View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Toulouse Capitole Publications

WORKING PAPERS

N° 17-773

March 2017

"Differential Taxation and Occupational Choice"

Renato Gomes, Jean-Marie Lozachmeur and Alessandro Pavan

Toulouse School of Economics

Differential Taxation and Occupational Choice*

Renato Gomes[†] **Jean-Marie Lozachmeur**[‡] **Alessandro Pavan**[§]

This version: December 2016

Abstract

We develop a framework to study optimal sector-specific taxation, where each agent chooses an occupation by comparing her skill differential with the tax burden differential across sectors. Because skills are not perfectly transferable, the Diamond-Mirrlees theorem (according to which the second-best entails production efficiency) fails: social welfare can be increased by inducing some agents to join the sector in which their productivity is not the highest. At the optimum, income taxes balance the marginal losses from inter-sector migration with the marginal gains from tailoring tax schedules to the distribution of productivities in each sector ("tagging"). A calibrated model indicates that sector-specific taxation generates substantive welfare gains when skill transferability decreases with income, as it enables the government to increase average taxes on high earners with large wage premia.

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

JEL classification: C72, D62.

*For comments and useful suggestions, we thank the Editor, anonymous referees, and seminar participants at LAGV 2012 (Marseille), SED 2013 (Seoul), PET 2013 (Lisbon), PSE 2013 (Munich), Université de Rennes, TTC 2014 (Köln), Warwick, University College London, Fundacao Getulio Vargas, Paris School of Economics, Southampton, the Federal Reserve Bank of Chicago and IIESE Stockholm. We are also grateful to Pierre Bachas, Gadi Barlevy, Marco Bassetto, Felix Bierbauer, Philippe Choné, Helmut Cremer, Guy Laroque, Etienne Lehmann, Vitor Farinha Luz, Miltos Makris, Casey Rothschild and Florian Scheuer for very helpful conversations, and to Laura Doval for excellent research assistance. The usual disclaimer applies.

[†]Toulouse School of Economics, CNRS, University of Toulouse Capitole, renato.gomes@tse-fr.eu.

[‡]Toulouse School of Economics, CNRS, University of Toulouse Capitole, jean-marie.lozachmeur@tse-fr.eu.

[§]Northwestern University, alepavan@northwestern.edu.

1 Introduction

The sharp rise in income inequality experienced by advanced economies in the last forty years has spurred intense research on how skills affect remuneration in different sectors of the economy. Recent empirical contributions document large differences on how various sectors/occupations affect inequality measures. For instance, Bell and Van Reenen (2014) show that the financial sector is responsible for most of the increase in the income share held by the top one percent in the UK.[1] A recent study by Bloom et al. (2016) based on a rich data set of administrative records confirms that the increase of income inequality in the US in the last three decades is primarily driven by an increase in the dispersion of compensation across firms and occupations. In contrast, pay differences within firms have remained virtually unchanged.[2]

The dynamics of compensation across different quantiles of the income distribution also vary significantly across sectors: in finance, for instance, the increase in compensation has been primarily concentrated at the very top.[3] As documented by Kaplan and Rauh (2010) and Bakija et al (2012), occupations in real state and legal services also experienced a significant growth in top incomes, whereas top incomes in manufacturing, transportation and construction grew at a more moderate rate.

These empirical observations accord with a renewed interest on taxation policies that discriminate according to the sector where income is generated.[4] One notable example is the so called "bonus tax" in the UK, according to which the bonuses received by financial employees would be taxed at a higher rate than wage income.[5] Similar proposals, involving, for instance, a different tax treatment of CEO pay or special tax schedules for certain sectors (such as finance or manufacturing) are often debated in the US. At the root of these policy proposals is the belief that certain sectors exceedingly remunerate occupation-specific skills. This may be due to technological reasons, labor market frictions, or economic rents stemming from imperfections in competition and regulation (which are typically out of the scope of the tax authority).[6]

It is perhaps surprising that, notwithstanding their natural appeal, policies advocating for differential taxation across sectors receive little support from the optimal taxation literature. At the heart of the matter lies the celebrated theorem of Diamond and Mirrlees (1971), which shows that,

[1] For similar evidence for the US and France, see, respectively, Bakija et al (2012) and Godechot (2012).

[2] See also Guvenen and Kuruscu (2012).

[3] See Philippon and Reshef (2012) for evidence for the US, and Denk (2015) for 18 European countries.

[4] Many countries engage in such policies. Algeria, for example, levies a corporate income tax of 25% on trade activities and services, and 19% on manufacturing, construction and tourism. Other countries employing sector-specific income taxes include Morocco, Tunisia, Luxembourg and Israel. The use of sector-specific sales taxes is even more common; France is a notable example for its very fine sectoral classification (resulting in many tax regimes).

[5] The bonus tax was levied in the 2009/10 tax year in the UK. This tax made employers in the financial sector pay a 50% tax rate for any (pay-for-performance) bonus paid to employees in excess of 25000 pounds. Proposed as a one-off event in the aftermath of the financial crisis, this tax collected 2.3 billion pounds.

[6] See Philippon and Reshef (2012) for a discussion of this point centered in the US financial sector.

when the government can levy differentiated (possibly nonlinear) taxes on all factors (input and output), the economy should lie at the production efficient frontier. Strikingly, at the optimum, distortions in consumption induced by income taxation do not translate into distortions in production. This result has important implications for the design of tax systems. For instance, it provides an intellectual justification for opposing the taxation of intermediate goods, as well as for the use of differential sales taxes, or sector-specific income tax regimes. Differential taxation would create a wedge between productivities and wages across sectors, thus leading to distortions in the allocation of labor across sectors and undesirable violations of production efficiency.

One key assumption of the Diamond-Mirrlees theorem is that skills are perfectly transferable across sectors. In this paper, we provide a framework for studying optimal differential taxation in settings where the degree of skill transferability is heterogeneous across individuals and occupations.[7] As our results show, this realistic feature has important implications for the optimality of production efficiency, and the design of income tax schedules.

Our analysis embeds a Mirrleesian taxation problem into an occupational choice model à la Roy (1951), where the agents' productivities are 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 in 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 they are most productive. The government wishes to implement a second best redistributive tax system à la Mirrlees (1971) using a rich set of sector-specific taxes. We allow the government's objective to be Ralwsian or concave Utilitarian.

We start with the general case in which the government can use sector-specific non-linear income tax schedules. The government can observe the income and the sector chosen by each individual, but cannot control the individual's choice of labor supply or of 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 occupational choice 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 to develop 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

[7]There is diverse empirical evidence documenting that skill transferability across sectors decreases with income, and varies from sector to sector. See, for example, Bakija et al (2012) and Dent (2015), among others.

determines the 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.[8]

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 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 turn, the solution to the dual problem delivers an Euler equation that determines the optimal allocation of workers across sectors. At the optimum, the marginal loss in tax revenue due to the migration of workers across sectors equalizes the marginal gains from tailoring tax schedules to the distribution of productivities in each sector ("tagging"). Importantly, when skills are imperfectly transferable across sectors and income taxes are sector-specific, the Diamond-Mirrlees theorem fails: Social welfare is increased by assigning some agents to a sector different from the one in which they are most productive. A similar conclusion holds in the (perhaps more realistic) scenario where the government is not able to tax labor income using a sector-specific schedule, but can levy different sales taxes across sectors. Our analysis then 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). Therefore, the use of differential taxation across sectors is strictly needed to implement the welfare-maximizing outcome.

The key to these results lies in how the tax system affects the informational costs of redistribution. Differential taxes allow the government to relax the incentive constraints of high-ability agents whose skills are poorly transferable, at the cost of allocating some low-ability agents to a sector different from the one in which they are most productive. In particular, at the production efficient outcome, by appropriately increasing taxes in some sector, the planner can obtain a first-order reduction in the informational costs of redistribution by incurring only second-order losses in total output.

The trade-off between reducing the informational rents of high-earners and inducing skill misallocation among low-earners is key to assessing which sector should be favored at the optimum. Our analysis identifies two independent (but related) conditions guaranteeing that one sector (let

[8]In Section 6, we discuss how this methodology can be applied to other settings, such as nonlinear pricing by a multi-product monopolist and managerial compensation with multiple career options.

us say, sector a) should be favored at the optimum. To describe these conditions, let us identify the *degree of skill transferability* of a worker with the loss of wage per hour as he moves away from his most productive sector. Similarly, we say that sector a is more *skill intensive* than sector b if the former sector has relatively more high-earners than the latter, under the assumption that workers choose the sector where they are most productive (i.e., production efficiency prevails).

