbauer, Philippe Choné, Helmut Cremer, Guy Laroque, Casey Rothschild and Florian Scheuer for very helpful conversations. The usual disclaimer applies.

1 Introduction

Following the seminal contribution of Diamond and Mirrlees (1971), production efficiency is often thought as a key property of optimal tax systems. The celebrated Diamond-Mirrlees theorem shows that when the government can use differentiated linear taxes on all factors (input and output), the economy should lie at the production efficient frontier. Strikingly, at the optimum, distortions in consumption produced by income taxation do not translate into distortions in production. This result has important consequences for the design of tax systems. For instance, the Diamond-Mirrlees theorem provides an intellectual justification for opposing the taxation of intermediate goods, as well as the use of differential sales taxes, or sector-specific income tax deductions. Such taxes would create a wedge between productivities and wages across sectors, thus leading to distortions on the allocation of labor across sectors, and undesirable violations of production efficiency.

This paper revisits the optimality of production efficiency in a Roy (1951) model in which agents' productivity is sector-specific. Agents compare wage levels and the tax burden across occupations, and then choose which sector to work in along with their labor supply. To isolate the impact of taxation on the production side of the economy, we assume that the goods produced in different sectors are perfect substitutes. The technology on each sector is described by a representative firm with a linear production function, which rules out general equilibrium effects or externalities across sectors. Accordingly, in our model, the notion of production efficiency coincides with that of occupational-choice efficiency: Agents should join the sector in which their hour of work is more productive. The government wishes to implement a second best redistributive tax system à la Mirrlees (1971) using a rich set of non-linear incomes taxes possibly complemented by differential sales taxes. We allow the government's welfare objective to be Rawlsian or Concave Utilitarian.

We first study the general case in which the government can use a sector-specific non-linear income tax schedules (in which case sales taxes are redundant). The government can observe the income and the sector chosen by each individual, but cannot control the individual's choice of labor supply or the sector of employment. Accordingly, the government maximizes welfare subject to the usual *intensive-margin* incentive constraints associated with the choice of labor supply by each individual, as well as an *extensive-margin* incentive constraint associated with the choice of sector by each individual. The multi-dimensionality of each agent's productivity plays a key role in the extensive margin constraint, as the agent's occupational choice is determined by how the agent' skill differential across sectors compares to the difference in the tax burden across sectors.

Our first contribution is in developing a methodology for solving multi-dimensional screening problems governed by intensive-margin (labor supply) and extensive-margin (sector choice) decisions. Namely, we proceed by first solving a *primal problem*, where the occupational choice rule (which determines sector choice as a function of the worker's productivity profile) is held fixed, and the tax system is chosen to maximize welfare *subject to* implementing that occupational choice rule (as well as satisfying the intensive-margin incentive constraints). Next, we solve a *dual problem*, where the tax schedule in a given sector is held fixed, and the tax schedule in the other sector (as well as the occupational choice rule) are chosen to maximize welfare.

The solution to the primal problem delivers a Mirrlees tax formula generalized to a multi-sector economy with endogenous occupational choice and multi-dimensional types. As in Mirrlees (1971), Diamond (1998), and Saez (2001), the tax schedule balances efficiency and redistributive considerations. Efficiency concerns are captured by elasticity (or behavioral) effects, that measure how individuals adjust labor supply in response to higher marginal taxes. Redistributive concerns are captured by a direct (or mechanical) effects, that measure how an increase in the marginal tax in a given income bracket increases tax collection in all higher income brackets. Our characterization reveals how the government optimally balances intensive-margin distortions in labor supply *across sectors*, as a function of the occupational choice rule to be implemented. In particular, we show that, depending on the occupational choice sought by the government, negative marginal taxes naturally arise at the optimum.

In turn, the solution to the dual problem delivers an Euler equation that determines the optimal allocation of agents across sectors. At the optimum, the marginal loss in tax revenue due to migration of workers across sectors equalizes the marginal gains from tailoring tax schedules to the distribution of productivities in each sector ("tagging"). Importantly, under sector-specific income taxation, the Diamond-Mirrlees theorem (according to which the second-best displays production efficiency) fails: social welfare is increased by assigning some agents to their least productive sector. The key to understanding this result is that, because the distribution of agents in each sector is endogenous, the informational costs of redistributions can be affected by the tax system employed by the government. As we show, by sacrificing production efficiency, the planner incurs second-order losses in total output, but obtains a first-order reduction in the informational costs of redistribution.

Our analysis shows that familiar insights from the standard Mirrlees model regarding the taxation of top earners are robust to multi-dimensional environments where occupational-choice distortions are present. Namely, when the support of productivities is bounded, the agents with the highest productivity on *both* sectors face zero marginal income taxes (revealing that distortions in occupational choice *do not translate* into distortions in labor supply). In turn, when the support of productivities is unbounded, we show that *marginal tax collection* vanishes at the top of the income distribution.

Next, we consider the (perhaps more realistic) scenario in which the government is not able to tax labor income using a sector-specific schedule, but can levy different sales taxes across sectors. Indirect taxation in the form of sales taxes is an imperfect substitute for direct sector-specific income taxation, as it *uniformly* affects the wages earned in equilibrium by all agents (regardless of their income). Most importantly, our methods are flexible enough to accommodate such constraints in the set of tax instruments available to the government.

When only sales taxes can be made sector-specific, we specialize the Mirrlees formula discussed above to show how sales taxes and income taxation optimally allocate intensive-margin distortions across sectors. Moreover, we show that production inefficiency is, as in the general case, a robust feature of the second-best. In this latter case, our analysis implies the failure of the Atkinson-Stiglitz theorem (according to which, when preferences over consumption and leisure are separable, as they are in our economy, the second-best can be implemented with zero sales taxes). Intuitively, sales taxes are useful by distorting equilibrium wages and indirectly reducing the informational costs of redistribution.

The rest of the paper is organized as follows. Below, we close the introduction by briefly reviewing the most pertinent literature. Section 2 describes the model. Section 3 characterizes optimal sector-specific income tax schedules. Section 4 contains results for economies where only sales taxes can be sector-specific. Section 5 discusses a few extensions and concludes.

1.1 Related literature

Our paper contributes to the literature on optimal taxation theory in the tradition of Mirrlees (1971). Our analysis is directly related to fundamental results in this literature. First, Diamond and Mirrlees (1971) show that the second-best optimum exhibits production efficiency in a general equilibrium setting where the government can use (linear) taxes on all inputs and outputs, and firms can be taxed in a lump-sum fashion.[1] In turn, Atkinson and Stiglitz (1976) show that differentiated sales taxes across goods are detrimental to welfare, provided the government can use a non-linear income tax schedule and preferences are weakly separable between consumption and leisure.[2] These results were first challenged by Naito (1999), who considers a two-sector model in which two goods are produced using skilled and unskilled labor in different intensities. Naito shows that a tax/subsidy on one good, implicitly creating a subsidy to low-skilled labor, is always desirable provided the government can use a non-linear income tax. This indirect form of wage subsidy (as opposed to a subsidy on total labor income) allows the government to ease redistribution without affecting incentive constraints. This result comes from the fact that the high skilled individuals cannot effectively claim the low skilled wage. Later, Saez (2004) discusses this assumption and argues that, in the long run, individuals choose their occupation (say, skilled or unskilled). As a consequence, the optimality of production efficiency and of uniform sales taxes is restored.

In turn, Saez (2002) derives the optimal tax system in a setting where labor supply responses involve an intensive margin (high or low-paying occupations) as well as an extensive margin (participation into the labor force). One important assumption in Saez (2002, 2004) is that every two workers are equally productive in all occupations, but differ in their tastes for each occupation (including tastes for not working). By contrast, in the spirit of Roy (1951), we assume that workers have heterogenous skills *across* occupations (extensive margin), and make intensive-margin choices

[1] See Hammond (2000) for a generalization of this result that allows for asymmetric information about workers' skills and nonlinear taxation.

[2] See Boadway (2012) for a unified treatment of these results.

within occupation (i.e., hours of work).

To the best of our knowledge, only Rothschild and Scheuer (2013, 2014) analyze optimal non-linear taxes in a Roy model with multiple occupations. They consider a very general production function allowing for spillovers and externalities between sectors, but assume uniform taxation (i.e., all workers face the same income tax schedule, and sale taxes are not allowed to be sector specific). The focus of their work is on the interplay between uniform income taxation and spillovers/externalities across sectors. By contrast, we rule out spillovers/externalities across sectors and focus on the role of differential taxation in reducing the informational costs of redistribution.

Related, Scheuer (2014) studies the role of differential taxation in a model in which individuals can choose between becoming workers or entrepreneurs. Each individual has the same skill in both occupations, but faces a cost of setting up a firm that enters additively in the agent's utility function. In this setting, production efficiency is optimal when the government can use differential income taxation, or when income taxation is uniform and the production function is linear (as assumed in the present paper).

Our paper is also related to the literature on "tagging," initiated by Akerlof (1978) and further developed by Cremer et al. (2010) and Mankiw and Weinzierl (2010) in the context of optimal non-linear income taxation. The idea of "tagging" is that the Government can increase efficiency and redistribute more by conditioning income taxes on observable characteristics, such as age, sex, or height. A fundamental difference with respect to our paper is that, in this literature, the tagging variable is *exogenous* (agents' cannot respond by changing sex, age, or height). In contrast, in our economy, workers are able to migrate across sectors in response to differential taxation (i.e., the tagging variable is endogenous).[3]

Allowing for endogenous occupational choice naturally leads to a multi-dimensional screening problem. Solving such problems is often challenging, as one cannot determine from the outset the direction in which incentive constraints bind (see Rochet and Choné (2003) and references therein). In our setting, the multi-dimensionality of workers' productivities only affects sector-choice (extensive-margin) decisions. This allows us to employ the primal-dual approach described above, which treats in isolation the role of taxation in shaping labor supply decisions and occupational choices, bringing considerable tractability to the analysis.

As our analysis reveals, the multi-dimensionality of workers' types has important implications for the design of optimal tax systems. Other recent studies share a similar view: Choné and Laroque (2010) study the optimality of negative marginal taxes in a model where workers have a bi-dimensional type comprising a skill level and an outside option (that determines participation in the labor force). More recently, Golosov et al. (2013) study optimal non-linear income and capital taxes in a model where individuals differ both in their skills and in their time preferences.

[3] The possibility of endogenous tagging is discussed but not solved in the seminal contribution of Akerlof (1978).

2 Model and Preliminaries

2.1 Set-up

We consider an economy with a unit-mass continuum of agents and two sectors indexed by $j \in \{a, b\}$. For simplicity, the goods produced in the two sectors are assumed to be perfect substitutes, and their prices are normalized to one. Each agent chooses which sector to work in, and the number of hours (or effort) to supply in the chosen sector. The productivity of an agent in sector $j \in \{a, b\}$ is denoted by $n_j \in N \equiv (\underline{n}, \overline{n})$, where $\underline{n} > 0$, $\overline{n} \in \mathbb{R}_{++} \cup \{+\infty\}$ and $\overline{n} > \underline{n}$. An agent's type is thus given by the vector $\mathbf{n} \equiv (n_a, n_b)$ describing the agent's productivity in each of the two sectors. Each agent's type is an independent draw from a distribution F with support $\mathbf{N} \equiv N^2$. We assume that F is absolutely continuous with respect to the Lebesgue measure and denote by F_j its marginal distribution with respect to the j-dimension (with bounded densities f_j). The conditional distributions are denoted by $F_{j|k}$, for $j, k \in \{a, b\}$, $j \neq k$ (with bounded densities $f_{j|k}$).

An agent with productivity n_j supplying $h_j \in \mathbb{R}_+$ hours in sector $j \in \{a, b\}$ produces $n_j h_j$ units of *effective labor*. The income generated by this agent is then $y_j = w_j n_j h_j$, where $w_j \in \mathbb{R}_+$ is the wage per unit of effective labor.

The government taxes labor income according to the (possibly) non-linear sector-specific tax schedule $T_j(y_j)$. Each agent's utility is quasilinear in consumption so that the utility of an agent of type $\mathbf{n}$ supplying h_j hours in sector j is:

$$w_j h_j n_j - T_j(w_j h_j n_j) - \psi(h_j), \tag{1}$$

where $\psi(h)$ is the disutility of labor, which is assumed to take the isoelastic form $\psi(h) = h^{\frac{1}{\xi}}$, with $\xi \in (0, 1)$. The elasticity of labor supply with respect to wages is then equal to $\frac{\xi}{1-\xi}$, which is increasing in ξ.

The production side in each sector is 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 amount of effective labor hired by the firm. Firm j's profits are then equal to

$$\pi_j = (1 - w_j - \tau_j) L_j, \tag{2}$$

where τ_j is the sales tax rate on good j. The wage rates $\mathbf{w} \equiv (w_a, w_b)$, the agents' labor supply in the two sectors, and the labor demand from the two representative firms are all simultaneously determined in equilibrium, as explained below.

2.2 Taxation equilibrium

The occupational choice of each agent is described by the *occupational choice rule* $\mathcal{C} : \mathbf{N} \to \{a, b\}$. This rule specifies for each type $\mathbf{n} = (n_a, n_b) \in \mathbf{N}$ the sector in which the agent works. In turn, the labor supply schedules $h_j : N_j \to \mathbb{R}_+$ determine the amount of labor supplied as a function of the agent's productivity, with the domain N_j of the function h_j denoting the set of productivity levels of those agents working in sector j.

Hereafter we will refer to an *allocation* as a triple $(\mathcal{C}, h_a, h_b)$. Implicit in the above definition is the assumption that the amount of labor supplied by each agent working in each of the two sectors is independent of the agent's productivity in the other sector. This assumption is without loss of generality, as it will become clear from the analysis below.

Next, we define a *tax system* $\mathcal{T} \equiv \{T_a, T_b, \tau_a, \tau_b\}$ as a collection of sector-specific income tax schedules $T_j : \mathbb{R}_+ \to \mathbb{R}$ along with sector-specific sales taxes (or, alternatively, subsidies) $\tau_j \in \mathbb{R}$. An allocation $(\mathcal{C}, h_a, h_b)$ is said to be *implementable at the wage rates* $\mathbf{w}$ if there exists a tax system $\mathcal{T}$ such that the following conditions jointly hold.

The first condition is a consistency property requiring that the domain N_j of each labor supply function h_j coincides with the set of productivity levels of those agents working in sector j, as determined by the occupational choice rule $\mathcal{C}$. That is,

$$N_a = \{n_a \in N : \exists n_b \in N \quad \text{such that} \;\; \mathcal{C}(n_a, n_b) = a\}$$

and symmetrically for sector b.

The second condition is the usual *incentive compatibility* condition on the *intensive margin of labor supply*. To describe this condition, let

$$\tilde{u}_j(n_j) \equiv \max_h \; \{w_j h n_j - T_j(w_j h n_j) - \psi(h)\} \qquad \text{for all} \qquad n_j \in N. \tag{3}$$

and

$$u_j(n_j) \equiv w_j h_j(n_j) n_j - T_j(w_j h_j(n_j) n_j) - \psi(h_j(n_j)) \qquad \text{for all} \qquad n_j \in N_j. \tag{4}$$

This condition then requires that $\tilde{u}_j(n_j) = u_j(n_j)$ for all $n_j \in N_j$.

In order to relate labor supply schedules and marginal taxes, it is convenient to consider the first-order condition associated with (3), which has to be satisfied at any interior point where the schedule T_j is differentiable:

$$w_j n_j - \psi'(h_j(n_j)) = w_j n_j T_j'(w_j h_j(n_j) n_j). \tag{5}$$

The third condition is an *incentive compatibility* condition on the *extensive margin of occupational choice*. It requires that each agent working in sector j would not be strictly better off by working in sector $k \neq j$:

$$\mathcal{C}(\mathbf{n}) = j \;\Rightarrow\; \tilde{u}_j(n_j) \geq \tilde{u}_k(n_k) \quad \text{for all } \mathbf{n} \in \mathbf{N}.$$

The forth and last condition requires that by employing the effective labor

$$L_j = \int_{\{\mathbf{n}:\mathcal{C}(\mathbf{n})=j\}} h_j(n_j) n_j dF(n_a, n_b)$$

each firm $j = a, b$ maximizes profits (2). We combine the above conditions into the definition of a *taxation equilibrium*.

Definition 1 ***(Taxation Equilibrium)*** *A taxation equilibrium $\mathcal{E} \equiv (\mathcal{C}, h_a, h_b, \mathcal{T}, \mathbf{w})$ consists of an allocation $(\mathcal{C}, h_a, h_b)$, a tax system $\mathcal{T}$, and a pair of wage rates $\mathbf{w}$ such that the following conditions jointly hold:*

1. *The allocation $(\mathcal{C}, h_a, h_b)$ is implementable at the wage rates $\mathbf{w}$ by the tax system $\mathcal{T}$;*

2. *The tax system $\mathcal{T}$ satisfies the Government budget constraint, i.e.,*

$$\sum_{j\in\{a,b\}} \int_{\{\mathbf{n}:\mathcal{C}(\mathbf{n})=j\}} \left(T_j(w_j h_j(n_j)n_j) + \tau_j h_j(n_j)n_j\right) dF(n_a, n_b) \geq G, \tag{6}$$

where G is the exogenous government budget requirement.

It is convenient to define the indirect utility of an agent with type $\mathbf{n}$ under the taxation equilibrium $\mathcal{E} \equiv (\mathcal{C}, h_a, h_b, \mathcal{T}, \mathbf{w})$ as

$$U(\mathbf{n}; \mathcal{E}) \equiv \max_{j\in\{a,b\}} \{u_j(n_j)\},$$

where the schedules u_j are given by (4).

The government chooses a taxation equilibrium $\mathcal{E} \equiv (\mathcal{C}, h_a, h_b, \mathcal{T}, \mathbf{w})$ to maximize some welfare function. We focus on two common specifications. The first is a Rawlsian objective, which consists in the utility of the worst-off individual:

$$\Phi^R [U(\cdot; \mathcal{E})] \equiv \min_{\mathbf{n}\in\mathbf{N}} \{U(\mathbf{n}; \mathcal{E})\}.$$

The second welfare function is a generalized Utilitarian one, which consists in a concave transformation of the agents' utilities:

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

where ϕ is a strictly increasing and weakly concave function.