Let sector a be the sector in which the degree of skill transferability is the lowest among low-productivity workers and the highest among high-productivity ones, and assume both sectors are equally skill-intensive. At the optimum, taxes in sector a should be lower than in sector b, inducing workers to migrate from the latter sector to the former. Intuitively, tilting the tax system in favor of sector a (by making the tax burden in sector b heavier than in sector a) entails (i) a lower opportunity cost from the migration of low-productivity agents away from the sector in which they are most productive, and (ii) higher gains in tax collection from high-ability agents remaining in the sector in which they are most productive. This is the *skill transferability motive* for differential taxation.

Alternatively, let both sectors have the same degree of skill transferability at all productivity levels, but assume that sector b is more skill-intensive than sector a. Once again, at the optimum, taxes should be higher in sector b than in sector a, inducing certain workers to migrate from the former sector to the latter. Intuitively, favoring sector a rather than b entails (i) a lower mass of low-productivity agents migrating to the sector in which they are least productive, and (ii) a higher volume of high-productivity agents paying larger taxes in the sector in which they are most productive. This is the *skill intensity motive* for differential taxation.

We quantitatively assess the effects discussed above by calibrating our model using data on wages and industry classification from the US Current Population Survey (CPS). To generate conservative estimates on the gains from differential taxation, we construct two large sectors: *manufacturing* and *services*. The manufacturing sector aggregates traditional industries, while the services sector contains finance, banking, legal and business services, as well as technology-intensive activities characterized by large returns to occupation-specific skills. We interpret the wage data as generated by a (sub-optimal) uniform tax system where production efficiency prevails, and construct different scenarios regarding the degree of skill transferability in the manufacturing and services sectors.

Our analysis delivers three main lessons: First, the welfare gains from differential taxation can be large (of the order of 1.5% of GDP). Moreover, most of these gains are due to the skill transferability motive: Accordingly, the services sector (displaying the lowest degree of skill transferability at the top) faces the largest tax burden. With a more complex tax system and a finer sectorial classification (involving more than two sectors), the importance of the skill intensity motive for differential taxation is expected to be higher.

Second, sales taxes (or, equivalently, payroll taxes) are able to generate roughly half of the welfare gains from sector-specific income tax schedules. This result suggests that the welfare gains from "simple" tax systems, while non-negligible, are far from the levels delivered by fully optimal sector-specific income taxation.

Third, we document the incidence of production inefficiencies and the shape of optimal marginal tax schedules. We show that differential taxation can increase significantly the average tax rate faced by high earners in occupations with large wage premia.

The rest of the paper is organized as follows. Below, we close the introduction by briefly reviewing the most pertinent literature. Section 2 previews the main themes of our analysis through a simple discrete-type example. Section 3 presents the continuum-type version of our model. Section 4 characterizes the optimal tax system under differential taxation, and discusses implications for production efficiency. Section 5 quantifies these results by means of a calibration exercise. Section 6 discusses a few extensions and concludes.

1.1 Related literature

Our paper contributes to the literature on optimal taxation 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.[9] In turn, Atkinson and Stiglitz (1976) show that differentiated sales taxes across goods are detrimental to welfare, when the government can use a non-linear income tax schedule and preferences are weakly separable between consumption and leisure.[10] 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 all 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 *within* occupation (i.e., hours of work).

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

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

Rothschild and Scheuer (2013) consider a two-sector Roy model with endogenous wages (as workers are paid their marginal product of labor in a constant-returns technology) and assume that taxation is uniform across sectors (as the income tax schedule is the same across sectors and sales taxes are not considered). In turn, Rothschild and Scheuer (2016) consider an economy where agents can work in a traditional sector (where private and social returns coincide) or in a rent-seeking sector (which imposes a negative externality on the traditional sector). The theme of these papers is how general equilibrium effects (determining relative wages), or externalities across sectors, shape the optimal tax system in a world with cross-sector migration and imperfect tagging (uniform taxation). Ales et al. (2015) consider an economy with a continuum of sectors, but where agents are endowed with a one-dimensional productivity type. They restrict attention to uniform taxation and simulate their model to assess the impact of technical change (which affects relative wages across sectors) on the optimal income tax schedule. In contrast to these contributions, our model allows for sector-specific taxation (in the form of income or sales taxes), but abstracts from general equilibrium effects (as technology is linear in our model) and externalities across sectors.

Another related contribution is Scheuer (2014), who considers an economy where agents have one-dimensional skills and choose between being workers or entrepreneurs (with the latter choice involving a setup cost that enters additively in the agents' utility function). This simple structure of heterogeneity implies that production efficiency is optimal when the government can tax the incomes from wages and profits differently. This feature eliminates the tension between "tagging" gains and inefficiency losses that is at the core of our work.[11]

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).

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 the references therein). In our setting, the multi-dimensionality of workers' productivity only affects sector-choice (extensive-margin) decisions. This allows us to employ the primal-dual approach described above, bringing considerable tractability to the analysis.

As our analysis reveals, the multi-dimensionality of workers' types has important implications

[11] See also Scheuer and Werning (2016) for the analysis of optimal taxation in economies with extensive and intensive margins as well as for a discussion of how the Mirrlees (1971) problem can be recast as a special case of Diamond and Mirrlees (1971). In this paper, as well as in all other papers cited above, taxation is uniform across sectors.

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 is responsible for participation in the labor force. 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. Jacquet and Lehmann (2016) study optimal income taxation when agents are heterogeneous in their skills and behavioral elasticities.

2 Preamble: An Illustrative Example

In order to introduce the main ideas in the simplest possible way, this section studies differential taxation in a stylized discrete-type example. Consider a unit-mass continuum of agents and two sectors indexed by $j \in \{a, b\}$. We identify the type of each agent with the pair (n_a, n_b) describing the agent's productivity in each of the two sectors. The utility of an agent with type (n_a, n_b) working h_j hours in sector j and paying t dollars in taxes is $h_j n_j - t - \psi(h_j)$, where $\psi(h)$ is the disutility of labor (which, in this example, we assume to be quadratic, i.e., $\psi(h) = h^2/2$).

The agents' types are independently drawn from a distribution with probability mass function f satisfying

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

where $0 < \underline{n} - \varepsilon_j \leq \underline{n} < \overline{n} - \delta_j \leq \bar{n}$ for $j \in \{a, b\}$. Figure 1 depicts the support of f. That is, an agent with sector-j productivity equal to $\underline{n}$ (alternatively, $\bar{n}$) loses ε_j (alternatively, δ_j) dollars per hour if he works in sector $k \neq j$. We refer to $-\varepsilon_j$ (alternatively, $-\delta_j$) as the *degree of skill transferability* among low-productivity (alternatively, high-productivity) agents in sector j. For convenience, we assume that the two sectors have equal sizes in case all agents work in the sector in which their productivity is the highest, i.e., $p_j + q_j = \frac{1}{2}$ for $j = a, b$. We refer to q_j/p_j as the *skill intensity* of sector j, that is, the ratio between high and low productivity agents in sector j that obtains when all agents choose the sector in which their productivity is the highest.

The social planner designs a budget-balanced income tax system to maximize the utility of the worst-off agent in the economy (i.e., the planner's objective function is Ralwsian). Agents compare productivity levels and the tax burden across sectors and then choose (a) which sector to work in and (b) the number of hours to supply in the chosen sector.

Foreshadowing the analysis in the next sections, we shall proceed in two steps. In the first step, we fix which types work in each sector (i.e., the occupational choice rule) and compute the tax system that maximizes social welfare among all tax systems that induce agents to sort themselves over the two sectors according to the given occupational choice rule. In the second step, we compare welfare across occupational choice rules.

Let us consider first the occupational choice rule according to which types $(\underline{n}, \underline{n} - \varepsilon_a)$ and $(\bar{n}, \bar{n} - \delta_a)$ work in sector a and types $(\underline{n} - \varepsilon_b, \underline{n})$ and $(\bar{n} - \delta_b, \bar{n})$ work in sector b. This rule respects *production efficiency*, as the labor supply of each agent is employed in the sector in which the agent is most productive.

Under the optimal tax system satisfying production efficiency, the "high" types $(\bar{n}, \bar{n} - \delta_a)$ and $(\bar{n} - \delta_b, \bar{n})$ supply labor at the first-best level. The government's ability to tax these individuals is constrained by their ability to mimic the "low" types $(\underline{n}, \underline{n} - \varepsilon_a)$ and $(\underline{n} - \varepsilon_b, \underline{n})$. In the typical case where the only binding incentive constraints are the ones regarding the provision of labor supply within each sector, a sector-j high type obtains the informational rent

$$u_j(\bar{n}) = u_j(\underline{n}) + \psi\left(h_j(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n}} h_j(\underline{n})\right), \tag{1}$$

where the schedules $u_j(\cdot)$ and $h_j(\cdot)$ describe the indirect utility and the labor supply of individuals working in sector $j = a, b$. This rent originates in the ability of the most productive agents to generate the same income as the least productive ones by working less, thus economizing on the disutility of labor. As a result, the rent for these most productive agents equals the utility of the least productive agents working in the same sector, augmented by a term that equals the differential in the disutility of labor from generating the same income as the least productive agents. In order to reduce these rents and foster redistribution from the most productive agents to the least productive ones, the government taxes the labor income of the least productive agents so as to induce them to work less. At the optimum, the government thus distorts the labor supplied by the least productive agents downwards relative to the first-best level. We denote by Φ_e the social welfare achieved by the optimal tax system under production efficiency.