We will use the index $x = R$ (alternatively, $x = CU$) to refer to the Rawlsian (alternatively, Concave Utilitarian) welfare objective. We will say that a taxation equilibrium is x-optimal if it solves the respective x-problem and refer to the tax system associated with an x-optimal taxation equilibrium as an x-optimal tax system. For future reference, we define the indicator function $\mathbf{1}_x^{CU}$, which equals zero if $x = R$ and one if $x = CU$.

2.3 Implementability

The next lemma characterizes the set of implementable allocations for given wage rates.

Lemma 1 ***(Implementability)*** *The allocation $(\mathcal{C}, h_a, h_b)$ is implemented at the wage rates $\mathbf{w}$ by the tax system $\mathcal{T}$ if and only if the following conditions jointly hold.*

1. *For every $j \in \{a,b\}$, the income schedule $y_j(n_j) \equiv w_j h_j(n_j) n_j$ is weakly increasing over N_j. Moreover, the indirect utility schedule $u_j(n_j)$ is Lipschitz continuous over N_j with derivative equal to*
$$u_j'(n_j) = \psi'\left(h_j(n_j)\right) \frac{h_j(n_j)}{n_j} \quad \text{for almost every} \quad n_j \in N_j. \tag{7}$$

2. *The occupational choice rule $\mathcal{C}$ can be described by an absolutely continuous and weakly increasing threshold function $c : N \to \bar{N}$ such that $\mathcal{C}(n_a, n_b) = a$ if $n_b < c(n_a)$ and $\mathcal{C}(n_a, n_b) = b$ if $n_b > c(n_a)$.*[4] *Furthermore, the threshold function c is such that $c(n_a) = \underline{n}$ if $\mathcal{C}(n_a, n_b) = b$ for all $n_b \in N$, $c(n_a) = \overline{n}$ if $\mathcal{C}(n_a, n_b) = a$ for all $n_b \in N$, and otherwise solves $u_a(n_a) = u_b(c(n_a))$.*

3. *For every $j \in \{a,b\}$, wages are given by*
$$w_j = 1 - \tau_j. \tag{8}$$

Condition (7) is the standard envelope formula that relates the agents' indirect utilities to their payoff-maximizing labor supply in each sector. Part 2 establishes that any occupational choice rule can be described by a weakly increasing threshold function c that maps n_a into the sector-b productivity threshold $c(n_a)$ such that an agent with type $\mathbf{n} = (n_a, c(n_a)) \in \mathbf{N}$ is indifferent between working in one sector or the other. Finally, Condition (8) is the standard labor market clearing condition: because the technology is linear, labor markets clear if and only if the marginal product of labor in each sector, net of sales taxes, equals its marginal cost. Accordingly, sales taxes affect equilibrium wages in a one-to-one fashion.

Because the payoff from working in a given sector strictly increases with the agent's productivity in that sector, the threshold $c(n_a)$ strictly increases in n_a at any interior point (i.e., at any point where $c(n_a) \in N$). We refer to the graph of c as the *locus of indifferent types*. The next remark shows how the envelope condition (7) can be used to relate the occupational choice rule to the labor supply schedules in the two sectors.

Remark 1 *Let $\mathcal{E} = (\mathcal{C}, h_a, h_b, \mathcal{T}, \mathbf{w})$ be a taxation equilibrium in which the occupational choice rule $\mathcal{C}$ is described by the threshold function c. For almost every point $n_a \in N$ such that $c(n_a) \in N$, the function c is strictly increasing and satisfies the following differential equation:*
$$\psi'\left(h_a(n_a)\right) \frac{h_a(n_a)}{n_a} = c'(n_a) \psi'\left(h_b(c(n_a))\right) \frac{h_b(c(n_a))}{c(n_a)}. \tag{9}$$
Equivalently, using the isoelastic specification of the disutility of labor, the labor supply schedules are such that, for almost every $n_a \in N$ such that $c(n_a) \in N$,
$$h_b(c(n_a)) = J_c[n_a] h_a(n_a), \tag{10}$$

[4] The notation $\bar{N}$ denotes the closure of the set N, i.e., $\bar{N} = [\underline{n}, \bar{n}]$.

where the operator J_c is defined by

$$J_c[n] \equiv \left(\frac{c(n)}{nc'(n)}\right)^{\xi}. \tag{11}$$

To get more intuition, it is worth rewriting the differential equation (9) in terms of marginal tax schedules. Assuming that income tax schedules are differentiable, we can combine the first-order condition of problem (3) with condition (9) to obtain that

$$c'\left(n_a\right) = \frac{u_a'(n_a)}{u_b'\left(c\left(n_a\right)\right)} = \frac{w_a h_a\left(n_a\right)\left[1 - T_a'\left(y_a\left(n_a\right)\right)\right]}{w_b h_b\left(c\left(n_a\right)\right)\left[1 - T_b'\left(y_b\left(c\left(n_a\right)\right)\right)\right]}. \tag{12}$$

Accordingly, the slope of the threshold function c describes by how much the productivity in sector b must increase following an increase in the productivity in sector a for an agent to remain indifferent as to which sector to work in. This slope is related to the ratio of the *marginal net incomes* in the two sectors evaluated along the locus of indifferent types.

2.4 Distribution of Productivities

A key feature of the model is that the distribution of productivities within each sector is endogenous (as agents choose in which sector to work in response to the tax system). It is convenient to describe these distributions in terms of the threshold function c associated with the occupational choice rule $\mathcal{C}$. In order to do so, we choose sector labels in the following way. We call sector a the sector (if one exists) for which there is a productivity threshold $n_a'' \in N$ such $c(n_a) = \bar{n}$ for all $n_a \geq n_a''$. In words, all agents whose sector-a productivity is above n_a'' work in sector a, irrespective of their sector-b productivity. If no sector satisfies this property, the choice of labels is arbitrary.[5]

It is also convenient to define the threshold $n_a' \in \bar{N}$ such that $c(n_a) > \underline{n}$ if and only if $n_a > n_a'$.[6] We will then say that the occupational choice rule $\mathcal{C}$ is *admissible* if its associated threshold function c is absolutely continuous and strictly increasing over a set (n_a', n_a''), equal to $\underline{n}$ for all $n_a < n_a'$ and equal to $\bar{n}$ for all $n_a > n_a''$.

Next, let $G_a(n_a|c)$ denote the mass of agents working in sector a whose productivity does not exceed n_a, as determined by the occupational choice rule corresponding to the threshold function c. This is given by

$$G_a(n_a|c) \equiv \int_{\underline{n}}^{n_a} \int_{\underline{n}}^{c(n_a)} f(\tilde{n}_a, \tilde{n}_b) d\tilde{n}_b d\tilde{n}_a = \int_{\underline{n}}^{n_a} f_a(\tilde{n}_a) F_{b|a}(c(\tilde{n}_a)|\tilde{n}_a) d\tilde{n}_a$$

with density $g_a(n_a|c) \equiv f_a(n_a) F_{b|a}(c(n_a)|n_a)$.

Denoting by

$$c^{-1}(n_b) \equiv \inf\{n_a \in \mathbf{N} : c(n_a) \geq n_b\}$$

[5] For example, consider the threshold function $c(n) = n$. In this case, no sector satisfies the property described above, and the choice of labels is arbitrary.

[6] We let $n_a' = \underline{n}$ if $\{n_a \in N : c(n_a) = \underline{n}\} = \varnothing$.

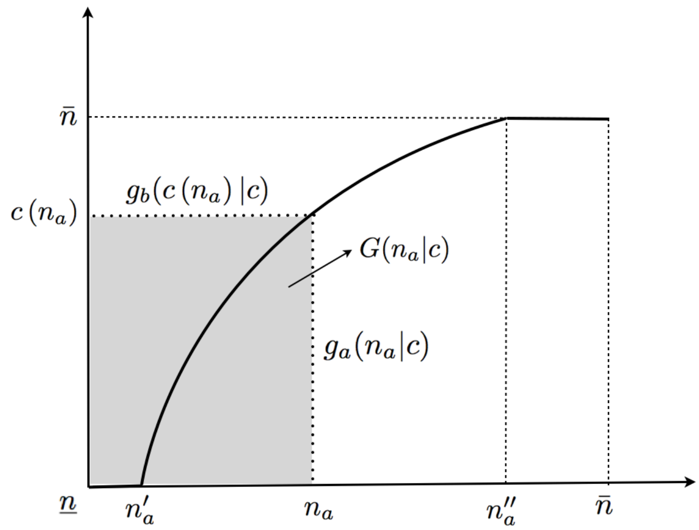

Figure 1: The threshold function and its induced distributions of productivities. The shaded area corresponds to the set of types whose mass is $G(n_a|c)$, while the dotted lines illustrate the sets of types associated with the densities $g_a(n_a|c)$ and $g_b(c\,(n_a)\,|c)$, respectively.

the generalized inverse of the threshold function c, we then have that the mass of agents working in sector b whose productivity does not exceed n_b is given by

$$G_b(n_b|c) \equiv \int_{\underline{n}}^{n_b} \int_{\underline{n}}^{c^{-1}(n_b)} f(\tilde{n}_a, \tilde{n}_b) d\tilde{n}_a d\tilde{n}_b = \int_{\underline{n}}^{n_b} f_b(\tilde{n}_b) F_{a|b}(c^{-1}(\tilde{n}_b)|\tilde{n}_b) d\tilde{n}_b.$$

with density $g_b(n_b|c) = f_b(n_b)F_{a|b}(c^{-1}(n_b)|n_b)$.

For future reference, it is also useful to define the distribution

$$G(n_a|c) \equiv F\,(n_a, c\,(n_a)) = G_a(n_a|c) + G_b(c\,(n_a)\,|c)$$

In words, $G(n_a|c)$ is the measure of types with side-a productivity less than n_a and side-b productivity less than $c\,(n_a)$. We can then write the density of $G(n_a|c)$ as

$$g(n_a|c) = g_a(n_a|c) + c'\,(n_a)\, g_b(c\,(n_a)\,|c).$$

Figure 1 illustrates the concepts presented above.

2.5 Characterization Procedure

We now describe how to find the x-optimal taxation equilibria, both for $x = R$ and $x = CU$. The characterization below proceeds in two steps. In the first step, we fix an arbitrary admissible occupational choice rule $\mathcal{C}$ and find the taxation equilibrium that maximizes the government x-objective among those that implement $\mathcal{C}$. We refer to this problem as the *primal problem*:

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

Clearly, the x-optimal taxation equilibrium $\mathcal{E}^x = (\mathcal{C}^x, h_a^x, h_b^x, \mathcal{T}^x, \mathbf{w}^x)$ must be such that the quadruple $(h_a^x, h_b^x, \mathcal{T}^x, \mathbf{w}^x)$ solves $\mathcal{P}_1^x(\mathcal{C}^x)$.

In the second step, we complete the characterization by considering a dual of problem to $\mathcal{P}_1^x(\mathcal{C})$. In this dual problem, which we call $\mathcal{P}_2^x(h_a)$, we fix some implementable sector-a labor supply schedule h_a and find the taxation equilibrium that maximizes the government's x-objective among those that implements h_a:

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

Clearly, the x-optimal taxation equilibrium $\mathcal{E}^x = (\mathcal{C}^x, h_a^x, h_b^x, \mathcal{T}^x, \mathbf{w}^x)$ must be such that the quadruple $(\mathcal{C}^x, h_b^x, \mathcal{T}^x, \mathbf{w}^x)$ solves $\mathcal{P}_2^x(h_a^x)$. As a consequence, the x-optimal taxation equilibrium $\mathcal{E}^x$ must satisfy the necessary optimality conditions associated to both problems $\mathcal{P}_1^x$ and $\mathcal{P}_2^x$.

2.6 Production Efficiency

We conclude this section by defining *production efficiency*. The definition below adapts the usual definition to the environment studied in this paper.

Definition 2 ***(Production Efficiency)*** *The equilibrium $\mathcal{E} = (\mathcal{C}, h_a, h_b, \mathcal{T}, \mathbf{w})$ exhibits production efficiency if and only if, holding fixed the labor supply of each worker (as specified by the equilibrium $\mathcal{E}$), there exists no reallocation of workers across sectors that yields a higher aggregate output. This is the case if and only if the threshold function c associated with the equilibrium occupational choice rule $\mathcal{C}$ is such that $c(n_a) = n_a$ for all $n_a \in N$.*

This definition is thus the standard one in public economics; simply notice that, in this economy, fixing the supply of inputs and changing their usage across firms/sectors is equivalent to holding fixed the labor supply (i.e., hours of work) of each individual and changing his occupation.

3 Sector-Specific Income Taxation

We now characterize properties of optimal tax systems when the Government is able to employ sector-specific income tax schedules. It should come as no surprise that the ability to tailor income taxes to occupational choice renders sales taxes redundant.

Remark 2 ***(Effective Tax Schedules)*** *Let $\mathcal{E} = (\mathcal{C}, h_a, h_b, \mathcal{T}, \mathbf{w})$ be a taxation equilibrium. There exists another taxation equilibrium $\hat{\mathcal{E}} = (\mathcal{C}, h_a, h_b, \hat{\mathcal{T}}, \hat{\mathbf{w}})$ implementing the same allocation $(\mathcal{C}, h_a, h_b)$ as in $\mathcal{E}$ and such that*

1. *the tax system $\hat{\mathcal{T}} = (\hat{T}_a, \hat{T}_b, \hat{\tau}_a, \hat{\tau}_b)$ satisfies*

$$\hat{T}_j(y) = \tau_j y + T_j((1-\tau_j) y), \text{ all } y \in \mathbb{R}_+, \tag{13}$$

and

$$\hat{\tau}_j = 0, \quad j = a, b,$$

2. *wages are given by*

$$\hat{w}_j = 1, \quad j = a, b,$$

3. *all agents' payoffs under $\hat{\mathcal{E}}$ are the same as under $\mathcal{E}$.*

Hereafter, we refer to $(\hat{T}_a, \hat{T}_b)$ as the "effective tax schedule" of the tax system $\mathcal{T}$.

Intuitively, if the government has enough flexibility to design sector-specific income tax schedules, it can then always replicate the effects of sales taxes with appropriately chosen income taxes. As a consequence, it is without loss of generality to consider taxation equilibria where $\tau_a = \tau_b = 0$. To lighten notation, in the remainder of this section, we thus drop the wage pair $\mathbf{w}$ from the description of taxation equilibria, and write the latter as $\mathcal{E} = (\mathcal{C}, h_a, h_b, , \mathcal{T})$, with the implicit understanding that $\mathbf{w} = (1, 1)$.

Below, we will thus characterize x-optimal taxation equilibria in terms of their effective tax schedules (and drop the qualification "effective" to lighten the exposition). For simplicity, and following the literature, we will abstract from bunching and corner solutions; that is, we will restrict attention to economies in which the optimality conditions described below identify income schedules $y_j(n_j)$ that are nondecreasing and such that $y_j(n_j) > 0$ for all $n_j \in N_j$ (equivalently, $h_j(n_j) > 0$ for all $n_j \in N_j$).

3.1 Optimal marginal tax rates

Denote by $m_j(n_j) \equiv \phi'(u_j(n_j))/\lambda$ the ratio of social marginal utility of all individuals with productivity n_j working in sector j to the marginal value of public funds for the government. The next proposition derives a necessary condition for problem $\mathcal{P}_1^x(\mathcal{C})$, showing how to compute x-optimal marginal tax rates implementing some admissible occupational choice rule $\mathcal{C}$.

Proposition 1 ***(Generalized Mirrlees Formula)*** *Let c be the threshold function corresponding to the admissible occupational choice rule $\mathcal{C}$. The x-optimal tax system implementing the choice rule $\mathcal{C}$ satisfies the following generalized Mirrlees formula for almost any $n_a \in (n'_a, n''_a)$:*

$$\underbrace{\xi \frac{T'_a(y_a(n_a))}{1 - T'_a(y_a(n_a))} n_a g_a(n_a|c)}_{E_a(n_a)} + \underbrace{\xi \frac{T'_b(y_b(c(n_a)))}{1 - T'_b(y_b(c(n_a)))} c(n_a) g_b(c(n_a)|c)}_{E_b(c(n_a))}$$

$$= \underbrace{\int_{n_a}^{\overline{n}} \left[1 - \mathbf{1}_x^{CU} m_a(\tilde{n}_a)\right] g(\tilde{n}_a|c) d\tilde{n}_a}_{D(n_a)}. \tag{14}$$

For $n_a > n''_a$, marginal taxes satisfy the standard Mirrlees formula:

$$\xi \frac{T'_a(y_a(n_a))}{1-T'_a(y_a(n_a))} n_a f_a(n_a) = \int_{n_a}^{\overline{n}} \left[1 - \mathbf{1}_x^{CU} m_a(\tilde{n}_a)\right] f_a(\tilde{n}_a) d\tilde{n}_a \tag{15}$$

Proposition 1 generalizes the Mirrlees formula to a multi-sector economy with endogenous occupational choice and multi-dimensional types. To obtain some intuition, suppose the government were to increase marginal taxes in sectors a and b by one dollar at income levels $y_a(n_a)$ and $y_b(c(n_a))$. This perturbation has two effects. The first one is the "direct effect", $D(n_a)$, represented by the integral in the right-hand-side of (14). This effect captures the additional tax revenue collected from all agents in sectors a and b whose incomes are above $y_a(n_a)$ and $y_b(c(n_a))$, respectively. Note that, for such agents, the change in tax schedules is equivalent to the introduction of a lump-sum tax equal to one dollar, given that, for such agents, labor supply is unaffected by the local increase in the marginal tax rates at the lower income levels. When the planner's objective is Rawlsian, the direct effect thus coincides with the total measure of those agents in sector a whose productivity exceeds n_a and of those agents in sector b whose productivity exceeds $c(n_a)$ (recall the definition of the density function $g(\cdot|c)$ in Section 2.4). When, instead, the planner's objective is Concave Utilitarian, the gains of raising this extra money from such agents must be discounted by the reduction in these agents' utility, as captured by the terms $m_j(n_j) = \phi'(u_j(n_j))/\lambda$.

The second effect is the "elasticity effect", which accounts for the intensive-margin distortions at the income levels $y_a(n_a)$ and $y_b(c(n_a))$ that result from the higher marginal tax rates. This effect corresponds to the sum of the terms $E_a(n_a)$ and $E_b(c(n_a))$ in the left-hand-side of (14). To understand these terms, note that the densities of those agents working in sector a with productivity n_a and of those agents working in sector b with productivity $c(n_a)$ are given by $g_a(n_a|c)$ and $g_b(c(n_a)|c)$, respectively. Next note that the terms $\xi\frac{T'}{1-T'}n$ in $E_a(n_a)$ and $E_b(c(n_a))$ capture the loss in tax revenues from those agents whose incomes are $y_a(n_a))$ and $y_b(c(n_a))$, due to the reduction in these agents' labor supply.[7] Figure 2 illustrates the effects discussed above by indicating the sets of agents affected by each of these effects.

Next, consider agents with productivity $n_a \in (n''_a, \bar{n})$ and recall that these agents are better off working in sector a than in sector b, irrespective of their sector-b productivity. For these agents, marginal taxes are given by the standard Mirrlees formula (15); this is because, for such agents, the extensive-margin incentive constraint associated with occupational choice is non-binding.

Note that the generalized Mirrlees formula (14), when combined with the incentive-compatibility constraint for occupational choice (12), pins down the marginal tax rates (and hence the labor supply) along the locus of indifferent types. Because the labor supply (and utility) of any agent whose type does not belong to this locus coincides with that of *some* type belonging to this locus, Proposition 1 delivers a complete characterization of the x-optimal taxation equilibrium implementing the choice rule $\mathcal{C}$.

[7] See Saez (2001) for an interpretation of the terms $\xi\frac{T'}{1-T'}n$ in terms of behavioral elasticities.

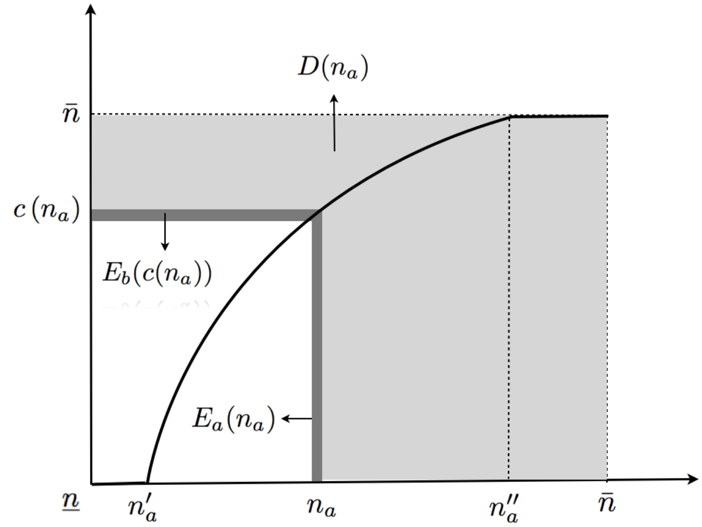

Figure 2: The sets of types affected by the elasticity and direct effects from the generalized Mirrlees formula (14).

More interestingly, Condition (12) determines the ratio of marginal taxes between sectors, while the generalized Mirrlees formula (14) determines their total magnitudes (as implied by the direct effect $D(n_a)$). Accordingly, the solution to the primal problem shows how the government optimally balances intensive-margin distortions in labor supply *across sectors*, as a function of the occupational choice rule to be implemented.

The next corollary shows how the marginal taxes in turn depend on the distribution of productivities and on the shape of the occupational choice rule.

Corollary 1 ***(Sign of Marginal Taxes)*** *Let c be the threshold function corresponding to the admissible occupational choice rule $\mathcal{C}$. Under the x-optimal taxation equilibrium implementing $\mathcal{C}$, the marginal income tax rates are such that, for any $n_a < n_a''$,*

$$sign\ \left\{T_a'(y_a(n_a))\right\} = sign\left\{\xi^{-1}\frac{\int_{n_a}^{\overline{n}}\left[1-\mathbf{1}_x^{CU}m_a\left(\tilde{n}_a\right)g(\tilde{n}_a|c)\right]d\tilde{n}_a}{c(n_a)g_b(c(n_a)|c)}+1-c'(n_a)J_c[n_a]\right\}. \tag{16}$$

In the standard Mirrlees model, marginal taxes are everywhere positive. This is true in our model if the choice rule C induces production efficiency, in which case $1-c'(n_a)J_c[n_a]=0$. This conclusion, however, is not necessarily true when production efficiency does not hold.

To get some intuition, consider a planner with a Rawlsian objective who wants to implement the linear occupational choice rule $c(n_a)=\rho n_a$, with $\rho>1$.[8] This rule reflects the desire to induce certain individuals to work in sector a, despite being more productive in sector b. According to Corollary 1, the marginal tax rate for those agents in sector a with income $y_a(n_a)$ is negative if and

[8]As will be clear in the next section, linear occupational choice rules play an important role when the government does not have the flexibility to employ sector-specific income tax schedules.

only if

$$\xi^{-1}\frac{1-G(n_a|c)}{\rho n_a g_b(\rho n_a|c)}+1<\rho.$$

The above inequality holds when ξ is high, in which case the elasticity of labor supply is high. Intuitively, as one can see from Condition (12), to induce an agent whose productivities are $(n_a, \rho n_a)$ to favor sector a over sector b, despite being more productive in sector b, the government must either subsidize sector a by setting $T_a'(y_a(n_a))$ negative, or penalize sector b by setting $T_b'(y_b(c(n_a)))$ sufficiently high. When the elasticity of labor supply is high, setting a high marginal tax rate $T_b'(y_b(c(n_a)))$ in sector b has a large negative impact on production due to its distortions on labor supply. At the optimum, the government then prefers to subsidize sector a by setting a negative marginal tax rate at income level $y_a(n_a)$.

3.2 Optimal occupational choice rule

We now turn to the dual problem $\mathcal{P}_2^x(h_a)$, where the side-a labor supply schedule h_a is held fixed, and the side-b labor supply h_b (or equivalently, the occupational choice rule $\mathcal{C}$) is chosen to maximize the planner's x-objective. This is the subject of the next proposition. Let

$$\varepsilon_{y_b}(n_b)\equiv\frac{dy_b(n_b)}{dn_b}\frac{n_b}{y_b(n_b)}$$

denote the elasticity of income with respect to productivity, in sector b.

Proposition 2 ***(Occupational Choice)*** *The x-optimal tax system implementing the side-a labor supply schedule h_a satisfies the following integral-form Euler equation at every point $n_a\in(n_a', n_a'')$:*

$$\mathbf{1}_x^{CU}W_b(c(n_a))=R_b(c(n_a))+M_a(n_a)+\underbrace{E_b(c(n_a))(1-T_b'(y_b(c(n_a)))y_b(c(n_a))}_{\textit{continuity correction: }\Delta_b(c(n_a))}\tag{17}$$

where

$$W_b(c(n_a))\equiv\int_{\underline{n}}^{c(n_a)}m_b(n_b)\left[1-T_b'(y_b(n_b))\right]y_b(n_b)dG_b(n_b|c)\tag{18}$$

is the "welfare effect,"

$$R_b(c(n_a))\equiv\int_{\underline{n}}^{c(n_a)}\left[1-T_b'(y_b(n_b))\varepsilon_{y_b}(n_b)\right]y_b(n_b)dG_b(n_b|c)\tag{19}$$

is the "revenue collection effect,"

$$M_a(n_a)\equiv\int_{n_a'}^{n_a}\left[T_a(y_a(\tilde{n}_a))-T_b(y_b(c(\tilde{n}_a)))\right]c(\tilde{n}_a)f(\tilde{n}_a,c(\tilde{n}_a))d\tilde{n}_a\tag{20}$$

is the "migration effect," and $E_b(c(n_a))$ is the elasticity effect described in (14).

The proof in the appendix provides a formal analysis of the dual problem $\mathcal{P}_2^x(h_a)$, and employs variational techniques to establish the necessity of the Euler equation. To help intuition, we present below an heuristic derivation for Condition (17).

Heuristic Derivation of the Euler Equation. To understand the Euler equation (17), consider a particular class of incremental tax reforms, which we call *payroll tax reforms*. Such reforms consist in introducing a new payroll tax that withholds a fraction $\alpha > 0$ of the sector-b workers' income and taxes the residual income $(1-\alpha)y_b$ according to the original income tax schedule T_b. Formally, an *α-payroll-tax reform applied to all income levels up to* $y_b(c(n_a))$ implies the following effective tax schedule in sector b:

$$T_b^\alpha(y) \equiv \begin{cases} \alpha y + T_b\left((1-\alpha)y\right) & \text{if} \quad y < y_b(c(n_a)) \\ T_b\left(y\right) & \text{if} \quad y \geq y_b(c(n_a)). \end{cases} \tag{21}$$

Now, let $(\mathcal{C}, h_b, \mathcal{T})$ be a solution to problem $\mathcal{P}_2^x(h_a)$, where h_a is an implementable labor supply schedule. To simplify the exposition, let us consider the case where $\mathcal{C}(\underline{n}, \underline{n}) = a$.[9] Optimality implies that no incremental payroll tax reform increases the government's x-objective. Accordingly, let T_b be the sector-b tax schedule under the tax system $\mathcal{T}$ and consider "perturbing" T_b by means of an α-payroll-tax reform up to income level $y_b(c(n_a))$. Under the effective tax schedule T_b^α, any worker in sector b with productivity $n_b < c(n_a)$ obtains a payoff

$$h_b n_b - \psi(h_b) - T_b^\alpha(h_b n_b) = (1-\alpha)h_b n_b - \psi(h_b) - T_b\left((1-\alpha)h_b n_b\right). \tag{22}$$

As one can see from (22), the problem faced by an agent with productivity $n_b < c(n_a)$ under the α-payroll-tax reform T_b^α is the same as the problem that an agent with sector-b productivity $(1-\alpha)n_b$ would have faced under the original tax schedule T_b. This implies that, under T_b^α, the indirect utility u_b^α of each agent with sector-b productivity n_b is given by

$$u_b^\alpha(n_b) \equiv \begin{cases} u_b\left((1-\alpha)n_b\right) & \text{if} \quad n_b < c(n_a) \\ u_b\left(n_b\right) & \text{if} \quad n_b \geq c(n_a). \end{cases} \tag{23}$$

where u_b is the indirect utility function under the original tax schedule T_b. As a consequence, the occupational choice rule under the perturbed schedule T_b^α, which we denote by $\mathcal{C}^\alpha$ can be described by a threshold function c^α that is a linear transformation

$$c^\alpha(\tilde{n}_a) = \frac{1}{1-\alpha}c(\tilde{n}_a) \tag{24}$$

of the threshold rule c under the original schedule T_b, for any $\tilde{n}_a < n_a$.

Remarkably, as we show below, the Euler equation (17) accounts for the gains and losses of α-payroll-tax reforms up to each income level $y_b(c(n_a))$ (for short, α-reforms). Hereafter, we discuss each of these effects.

[9] The heuristic derivation can be easily adapted for the case where $\mathcal{C}(\underline{n}, \underline{n}) = b$.

- **Welfare effect.** The first effect is the impact of the reform on the agents' utility. From (23), it is easy to see that, at $\alpha = 0$, the marginal effect of an α-reform up to income level $y_b(c(n_a))$ on the indirect utility of any worker whose sector-b productivity is $n_b < c(n_a)$ is equal to $-n_b u_b'(n_b)$. When the government's objective is concave-utilitarian, the importance assigned to this effect, adjusted for the opportunity cost of raising money, is given by

$$-m(n_b)\, u_b'(n_b)\, n_b = -m(n_b)\left[1 - T_b'(y_b(n_b))\right] y_b(n_b),$$

where the equality follows from the incentive-compatibility constraint (7) along with the first-order condition (5). Integrating the expression above for all $n_b < c(n_a)$ leads to the welfare effect $W_b(c(n_a))$ in the Euler equation, as defined in (18). In the case of a Rawlsian government, this effect is zero, given that the effect of tax reforms on the indirect utility of all agents but the worst-off individuals are disregarded by the planner.

- **Revenue collection effect.** The second effect is the impact of the reform on the tax revenues collected by the government. From the definition of the perturbed tax system in (21), it is easy to see that, under the α-payroll-tax reform, the tax revenue collected from each agent working in sector b with productivity $n_b < c(n_a)$ is given by

$$\begin{aligned} &\alpha n_b h_b^{\alpha}(n_b) + T_b\left((1-\alpha) n_b h_b^{\alpha}(n_b)\right) \\ &= \alpha n_b h_b((1-\alpha)n_b) + T_b\left((1-\alpha) n_b h_b((1-\alpha)n_b)\right), \end{aligned} \tag{25}$$

where h_b^{α} is the sector-b labor supply schedule under T_b^{α}, and where the equality in (25) follows from the fact the labor supply of each agent with productivity $n_b < c(n_a)$ under the schedule T_b^{α} coincides with the labor supply of an agent with productivity $(1-\alpha)n_b$ under the original schedule T_b. Differentiating the right-hand-side in (25) with respect to α and evaluating the expression at $\alpha = 0$, we obtain that the marginal effect of the reform on the revenues collected from each agent whose sector-b productivity is $n_b < c(n_a)$ is equal to

$$\left[1 - T_b'(y_b(n_b)\varepsilon_{y_b}(n_b)\right] y_b(n_b).$$

Integrating the expression above for all $n_b < c(n_a)$ leads to the revenue collection effect $R(n_a)$ in the Euler equation, as defined in (19).

- **Migration effect.** The third effect accounts for the fact that agents change occupation in response to the tax reform. Differentiating equation (24) with respect to α and evaluating the derivative at $\alpha = 0$, we have that the occupational choice rule shifts at a rate $c(\tilde{n}_a)$, at each productivity level $\tilde{n}_a < n_a$ in response to an incremental α-reform. Accordingly, for any $\tilde{n}_a < n_a$, the mass of agents whose sector-a productivity is $\tilde{n}_a$ and who change occupations is given by $c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a))$. As a consequence, the impact on tax revenues from the migration of these agents is equal to

$$\left[T_a(y_a(\tilde{n}_a)) - T_b(y_b(c(\tilde{n}_a)))\right] c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)).$$

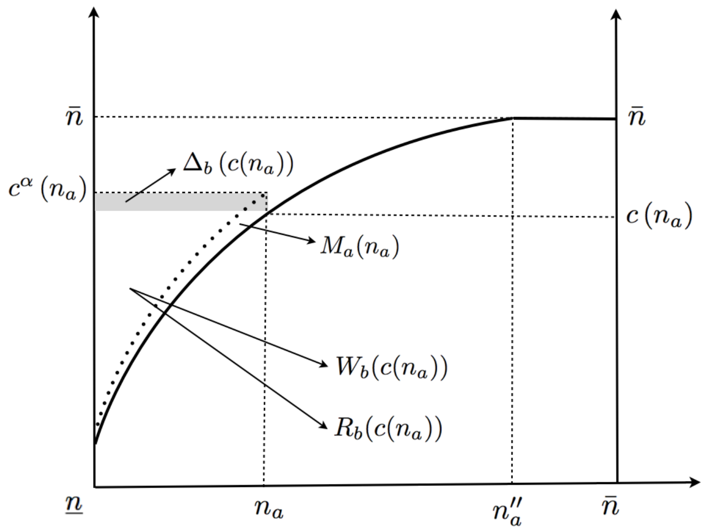

Figure 3: The figure illustrates the types affected by the welfare, revenue collection, migration, and continuity correction effects discussed in the main text. The dotted curve corresponds the occupational choice rule under the α-payroll tax reform.

Integrating the above expression for all $\tilde{n}_a < n_a$ leads to the migration effect in the Euler equation, as defined in (20).

- **Continuity correction.** Finally, consider the last term in the right-hand side of the Euler equation (17), $\Delta_b\,(c(n_a))$. As can be seen from equation (23), an α-reform leads to a sector-b indirect utility schedule that has a (single) discontinuity point at $c(n_a)$. Indeed, $u_b^{\alpha}(\cdot)$ is continuous at any $n_b < c(n_a)$ and at any $n_b > c(n_a)$, but

$$\lim_{n_b \to c(n_a)^-} u_b^{\alpha}(n_b) = u_b((1-\alpha)c(n_a)) < u_b(c(n_a)) = \lim_{n_b \to c(n_a)^+} u_b^{\alpha}(n_b),$$

for any $\alpha > 0$. Accordingly, for an α-reform to lead to an implementable allocation, it has to be coupled with transfers to sector-b agents with productivities in a neighborhood of $c(n_a)$, so as to restore the continuity of the indirect utility schedule. For incremental α-reforms (i.e., $\alpha \approx 0$) only sector-b agents with productivity $c(n_a)$ need receive such transfers. In order to reduce the indirect utility of those agents whose sector-b productivity is equal to $c(n_a)$ to its "continuity level" $\lim_{n_b \to c(n_a)^-} u_b^{\alpha}(n_b)$, the planner charges a lump-sum tax to such agents equal to the extra taxes that these agents would pay were they subject to the reform. This lump-sum charge is the term $\Delta_b\,(c(n_a))$ in the right-hand side of the Euler equation (17). It is equal to the product of (a) the elasticity effect $E_b(c(n_a))$ (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 their variation in indirect utility, which is equal to $[1 - T_b'(y_b(c(n_a)))]\, y_b(c(n_a))$, as shown in the derivation of the Welfare effect.

Figure 3 illustrates the sets of types affected by each of the effects discussed above.

Note that the welfare, revenue collection, and continuity correction effects account for net impact of perturbing the sector-b tax schedule on the utilities and tax revenues from sector-b workers. As such, taken together, these effects measure the marginal gain of better tailoring the taxation of sector-b workers to the distribution of productivities on *that* sector (tagging). The marginal gains from tagging, at the optimum, equalize the marginal losses due to the migration of workers across sectors (as captured by the migration effect).

3.3 On the optimality of production inefficiency

Using the characterization in the previous two propositions, we can now establish two key properties of optimal taxation equilibria. To this end, the following definition is instrumental.