Let us now consider the occupational choice rule according to which types $(\underline{n}, \underline{n} - \varepsilon_a)$, $(\underline{n} - \varepsilon_b, \underline{n})$ and $(\bar{n}, \bar{n} - \delta_a)$ work in sector a, and type $(\bar{n} - \delta_b, \bar{n})$ works in sector b. A tax system implementing such a rule is said to *favor sector a* (see Figure 1 for an illustration).[12]

Relative to the production efficiency benchmark, this occupational choice rule moves type $(\underline{n} - \varepsilon_b, \underline{n})$ from sectors b to a. While this assignment entails an opportunity cost of ε_b per hour of work (which corresponds to the productivity loss from having this type working in the "wrong" sector), it allows the government to increase tax collection from high-productivity agents working in sector b (whose type is $(\bar{n} - \delta_b, \bar{n})$).

To understand why, note that, at the Ralwsian optimum, type $(\bar{n} - \delta_b, \bar{n})$ has to be indifferent between (i) working on sector b with productivity $\bar{n}$, and (ii) migrating to sector a and working with productivity $\bar{n} - \delta_b$. In case he decides to migrate, the best that this type can do is to mimic agents with productivity $\underline{n}$. The optimal tax system leaves type $(\bar{n} - \delta_b, \bar{n})$ perfectly indifferent between these two options:

$$u_b(\bar{n}) = u_a(\underline{n}) + \psi\left(h_a(\underline{n})\right) - \psi\left(\frac{\underline{n}}{\bar{n} - \delta_b} h_a(\underline{n})\right). \tag{2}$$

[12]Analogously, we say that a tax system favors sector b when the only type to work in sector a is $(\bar{n}, \bar{n} - \delta_a)$.

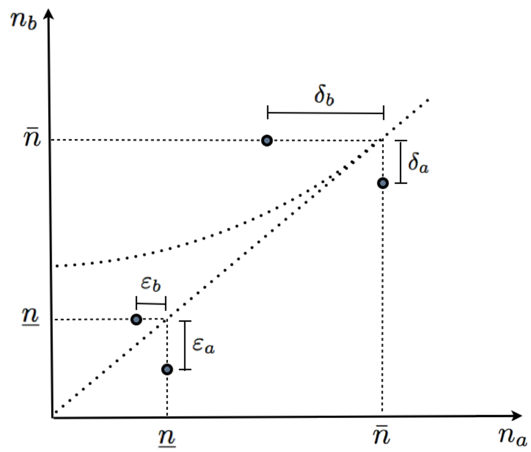

Figure 1: The shaded points describe the support of agent's productivity pairs. Under production efficiency, types above the 45-degree line work in sector b, while those below it work in sector a. When the tax system favors sector a, all types below the dotted curve work in sector a.

As the comparison with equation (1) reveals, the ability to mimic agents whose productivity is $\underline{n}$ is hindered by the fact that skill is not perfectly transferable across sectors (i.e., $\delta_b > 0$). As a result, the government is able to levy higher taxes from type $(\bar{n} - \delta_b, \bar{n})$ than under production efficiency (and the more so the higher is δ_b).

Of course, this reduction in the informational costs of redistribution comes at the cost of misallocating the hours of work provided by type $(\underline{n} - \varepsilon_b, \underline{n})$. Moreover, this misallocation cost is decreasing in $-\varepsilon_b$, which is the degree of skill transferability among low-productivity agents in sector b. Denoting by Φ_j the social welfare achieved by the optimal tax system that favors sector j and recalling that Φ_e is social welfare under the optimal tax system consistent with production efficiency, we then have the following result:

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

The proof for both this result and the next one are in the Supplementary Material. Intuitively, the optimal tax system exhibits production inefficiency whenever favoring some sector (and therefore reducing the informational costs of redistribution) entails low costs in terms of skill misallocation. As our analysis in the next sections reveals, the analog of this condition in the continuum-type case has little bite, and production inefficiency is a robust (or "generic") feature of the second-best.

In case production inefficiency prevails, which sector should be favored? The trade-off discussed above is key to answering this question. To see why, let both sectors be equally skill intensive, but assume that sector a simultaneously exhibits the lowest degree of skill transferability among low-productivity agents (i.e., $\varepsilon_a > \varepsilon_b$) and the highest among high-productivity agents (i.e., $\delta_a < \delta_b$). In this case, welfare is greater by inducing low-productivity agents in sector b to migrate to sector a, relative to the opposite migration pattern. The reason is that favoring sector a, rather than b, entails (i) lower opportunity costs of skill misallocation for each low-productivity agent that migrates, and (ii) higher gains in tax collection from each high-ability agent that stays in the unfavored sector. This is the skill transferability motive for differential taxation.

Another possibility is that both sectors enjoy the same degree of skill transferability for all productivity levels, but a sector, let us say a, is less skill-intensive than the other ($\frac{q_a}{p_a} < \frac{q_b}{p_b}$). In this case, taxes should be lower in sector a than in sector b. The reason is that favoring sector a, as opposed to favoring b, entails (i) a lower mass of low-productivity agents migrating to their least-productive sector, and (ii) a higher mass of high-productivity agents paying large taxes in their most productive sector. This is the skill intensity motive for differential taxation.

This discussion is summarized in the next result.

Result 2 ***(Sectorial Bias)*** *Social welfare is higher by favoring sector j rather than sector $k \neq j$, that is, $\Phi_j > \Phi_k$, whenever one of the following mutually exclusive conditions hold:*

1. *Sector j enjoys the lowest degree of skill transferability among low-productivity agents and the highest among high-productivity agents (that is, $\varepsilon_j > \varepsilon_k$ and $\delta_j < \delta_k$), and both sectors are equally skill-intensive (that is, $\frac{q_j}{p_j} = \frac{q_k}{p_k}$).*

2. *Sector j is less skill-intensive than sector k (that is, $\frac{q_j}{p_j} < \frac{q_k}{p_k}$), and both sectors enjoy the same degree of skill transferability among low and high productivity agents (that is, $\varepsilon_j = \varepsilon_k$ and $\delta_j = \delta_k$).*

Of course, the case for favoring sector j is only strengthened if the skill transferability motive ($\varepsilon_j > \varepsilon_k$ and $\delta_j < \delta_k$) and the skill intensity motive ($\frac{q_j}{p_j} < \frac{q_k}{p_k}$) are present simultaneously. By contrast, it is a priori unclear which sector should be favored if the degree of skill transferability is uniformly greater in some sector (e.g., $\varepsilon_a > \varepsilon_b$ and $\delta_a > \delta_b$), or if the skill transferability motive and the skill intensity motive point in different directions (e.g., $\varepsilon_j > \varepsilon_k$ and $\delta_j < \delta_k$, but $\frac{q_j}{p_j} > \frac{q_k}{p_k}$). In these cases, the welfare level attained from favoring either sector depends on the magnitudes of the "informational cost" and "skill misallocation" effects discussed above, and a quantitative analysis is needed to assess the sectorial bias at the optimum. This often occurs in the continuum-type case, where the bivariate distribution of skills is unlikely to satisfy stringent order relations across all productivity levels. We shall come back to this important issue in Section 5, where we simulate our model using US income data. Before doing so, we first extend the model to a continuum of types.

3 Model and Preliminaries

3.1 Set-up

We consider an economy with a unit-mass continuum of agents and two sectors indexed by $j \in \{a, b\}$.[13] 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 density f_j). The conditional distributions are denoted by $F_{j|k}$, for $j, k \in \{a, b\}$, $j \neq k$ (with bounded density $f_{j|k}$).

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 $T_j(y_j)$. For simplicity, we assume that 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 given by

$$w_j h_j n_j - T_j(w_j h_j n_j) - \psi\left(h_j\right), \tag{3}$$

where $\psi\left(h\right)$ is the disutility of labor, which we assume takes the isoelastic form $\psi\left(h\right) = h^{\frac{1}{\xi}}$, with $\xi \in (0, 1)$. The elasticity of labor supply with respect to wages is then equal to $\xi/(1-\xi)$, 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

$$\pi_j = (1 - w_j - \tau_j) L_j, \tag{4}$$

where τ_j is the sales tax rate on good j.[14] The wage rates $\mathbf{w} \equiv (w_a, w_b)$, the agents' labor supply, and the labor demand from the representative firms are all simultaneously determined in equilibrium, as explained below.

[13] In the Supplementary Material we extend the analysis model to an arbitrary (finite) number of sectors.

[14] In an alternative interpretation, τ_j is a payroll tax levied on employers on sector j. For consistency, we shall favor the sales tax interpretation.

3.2 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 by the agents working in sector j as a function of their sector-j productivity, with the domain N_j of each function h_j denoting the set of productivity levels of those agents working in sector j.[15] For future reference, for any set $N_j \subset N$, we denote by $\bar{N}_j$ the closure of the set, with $\bar{N} = [\underline{n}, \overline{n}]$.

Hereafter we will refer to an *allocation* as a triple $(\mathcal{C}, h_a, h_b)$. 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 four 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

$$\tilde{u}_j(n_j) \equiv \max_h \ \{w_j h n_j - T_j(w_j h n_j) - \psi(h)\} \quad \text{for all} \quad n_j \in N, \tag{5}$$

and

$$u_j(n_j) \equiv w_j h_j(n_j) n_j - T_j(w_j h_j(n_j) n_j) - \psi(h_j(n_j)) \quad \text{for all} \quad n_j \in N_j. \tag{6}$$

This condition then requires that $u_j(n_j) = \tilde{u}_j(n_j)$ 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 (5), which has to be satisfied at any interior point where the schedule T_j is differentiable:

$$w_j n_j [1 - T'_j(w_j h_j(n_j) n_j)] = \psi'(h_j(n_j)) \tag{7}$$

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 $k \neq j$:

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

[15] Because, if an agent works in sector j, his productivity in sector $k \neq j$ does not affect his utility function, the schedules h_j do not depend on n_k for $k \neq j$.

Finally, the forth and final 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 (4). We incorporate the above four 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.,*

$$\sum_{j\in\{a,b\}} \int_{\{\mathbf{n}:\mathcal{C}(\mathbf{n})=j\}} \left(T_j(w_j h_j(n_j) n_j) + \tau_j h_j(n_j) n_j\right) dF(n_a, n_b) \geq B, \tag{8}$$

where B *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

$$U(\mathbf{n}; \mathcal{E}) \equiv u_{\mathcal{C}(\mathbf{n})}(n_{\mathcal{C}(\mathbf{n})}) = \max_{j\in\{a,b\}} \left\{\tilde{u}_j(n_j)\right\},$$

where the schedules $\tilde{u}_j$ and u_j are given by (5) and (6), respectively.

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

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

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

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

where ϕ is a strictly increasing and weakly concave function reflecting the government preferences for redistribution.

We will use the index $x = R$ (alternatively, $x = CU$) to refer to the Ralwsian (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 $x = CU$.

3.3 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 $\mathcal{T}$ only if the following conditions jointly hold:*

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

2. *For every $j \in \{a, b\}$, the income schedule $y_j(n_j) \equiv w_j h_j(n_j) n_j$ is nondecreasing over N_j. Moreover, the indirect utility schedule $u_j(n_j)$ is Lipschitz continuous over N_j with derivative equal to*
$$u_j'(n_j) = \psi'\left(h_j(n_j)\right) \frac{h_j(n_j)}{n_j} \quad \text{for almost every } n_j \in N_j. \tag{9}$$

3. *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 $n_b > c(n_a)$.[16] 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 $u_a(n_a) = u_b(c(n_a))$.*

Conversely, suppose the allocation $(\mathcal{C}, h_a, h_b)$, along with the wage rates $\mathbf{w}$ and the tax system $\mathcal{T}$, satisfy the properties in parts 1-3 above. Then there exists a tax system $\mathcal{T}'$ such that the allocation $(\mathcal{C}, h_a, h_b)$ is implemented at the wage rates $\mathbf{w}$ by the tax system $\mathcal{T}'$.

Part 1 shows that, 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 to the firm. Accordingly, sales taxes affect equilibrium wages in a one-to-one fashion. Part 2 is the standard characterization of incentive compatibility on the intensive margin. The envelope condition (9) relates the agents' indirect utilities to their utility-maximizing labor supply in each sector.

Part 3, in turn, offers a convenient characterization of the extensive margin of occupational choice. It establishes that any occupational choice rule can be described by a continuous and nondecreasing 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. We refer to the graph of c as the *locus of indifferent types*. Because the payoff from working in a given sector strictly increases with the agent's productivity in that sector, the threshold $c(n_a)$ is strictly increasing at any interior point (i.e., at any point where $c(n_a) \in N$). This function is instrumental in describing the distribution of productivities in each sector of the economy.

[16] Recall that $\bar{N}$ denotes the closure of the set N, i.e., $\bar{N} = [\underline{n}, \bar{n}]$.

3.4 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 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.[17] It is also convenient to define the threshold $n_a' \in \bar{N}$ such that $c(n_a) > \underline{n}$ if and only if $n_a > n_a'$.[18] 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 $n_a > n_a''$.

For such an admissible rule, what is the mass of agents working in sector a whose productivity does not exceed n_a? As illustrated in Figure 2, this mass corresponds to the probability of the shaded area *below* the locus of indifferent types and to the *left of* n_a. This is given by

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

with density $g_a(n_a|c) \equiv f_a(n_a)F_{b|a}(c(n_a)|n_a)$. An analogous expression determines the mass of agents working in sector b whose productivity does not exceed n_b, denoted by $G_b(n_b|c)$. When we evaluate this function at $n_b = c(n_a)$, we obtain the probability of the shaded area to the *left of* the locus of indifferent types and *below* $c(n_a)$, as illustrated in Figure 2. The density of $G_b(n_b|c)$ is denoted by $g_b(n_b|c)$.

3.5 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

[17] 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.

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

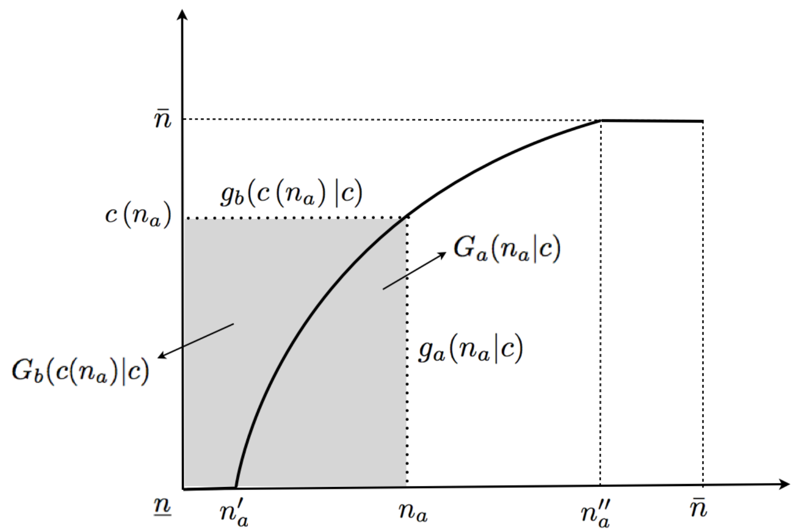

Figure 2: The threshold function and its induced distributions of productivities. The shaded area corresponds to the set of types whose sector-a productivity is smaller or equal to n_a and whose sector-b productivity is smaller or equal to $c(n_a)$. The dotted lines illustrate the sets of types associated with the densities $g_a(n_a|c)$ and $g_b(c(n_a)|c)$, respectively.

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 $\mathcal{P}_2^x$.

3.6 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 agent (as specified by the equilibrium $\mathcal{E}$), there exists no reallocation of agents 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;[19] 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

[19] For a textbook treatment, see chapter 5 of Mas-Colell, Whinston and Green (1995).

supply (i.e., hours of work) of each individual and changing his occupation.

4 Optimal Differential Taxation

We study first optimal differential taxation 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 1 ***(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 $(\mathcal{C}, h_a, h_b)$ and producing the same payoffs under $\hat{\mathcal{E}}$ such that, for $j = a, b$,*

1. *the income tax schedules satisfy $\hat{T}_j(y) = \tau_j y + T_j((1 - \tau_j) y)$,*
2. *sales taxes and wages are given by $\hat{\tau}_j = 0$ and $\hat{w}_j = 1$.*

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 in designing 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 optimality to consider taxation equilibria where $\tau_a = \tau_b = 0$. To lighten notation, 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 $n_j \in N_j$).

4.1 Optimal marginal tax rates

Ley λ denote the multiplier associated with the government budget constraint (8) and 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 an 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 $n_a \geq n_a'$:*

$$\underbrace{\xi \frac{T_a'(y_a(n_a))}{1-T_a'(y_a(n_a))} n_a g_a(n_a|c)}_{E_a(n_a)} \quad + \quad \underbrace{\xi \frac{T_b'(y_b\left(c(n_a)\right))}{1-T_b'(y_b(c(n_a)))} c(n_a) g_b(c(n_a)|c) \mathbf{1}\left\{n_a \leq n_a''\right\}}_{E_b(c(n_a))}$$

$$= \underbrace{\int_{n_a}^{\overline{n}} \left[1 - \mathbf{1}_x^{CU} m_a\left(\tilde{n}_a\right)\right] dF\left(\tilde{n}_a, c(\tilde{n}_a)\right),}_{D(n_a)} \tag{10}$$

together with the occupational choice constraint

$$c'\left(n_a\right) = \frac{h_a\left(n_a\right)\left[1 - T_a'\left(y_a\left(n_a\right)\right)\right]}{h_b\left(c\left(n_a\right)\right)\left[1 - T_b'\left(y_b\left(c\left(n_a\right)\right)\right)\right]} \tag{11}$$

for all $n_a \in (n_a', n_a'')$.