Definition 3 ***(Non-generic Distributions)*** *The distribution of productivities F is non-generic if there exists $\delta > 0$ such that*

$$f_a(n)F_{b|a}(n|n) = \delta f_b(n)F_{a|b}(n|n) \quad \textit{for almost every} \quad n \in N. \tag{26}$$

The distribution F is generic if the above property does not hold.

Note that symmetric distributions, i.e., those for which $F(n_a, n_b) = F(n_b, n_a)$, are non-generic (as they satisfy the Condition in (26) with $\delta = 1$). The next proposition shows that production inefficiencies are a robust feature of optimal taxation equilibria.

Proposition 3 ***(Equilibrium Properties)*** *Let $\mathcal{E} = (\mathcal{C}, h_a, h_b, \mathcal{T})$ be any x-optimal taxation equilibrium. The following properties hold under $\mathcal{E}$.*

1. *If the distribution of productivities F is generic, then production efficiency fails: there exists a subset of N (of positive Lebesgue measure) such that*

$$c(n_a) \neq n_a.$$

2. *The marginal tax collection vanishes at the top of the distribution, in each sector:*

$$\lim_{n_j \to \overline{n}} T_j'(y_j(n_j))g_j(n_j|c) = 0 \tag{27}$$

for $j = a, b$.

Part 1 of the proposition shows that, whenever the distribution of productivities is generic in the sense of Definition 3, then any x-optimal taxation equilibrium fails to satisfy production efficiency. Below we sketch the main idea — the details are in formal proof in the Appendix.

Sketch of the Proof of Proposition 3 - Part 1. Let $\mathcal{E}$ be an x-optimal taxation equilibrium and suppose that it satisfies production efficiency. Then let c denote the threshold function associated with the efficient occupational choice rule $\mathcal{C}$ and recall that the latter coincides with the identity function, that is, $c(n) = n$ for all $n \in N$ (in which case the choice of sector labels is arbitrary). Conditions (5) and (9) then imply that the income and tax schedules must coincide in the two sectors. Let these schedules be $y(\cdot)$ and $T'(\cdot)$, respectively. Proposition 2 then implies that, for every $n \in N$ and any $j \in \{a, b\}$, the following condition must hold:

$$\mathbf{1}_x^{CU} W_j(n) = R_j(n) + E_j(n)(1 - T'(y(n)))y(n),$$

where the welfare, revenue, and elasticity effects are evaluated at the equilibrium schedules $y(\cdot)$, $T'(\cdot)$, and $c(\cdot)$.

Differentiating the equation above with respect to n for $j = a, b$ leads to two linear *homogenous* differential equations in $g_b(n|c)$ and $g_a(n|c)$, respectively, with boundary conditions $g_a(\underline{n}|c) = g_b(\underline{n}|c) = 0$. Because these homogenous differential equations are identical, the Picard-Lindelof theorem implies that their solutions must satisfy

$$g_a(n|c) = \delta g_b(n|c) \text{ for all } n \in N. \tag{28}$$

Under production efficiency, $g_a(n|c) = f_a(n)F_{b|a}(n|n)$ and $g_b(n|c) = f_b(n)F_{a|b}(n|n)$. For (28) to hold, it must then be the case that the distribution of productivities is non-generic. Q.E.D.

Intuitively, the densities $f_j(n)F_{j|k}(n|n)$, for $j, k \in \{a, b\}$ $k \neq j$, capture the informational costs of redistribution in the two sectors. Whenever such costs differ across the two sectors, the planner can improve upon any equilibrium satisfying production efficiency by distorting occupational choice away from $c(n) = n$. Doing so yields a first-order reduction in the informational costs of redistribution and only a second-order efficiency loss from the misallocation of talent across the two sectors (as the migration effect is zero under the efficient occupational choice rule). At the optimum, the planner then distorts occupational choice up to the point where the marginal losses in tax revenue due to the migration effect are equalized to the marginal gains from tailoring the tax schedule in each sector to the endogenous distribution of talent (tagging), as required by the Euler equation (17).

Turning to Part 2, the result in the proposition says that, under any x-optimal taxation equilibrium, *marginal tax collection vanishes at the top.* This is either because top earners face vanishing marginal tax rates (which happens when the support of the productivity distribution is bounded, i.e. $\bar{n} < \infty$, and the density is bounded away from zero in a neighborhood of $(\bar{n}, \bar{n})$), or because the density of top earners vanishes (when $\bar{n} = \infty$, marginal taxes do not necessarily vanish at the "top", but (27) necessarily holds). The result thus extends familiar findings on the taxation of top earners (e.g., Mirrlees (1971), Diamond and Mirrlees (1971), Saez (2002), among others) to the economy with multi-dimensional productivity and endogenous occupational choice under examination here. In particular, when $\bar{n} < \infty$, Proposition 3 reveals that distortions in occupational choice do *not* translate into distortions in labor supply for those agents at the top of the income distribution in each of the two sectors.

4 Sales Taxes under Uniform Income Taxation

The results above are developed under the assumption that the government can employ sector-specific income tax schedules. While this possibility appears plausible (e.g., business owners face a different tax schedule than earners whose income comes through wages), it is worth investigating the validity of the above results in settings in which the government is unable to use sector-specific income tax schedules, so that $T_a = T_b$. In this case, the tax treatment of the two sectors can differ only through the sales taxes τ_a and τ_b, which we now reintroduce (recall that these taxes play no role when income taxation is allowed to be sector-specific).

The next lemma shows how the restriction to uniform income taxation translates into a restriction over the effective tax schedules. This lemma will allow us to employ the general methodology developed in the previous section to the case of uniform income taxation.

Lemma 2 ***(Effective Tax Schedules)*** *Let $\mathcal{E} = (\mathcal{C}, h_a, h_b, \mathcal{T}, \mathbf{w})$ be a taxation equilibrium with tax system $\mathcal{T} = \{T, T, \tau_a, \tau_b\}$ featuring uniform income taxation.*

1. *There exists another taxation equilibrium $\hat{\mathcal{E}} = (\mathcal{C}, h_a, h_b, \hat{\mathcal{T}}, \hat{\mathbf{w}})$, implementing the same allocation $(\mathcal{C}, h_a, h_b)$ as $\mathcal{E}$, such that the tax system $\hat{\mathcal{T}}$ employs no sales taxes in either sector (i.e., $\hat{\tau}_a = \hat{\tau}_b = 0$ and $\hat{\mathbf{w}} = (1,1)$) and its effective tax schedules satisfy the following constraint:*

$$\hat{T}_a(y) = \alpha(\tau_a, \tau_b) \cdot y + \hat{T}_b\left((1 - \alpha(\tau_a, \tau_b)) \cdot y\right) \text{ for all } y \in \mathbb{R}_+, \tag{29}$$

   where

$$\alpha(\tau_a, \tau_b) \equiv \frac{\tau_a - \tau_b}{1 - \tau_b}.$$

2. *For any taxation equilibrium $\hat{\mathcal{E}} = (\mathcal{C}, h_a, h_b, \hat{\mathcal{T}}, \hat{\mathbf{w}})$ where $\hat{\tau}_a = \hat{\tau}_b = 0$, $\hat{\mathbf{w}} = (1,1)$ and the tax schedules $(\hat{T}_a, \hat{T}_b)$ satisfy (29), there exists another taxation equilibrium $\mathcal{E} = (\mathcal{C}, h_a, h_b, \mathcal{T}, \mathbf{w})$ featuring uniform income taxation and sales taxes (τ_a, τ_b) that implements the same allocation $(\mathcal{C}, h_a, h_b)$ as in $\hat{\mathcal{E}}$.*

The result in the lemma says that equilibrium outcomes sustained with uniform income taxation but sector-specific sales taxes can always be sustained as equilibrium outcomes with no sales taxes and sector-specific income taxation. The opposite is not generally true. Because sales taxes are linear in output, the planner's problem when only sales taxes can be made sector-specific is more constrained than when income taxes can be sector-specific. In particular, when only sales taxes can be made sector specific, the effective tax schedules must satisfy the additional constraint (29). Interestingly, this constraint reveals that the effective tax schedule in sector a is an $\alpha(\tau_a, \tau_b)$-payroll tax reform of the sector-b tax schedule up to the highest income level.

To characterize the x-optimal taxation equilibria under uniform income taxation, we then proceed as in the previous section, but now impose that the effective tax schedules satisfy the constraint in (29). In particular, we start by identifying necessary and sufficient conditions for the implementability of a given occupational choice rule $\mathcal{C}$ under uniform income taxation. We then proceed by solving problem $\mathcal{P}_1^x(\mathcal{C})$, which delivers the x-optimal (effective) income tax schedules among those implementing the desired occupational choice rule $\mathcal{C}$. We then translate the optimal effective tax schedules into their corresponding sales and income tax schedules. Finally, we look at the dual problem $\mathcal{P}_2^x(h_a)$ where we derive the x-optimal taxation equilibrium implementing the sector-a labor supply schedule h_a. To this end, the following lemma is instrumental.

Lemma 3 ***(Sales Taxes and Occupational Choice)*** *Consider a taxation equilibrium* $\hat{\mathcal{E}} = (\mathcal{C}, h_a, h_b, \hat{\mathcal{T}}, \hat{\mathbf{w}})$ *in which the tax system* $\hat{\mathcal{T}}$ *employs no sales taxes in either sector (i.e.,* $\tau_a = \tau_b = 0$ *and* $\hat{\mathbf{w}} = (1,1)$*). The income tax schedules* $(\hat{T}_a, \hat{T}_b)$ *satisfy Condition (29) if and only if the occupational choice rule* $\mathcal{C}$ *can be described by a threshold function of the form*

$$c(n_a) = (1 - \alpha(\tau_a, \tau_b))n_a \tag{30}$$

at every point where $c(n_a) \in N$.

Together, Lemmas 2 and 3 imply that any occupational choice rule sustained with uniform income taxation must be described by a linear threshold function. The linearity of threshold functions follows from the fact that, when income is taxed uniformly across sectors, the only instruments that can possibly create heterogeneity in the effective tax schedules are the sales taxes. Because the latter are linear in output (and hence in income), so are the corresponding occupational choice rule sustained in equilibrium, *irrespective* of the shape of the common income tax schedule T.

For convenience, we will choose sector labels such that $\tau_a \leq \tau_b$. This is the same labeling convention of subsection 2.4, where sector a was chosen to be the sector for which there is a productivity threshold $n_a'' \in N$ such that $c(n_a) = \bar{n}$ for all $n_a \geq n_a''$; that is, all agents whose sector-a productivity exceeds n_a'' work in sector a, irrespective of their sector-b productivity.

The next proposition provides a partial characterization of the properties of optimal taxation equilibria by fixing the occupational choice rule and then deriving the tax system that maximizes the government's objective subject to implementing the desired occupational rule.

Proposition 4 ***(Generalized Mirrlees Formula: Sales Taxes)*** *Let c be the linear threshold function associated with the occupational choice rule $\mathcal{C}$. The x-optimal tax system implementing the choice rule $\mathcal{C}$ satisfies the following generalized Mirrlees formula for almost any $n_a \in (n_a', n_a'')$*

$$\begin{aligned} &\xi \frac{T'(y_a(n_a)) + \frac{\tau_a}{1-\tau_a}}{1 - T'(y_a(n_a))} n_a g_a(n_a|c) + \xi \frac{T'(y_b(c(n_a))) + \frac{\tau_b}{1-\tau_b}}{1 - T'(y_b(c(n_a)))} c(n_a) g_b(c(n_a)|c) \\ &= \int_{n_a}^{\overline{n}} \left[1 - \mathbf{1}_x^{CU} m_a\left(\tilde{n}_a\right)\right] g(\tilde{n}_a|c) d\tilde{n}_a. \end{aligned} \tag{31}$$

and the following Mirrlees formula

$$\xi \frac{T_a'(y_a(n_a)) + \frac{\tau_a}{1-\tau_a}}{1 - T_a'(y_a(n_a))} n_a f_a(n_a) = \int_{n_a}^{\overline{n}} \left[1 - \mathbf{1}_x^{CU} m_a\left(\tilde{n}_a\right)\right] f_a(\tilde{n}_a) d\tilde{n}_a$$

for any $n_a \geq n_a''$.

Proof of Proposition 4. The generalized Mirrlees formula (31) follows from Proposition 1 by observing that, under uniform labor income taxation, the effective marginal tax rates in sectors a and b are equal to $\tau_a + T'(y)(1 - \tau_a)$ and $\tau_b + T'(y)(1 - \tau_b)$, respectively. Q.E.D.

The idea behind the formula in (31) parallels the one behind the generalized Mirrlees formula in (14): at the optimum, marginal taxes balance redistributive and efficiency considerations, as captured by the elasticity and direct effects identified in Proposition 1.

Notice that the formula in (31) does not pin down a unique tax system. As discussed in Remark 2, the government can change sale taxes across sectors and adjust the common income tax schedule in a way that leaves the effective tax schedules unchanged. Yet, the equilibrium allocation is unique. Accordingly, and similarly to the previous section, the generalized Mirrlees formula (31), together with Condition (12), determines how the government optimally balances intensive-margin distortions in labor supply *across sectors*, as a function of the occupational choice rule to be implemented.

We now turn to the dual problem $\mathcal{P}_2^x(h_a)$, which consists in finding the taxation equilibrium that maximizes the government's objective, among those that implement a given sector-a labor supply schedule h_a, but now imposing that the tax system satisfy uniform income taxation (equivalently, restricting the effective tax schedules to satisfy constraint (29)).

Proposition 5 ***(Occupational Choice: Sales Taxes)*** *Suppose the government is constrained to tax labor income homogeneously across sectors. The x-optimal tax system implementing the labor supply schedule h_a satisfies the following condition*

$$\lim_{n_b \to \bar{n}} \left\{\mathbf{1}_x^{CU} \cdot W_b\left(n_b\right) - R_b\left(n_b\right)\right\} = \lim_{n_a \to \bar{n}\frac{1-\tau_b}{1-\tau_a}} M_a\left(n_a\right) \tag{32}$$

where W_b, R_b, and M_b are, respectively, the welfare, the revenue collection, and the migration effects defined in Proposition 2, evaluated at the occupational choice rule (30).

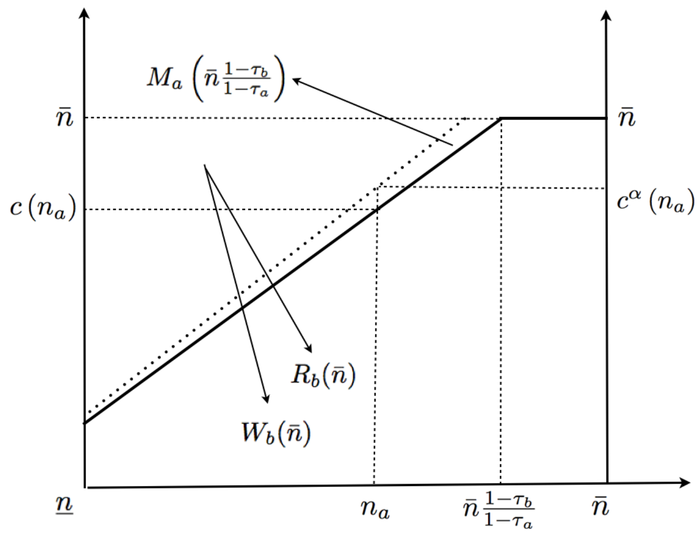

Figure 4: The dotted line corresponds the occupational choice rule after the α-payroll tax reform. The set of types affected by each of the welfare, revenue collection and migration effects are indicated above.

The formula in (32) is closely related to the general Euler condition (17) of Proposition 2. The intuition for this formula can thus be obtained by considering α-payroll tax reforms similar to those considered in the previous section, but now applied to *all* individuals in sector b. The reason why the welfare, revenue collection, and migration effects must now be evaluated over *all* sector-b productivity levels is the limited flexibility of the government's tax instruments under uniform income taxation. In particular, the fact that sales taxes impact *uniformly* all income levels precludes the possibility of restricting the α-payroll reform to a subset of the income levels in sector b. As one should expect, this implies that the x-optimal tax system under uniform income taxation is in general welfare-inferior to the x-optimal tax system with sector-specific income taxation. Moreover, (32) displays no continuity correction, given that any α-payroll-tax reform *up to the highest income level* generates no discontinuities in the schedule of indirect utilities. Figure 4 illustrates the sets of types affected by each of the effects discussed above.

Under uniform income taxation, the migration effect in condition (32) can be expressed in familiar terms. Using Lemma 2, this effect equals

$$\lim_{n_a \to \bar{n}\frac{1-\tau_b}{1-\tau_a}} M(n_a) = (\tau_a - \tau_b) \cdot \lim_{n_a \to \bar{n}\frac{1-\tau_b}{1-\tau_a}} \int_{\underline{n}}^{n_a} y_a(\tilde{n}_a) c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)) d\tilde{n}_a,$$

which is simply the loss in the collection of sales taxes among all individuals that migrate from sector b to sector a as the wedge $\tau_a - \tau_b$ decreases.

Finally, as in the case of differential taxation, production efficiency generically fails under x-optimal taxation equilibria. To see why, consider a planner whose objective is Rawlsian. Suppose there exists an x-optimal taxation equilibrium featuring production efficiency. It then follows from the Mirrlees formula (31) that[10]

$$\xi \frac{T'(y(n))}{1-T'(y(n))} n = \frac{1-F(n,n)}{f_a(n)F_{b|a}(n|n) + f_b(n)F_{a|b}(n|n)}, \tag{33}$$

where $y(n)$ is the income schedule, which is common to the two sectors. Importantly, the marginal tax schedule $T'(\cdot)$ only depends on the "diagonal" distribution $\hat{F}(n) \equiv F(n,n)$ (whose density is the denominator in the right-hand side of the equality above). By contrast, the characterization of Proposition 5 implies that the "off-diagonal" densities $f_i(n)F_{i|j}(n|n)$ satisfy

$$\int_{\underline{n}}^{\bar{n}} \left[1 - T'(y(n))\varepsilon_y(n)\right] y(n) f_j(n) F_{k|j}(n|n) dn = 0$$

for all $j, k \in \{a, b\}$ with $k \neq j$. The equality above can be satisfied by both $f_a(n)F_{b|a}(n|n)$ and $f_b(n)F_{a|b}(n|n)$ only in knife-edge cases, such as when the distribution F is non-generic (in the sense of Definition 3).