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 $y_b(c(n_a))$, for some $n_a \in (n_a', n_a'')$. This perturbation has two effects. The first one is the "direct effect", $D(n_a)$, represented by the integral on the right-hand-side of (10). 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 Ralwsian, 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 $c(n_a)$ (which is, simply, $1 - F\left(n_a, c(n_a)\right)$). 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 $m_j\left(n_j\right) = \phi'\left(u_j(n_j)\right)/\lambda$.

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 (10). 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.[20] Figure 3 illustrates the effects discussed above by indicating the sets of agents affected by each of these effects.

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

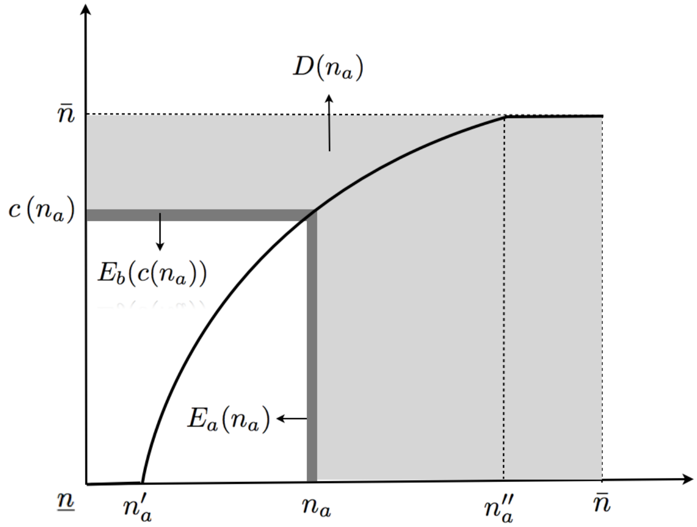

Figure 3: The sets of types affected by the elasticity and direct effects from the generalized Mirrlees formula (10).

In turn, the incentive compatibility constraint (11) describes how marginal taxes relate to the slope of the threshold function c. For agents with types $(n_a, c(n_a)) \in \mathbf{N}$ to remain indifferent as to which sector to work in, the ratio of marginal net incomes (with respect to productivity) across sectors has to equal $c'\,(n_a)$, as implied by the characterization of Lemma 1. Combined with this condition, the generalized Mirrlees formula (10) then determines 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 $\mathcal{C}$.

4.2 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 $\varepsilon_{y_b}(n_b) \equiv y_b'(n_b)n_b/y_b(n_b)$ 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 $n_a \in (n'_a, n''_a)$:*

$$\mathbf{1}_x^{CU} W_b(c(n_a)) = R_b(c(n_a)) + M_a(n_a) + \underbrace{E_b(c(n_a))(1 - T_b'(y_b(c(n_a)))y_b(c(n_a))}_{\textit{continuity correction: } \Delta_b\,(c(n_a))} \tag{12}$$

where

$$W_b(c(n_a)) \equiv \int_{c(n'_a)}^{c(n_a)} m_b\,(n_b)\left[1 - T_b'(y_b(n_b))\right] y_b(n_b) dG_b(n_b|c) \tag{13}$$

is the "welfare effect",[21]

$$R_b(c(n_a)) \equiv \int_{c(n_a')}^{c(n_a)} \left[1 - T_b'(y_b(n_b))\varepsilon_{y_b}(n_b)\right] y_b(n_b) dG_b(n_b|c) \tag{14}$$

is the "revenue collection effect",

$$M_a(n_a) \equiv \int_{n_a'}^{n_a} \left[T_a(y_a(\tilde{n}_a)) - T_b(y_b(c(\tilde{n}_a)))\right] c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)) d\tilde{n}_a \tag{15}$$

is the "migration effect", and $E_b(c(n_a))$ is the elasticity effect defined in (10).

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 (12).

Heuristic Derivation of the Euler Equation. To understand the Euler equation (12), 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 agents' 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 (for short, an α-reform) *applied to all income levels up to* $y_b(c(n_a)) = c(n_a)h_b(c(n_a))$, for some $n_a \in (n_a', n_a'')$, implies the following effective tax schedule in sector b:

$$T_b^\alpha(y) \equiv \begin{cases} \alpha y + T_b\left((1-\alpha)y\right) & \text{if} \quad y < y_b(c(n_a)) \\ T_b\left(y\right) & \text{if} \quad y \geq y_b(c(n_a)). \end{cases} \tag{16}$$

Now, let $(\mathcal{C}, h_b, \mathcal{T})$ be a solution to the dual 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 $\mathcal{C}(\underline{n}, \underline{n}) = a$.[22] 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 $y_b(c(n_a)) = c(n_a)h_b(c(n_a))$, for some $n_a \in (n_a', n_a'')$. Under the effective tax schedule T_b^α, the utility that any agent with sector-b productivity equal to n_b obtains from supplying $h_b < y_b(c(n_a))/n_b$ hours of labor in sector b is equal to

$$h_b n_b - \psi(h_b) - T_b^\alpha(h_b n_b) = (1-\alpha)h_b n_b - \psi(h_b) - T_b\left((1-\alpha)h_b n_b\right). \tag{17}$$

As one can see from (17), the utility that any such agent obtains under the α-payroll-tax reform T_b^α is the same as the utility that an agent with sector-b productivity equal to $(1-\alpha)n_b$ would have obtained under the original tax schedule T_b. This implies that, for α small enough, under the

[21] Recall that $m_j(n_j) \equiv \phi'(u_j(n_j))/\lambda$ is the ratio of social marginal utility of all individuals with sector-j productivity n_j working in sector j to the marginal value of public funds for the government.

[22] Recall that this means that $n_a' = \underline{n}$. The heuristic derivation can be easily adapted for the case where $\mathcal{C}(\underline{n}, \underline{n}) = b$ (i.e., to $n_a' > \underline{n}$).

α-payroll tax reform T_b^α, the indirect utility u_b^α of each agent with sector-b productivity equal to n_b is given by[23]

$$u_b^\alpha(n_b) \equiv \begin{cases} u_b\left((1-\alpha)n_b\right) & \text{if} \quad n_b \in \left(\frac{c(n_a')}{1-\alpha}, c(n_a)\right) \\ u_b\left(n_b\right) & \text{if} \qquad n_b > c(n_a). \end{cases} \tag{18}$$

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

$$c^\alpha(\tilde{n}_a) = \frac{1}{1-\alpha} c(\tilde{n}_a) \tag{19}$$

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 (12) accounts for the gains and losses of α-payroll-tax reforms up to income level $y_b(c(n_a))$. Hereafter, we discuss each of these effects.

- **Welfare effect.** The first effect is the impact of the reform on the agents' utility. From (18), 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 agent 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 shadow cost of raising money, is given by

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

  where the equality follows from the incentive-compatibility constraint (9) along with the fact that at any point of differentiability of the tax schedule T_b, the optimal choice of labor supply must satisfy the first-order condition (7). 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 (13). In the case of a Ralwsian objective, this effect is zero, given that the effect of tax reforms on the indirect utility of all agents but the worst-off individuals is 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 (16), it is easy to see that, under the α-reform, the tax revenue collected from each agent working in sector b with productivity $n_b \in (c(n_a')/(1-\alpha), c(n_a))$ is given by

$$\begin{aligned} &\alpha n_b h_b^\alpha(n_b) + T_b\left((1-\alpha) n_b h_b^\alpha(n_b)\right) \\ &= \alpha n_b h_b((1-\alpha)n_b) + T_b\left((1-\alpha)n_b h_b((1-\alpha)n_b)\right), \end{aligned} \tag{20}$$

  where h_b^α is the sector-b labor supply schedule under T_b^α, and where the equality in (20) follows from the fact that the labor supply of each such agent under the schedule T_b^α coincides with

[23] That α is small guarantees that agents whose sector-b productivity is above $c(n_a)$ continue to prefer generating incomes $y(n_b) > y(c(n_a))$ to generating incomes $y < y(c(n_a))$ and that agents with productivity $n_b < c(n_a)$ prefer generating income $y_b((1-\alpha)n_b)$ to any other income.