The above result thus contrasts with the Atkinson-Stiglitz theorem, according to which, when preferences over consumption and leisure are separable, as they are in our economy, the second-best can be implemented with zero sales taxes. The reason why this theorem does not hold in the economy under examination here is that, when occupational choice is endogenous, the informational costs of redistribution are also endogenous and can be effectively manipulated by differential taxation. When the government is unable to levy sector-specific income taxes, sector-specific sales taxes then become strictly optimal.

5 Discussion and Conclusions

This paper studies optimal differential taxation in a setting in which agents' productivity is sector-specific and in which agents choose which sector to work in. We show how properties of optimal taxation equilibria can be identified by first considering a primal problem, where the occupational choice rule (describing the allocation of workers across sectors) is held fixed, and where the government chooses a tax system to maximize welfare *subject to* implementing that occupational choice rule. Next, one considers a dual problem, where the labor supply of a given sector is held fixed, and where the government chooses a tax system, along with an occupational choice rule, to maximize welfare *subject to* implementing that labor supply schedule. A welfare-maximizing taxation equilibrium, comprising a tax system, an occupational choice rule, and a collection of sector-specific labor supply schedules, must be a solution to each of the above problems.

The primal-dual approach described above delivers a number of important insights. First, it delivers a formula for marginal tax rates that generalizes the well-known Mirrlees formula to a

[10] That $\tau_a = \tau_b$ in the formula in (33) is without loss of generality. In fact, given any taxation equilibrium $\mathcal{E}$ featuring (a) uniform income taxation, (b) production efficiency, and (c) $\tau_a = \tau_b \neq 0$, there exists another taxation equilibrium $\mathcal{E}'$ also featuring (a) and (b) in which $\tau_a = \tau_b = 0$ and such that the allocation implemented under $\mathcal{E}'$ is the same as under $\mathcal{E}$. This follows directly from the result in Remark 1.

multi-dimensional setting with endogenous occupational choice. This generalization shows how the government optimally balances intensive-margin distortions in labor supply *across sectors*, for any desired occupational choice rule. The formula shows that, for certain occupational choice rules sought by the government, negative marginal taxes may be optimal. The approach also delivers an Euler equation that determines the optimal allocation of agents across sectors. We provide an heuristic derivation of this equation using an important class of tax perturbations which we labelled payroll-tax reforms. Most importantly, this Euler equation reveals that, under sector-specific income taxation, the Diamond-Mirrlees theorem (according to which the second-best entails production efficiency) fails: social welfare can be increased by inducing certain agents to work in the sector in which they are least productive. Intuitively, by sacrificing production efficiency, the planner incurs second-order losses in total output, but obtains a first-order reduction in the informational costs of redistribution. The same logic applies when the government is constrained to a uniform income tax schedule, as long as sales taxes can be made sector-specific. In this latter case, the analysis shows the failure of the Atkinson-Stiglitz theorem (according to which, when preferences over consumption and leisure are separable, as they are in our economy, the second-best can be implemented with zero sales taxes).

Our analysis is conducted in the context of a multi-sector economy where workers choose their occupation in response to the tax system. One alternative, and equally appealing, application of our results pertains to the design of tax systems in a federation of states. In this application, the type of each worker describes his productivity in the different member states. That a worker's productivity varies across geographical areas may reflect technological, cultural, and linguistic differences across member states. After observing the tax schedules and the wages in each member state, workers decide where to locate themselves, taking into account their differences in productivities.

The planner's problem studied in this paper (with either a Rawlsian or Concave Utilitarian objective) coincides with the problem of a federal authority designing the tax system of each of its member states so as to maximize aggregate welfare over the entire federation. Our results can then be directly applied to this problem, and imply that differential tax treatments across member states are a robust feature of the optimal *centralized* tax system.[11]

Another application of the methods developed in the present paper pertains to the design of optimal tax systems in economies with a large informal sector. Economies plagued by a large degree of informality in the labor market display a somewhat extreme form of differential taxation: Workers in the formal sector face income taxes, while workers in the informal sector are able to evade such taxes. Yet, wages in both sectors are affected by sales taxes (and other forms of indirect

[11] For models of competing tax authorities, see Hamilton and Pestieau (2005) and the references therein. In these models, it is typically assumed that workers are (i) equally productive in the various member states, and (ii) heterogenous in their mobility cost, which determines their location choice. In this setup, the centralized optimum always exhibits production efficiency. By contrast, the richer heterogeneity considered in the present paper reveals that differential taxation is a robust feature of centralized optimal tax systems.

taxation), which are typically easier to enforce than income taxes.[12] More broadly, the interplay between tax enforceability and occupational choice, and its implications for the design of optimal tax systems, is an exciting topic for future research.

Applying the optimal tax formulas obtained in the present paper to quantitative exercises is non-trivial, but of first-order importance. The main difficulty pertains to the multi-variate distribution of skills that governs the workers' sectorial choices. Estimating such distribution in a Roy model is subject to well-known identification problems (see Heckman and Honore (1990) and the references therein). Yet, obtaining reliable estimates of this distribution is essential given that the sign and magnitude of the optimal tax wedges across sectors naturally depend on the migration patterns of workers across occupations.

Finally, the methods developed in this paper can be applied to other multi-dimensional screening problems in which agents face rival choices. For example, in the context of nonlinear pricing, consider a car seller designing price-quality schedules for various car categories (sport, SUV, family, minivans). Buyers, given their preferences for each category and the price-quality schedules offered by the seller, then decide which type of car to buy and then select, within the chosen category, the desired model (identified by a combination of price and quality).

[12] Indeed, that sales taxes are easier to enforce than income taxes is widely recognized as a justification for the heavy reliance on such taxes (as well as other modes of indirect taxation) in underdeveloped countries.

6 Appendix: Omitted Proofs

Proof of Lemma 1. *Part 1.* For necessity, let $\hat{u}_j(y;n_j) \equiv y - T_j(y) - \psi\left(y/(w_j n_j)\right)$ denote the utility that an agent with productivity n_j obtains by generating income y in sector j. Because the function satisfies the strict increasing difference property, the optimal income choice $y_j(n_j)$, and hence the supply of effective labor $n_j h_j(n_j)$, must be weakly increasing over N_j. Next note that $\hat{u}_j(y;n_j)$ is differentiable and Lipschitz continuous in n_j. Standard envelope theorems (e.g., Milgrom and Segal (2002)) then imply that the value function $u_j(n_j) = \max_h\{w_j h n_j - T_j(w_j h n_j) - \psi(h)\} = \max_y\{\hat{u}_j(y;n_j)\}$ must be Lipschitz continuous over N with derivative equal to

$$u_j'(n_j) = \psi'\left(\frac{\hat{y}_j(n_j)}{w_j n_j}\right)\frac{\hat{y}_j(n_j)}{w_j n_j^2}$$

for almost every $n_j \in N$, where $\hat{y}_j : N \to \arg\max_y \left\{y - T_j(y) - \psi(\frac{y}{w_j n_j})\right\}$ is an arbitrary selection. Using the fact that, for any $n_j \in N_j$,

$$h_j(n_j) = \frac{\hat{y}_j(n_j)}{w_j n_j}$$

for some selection $\hat{y}_j(\cdot)$, we then arrive to the result in Part 1. Sufficiency follows from standard arguments (e.g., in Milgrom (2004)).

Part 2. We establish necessity first. Because u_j is absolutely continuous and strictly increasing, $j = a, b$, any occupational choice rule $\mathcal{C}$ must be described by an absolutely continuous and weakly increasing threshold function $c : N \to \bar{N}$ such that $\mathcal{C}(n_a, n_b) = a$ if $n_b < c(n_a)$ and $\mathcal{C}(n_a, n_b) = b$ if $n_b > c(n_a)$. Furthermore at any point $n_a \in N$ in which $c(n_a) \in N$, the threshold $c(n_a)$ must satisfy $u_a(n_a) = u_b(c(n_a))$. Sufficiency follows by construction, after noting that u_a and u_b are strictly increasing in n_a and n_b, respectively.

Part 3. That Condition (8) is necessary and sufficient for the labor market to clear follows directly from the fact that the production function exhibits constant returns to scale. Q.E.D.

Proof of Remark 1. Because u_a and u_b are Lipschitz continuous and strictly increasing, for almost every $n_a \in N$ such that $c(n_a) \in N$, condition $u_a'(n_a) = u_b'(c(n_a))c'(n_a)$ must hold. Using Condition (7), we then have that, for almost every $n_a \in N$ such that $c(n_a) \in N$, Condition (9) must hold. Q.E.D.

Proof of Remark 2. From Condition (8), under the original tax system $\mathcal{T}$, wages are given by $w_j = 1 - \tau_j$. Faced with these wages, under the original tax system, the optimal choice of effective labor $\hat{h} = nh$ for an agent with sector-j productivity n_j who chooses to work in sector j is given by

$$\arg\max_{\hat{h}}\left\{(1-\tau_j)\hat{h} - \psi(\hat{h}/n_j) - T_j\left((1-\tau_j)\hat{h}\right)\right\}.$$

Under the new tax system $\hat{\mathcal{T}}$, wages are equal to $\hat{w}_j = 1$, $j = a, b$, and the optimal choice of effective labor by the same agent working in sector j is given by

$$\arg\max_{\hat{h}} \left\{ \hat{h} - \psi(\hat{h}/n_j) - \hat{T}_j(\hat{h}) \right\} = \arg\max_{\hat{h}} \left\{ \hat{h} - \psi(\hat{h}/n_j) - \tau_j \hat{h} - T_j\left((1-\tau_j)\hat{h} \right) \right\}.$$

It is then easy to see that the original allocation $(\mathcal{C}, h_a, h_b)$ can be implemented under the wages $\hat{\mathbf{w}} = (1,1)$ by the new tax system $\hat{\mathcal{T}}$. It is also easy to see that all agents' payoffs (as well as the government's tax income) under $\hat{\mathcal{E}}$ are the same as under $\mathcal{E}$. Q.E.D.

Proof of Proposition 1. The government's problem consists in choosing labor supply schedules $h_a : N_a \to \mathbb{R}_+$, $h_b : N_b \to \mathbb{R}_+$ along with tax schedules $T_a : \mathbb{R}_+ \to \mathbb{R}$ and $T_b : \mathbb{R}_+ \to \mathbb{R}$ so as to maximize its x-objective:[13]

$$\mathbf{1}_x^{CU} \int_{N_a} \phi(u_a(n_a)) dG_a(n_a|c) + \mathbf{1}_x^{CU} \int_{N_b} \phi(u_b(n_b)) dG_b(n_b|c) + [1 - \mathbf{1}_x^{CU}] u_a(n_a'),$$

where

$$u_j(n_j) = h_j(n_j) n_j - \psi\left(h_j(n_j)\right) - T_j(n_j h_j(n_j)) \text{ for every } n_j \in N_j,\ j = a, b$$

subject to (i) the budget constraint:

$$\int_{N_a} T_a(n_a h_a(n_a)) dG_a(n_a|c) + \int_{N_b} T_b(n_b h_b(n_b))) dG_b(n_b|c) \geq G,$$

(ii) the labor-supply incentive-compatibility constraints:

$$u_a'(n_a) = \psi'\left(h_a(n_a)\right) \frac{h_a(n_a)}{n_a} \text{ for almost every } n_a \in N_a,$$

$$u_b'(n_b) = \psi'\left(h_b(n_b)\right) \frac{h_b(n_b)}{n_b} \text{ for almost every } n_b \in N_b,$$

(iii) the occupational-choice incentive-compatibility constraints:

$$h_b(c(n_a)) = J_c[n_a] h_a(n_a), \text{ for all } n_a \in (n_a', n_a''),$$

and (iv) the monotonicity constraints

$$y_j(n_j) = h_j(n_j) n_j \text{ nondecreasing over } N_j,\ j = a, b.$$

As mentioned in the main text, hereafter, we proceed by abstracting from the monotonicity constraints (iv), which is consistent with the practice commonly followed in the literature.

Using the fact that (i) for any $n_a \in (n_a', \bar{n})$,

$$T_a(n_a h_a(n_a))) = h_a(n_a) n_a - \psi\left(h_a(n_a)\right) - u_a(n_a),$$

[13]Note that, in case of a Rawlsian objective, i.e., for $x = R$, the lowest-utility agent is always an agent whose sector-a productivity is n_a'.

along with the fact that (ii) for $n_b \in N_b$, $h_b(n_b) = J_c[c^{-1}(n_b)]h_a(c^{-1}(n_b))$ and $u_b(n_b) = u_a(c^{-1}(n_b))$, it follows that

$$\begin{aligned}T_b(n_b h_b(n_b)) &= h_b(n_b)n_b - \psi\left(h_b(n_b)\right) - u_b(n_b) \\ &= J_c[c^{-1}(n_b)]h_a(c^{-1}(n_b))n_b - \psi\left(J_c[c^{-1}(n_b)]h_a(c^{-1}(n_b))\right) - u_a(c^{-1}(n_b)).\end{aligned}$$

Using the definition of the density

$$\begin{aligned}g(n_a|c) &= g_a(n_a|c) + c'\left(n_a\right) g_b(c\left(n_a\right)|c) \\ &= f_a(n_a)F_{b|a}(c(n_a)|n_a) + c'\left(n_a\right) f_b(c(n_a))F_{a|b}(n_a|c(n_a)),\end{aligned}$$

we can then rewrite the government's problem as that of choosing functions $u_a : N_a \to \mathbb{R}$, $h_a : N_a \to \mathbb{R}_+$ so as to maximize

$$\int_{N_a} \left\{\mathbf{1}_x^{CU}\phi(u_a(n_a))g(n_a|c) + [1-\mathbf{1}_x^{CU}]f_a(n_a)u_a(n_a')\right\} dn_a \tag{34}$$

subject to the budget constraint

$$\begin{aligned}&\int_{N_a} \left\{[h_a(n_a)n_a - \psi\left(h_a(n_a)\right) - u_a(n_a)]\, f_a(n_a)F_{b|a}(c(n_a)|n_a)\right\} dn_a \\ &+ \int_{N_a} \left\{[J_c[n_a]h_a(n_a)c(n_a) - \psi\left(J_c[n_a]h_a(n_a)\right) - u_a(n_a)]\, c'(n_a)f_b(c(n_a))F_{a|b}(n_a|c(n_a))\right\} dn_a, \\ &\geq G\end{aligned} \tag{35}$$

and the IC constraints

$$u_a'(n) = \psi'\left(h_a(n_a)\right)\frac{h_a(n_a)}{n_a} \text{ for almost every } n_a \in N_a.$$

This is a standard optimal control problem with control variable h_a and state variable u_a. The Hamiltonian associated to this problem is:

$$\begin{aligned}H = &\mathbf{1}_x^{CU}\phi(u_a(n_a))g(n_a|c) + [1-\mathbf{1}_x^{CU}]f_a(n_a)u_a(n_a') \\ &+ \lambda\left\{[h_a(n_a)n_a - \psi\left(h_a(n_a)\right) - u_a(n_a)]\, f_a(n_a)F_{b|a}(c(n_a)|n_a)\right\} \\ &+ \lambda\left\{J_c[n_a]h_a(n_a)c(n_a) - \psi\left(J_c[n_a]h_a(n_a)\right) - u_a(n_a)\right\} c'(n_a)f_b(c(n_a))F_{a|b}(n_a|c(n_a)) \\ &+ \mu(n_a)\cdot\psi'\left(h_a(n_a)\right)\frac{h_a(n_a)}{n_a} - \lambda G,\end{aligned}$$

where λ is the Lagrange multiplier associated to the common budget constraint (35) and where μ is the co-state variable associated with the law of motion of u_a. The transversality conditions are:

$$\mu(n_a') = \mu(\bar{n}) = 0. \tag{36}$$

From the Pontryagin Maximum Principle,

$$\mu'(n_a) = -\frac{\partial H}{\partial u_a} = \left[\lambda - \mathbf{1}_x^{CU}\phi'(u_a(n_a))\right] g(n_a|c). \tag{37}$$

Integrating the right-hand side of (37) and using the transversality condition (36) we have that

$$\mu(n_a) = -\lambda \int_{n_a}^{\overline{n}} \left[1 - m_a(\tilde{n}_a)\mathbf{1}_x^{CU}\right] g(\tilde{n}_a|c)d\tilde{n}_a, \tag{38}$$

where we used the definition of

$$m_a(n_a) \equiv \frac{\phi'\left(u_a(n_a)\right)}{\lambda}.$$

Furthermore, for any n_a such that $h_a\left(n_a\right) > 0$, the following first order condition must hold:

$$\begin{aligned}
&\frac{\partial H}{\partial h_a} = \lambda\left[n_a - \psi'\left(h_a(n_a)\right)\right] f_a(n_a)F_{b|a}(c(n_a)|n_a) + \\
&\lambda\left\{J_c[n_a]c(n_a) - \psi'\left(J_c[n_a]h_a(n_a)\right)J_c[n_a]\right\} c'(n_a)f_b(c(n_a))F_{a|b}(n_a|c(n_a)) + \\
&\mu(n_a)\frac{\psi'\left(h_a(n_a)\right) + \psi''\left(h_a(n_a)\right)h_a(n_a)}{n_a} = 0
\end{aligned} \tag{39}$$

Combining (38) with (39) and using the definitions of the densities

$$g_a(n_a|c) = f_a(n_a)F_{b|a}(c(n_a)|n_a) \text{ and } g_b(c(n_a)|c) = f_b(c(n_a))F_{a|b}(n_a|c(n_a))$$

we obtain that:

$$\begin{aligned}
&\left[n_a - \psi'\left(h_a(n_a)\right)\right] g_a(n_a|c) + \left\{J_c[n_a]c(n_a) - \psi'\left(J_c[n_a]h_a(n_a)\right)J_c[n_a]\right\} c'(n_a)g_b(c(n_a)|c) \\
&= \left\{\frac{\psi'\left(h_a(n_a)\right) + \psi''\left(h_a(n_a)\right)h_a(n_a)}{n_a}\right\} \int_{n_a}^{\overline{n}} \left[1 - m_a(\tilde{n}_a)\mathbf{1}_x^{CU}\right] g(\tilde{n}_a|c)d\tilde{n}_a
\end{aligned} \tag{40}$$

From (5),

$$n_j - \psi'\left(h_j(n_j)\right) = T_j'(y_j(n_j))n_j, \; j = a, b$$

and

$$\begin{aligned}
\frac{\psi'\left(h_a(n_a)\right) + \psi''\left(h_a(n_a)\right)h_a(n_a)}{n_a} &= \frac{\psi'\left(h_a(n_a)\right)}{n_a}\left\{1 + \frac{\psi''\left(h_a(n_a)\right)h_a(n_a)}{\psi'\left(h_a(n_a)\right)}\right\} \\
&= \left[1 - T_a'(y_a(n_a))\right]\xi^{-1},
\end{aligned}$$

where $y_a(n_a) = n_a h_a(n_a)$. Hence, for any $n_a > n_a'' = c^{-1}(\bar{n})$, the optimality condition (40) can be rewritten as the usual Mirrlees condition

$$\xi\frac{T_a'(y_a(n_a))}{1 - T_a'(y_a(n_a))}n_a f_a(n_a) = \int_{n_a}^{\overline{n}} \left[1 - \mathbf{1}_x^{CU}m_a(\tilde{n}_a)\right] f_a(\tilde{n}_a)d\tilde{n}_a, \tag{41}$$

where we also used the fact that, for $n_a > c^{-1}(\bar{n})$, $g_a(n_a|c) = f_a(n_a)$. Next, consider any $n_a \in (n_a', n_a'')$. Using the fact that $J_c[n_a]h_a(n_a) = h_b(c(n_a))$, equation (40) can be rewritten as

$$\begin{aligned}
&\frac{T_a'(y_a(n_a))}{1 - T_a'(y_a(n_a))}n_a g_a(n_a|c) + \frac{T_b'(y_b(c(n_a)))}{1 - T_a'(y_a(n_a))}J_c[n_a]c'(n_a)c(n_a)g_b(c(n_a)|c) \\
&= \xi^{-1}\int_{n_a}^{\overline{n}} \left[1 - m_a(\tilde{n}_a)\mathbf{1}_x^{CU}\right] g(\tilde{n}_a|c)d\tilde{n}_a.
\end{aligned} \tag{42}$$

From Condition (9) we have that, for any $n_a \in (n'_a, n''_a)$,

$$c'(n_a) = \frac{\psi'\left(h_a(n_a)\right)\frac{h_a(n_a)}{n_a}}{\psi'\left(h_b(c(n_a))\right)\frac{h_b(c(n_a))}{c(n_a)}},$$

which, using (5), can be rewritten as

$$c'(n_a) = \frac{\left[1 - T'_a(y_a(n_a))\right]h_a(n_a)}{\left[1 - T'_b(y_b(n_b))\right]h_b(c(n_a))}. \tag{43}$$

Replacing (43) into (42) and using again the fact that $J_c[n_a]h_a(n_a) = h_b(c(n_a))$, we then have that, for any $n_a \in (n'_a, n''_a)$,

$$\begin{aligned}&\xi\frac{T'_a(y_a(n_a))}{1 - T'_a(y_a(n_a))}n_a g_a(n_a|c) + \xi\frac{T'_b(y_b(c(n_a)))}{1 - T'_b(y_b(c(n_a)))}c(n_a)g_b(c(n_a)|c) \\ &= \int_{n_a}^{\overline{n}}\left[1 - \mathbf{1}_x^{CU}m_a(\tilde{n}_a)\right]g(\tilde{n}_a|c)d\tilde{n}_a.\end{aligned} \tag{44}$$

Combining the results establishes the proposition. Q.E.D.

Proof of Corollary 1. From (14), we have that

$$\begin{aligned}\frac{T'_a(y_a(n_a))}{1 - T'_a(y_a(n_a))}M &= \frac{\xi^{-1}\int_{n_a}^{\overline{n}}\left[1 - \mathbf{1}_x^{CU}m_a\left(\tilde{n}_a\right)\right]g(\tilde{n}_a|c)d\tilde{n}_a}{c(n_a)g_b(c(n_a)|c)} - \frac{T'_b(y_b\left(c(n_a)\right))}{1 - T'_b(y_b(c(n_a)))} \\ &= \frac{\xi^{-1}\int_{n_a}^{\overline{n}}\left[1 - \mathbf{1}_x^{CU}m_a\left(\tilde{n}_a\right)\right]g(\tilde{n}_a|c)d\tilde{n}_a}{c(n_a)g_b(c(n_a)|c)} + 1 - c'(n_a)J_c[n_a]\frac{1}{1 - T'_a(y_a(n_a))},\end{aligned}$$

where

$$M \equiv \frac{n_a g_a(n_a|c)}{c(n_a)g_b(c(n_a)|c)} > 0,$$

and where the second equality uses Condition (12). Rearranging, we have that

$$\begin{aligned}&T'_a(y_a(n_a))\left\{M + \frac{\xi^{-1}\int_{n_a}^{\overline{n}}\left[1 - \mathbf{1}_x^{CU}m_a\left(\tilde{n}_a\right)g(\tilde{n}_a|c)\right]d\tilde{n}_a}{c(n_a)g_b(c(n_a)|c)} + 1\right\} \\ &= \frac{\xi^{-1}\int_{n_a}^{\overline{n}}\left[1 - \mathbf{1}_x^{CU}m_a\left(\tilde{n}_a\right)g(\tilde{n}_a|c)\right]d\tilde{n}_a}{c(n_a)g_b(c(n_a)|c)} + 1 - c'(n_a)J_c[n_a].\end{aligned}$$

Because the term in curly brackets is strictly positive, the above condition implies the result in the corollary. Q.E.D.

Proof of Proposition 2. Fix the sector-a labor supply schedule h_a (with domain N_a). The planner's problem is as in the proof of Proposition 1, except that the control policies are now (i) the threshold function $c : N \to \bar{N}$ defining the occupational choice rule, with c continuous over N, strictly increasing over (n'_a, n''_a) for some $n''_a \leq \bar{n}$, and such that $c(n_a) = \underline{n}$ for all $n_a \leq n'_a$ and $c(n_a) = \bar{n}$ for all $n_a \geq n''_a$, (ii) the sector-b labor supply schedule $h_b : N_b \to \mathbb{R}_+$, and the tax schedules $T_j : \mathbb{R}_+ \to \mathbb{R}$, $j = a, b$.

The planner's problem can be conveniently rewritten by letting

$$R_j^x(n_j) \equiv \mathbf{1}_x^{CU}\phi(u_j(n_j)) + \lambda\{h_j(n_j)n_j - \psi(h_j(n_j)) - u_j(n_j)\}, \tag{45}$$

denote the value the planner assigns to the utility of an agent whose sector-j productivity is n_j, adjusted for the opportunity cost or 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 threshold function $c : N \to \bar{N}$ satisfying the properties above, along with a sector-b labor supply schedule $h_b : N_b \to \mathbb{R}_+$ and a pair of utility functions $u_a : N_a \to \mathbb{R}$, $u_b : N_b \to \mathbb{R}$ that jointly maximize

$$\textstyle\sum_{j=a,b} \displaystyle\int_{\underline{n}}^{\bar{n}} R_j^x(n)\, g_j(n|c)dn + [1 - \mathbf{1}_x^{CU}]u_a(n_a') \tag{46}$$

subject to the incentive compatibility constraints for labor supply

$$u_a'(n_a) = \psi'(h_a(n_a))\frac{h_a(n_a)}{n_a} \text{ for almost every } n_a \in N_a, \tag{47}$$

$$u_b'(n_b) = \psi'(h_b(n_b))\frac{h_b(n_b)}{n_b} \text{ for almost every } n_b \in N_b, \tag{48}$$

and the occupational choice constraint

$$h_b(c(n_a)) = J_c[n_a]h_a(n_a),$$

for all $n_a \in (n_a', n_a'')$ where $n_a'' = c^{-1}(\bar{n})$.

Using the fact that, for any $n_b \in \text{int}(N_b)$,

$$h_b(n_b) = J_c[c^{-1}(n_b)]h_a(c^{-1}(n_b)), \tag{49}$$

and

$$u_b(n_b) = u_a(c^{-1}(n_b)), \tag{50}$$

along with the change in variables $n_b = c(n_a)$, we have that the planner's objective can be rewritten as

$$\int_{n_a'}^{\bar{n}} \left[R_a^x(n_a)\, g_a(n_a|c) + \hat{R}_b^x(n_a|c)\, g_b(c(n_a)|c)c'(n_a)\right] dn_a + [1 - \mathbf{1}_x^{CU}]u_a(n_a') \tag{51}$$

where, for any $n_a \in (n_a', n_a'')$,

$$\begin{aligned}\hat{R}_b^x(n_a|c) &= R_b^x(c(n_a)) \\ &= \mathbf{1}_x^{CU}\phi(u_b(c(n_a))) + \lambda\{h_b(c(n_a))c(n_a) - \psi(h_b(c(n_a))) - u_b(c(n_a))\} \\ &= \mathbf{1}_x^{CU}\phi(u_a(n_a)) + \lambda\{J_c[n_a]h_a(n_a)c(n_a) - \psi(J_c[n_a]h_a(n_a)) - u_a(n_a)\}\end{aligned}$$

and $\hat{R}^x_b\left(n_a|c\right) = 0$ if $n_a \geq n''_a$.

The planner's problem can then be thought of as choosing (i) a scalar $u_a(n'_a)$, and (ii) an absolutely continuous function $c : N \to \bar{N}$, strictly increasing over (n'_a, n''_a) for some $0 \leq n'_a$ and $n''_a \leq \bar{n}$ and satisfying $c(n_a) = 0$ if $n_a \leq n'_a$ and $c(n_a) = \bar{n}$ if $n_a \geq n''_a$, so as to maximize (51). Given $u_a(n'_a)$, because $h_a : (n'_a, \bar{n}) \to \mathbb{R}_+$ is fixed, the function $u_a : (n'_a, \bar{n}) \to \mathbb{R}$ is then uniquely determined by (47). The labor supply schedule $h_b : N_b \to \mathbb{R}_+$ and the utility schedule $u_b : N_b \to \mathbb{R}$ are then given by (49) and (50), respectively. Finally, the tax schedules in the two sectors are given by

$$T_j(n_j h_j(n_j))) = h_j(n_j)n_j - \psi\left(h_j(n_j)\right) - u_j(n_j),\ j = a, b.$$

As a first step, let us fix the scalar $u_a(n'_a)$ — and hence the entire utility function $u_a : (n'_a, \bar{n}) \to \mathbb{R}$ — as well as the thresholds n'_a and n''_a and then look at the optimality conditions for the threshold function $c : [n'_a, n''_a] \to \bar{N}$. To ease the exposition, let $\tilde{R}^x_b : \mathbb{R}^3 \to \mathbb{R}$ be the function defined by

$$\tilde{R}^x_b\left(n_a, c, J\right) \equiv \mathbf{1}^{CU}_x \phi\left(u_a\left(n_a\right)\right) + \lambda\left\{J h_a(n_a)c - \psi\left(J h_a(n_a)\right) - u_a\left(n_a\right)\right\}, \tag{52}$$

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

$$J(n, c, c') = \left(\frac{c}{nc'}\right)^{\xi}, \tag{53}$$

and note that, for $n_a \in (n'_a, n''_a)$, $J(n_a, c(n_a), c'(n_a)) = J_c[n_a] = \left(c(n_a)/\left(n_a \cdot c'(n_a)\right)\right)^{\xi}$. 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 the densities

$$g_a(n_a|c) \equiv f_a(n_a)F_{b|a}(c(n_a)|n_a) = \int_{\underline{n}}^{c(n_a)} f(n_a, x)dx \tag{54}$$

and

$$g_b(c(n_a)|c) = f_b(c(n_a))F_{a|b}(n_a|c(n_a)) = \int_{\underline{n}}^{n_a} f(x, c(n_a))dx \tag{55}$$

depend on the entire function c only through the value that this function takes at n_a. In other words, $g_a(n_a|c)$ and $g_b(c(n_a)|c)$ can be thought of as functions of n_a, and $c(n_a)$. This means that the optimality conditions for the threshold function c can be obtained as a solution to a calculus of variations problem with control c and objective

$$\int_{n'_a}^{n''_a} \left[R^x_a\left(n_a\right) g_a(n_a|c) + \tilde{R}^x_b\left(n_a, c(n_a), J(n_a, c(n_a), c'(n_a))\right) \cdot c'(n_a) \cdot g_b(c(n_a)|c)\right] dn_a + [1 - \mathbf{1}^{CU}_x] u_a(n'_a).$$

Dropping the arguments from $\tilde{R}^x_b$ and J to facilitate the writing, we then have that, at any

$n_a \in (n_a', n_a'')$, the point-wise Euler equation of this problem is given by

$$R_a^x(n_a)\, f(n_a, c(n_a)) + \frac{\partial \tilde{R}_b^x}{\partial c} c'(n_a) g_b(c(n_a)|c) + \frac{\partial \tilde{R}_b^x}{\partial J} J_c c'(n_a) g_b(c(n_a)|c) + \tilde{R}_b^x c'(n_a) \frac{\partial}{\partial c}\left[g_b(c(n_a)|c)\right]$$
$$= \frac{d}{dn_a}\left[\frac{\partial \tilde{R}_b^x}{\partial J} J_{c'} c'(n_a) g_b(c(n_a)|c) + \tilde{R}_b^x g_b(c(n_a)|c)\right]. \tag{56}$$

Use (55) to note that the fourth term of the left-hand-side of (56) can be developed as follows:

$$\begin{aligned}\tilde{R}_b^x c'(n_a) \frac{\partial}{\partial c}\left[g_b(c(n_a)|c)\right] &= \tilde{R}_b^x c'(n_a) \left(\int_{\underline{n}}^{n_a} \frac{\partial}{\partial n_b} f(x, c(n_a)) dx\right) \\ &= \tilde{R}_b^x \frac{d}{dn_a}\left[g_b(c(n_a)|c)\right] - \tilde{R}_b^x f(n_a, c(n_a)) \\ &= \frac{d}{dn_a}\left[\tilde{R}_b^x g_b(c(n_a)|c)\right] - \frac{d\tilde{R}_b^x}{dn_a} g_b(c(n_a)|c) - \tilde{R}_b^x f(n_a, c(n_a)).\end{aligned} \tag{57}$$

Substituting (57) into (56) and simplifying, we can rewrite the point-wise Euler equation (56) as follows

$$\left[R_a^x(n_a) - \tilde{R}_b^x\right] f(n_a, c(n_a)) + \frac{\partial \tilde{R}_b^x}{\partial c} c'(n_a) g_b(c(n_a)|c) + \frac{\partial \tilde{R}_b^x}{\partial J} J_c g_b(c(n_a)|c) c'(n_a) - \frac{d\tilde{R}_b^x}{dn_a} g_b(c(n_a)|c)$$
$$= \frac{d}{dn_a}\left[\frac{\partial \tilde{R}_b^x}{\partial J} J_{c'} c'(n_a) g_b(c(n_a)|c)\right]. \tag{58}$$

Multiplying both sides of (58) by $c(n_a)$ and rearranging terms, we obtain that

$$\left[R_a^x(n_a) - \tilde{R}_b^x\right] f(n_a, c(n_a)) c(n_a) + \frac{\partial \tilde{R}_b^x}{\partial c} c'(n_a) c(n_a) g_b(c(n_a)|c) - \frac{d\tilde{R}_b^x}{dn_a} c(n_a) g_b(c(n_a)|c)$$
$$= \frac{d}{dn_a}\left[\frac{\partial \tilde{R}_b^x}{\partial J} J_{c'} c'(n_a) g_b(c(n_a)|c)\right] c(n_a) - \frac{\partial \tilde{R}_b^x}{\partial J} J_c g_b(c(n_a)|c) c(n_a) c'(n_a). \tag{59}$$

Next, note that, for any $n_a \in (n_a', n_a'')$,

$$J_c = J_c(n_a, c(n_a), c'(n_a)) = \xi \left(\frac{c(n_a)}{n_a c'(n_a)}\right)^{\xi-1} \frac{1}{n_a c'(n_a)} = \xi J \frac{1}{c(n_a)},$$
$$J_{c'} = J_{c'}(n_a, c(n_a), c'(n_a)) = -\xi \left(\frac{c(n_a)}{n_a c'(n_a)}\right)^{\xi-1} \frac{c(n_a)}{n_a c'(n_a)} \frac{1}{c'(n_a)} = -\xi J \frac{1}{c'(n_a)}.$$

Note that the expressions above are always well-defined, as $c'(n_a) > 0$ for all $n_a \in (n_a', n_a'')$ (by virtue of (12)).