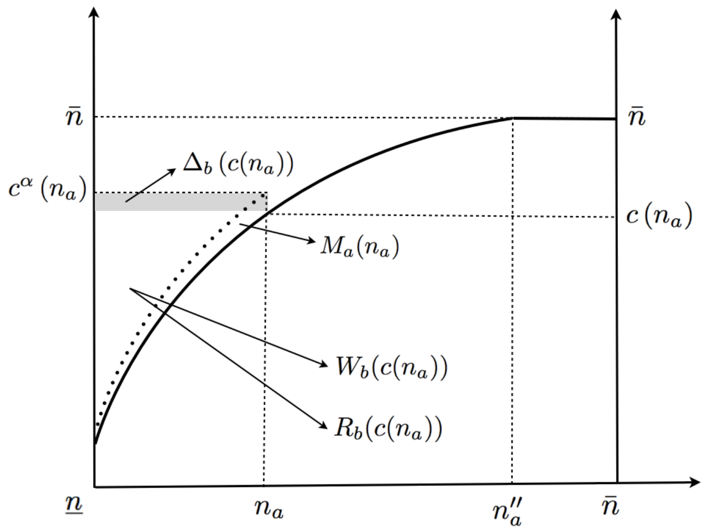

Figure 4: The figure illustrates the types affected by the welfare, revenue collection, migration, and continuity correction effects discussed in the main text. The dotted curve corresponds the occupational choice rule under the α-payroll tax reform.

the labor supply of an agent with productivity $(1-\alpha)n_b$ under the original schedule T_b. Differentiating the right-hand-side in (20) 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 $R(c(n_a))$ in the Euler equation, as defined in (14).

- **Migration effect.** The third effect accounts for the fact that agents change occupation in response to the tax reform. After differentiating equation (19) with respect to α and evaluating the derivative at $\alpha = 0$, we obtain 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

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

Integrating the above expression for all $\tilde{n}_a < n_a$ leads to the migration effect in the Euler equation, as defined in (15).

- **Continuity correction.** Finally, consider the last term in the right-hand side of the Euler equation (12), $\Delta_b\left(c(n_a)\right)$. As can be seen from equation (18), 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 $n_b \in (c(n_a')/(1-\alpha), c(n_a))$ 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 to 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 $\Delta_b\left(c(n_a)\right)$ in the right-hand side of the Euler equation (12). 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 $\left[1 - T_b'(y_b(c(n_a)))\right] y_b(c(n_a))$, as shown in the derivation of the Welfare effect.

Figure 4 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 agents. As such, taken together, these effects measure the marginal gain of better tailoring the taxation of sector-b agents 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 agents across sectors (as captured by the migration effect).

4.3 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*

$$f_a(n)F_{b|a}(n|n) = \delta f_b(n)F_{a|b}(n|n) \quad \text{for almost every} \quad n \in N. \tag{21}$$

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 (21) with $\delta = 1$). Equipped with this definition, we can state the following proposition.

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:*

$$\lim_{n_j \to \bar{n}} T'_j(y_j(n_j))g_j(n_j|c) = 0 \tag{22}$$

for $j = a, b$.

Part 1 establishes that production inefficiency is a robust feature of optimal taxation equilibria. 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.[24] 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 (12).[25]

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 (22) necessarily holds). The result thus extends familiar findings on the taxation of top earners

[24]These terms are the continuum-types analogs of the combination of the skill-transferability and and skill-intensity effects in the discrete type model of Section 2.

[25]Note that when the c.d.f. F has full support over the type space $[\underline{n}, \bar{n}]^2$, as assumed here, the continuum-type analog of the condition in Result 1 (namely, that there are low types sufficiently close to the 45-degree line) always holds. In this sense, aside from knife-edge cases such as distributions F that are symmetric around the 45-degree line, the genericity condition in Definition 3 and the condition in Result 1 in the discrete-type model coincide.

(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.

Before exploring the quantitative implications of the above characterization, we shall briefly discuss the case where sales taxes are the only instrument the government can employ to differentiate taxes across sectors. As we shall see in Section 5, this exercise is useful to inform policy-makers on what percentage of the welfare gains from differential taxation can be obtained by simple instruments such as sales taxes.

4.4 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 employees whose income comes through wages[26]), it is worth extending the above results to 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).

Consider an agent with type (n_a, n_b) facing the tax system $\mathcal{T} = \{T, T, \tau_a, \tau_b\}$, which features uniform income taxation. For this agent to be indifferent between working in one sector or the other, the maximal utility the agent can derive from working in either sector must be equalized, that is,

$$\max_h \{(1-\tau_a)hn_a - T((1-\tau_a)hn_a) - \psi(h)\} = \max_h \{(1-\tau_b)hn_b - T((1-\tau_b)hn_b) - \psi(h)\}$$

where we used the fact that $w_j = 1 - \tau_j$. As inspecting the problems above reveals, this occurs if and only if $(1-\tau_a)n_a = (1-\tau_b)n_b$. Therefore, whenever income taxation is uniform, the threshold function is linear and given by[27]

$$c(n_a) = \frac{1-\tau_a}{1-\tau_b} n_a. \tag{23}$$

The converse is also true: Whenever the occupational choice rule of an equilibrium $\mathcal{E}$ is described by a linear threshold function, there exists a taxation equilibrium featuring uniform income taxation that implements the same allocation as in $\mathcal{E}$ and yields the same utility to all agents. The reason is that, whenever the threshold function is linear (e.g., $c(n_a) = \kappa n_a$), the effective tax schedules must satisfy the relation

$$\hat{T}_a(y) = (1-\kappa)y + \hat{T}_b(\kappa y),$$

[26] See also the discussion in the Introduction about the use of sector-specific income taxation across countries.

[27] Consistently with the analysis in the previous section, we continue to assume that sector a is the one for which there exists a n''_a such that $c(n_a) = \bar{n}$ for all $n_a \geq n''_a$. This is equivalent to assuming that $\tau_a < \tau_b$.

as implied by the indifference condition $u_a(n_a) = u_b(c(n_a))$.[28] Such effective taxes can be generated, for example, by the tax system $\hat{\mathcal{T}} = \{\hat{T}_b, \hat{T}_b, 1-\kappa, 0\}$, which features uniform income taxes.

The equivalence between linear threshold functions and uniform income taxes (with possibly differentiated sales taxes) greatly simplifies the task of finding the optimal tax system. Regarding the primal problem $\mathcal{P}_1^x(\mathcal{C})$, Proposition 1 applies, as one only needs to set $c(n_a)$ according to (23). In turn, to solve the dual problem $\mathcal{P}_2^x(h_a)$, one only needs to maximize welfare over the linear coefficient that describes the threshold function. The result of this simple one-dimensional optimization problem is presented in the next proposition.

Proposition 4 ***(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*

$$\lim_{n_b \to \bar{n}} \left\{ \mathbf{1}_x^{CU} \cdot W_b(n_b) - R_b(n_b) \right\} = \lim_{n_a \to \bar{n}\frac{1-\tau_b}{1-\tau_a}} M_a(n_a) \tag{24}$$

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 (23).

The formula in (24) is closely related to the general Euler condition (12) of Proposition 2. The intuition for this formula can thus be obtained by considering α-payroll tax reforms similar to those considered above, 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 a result, (24) 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.

As in the case of sector-specific income taxes, production efficiency generically fails under x-optimal taxation equilibria. Once again, the reason is that talent is not perfectly transferable across sectors. Therefore, a small departure from production efficiency typically generates first-order gains in tax collection, while entailing only second-order losses in production. This 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 x-optimal tax system under uniform income taxation is in general welfare-inferior to the x-optimal tax system with sector-specific income taxation. The next section assesses quantitatively the magnitude of this discrepancy.

[28] Indeed, because $c(n_a) = \kappa n_a$, it follows that, at any $(n_a, \kappa n_a) \in \mathbf{N}$,

$$\max_h \left\{ hn_a - \hat{T}_a(hn_a) - \psi(h) \right\} = \max_h \left\{ h\kappa n_a - \hat{T}_b(h\kappa n_a) - \psi(h) \right\}$$

The right-hand side can be rewritten as $\max_h \left\{ hn_a - \left[(1-\kappa)hn_a + \hat{T}_b(\kappa h n_a) \right] - \psi(h) \right\}$, which implies the result.

5 Numerical Illustration

In this section, we develop optimal policy simulations calibrated to the U.S. economy. Our goal is threefold. First, we quantitatively assess the welfare gains from sector-specific taxation relative to a uniform tax system where production efficiency prevails. Second, we evaluate the welfare impact from employing sector-specific income taxes relative to the case where only sales taxes (or, equivalently, payroll taxes) can differ across sectors, therefore assessing the welfare costs generated by simpler tax systems. Third, we complement our analytical results by studying the incidence of production inefficiencies, and the shape of marginal and average tax schedules.