Replacing these expressions into the right-hand side of (59), we obtain that for any $n_a \in$

(n_a', n_a''), the right-hand side of the Euler equation becomes

$$\frac{d}{dn_a}\left[\frac{\partial \tilde{R}_b^x}{\partial J} J_{c'} c'(n_a) g_b(c(n_a)|c)\right] c(n_a) - \frac{\partial \tilde{R}_b^x}{\partial J} J_c g_b(c(n_a)|c) c(n_a) c'(n_a)$$
$$= -\xi \left\{ \frac{d}{dn_a}\left[J \frac{\partial \tilde{R}_b^x}{\partial J} g_b(c(n_a)|c)\right] c(n_a) + J \frac{\partial \tilde{R}_b^x}{\partial J} g_b(c(n_a)|c) c'(n_a) \right\}$$
$$= -\xi \frac{d}{dn_a}\left[J \frac{\partial \tilde{R}_b^x}{\partial J} g_b(c(n_a)|c) c(n_a)\right]. \tag{60}$$

Substituting (60) into (59), we then have that, for any $n_a \in (n_a', n_a'')$, the Euler equation becomes

$$\left[R_a^x(n_a) - \tilde{R}_b^x\right] f(n_a, c(n_a)) c(n_a) + \frac{\partial \tilde{R}_b^x}{\partial c} c'(n_a) c(n_a) g_b(c(n_a)|c) - \frac{d\tilde{R}_b^x}{dn_a} c(n_a) g_b(c(n_a)|c)$$
$$= -\xi \frac{d}{dn_a}\left[J \frac{\partial \tilde{R}_b^x}{\partial J} g_b(c(n_a)|c) c(n_a)\right] \tag{61}$$

Integrating (61) from n_a' to $n_a \in (n_a', n_a'')$ we then obtain that

$$\int_{n_a'}^{n_a} \left[R_a^x(\tilde{n}_a) - \tilde{R}_b^x(\tilde{n}_a)\right] c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)) d\tilde{n}_a \tag{62}$$
$$+ \int_{n_a'}^{n_a} \frac{\partial \tilde{R}_b^x(\tilde{n}_a)}{\partial c} c(\tilde{n}_a) c'(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a - \int_{n_a'}^{n_a} \frac{d\tilde{R}_b^x(\tilde{n}_a)}{dn_a} c(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a$$
$$= -\xi J(n_a) \frac{\partial \tilde{R}_b^x(n_a)}{\partial J} g_b(c(n_a)|c) c(n_a) + \lim_{n_a \to n_a'} \xi J(n_a) \frac{\partial \tilde{R}_b^x(n_a)}{\partial J} g_b(c(n_a)|c) c(n_a)$$

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

Consider the second term in the right-hand side of (62). This term is zero if $n_a' = \underline{n}$, as in this case

$$\lim_{n_a \to \underline{n}} g_b(c(n_a)|c) = \lim_{n_a \to \underline{n}} f_b(c(n_a)) F_{a|b}(n_a|c(n_a)) = 0,$$

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

$$\lim_{n_a \to n_a'} \xi J(n_a) \frac{\partial \tilde{R}_b^x(n_a)}{\partial J} g_b(c(n_a)|c) c(n_a) = 0,$$

which is exactly the second term in the right-hand side of (62).

We now express each term in (62) as a function of the income tax schedules. By definition of $R_a^x(\tilde{n}_a)$ and $\tilde{R}_b^x(\tilde{n}_a)$ in (45) and (52), the first term in (62) is simply:

$$\int_{n_a'}^{n_a} \left[R_a^x(\tilde{n}_a) - \tilde{R}_b^x(\tilde{n}_a)\right] c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)) d\tilde{n}_a$$
$$= \lambda \int_{n_a'}^{n_a} \left[T_a(y_a(\tilde{n}_a)) - T_b(y_b(c(\tilde{n}_a)))\right] c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)) d\tilde{n}_a. \tag{63}$$

The second term in (62) is obtained by differentiating (52) with respect to c, which yields

$$\int_{n_a'}^{n_a} \frac{\partial \tilde{R}_b^x(\tilde{n}_a)}{\partial c} c(\tilde{n}_a) c'(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a = \int_{n_a'}^{n_a} \lambda J(\tilde{n}_a) h_a(\tilde{n}_a) c(\tilde{n}_a) c'(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a$$

$$= \lambda \int_{n_a'}^{n_a} y_b(c(\tilde{n}_a)) c'(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a = \lambda \int_{c(n_a')}^{c(n_a)} y_b(n_b) g_b(n_b|c) dn_b. \tag{64}$$

where the last equality follows from changing the variable of integration from $\tilde{n}_a$ to n_b (using the relation $n_b = c(n_a)$).

The third term in (62) is obtained by totally differentiating (52) with respect to n_a, which gives

$$\int_{n_a'}^{n_a} \left(\frac{d\tilde{R}_b^x(\tilde{n}_a)}{dn_a} c(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) \right) d\tilde{n}_a$$

$$= \int_{n_a'}^{n_a} \left\{ \frac{d}{dn_a} \left[\mathbf{1}_x^{CU} \phi\left(u_a\left(\tilde{n}_a\right)\right) + \lambda T_b(y_b(c(\tilde{n}_a))) \right] \right\} c(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a$$

$$= \int_{n_a'}^{n_a} \left\{ \mathbf{1}_x^{CU} \phi'\left(u_a\left(\tilde{n}_a\right)\right) u_a'\left(\tilde{n}_a\right) + \lambda T_b'(y_b(c(n_a))) \frac{dy_b(c(\tilde{n}_a))}{dn_b} c'(\tilde{n}_a) \right\} c(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a$$

$$= \int_{n_a'}^{n_a} \left\{ \begin{array}{c} \mathbf{1}_x^{CU} \phi'\left(u_a\left(\tilde{n}_a\right)\right) \psi'\left(h_b(c(\tilde{n}_a))\right) h_b(c(\tilde{n}_a)) \\ +\lambda T_b'(y_b(c(n_a))) \frac{dy_b(c(\tilde{n}_a))}{dn_b} c(\tilde{n}_a) \end{array} \right\} c'(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a$$

$$= \int_{n_a'}^{n_a} \left\{ \begin{array}{c} \mathbf{1}_x^{CU} \phi'\left(u_a\left(\tilde{n}_a\right)\right) \left[1 - T_b'(y_b(c(\tilde{n}_a)))\right] y_b(c(\tilde{n}_a)) \\ +\lambda T_b'(y_b(c(n_a))) \frac{dy_b(c(\tilde{n}_a))}{dn_b} c(\tilde{n}_a) \end{array} \right\} c'(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a$$

where the last two equalities use (5), (7), and (9). Changing again the variables of integration using the relation $n_b = c(n_a)$, we then obtain that the third term in (62) is equal to

$$\int_{n_a'}^{n_a} \left(\frac{d\tilde{R}_b^x(\tilde{n}_a)}{dn_a} c(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) \right) d\tilde{n}_a$$

$$= \int_{c(n_a')}^{c(n_a)} \left\{ \mathbf{1}_x^{CU} \phi'\left(u_b\left(n_b\right)\right) \left[1 - T_b'(y_b(n_b))\right] + \lambda T_b'(y_b(n_b)) \varepsilon_{y_b}(n_b) \right\} y_b(n_b) g_b(n_b|c) dn_b,$$

where

$$\varepsilon_{y_b}(n_b) \equiv \frac{dy_b(n_b)}{dn_b} \frac{n_b}{y_b(n_b)}$$

Finally, the right-hand-side in (62) is obtained by differentiating (52) with respect to J which yields

$$\frac{\partial \tilde{R}_b^x}{\partial J} = \lambda h_a(n_a) \left[c(n_a) - \psi'\left(J(n_a) h_a(n_a)\right) \right] = \lambda h_a(n_a) c(n_a) T_b'(J(n_a) h_a(n_a) c(n_a)).$$

We then have that the right-hand-side in (62) can be rewritten as

$$-\xi J(n_a) \frac{\partial \tilde{R}_b^x(n_a)}{\partial J} g_b(c(n_a)|c) c(n_a) = -\xi \lambda y_b(c(n_a)) T_b'(y_b(c(n_a))) g_b(c(n_a)|c) c(n_a)$$

$$= \xi \frac{T_b'(y_b\left(c(n_a)\right))}{1 - T_b'(y_b(c(n_a)))} c(n_a) g_b(c(n_a)|c) \cdot \left\{ (1 - T_b'(y_b(c(n_a)))) y_b(c(n_a)) \right\}. \tag{65}$$

Substituting (63)-(65) into (62) and rearranging yields (17). Q.E.D.

Proof of Proposition 3 The proof has two parts, each establishing the result in the corresponding part in the proposition.

Part 1. We establish the result by showing that, when the distribution is generic, a taxation equilibrium sustaining production efficiency (that is, inducing an efficient occupational choice) fails to satisfy the necessary optimality conditions, as implied by (17), over a positive measure set of types. To see this, first use (5) and (9) to observe that, in any equilibrium sustaining production efficiency, $h_a(n) = h_b(n) = h(n)$ for all $n \in N$. Then use (7) and (12) to verify that, in any such equilibrium, $T_a(y) = T_b(y) = T(y)$ and hence $u_a(n) = u_b(n) = u(n)$ for all $n \in N$.

Next, observe that, for production efficiency to be optimal, the Euler equations in Proposition 2 must hold for each sector. Using the symmetry properties described above, we can rewrite these equations, for any $n \in N$, as follows

$$\mathbf{1}_x^{CU} \int_{\underline{n}}^{n} m(\tilde{n}) \left[1 - T'(y(\tilde{n}))\right] y(\tilde{n}) dG_b(\tilde{n}|c) = \int_{\underline{n}}^{n} \left[1 - T'(y(\tilde{n})\varepsilon_y(\tilde{n})\right] y(\tilde{n}) dG_b(\tilde{n}|c) + \xi T'(y(n)) y(n) n g_b(n|c) \tag{66}$$

$$\mathbf{1}_x^{CU} \int_{\underline{n}}^{n} m(\tilde{n}) \left[1 - T'(y(\tilde{n}))\right] y(\tilde{n}) dG_a(\tilde{n}|c) = \int_{\underline{n}}^{n} \left[1 - T'(y(\tilde{n})\varepsilon_y(\tilde{n})\right] y(\tilde{n}) dG_a(\tilde{n}|c) + \xi T'(y(n)) y(n) n g_a(n|c) \tag{67}$$

where we used he fact that $n'_a = n'_b = \underline{n}$ (which also implies that $g_a(\underline{n}|c) = g_b(\underline{n}|c) = 0$) along with the fact that, for all $n \in N$, $c(n) = n$, $m_a(n) = m_b(n) = m(n)$, $\varepsilon_{y_a}(n) = \varepsilon_{y_b}(n) = \varepsilon_y(n)$ with

$$\varepsilon_y(n) \equiv \frac{dy(n)}{dn} \frac{n}{y(n)} \text{ and } y(n) = nh(n).$$

Note that the Lagrange multiplier on the planner's budget constraint is the same in both equations. The two Euler equations (66) and (67) define two linear homogenous differential equations in $g_b(n|c)$ and $g_a(n|c)$, respectively, with boundary conditions $g_a(\underline{n}|c) = g_b(\underline{n}|c) = 0$. Because these homogenous differential equations are identical and the usual Lipschitz conditions hold (as implied by Lemma 1), the Picard-Lindelof theorem implies that their solutions must satisfy $g_b(n|c) = \delta g_a(n|c)$ for some δ. Because production efficiency implies that

$$g_b(n|c) = f_b(n) F_{a|b}(n|n) \text{ and } g_a(n|c) = f_a(n) F_{b|a}(n|n)$$

for all $n \in N$, we then conclude that, for production efficiency to be optimal, there must exist $\delta > 0$ such that, for all $n \in N$

$$f_b(n) F_{a|b}(n|n) = \delta f_a(n) F_{b|a}(n|n).$$

As a consequence, for any generic F, any x-optimal taxation equilibrium entails production inefficiency for a positive-measure subset of types.

Part 2. Consider the calculus of variations problem described in the proof of Proposition 2 with objective function (51). The optimal choice of n_a'' implies that the following transversality condition holds:

$$\lim_{n_a \to n_a''} \xi J(n_a) \frac{\partial \tilde{R}_b^x(n_a)}{\partial J} g_b(c(n_a)|c)c(n_a) = 0.$$

By the arguments in the proof of Proposition 2, this condition is equivalent to

$$0 = \lim_{n_a \to n_a''} E_b(c(n_a))(1 - T_b'(y_b(c(n_a)))y_b(c(n_a)) = \lim_{n_b \to \bar{n}} n_b y_b(n_b) T_b'(y_b(n_b)) g_b(n_b|c).$$

Because $n_b y_b(n_b)$ is strictly increasing and positive, it then follows that condition (27) has to hold for sector b. Finally, that condition (27) holds for sector a follows from the standard Mirrlees formula (15). Q.E.D.

Proof of Lemma 2. *Part 1*. Recall the definition of "effective income tax schedules", as given in (13). Using the fact that labor income taxes are uniform across the two sectors, we then have that, for any $y \in \mathbb{R}_+$, and any $\tau_a, \tau_b < 1$,

$$\hat{T}_a(y) = \tau_a y + T((1 - \tau_a) y)$$
$$\hat{T}_b\left(\frac{1 - \tau_a}{1 - \tau_b} y\right) = \tau_b \frac{1 - \tau_a}{1 - \tau_b} y + T((1 - \tau_a) y).$$

Combining the two expressions above with the result in Remark 2 leads to the result in the lemma.

Part 2. Let the tax schedule T be such that, for any y,

$$\tau_a y + T((1 - \tau_a)y) = \hat{T}_a(y).$$

Together with Condition (29), the equation above implies that

$$\tau_b y + T((1 - \tau_b)y) = \hat{T}_b(y).$$

The result then follows from the fact that the tax system $\mathcal{T} = (T, T, \tau_a, \tau_b)$ induces the same effective labor tax schedules and the same occupational choices as $\hat{\mathcal{T}}$. Q.E.D.

Proof of Lemma 3. Under the tax system $\hat{\mathcal{T}}$, the utility that an agent with productivity n_a obtains by working in sector a is given by

$$\begin{aligned} u_a(n_a) &= \max_h \left\{ n_a h - \psi(h) - \hat{T}_a(n_a h) \right\} = \max_h \left\{ (1 - \alpha) n_a h - \psi(h) - \hat{T}_b((1 - \alpha) n_a h) \right\} \\ &= u_b((1 - \alpha) n_a), \end{aligned}$$

where the equality follows from the fact that the tax system satisfies Condition (29). This in turn implies that, for any n_a for which $c(n_a) \in N$,

$$c(n_a) = (1 - \alpha) n_a$$

thus establishing the result. Q.E.D.

Proof of Proposition 5. Note that, with uniform labor income taxation and our convention about the labeling of the two sectors (which consists in assuming that $\tau_a \leq \tau_b$), $n'_a = \underline{n}$. From the proof of Proposition 4 observe that, for any pair of effective tax schedules $\hat{T}_a$ and $\hat{T}_b$ that are consistent with uniform labor income taxation (that is, that satisfy Condition (29)), there exist infinitely many combinations of sale taxes τ_a, τ_b along with a common labor income tax schedule T such that the effective income tax schedules corresponding to the system $\mathcal{T} = (\tau_a, \tau_b, T, T)$ are $\hat{T}_a$ and $\hat{T}_b$. In deriving the optimality conditions below, for convenience we will set the sale tax in sector b to $\tau_b = 0$. Once these conditions are identified, we will show how they can be expressed for arbitrary combinations of τ_a and τ_b.

Following arguments similar to those in the proof of Proposition 2, the dual problem $\mathcal{P}_2^x(h_a)$ can be recast as consisting in choosing a level of utility $u_a(\underline{n}) \geq 0$, along with a sale subsidy $\tau_a \leq 0$ so as to maximize

$$\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \mathbf{1}_x^{CU} \phi\left(u_a(n_a)\right)\left[g_a(n_a|c) + (1-\tau_a)g_b((1-\tau_a)n_a|c)\right] dn_a + \int_{\frac{\overline{n}}{1-\tau_a}}^{\overline{n}} \mathbf{1}_x^{CU} \phi\left(u_a(n_a)\right) f_a\left(n_a\right) dn_a$$
$$+ [1 - \mathbf{1}_x^{CU}] u_a(\underline{n})$$

subject to the budget constraint

$$\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left[(1-\tau_a)n_a h_a(n_a) - \psi\left(h_a(n_a)\right) - u_a\left(n_a\right)\right]\left[g_a(n_a|c) + (1-\tau_a)g_b((1-\tau_a)n_a|c)\right] dn_a +$$
$$\int_{\frac{\overline{n}}{1-\tau_a}}^{\overline{n}} \left[(1-\tau_a)h_a(n_a)n_a - \psi\left(h_a(n_a)\right) - u_a\left(n_a\right)\right] f_a\left(n_a\right) dn_a + \tau_a \int_{\underline{n}}^{\bar{n}} n_a h_a(n_a) g_a(n_a|c) dn_a \geq \mathcal{G}$$

where

$$u_a(n_a) = u_a(\underline{n}) + \int_{\underline{n}}^{n_a} \psi'\left(h_a(\tilde{n}_a)\right) \frac{h_a(\tilde{n}_a)}{\tilde{n}_a} d\tilde{n}_a \tag{68}$$

is determined by (7) and where the densities under the occupational choice rule corresponding to the sale tax τ_a are given by

$$g_a(n_a|c) = f_a(n_a) F_{b|a}((1-\tau_a)n_a)|n_a) \tag{69}$$

and

$$g_b(n_b|c) = f_b(n_b) F_{a|b}((1-\tau_a)^{-1} n_b)|n_b). \tag{70a}$$

Note that in writing the above program, we used the fact that, for any $n_a \in \left(\underline{n}, \frac{\overline{n}}{1-\tau_a}\right)$, (i) $c(n_a) = (1-\tau_a)n_a$, (ii) $h_b(c(n_a)) = h_b((1-\tau_a)n_a) = h_a(n_a)$, and (iii) $y_b(c(n_a)) = c(n_a)h_b(c(n_a)) = (1-\tau_a)n_a h_a(n_a)$. We also used the fact that, once $u_a(\underline{n})$ and τ_a are chosen, because h_a is given, the common labor income tax schedule T is then given by

$$T((1-\tau_a)n_a h_a(n_a)) = (1-\tau_a)n_a h_a(n_a) - \psi(h_a(n_a)) - u_a(n_a)$$

for all $n_a \in N$. To put it differently, once $u_a(\underline{n})$ is chosen, different choices of the sale tax τ_a translate into different common income tax schedules T while leaving the sector-a effective tax schedule

$$\hat{T}_a(y_a(n_a)) \equiv \tau_a y_a(n_a) + T\left((1-\tau_a)y_a(n_a)\right) = y_a(n_a) - \psi\left(\frac{y_a(n_a)}{n_a}\right) - u_a(n_a)$$

fixed, for any level of effective income $y = y_a(n_a) = n_a h_a(n_a)$, $n_a \in N_a$. This also means that the budget constraint in the above program can be rewritten as

$$\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \left[n_a h_a(n_a) - \psi\left(h_a(n_a)\right) - u_a\left(n_a\right)\right] g_a(n_a|c) dn_a + \int_{\frac{\bar{n}}{1-\tau_a}}^{\bar{n}} \left[n_a h_a(n_a) - \psi\left(h_a(n_a)\right) - u_a\left(n_a\right)\right] f_a(n_a) dn_a$$
$$+ \int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \left[(1-\tau_a)n_a h_a(n_a) - \psi\left(h_a(n_a)\right) - u_a\left(n_a\right)\right](1-\tau_a) g_b((1-\tau_a)n_a|c) dn_a \geq \mathcal{G}$$