5.1 Constructing the Distribution of Skills

We use data from the Current Population Survey (CPS) on wages and industry classification. In line with the tax practices followed by many countries, we assign industries to either one of the following two sectors: *manufacturing* and *services*. The manufacturing sector (indexed by a) comprises natural extraction, utilities, construction, transportation and all manufacturing industries. The services sector (indexed by b) comprises trade, the financial and banking industry, legal and business services, real state, arts, sports, and technology-intensive activities characterized by large returns to occupation-specific skills. Because these sectors are large and heterogenous, our analysis produces a conservative estimate of the gains from sector-specific taxation. A finer classification together with a more complex tax system can only magnify the welfare effects discussed below.[29]

We interpret the wage data as generated by a (sub-optimal) uniform tax system where production efficiency prevails. As such, each agent chooses the sector in which his productivity is the highest. We refer to this productivity as the *effective productivity* of an agent. We follow the literature in assuming a constant wage-elasticity of labor supply equal to 0.25. We then employ standard methods (as in Saez (2001)) to obtain the distribution of effective productivities using as inputs (i) the workers' earnings data, and (ii) the U.S. schedule of marginal tax rates. We denote by $\hat{G}$ the estimated distribution of effective productivities. As sector affiliation is observed in the data, we denote by $\tau_j(n)$ the fraction of agents with effective productivity n working on sector j.

Estimating the distribution of effective productivities in each sector is not enough to recover the bi-dimensional skill distribution F, which is needed for computing the optimal tax system.[30] Crucially, we cannot directly estimate the distribution of each agent's lowest productivity (which we call *latent productivity*), since we do not observe the earnings that an agent *would have* obtained had he worked in the sector where his productivity is the lowest.

We deal with this problem by constructing alternative scenarios on how latent productivities are distributed. In this respect, notice that, for an agent with effective productivity n, his latent productivity belongs to the support $[\underline{n}, n]$. We then denote by $H_k(\cdot|n)$ the distribution of the

[29] For further details on the data and the definition of sectors, see the Supplementary Material.

[30] This observation echoes the non-identification results of the Roy model discussed in Heckman and Honoré (1990).

latent productivity in sector k conditional on the effective productivity in sector $j \neq k$ being equal to n. In our simulations, we discretize the support of productivities by imposing a uniform grid. Productivity levels are indexed in increasing order by $\omega \in \Omega \equiv \{1, \ldots, |\Omega|\}$. As such, $\tilde{n}(\omega)$ is the ω^{th} lowest productivity in the grid. For its flexibility, we use the binomial c.d.f. to parametrize the conditional distributions of latent productivities. Accordingly, for any $\omega \leq \hat{\omega}$,

$$H_k\left(\tilde{n}(\omega)|\tilde{n}(\hat{\omega})\right) = \sum_{i=1}^{\omega} \begin{pmatrix} \hat{\omega} \\ i \end{pmatrix} p_j(\hat{\omega})^i (1 - p_j(\hat{\omega}))^{\hat{\omega}-i}.$$

Because the expected value of a random variable distributed according to $H_k\left(\tilde{n}(\omega)|\tilde{n}(\hat{\omega})\right)$ equals $p_j(\hat{\omega})\tilde{n}(\hat{\omega})$, the parameter $p_j(\hat{\omega}) \in [0, 1]$ is a natural measure of the *degree of skill transferability*, as defined in the discrete example of Section 2. Specifically, $p_j(\hat{\omega})$ is the (average) fraction of the effective productivity $\tilde{n}(\hat{\omega})$ that sector-j agents can transfer to sector k. When $p_j(\hat{\omega})$ is close to zero, sector-j agents are essentially unproductive as they move to sector k, and the opposite is true when $p(\hat{\omega})$ is close to one.

We employ the structure above to construct four scenarios:

Scenario 1 ***(high skill transferability)*** *$p_j(\omega) = 0.9$ for all $\omega \in \Omega$ and $j \in \{a, b\}$.*

In Scenario 1, all agents in either sector retain 90% of their productive capacity as they move away from the sector in which their productivity is the highest. This scenario thus captures an economy similar to those considered by Diamond and Mirrlees, where skills are perfectly transferable across sectors. In light of the Diamond and Mirrlees Theorem, under Scenario 1, one should expect sector-specific taxation to produce modest welfare gains relative to uniform taxation.

Scenario 2 ***(low skill transferability)*** *$p_j(\omega) = 0.1$ for all $\omega \in \Omega$ and $j \in \{a, b\}$.*

In Scenario 2, all agents in either sector lose 90% of their productive capacity as they move away from the sector in which their productivity is the highest. This scenario thus captures an economy similar to those studied in the tagging literature initiated by Akerlof (1978), where migration across "tags" is not feasible (which is equivalent to skills being perfectly untransferable). In this context, differential taxation is expected to improve significantly welfare only when the skill intensity is sufficiently different across sectors (recall the discussion on the skill intensity motive from Result 2 in Section 2).

Scenario 3 ***(high-low skill transferability)*** *$p_j(\omega) = 0.9$ if $\hat{G}\left(\tilde{n}(\omega)\right) < 0.5$, and $p_j(\omega) = 0.1$ otherwise, for $j \in \{a, b\}$.*

Inspired by Result 1 from the discrete-type example, Scenario 3 combines the two scenarios above so as to illustrate the potential gains from production inefficiency. It makes the (arguably, extreme) assumption that low-ability agents have a degree of skill transferability 0.9, while high-ability agents have a degree of skill transferability of 0.1 (with a discontinuous change at the

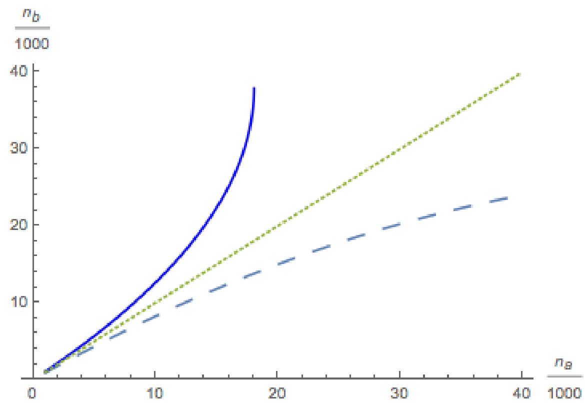

Figure 5: The figure depicts average latent productivities in Scenario 4. The dashed curve depicts the average latent productivity (in sector b) of an agent whose effective productivity is n_a. The thick curve depicts the average latent productivity (in sector a) of an agent whose effective productivity is n_b (thus mapping the Y axis into the X axis). The dotted line is the 45-degree line.

median skill level). This scenario captures in a stark way an economy in which agents of different productivity face different degrees of skill transferability. In the parlance of Section 2, this is an economy in which the skill transferability motive calls for sector-specific taxation.

Scenario 4 ***(smoothly decreasing skill transferability)*** $p_j(\omega) = \alpha_j - \beta_j(\omega - 1)$ *for all* $\omega \in \Omega$ *and* $j \in \{a, b\}$, *where* $\alpha_a > \alpha_b$ *and* $\beta_a < \beta_b$.

In Scenario 4, skill transferability is heterogeneous across agents and across sectors. In line with recent empirical evidence, we make the plausible assumption that the degree of skill transferability smoothly declines with earnings (in which case the *percentage losses* on earnings due to sectorial migration are the highest at the top of the income distribution).[31] As the services sector comprises industries with high returns to occupation-specific skills (such as banking, finance and business and legal services), we assume that it exhibits the lowest degree of skill transferability at the top, and the highest at the bottom of the income distribution. Accordingly, we let $\alpha_a < \alpha_b$ and $\beta_a < \beta_b$.

Figure 5 describes for Scenario 4 the expected latent productivity as a function of effective productivities (being therefore the continuum-type analog to Figure 1 in the Illustrative Example of Section 2). That sector b has a lower degree of skill transferability at the top than sector a is reflected in the fact that the thick curve is farther to the diagonal than the dashed curve for high values of n_b. Conversely, that sector b has a greater degree of skill transferability at the bottom than sector a is reflected in the fact that the dashed curve is farther from the diagonal than the thick curve for low values of n_a.[32]

[31] For evidence of this point on the finance and banking sector, see Bell and Van Reenen (2014) and Dent (2015). More generally, the labor literature finds that the percentage wage losses from switching sectors are greatest for workers with higher incomes (see for example Neal (1995)). This reflects the fact that larger incomes come with longer tenures, during which workers accumulate sector-specific (typically untransferable) human capital.

[32] To have a sense of magnitudes, the degree of skill transferability is 90% (resp., 95%) in sector a (resp., b) for the bottom effective skill, and 60% (resp., 45%) in sector a (resp., b) for the top effective skill.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
sector-specific income taxes	0.96%	0.76%	1.98%	1.42%
sector-specific sales taxes	0.40%	0.17%	0.72%	0.75%
sales tax $\tau_b(\tau_a \equiv 0)$	4.6%	4.2%	5.3%	10.8%

Table 1. In the top two rows, the percentage welfare gains achieved by the optimal tax system relative to uniform taxation. In the bottom row, the optimal sales tax in sector b when income taxes are uniform and τ_a is normalized to zero.