Then, for any $n_a \in N_a$ let

$$R_a^x\left(n_a\right) \equiv \mathbf{1}_x^{CU} \phi\left(u_a\left(n_a\right)\right) + \lambda\left[h_a\left(n_a\right) n_a - \psi\left(h_a\left(n_a\right)\right) - u_a\left(n_a\right)\right], \tag{71}$$

while for any $n_a \in \left(\underline{n}, \frac{\bar{n}}{1-\tau_a}\right)$ let

$$\begin{aligned}
\hat{R}_b^x\left(n_a|c\right) &\equiv R_b^x\left(c(n_a)\right) \qquad (72)\\
&= \mathbf{1}_x^{CU} \phi\left(u_b(c(n_a))\right) + \lambda\left\{h_b(c(n_a))c(n_a) - \psi\left(h_b(c(n_a))\right) - u_b\left(c(n_a)\right)\right\} \\
&= \mathbf{1}_x^{CU} \phi\left(u_a\left(n_a\right)\right) + \lambda\left\{J_c[n_a] h_a(n_a) c(n_a) - \psi\left(J_c[n_a] h_a(n_a)\right) - u_a\left(n_a\right)\right\} \\
&= \mathbf{1}_x^{CU} \phi\left(u_a\left(n_a\right)\right) + \lambda\left[h_a\left(n_a\right)(1-\tau_a) n_a - \psi\left(h_a\left(n_a\right)\right) - u_a\left(n_a\right)\right],
\end{aligned}$$

where λ is the Lagrangian multiplier associated with the government's budget constraint. The Lagrangian for the above program then becomes

$$\int_{\underline{n}}^{\bar{n}} R_a^x\left(n_a\right) g_a(n_a|c) dn_a + \int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a|c\right)(1-\tau_a) g_b((1-\tau_a)n_a|c) dn_a$$
$$+ [1-\mathbf{1}_x^{CU}] u_a(\underline{n}) - \lambda \mathcal{G}$$

Fixing $u_a(\underline{n})$, we then have that the first order condition with respect to τ_a is

$$\frac{d}{d\tau_a} \int_{\underline{n}}^{\bar{n}} R_a^x\left(n_a\right) g_a(n_a|c) dn_a$$
$$+ \frac{d}{d\tau_a} \int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a|c\right)(1-\tau_a) g_b((1-\tau_a)n_a|c) dn_a = 0. \tag{73}$$

The latter condition can be rewritten as

$$\begin{aligned}\frac{d}{d\tau_a}\int_{\underline{n}}^{\bar{n}} R_a^x\left(n_a\right) g_a(n_a|c)dn_a+ \\ +\frac{\bar{n}}{1-\tau_a}\hat{R}_b^x\left(\frac{\bar{n}}{1-\tau_a}|c\right) g_b(\bar{n}|c) \\ +\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a|c\right)\frac{d}{d\tau_a}\left\{(1-\tau_a)g_b((1-\tau_a)n_a|c)\right\} dn_a \\ +\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \frac{d\hat{R}_b^x\left(n_a|c\right)}{d\tau_a}(1-\tau_a)g_b((1-\tau_a)n_a|c)dn_a=0.\end{aligned} \tag{74}$$

Consider the first term in (74) and note that it is equal to

$$\begin{aligned}\frac{d}{d\tau_a}\int_{\underline{n}}^{\bar{n}} R_a^x\left(n_a\right) g_a(n_a|c)dn_a &= \frac{d}{d\tau_a}\left\{\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} R_a^x\left(n_a\right) g_a(n_a|c)dn_a+\int_{\frac{\bar{n}}{1-\tau_a}}^{\bar{n}} R_a^x\left(n_a\right) f_a(n_a)dn_a\right\} \\ &= \int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} R_a^x\left(n_a\right)\frac{d}{d\tau_a}[g_a(n_a|c)]dn_a \\ &= -\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} R_a^x\left(n_a\right) n_a f(n_a,(1-\tau_a)n_a)dn_a\end{aligned} \tag{75}$$

where we used the fact that $g_a(n_a|c)=f_a(n_a)$ for all $n_a\geq\frac{\bar{n}}{1-\tau_a}$ along with (69).

Next, consider the third term in (74) and use (70a) to note that

$$\frac{d}{d\tau_a}\left\{(1-\tau_a)g_b((1-\tau_a)n_a|c)\right\}=-\frac{d}{dn_a}\left[n_a g_b((1-\tau_a)n_a|c)\right]+n_a f\left(n_a,(1-\tau_a)n_a\right).$$

The the third term in (74) is thus equal to

$$\begin{aligned}&\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a;c\right)\frac{d}{d\tau_a}\left\{(1-\tau_a)g_b((1-\tau_a)n_a|c)\right\} dn_a \\ &= -\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a;c\right)\frac{d}{dn_a}\left[n_a g_b((1-\tau_a)n_a|c)\right] dn_a+\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a;c\right) n_a f\left(n_a,(1-\tau_a)n_a\right) dn_a\end{aligned}$$

Now observe that

$$\begin{aligned}&-\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a;c\right)\frac{d}{dn_a}\left[n_a g_b((1-\tau_a)n_a|c)\right] dn_a \\ &= -\left[\hat{R}_b^x\left(n_a;c\right) n_a g_b((1-\tau_a)n_a|c)\right]_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \\ &+\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \frac{d}{dn_a}\left[\hat{R}_b^x\left(n_a;c\right)\right] n_a g_b((1-\tau_a)n_a|c)dn_a \\ &= -\lim_{n_a\to\bar{n}}\frac{n_a}{1-\tau_a}\hat{R}_b^x\left(\frac{n_a}{1-\tau_a}|c\right) g_b(n_a|c) \\ &+\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \left\{\left[\mathbf{1}_x^{CU}\phi'\left(u_a(n_a)\right)-\lambda\right] u_a'\left(n_a\right)+\lambda\left[h_a'\left(n_a\right) n_a\left(1-\tau_a\right)+h_a\left(n_a\right)\left(1-\tau_a\right)-\psi'\left(h_a\left(n_a\right)\right) h_a'\left(n_a\right)\right]\right\}\cdot \\ &\cdot n_a g_b((1-\tau_a)n_a|c)dn_a\end{aligned}$$

where the first equality follows from integration by parts, whereas the third equality follows from the fact that $g_b((1-\tau_a)\underline{n}|c)=0$ along with the fact that

$$\frac{d}{dn_a}\left[\hat{R}_b^x(n_a;c)\right]=\left[\mathbf{1}_x^{CU}\phi'(u_a(n_a))-\lambda\right]u_a'(n_a)$$
$$+\lambda\left[h_a'(n_a)\,n_a(1-\tau_a)+h_a(n_a)(1-\tau_a)-\psi'(h_a(n_a))\,h_a'(n_a)\right].$$

We conclude that the third term in (74) is thus equal to

$$\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\hat{R}_b^x(n_a;c)\frac{d}{d\tau_a}\left\{(1-\tau_a)g_b((1-\tau_a)n_a|c)\right\}dn_a \tag{76}$$
$$=-\lim_{n_a\to\bar{n}}\frac{n_a}{1-\tau_a}\hat{R}_b^x\left(\frac{n_a}{1-\tau_a}|c\right)g_b(n_a|c)$$
$$+\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\left\{\left[\mathbf{1}_x^{CU}\phi'(u_a(n_a))-\lambda\right]u_a'(n_a)+\lambda\left[h_a'(n_a)\,n_a(1-\tau_a)+h_a(n_a)(1-\tau_a)-\psi'(h_a(n_a))\,h_a'(n_a)\right]\right\}\cdot$$
$$\cdot\,n_a g_b((1-\tau_a)n_a|c)dn_a$$
$$+\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\hat{R}_b^x(n_a;c)\,n_a f(n_a,(1-\tau_a)n_a)\,dn_a.$$

Finally, consider the forth term in (74). Differentiating (72) with respect to τ_a, we can show that this term is equal to

$$\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\frac{d\hat{R}_b^x(n_a;c)}{d\tau_a}(1-\tau_a)g_b((1-\tau_a)n_a|c)dn_a \tag{77}$$
$$=-\lambda\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}h_a(n_a)\,n_a(1-\tau_a)g_b((1-\tau_a)n_a|c)dn_a.$$

Substituting (75), (76) and (77) into (74) and simplifying, we obtain that the optimality condition can be rewritten as

$$\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\left\{\left[\mathbf{1}_x^{CU}\phi'(u_a(n_a))-\lambda\right]u_a'(n_a)+\lambda\left[h_a'(n_a)\,n_a(1-\tau_a)-\psi'(h_a(n_a))\,h_a'(n_a)\right]\right\}\cdot \tag{78}$$
$$\cdot\,n_a g_b((1-\tau_a)n_a|c)dn_a$$
$$=\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\left\{R_a^x(n_a)-\hat{R}_b^x(n_a;c)\right\}n_a f(n_a,(1-\tau_a)n_a)dn_a$$

Using the fact that

$$R_a^x(n_a)-\hat{R}_b^x(n_a;c)=\lambda[\hat{T}_a(y_a(n_a)-\hat{T}_b(y_b(c(n_a)))]$$

we then have that (78) can be rewritten as

$$\begin{aligned}
&\mathbf{1}_x^{CU} \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} m_a(n_a)\, u_a'(n_a)\, n_a g_b((1-\tau_a)n_a|c) dn_a \\
&+ \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{-u_a'(n_a) + h_a'(n_a)\, n_a(1-\tau_a) - \psi'(h_a(n_a))\, h_a'(n_a)\right\} n_a g_b((1-\tau_a)n_a|c) dn_a \\
&= \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} [\hat{T}_a(y_a(n_a) - \hat{T}_b(y_b(c(n_a)))] n_a f(n_a, (1-\tau_a)n_a) dn_a
\end{aligned} \tag{79}$$

where we used the fact that $m_a(n_a) \equiv \frac{\phi'(u_a(n_a))}{\lambda}$.

Using (5) and (7), we can then rewrite the first integral in (79) as follows:

$$\begin{aligned}
&\mathbf{1}_x^{CU} \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} m_a(n_a)\, u_a'(n_a)\, n_a g_b((1-\tau_a)n_a|c) dn_a \\
&= \mathbf{1}_x^{CU} \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} m_a(n_a)\left[1 - T'((1-\tau_a)y_a(n_a))\right](1-\tau_a)y_a(n_a) g_b((1-\tau_a)n_a|c) dn_a
\end{aligned} \tag{80}$$

where $y_a(n_a) = n_a h_a(n_a)$ is the effective labor supply by an agent working in sector a with productivity n_a.

Likewise, using (5) and (7), we can rewrite the second integral in (79) as follows:

$$\begin{aligned}
&\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{-u_a'(n_a) + h_a'(n_a)\, n_a(1-\tau_a) - \psi'(h_a(n_a))\, h_a'(n_a)\right\} n_a g_b((1-\tau_a)n_a|c) dn_a \\
&= \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{ \begin{array}{c} -[1 - T'((1-\tau_a)y_a(n_a))](1-\tau_a)h_a(n_a) \\ +h_a'(n_a)\, n_a(1-\tau_a) - \left[1 - T'((1-\tau_a)y_a(n_a))\right] n_a(1-\tau_a)h_a'(n_a) \end{array} \right\} \cdot \\
&\cdot\, n_a g_b((1-\tau_a)n_a|c) dn_a
\end{aligned} \tag{81}$$

Using the fact that

$$y'(n_a) = h_a(n_a) + h_a'(n_a)n_a.$$

we can rewrite (81) as follows:

$$\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{-1 + T'((1-\tau_a)y_a(n_a))\frac{y_a'(n_a)\, n_a}{y_a(n_a)}\right\}(1-\tau_a)y_a(n_a) g_b((1-\tau_a)n_a|c) dn_a$$

We conclude that (79) can be rewritten as

$$\begin{aligned}
&\mathbf{1}_x^{CU} \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} m_a(n_a)\left[1 - T'((1-\tau_a)y_a(n_a))\right](1-\tau_a)y_a(n_a) g_b((1-\tau_a)n_a|c) dn_a \\
&= \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{1 - T'((1-\tau_a)y_a(n_a))\frac{y_a'(n_a)\, n_a}{y_a(n_a)}\right\}(1-\tau_a)y_a(n_a) g_b((1-\tau_a)n_a|c) dn_a \\
&+ \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} [\hat{T}_a(y_a(n_a) - \hat{T}_b(y_b(c(n_a)))] n_a f(n_a, (1-\tau_a)n_a) dn_a
\end{aligned} \tag{82}$$

Changing the variable of integration to $n_b = (1-\tau_a)n_a$, and multiplying both sides by $(1-\tau_a)$ we can then rewrite (82) as

$$\begin{aligned} &\mathbf{1}_x^{CU} \int_{\underline{n}(1-\tau_a)}^{\overline{n}} m_b\left(n_b\right)\left[1 - T'(y_b(n_b))\right] y_b(n_b) g_b(n_b|c) dn_b \\ &= \int_{\underline{n}(1-\tau_a)}^{\overline{n}} \left\{1 - T'(y_b(n_b))\varepsilon_{y_b}(n_b)\right\} y_b(n_b) g_b(n_b|c) dn_b \\ &+ \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} [\hat{T}_a(y_a(n_a) - \hat{T}_b(y_b(c(n_a)))](1-\tau_a) n_a f(n_a, (1-\tau_a)n_a) dn_a \end{aligned} \tag{83}$$

where we used the fact that, for any $n_a \in \left(\underline{n}, \frac{\overline{n}}{1-\tau_a}\right)$

$$\varepsilon_{y_b}(c(n_a)) \equiv \frac{y_b'\left(c(n_a)\right) c(n_a)}{y_b(c(n_a))} = \frac{y_a'\left(n_a\right) n_a}{y_a(n_a)}$$

Using the definition of "welfare effect", "revenue collection effect", and "migration effects" we have that (83) can be rewritten as

$$\lim_{n_b \to \bar{n}} \left\{\mathbf{1}_x^{CU} \cdot W_b\left(n_b\right) - R_b\left(n_b\right)\right\} = \lim_{n_a \to \bar{n}\frac{1-\tau_b}{1-\tau_a}} M_a\left(n_a\right),$$

where the functionals above are evaluated at the threshold function $c(n_a) = (1-\tau_a)n_a$.

Finally, after reintroducing τ_b by eliminating the normalization to $\tau_b = 0$, and replacing $\frac{1-\tau_a}{1-\tau_b} n_a$ for $(1-\tau_a)n_a$, we obtain the formula in the proposition. Q.E.D.

References

[1] Akerlof, G. (1978), The economics of "tagging" as applied to the optimal income tax, welfare programs and manpower planning, *The American Economic Review*, 68, 8-19.

[2] Atkinson, A. and J. Stiglitz (1976), The design of the tax structure: direct versus indirect taxation, *Journal of Public Economics*, 6, 55-75.

[3] Boadway, R. (2012), From Optimal Tax Theory to Tax Policy: Retrospective and Prospective Views, The Munich Lectures, MIT Press.

[4] Choné, P. and G. Laroque (2010), Negative Marginal Tax Rates and Heterogeneity, *American Economic Review*, 100, 2532-47.

[5] Cremer H., F. Gahvari and J.-M. Lozachmeur (2010), Tagging and income taxation: theory and an application, *American Economic Journal: Economic Policy*, 2, 31-50.

[6] Diamond, P. (1998), Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Tax Rates, *American Economic Review*, 88, 83-95.

[7] Diamond, P. and J. Mirrlees (1971), Optimal taxation and public production 1: production efficiency, *American Economic Review* 61, 8-27.

[8] Golosov M., M. Troshkin, A. Tsyvinski, and M. Weinzierl (2013), Preference heterogeneity and optimal capital income taxation, *Journal of Public Economics*, 97, 160–175.

[9] Hammond, P. (2000), Reassessing the Diamond/Mirrlees Efficiency Theorem, in *Incentives, Organization, and Public Economics: Papers in Honour of Sir James Mirrlees*, Oxford University Press.

[10] Hamilton, J. and P. Pestieau (2005), Optimal Income Taxation and the Ability Distribution: Implications for Migration Equilibria, *International Tax and Public Finance*, 12, 29-45.

[11] Heckman, J. and B. Honore (1990), The empirical content of the Roy model, *Econometrica*, 42, 1121-1149.

[12] Mankiw G. and M. Weinzierl (2010), The Optimal Taxation of Height: A Case Study in Utilitarian Income Redistribution, *American Economic Journal: Economic Policy*, 2, 155-176.

[13] Milgrom, P. (2004), Putting Auction Theory to Work, Cambridge University Press.

[14] Milgrom, P. and I. Segal (2002), Envelope Theorems for Arbitrary Choice Sets, *Econometrica*, 70, 583-601.

[15] Mirrlees, J. (1971), An exploration in the theory of optimum income taxation, *Review of Economic Studies*, 38, 175-208.

[16] Naito, H. (1999), Re-examination of uniform commodity taxes under non-linear income tax system and its implication for production efficiency, *Journal of Public Economics*, 71, 165-188.

[17] Rochet, J.-C. and P. Choné (1998), Ironing, Sweeping and Multidimensional Screening, *Econometrica*, 66, 783-826.

[18] Roy, A. (1951), Some thoughts on the distribution of earnings, *Oxford Economic Papers*, 3, 135-146.

[19] Saez, E. (2001), Using elasticities to derive optimal tax rates, *Review of Economic Studies*, 68, 205-229.

[20] Saez, E. (2002), Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses, *Quarterly Journal of Economics*, 117, 1039-1073.

[21] Saez, E. (2004), Direct or indirect tax instruments for redistribution: short-run versus long-run, *Journal of Public Economics*, 88, 503-518.

[22] Rothschild, C. and F. Scheuer (2013), Redistributive taxation in the Roy model, *Quarterly Journal of Economics*, 128, 623-668.

[23] Rothschild, C. and F. Scheuer (2014), A Theory of Income Taxation under Multidimensional Skill Heterogeneity, *Working Paper*.

[24] Scheuer, F. (2014), Entrepreneurial Taxation with Endogenous Entry, *American Economic Journal: Economic Policy*, 6 , 126-163.