In light of the skill transferability motive from Section 2, we should expect the manufacturing sector to be favored by the optimal tax system under Scenario 4.

Equipped with the conditional distributions of latent productivities, we can readily recover the bi-dimensional distribution F. Letting $\omega_a, \omega_b \in \Omega$ and denoting by $\hat{g}$ the probability mass function associated with $\hat{G}$, it follows that

$$F(\tilde{n}(\omega_a), \tilde{n}(\omega_b)) = \hat{G}(\tilde{n}(\omega_k)) + \sum_{i=\omega_k+1}^{\omega_j} H_k\left(\tilde{n}(\omega_k)|\tilde{n}(i)\right) \hat{g}(\tilde{n}(i))\tau_j(\tilde{n}(i)),$$

where we choose indexes so that $\omega_j \equiv \max\{\omega_a, \omega_b\}$ and $\omega_k \equiv \min\{\omega_a, \omega_b\}$. The formula above decomposes the probability $F(\tilde{n}(\omega_a), \tilde{n}(\omega_b))$ into two terms: The first term is the probability that the effective probability is less than $\tilde{n}(\omega_k)$. The second term is the probability that an agent works on sector j, enjoys an effective probability between $\tilde{n}(\omega_k)$ and $\tilde{n}(\omega_j)$, *and* has a latent productivity smaller than $\tilde{n}(\omega_k)$. This decomposition allows us to express the c.d.f. F in terms of the c.d.f.'s $\hat{G}$ and $\{H_a(\cdot|n), H_b(\cdot|n)\}$ at any productivity pair $(\tilde{n}(\omega_a), \tilde{n}(\omega_b))$. Employing the results from the previous section, we use this distribution to compute the Ralwsian-optimal tax system. Under this welfare objective, the percentage gains in welfare can be conveniently interpreted also as percentage variations in total tax collection.

5.2 Results

The impact of sector-specific taxation on welfare is moderate when the degree of skill transferability is high across all income levels in both sectors, as considered in Scenario 1. In the case where income tax schedules can differ across sectors, these gains amount to 0.96%, and to 0.40% when only sales taxes are sector-specific (see Table 1). The reason for such a limited impact is that the migration effect is large when the distribution of productivities F concentrates most mass around the 45-degree line. As a result, the optimal differential tax schedules are close to uniform. Scenario 1, therefore, supports the conclusions of the Diamond-Mirrlees theorem when skill transferability is imperfect but high.

The welfare gains from sector-specific taxation are even smaller in Scenario 2, where skill transferability is low across all income levels in both sectors. In this scenario, the migration effect is

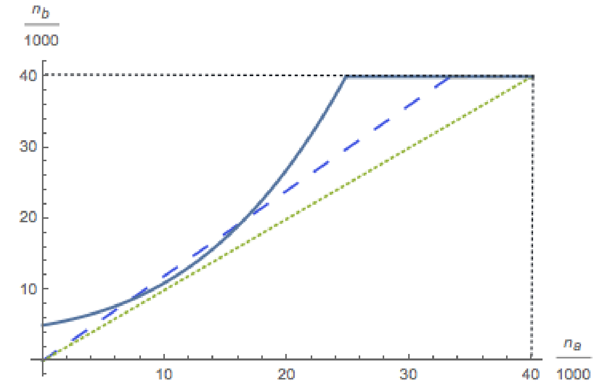

Figure 6: The figure depicts the optimal occupational choice rules in Scenario 4. The dashed line corresponds to the case where only sales taxes are sector-specific, and the thick curve to the case where income taxes are sector-specific.

small, and most welfare gains from differential taxation arise from the government's ability to tailor marginal taxes to the distributions of productivities (under production efficiency) among agents in each sector. However, because we purposely considered two large sectors, these distributions are rather similar. In the parlance of Section 2, the skill intensity does not vary much across sectors. As a result, the gains from sectorial taxation are limited, and optimal taxes are close to uniform.[33] We conjecture that by expanding the number of sectors, the gains from differential taxation are likely to increase, as the shape of informational rents should differ across sectors.

Scenario 3, where skill transferability sharply decreases with income, exhibits the highest gains from sectorial taxation. When sector-specific income taxation is possible, the welfare gains of moving from the existing tax code to the optimal one are close to 2%. Interestingly, only a third of these gains are realized by sector-specific sales taxes. This highlights the importance of non-linear instruments in tailoring marginal taxes to different income levels.

The last scenario is likely to be the most realistic, as skill transferability decreases smoothly with income in both sectors (but more markedly in the services sector). In this scenario, differential income taxation yields welfare gains of 1.42%, while differential sales taxes of 0.75%. These results indicate that "simplicity" is not costless: Another 0.67 percentage point in welfare gains could be obtained by employing income taxes that are sector-specific (therefore almost doubling the welfare gains from sector-specific sales taxes).

In scenario 4, the optimal tax system favors the manufacturing sector, as reflected by the occupational choice rule being uniformly above the 45-degree line (and sales taxes satisfying $\tau_b > \tau_a$) - see Figure 6.[34] Moreover, most of the migration from services to manufacturing occurs at

[33] More precisely, the distribution of productivities (under uniform taxation) in the services sector is slightly more right-skewed than that on manufacturing (see the Supplementary Material for further details). This property implies the services sector is overtaxed (e.g., the optimal sales tax sets $\tau_b > \tau_a$).

[34] Note that, conditional on the optimal tax schedule favoring sector a, the distributions of latent productivities in sector a do not affect the optimal tax system (since no agent in manufacturing migrates to services). As such, the welfare results from Table 1 and the tax schedules discussed below are robust to other choices of $H_a(\cdot|n_b)$.

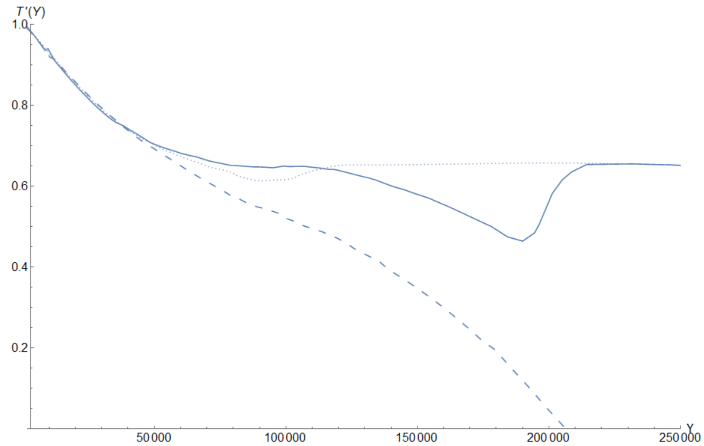

Figure 7: The figure depicts marginal tax schedules. The dotted curve refers to the the uniform-taxation benchmark. The full (resp., dashed) curves refer to the services (resp., manufacturing) sector in the optimal tax system under Scenario 4.

intermediate or low skill levels. This occurs because, at skills close to the top, latent productivities are so low (see Figure 5 for the schedule of average latent productivities) that few agents actually migrate (i.e., the area between the threshold function and the 45-degree line contains a small mass of agents when n_b is close to the top).

This result is expected in light of the discussion in Section 2: Because the services sector has a lower (resp., higher) degree of skill transferability at the top (resp., bottom) of the skill distribution than the manufacturing sector, the skill transferability motive for differential taxation calls for taxing more heavily the services sector. By favoring the manufacturing sector, the government collects high taxes from high-income agents in the services sector while generating small production losses from the misallocation of talent among low-income agents.

The schedules of marginal taxes in both sectors illuminate the logic behind the optimal tax system. Figure 7 depicts the optimal marginal tax schedules in both sectors, as well as the optimal marginal taxes when taxation is uniform across sectors. It is interesting to note that, relative to the uniform-taxation benchmark, marginal taxes in the services sector are higher for agents with income ranging from \$50,000 to \$120,000, and (weakly) smaller for agents earning more than \$120,000. The reason is simple: Because of migration to manufacturing, the distribution of income in the services sector contains a smaller proportion of agents with incomes between \$50,000 and \$120,000. As a result, the government can set *higher* marginal tax rates at these income levels while incurring distortions in labor supply which are lower than under uniform taxation.

This implies that the *level* of taxes faced by high-income individuals in the services sector is higher than under uniform taxation. Because the degree of skill transferability is decreasing with income, these agents are the ones to lose the most (in terms of productivity) from changing occupation so as to avoid high taxes (as illustrated in Figure 5). As a result, the marginal rates faced by these agents are *lower* at the optimum than in the uniform taxation benchmark. This

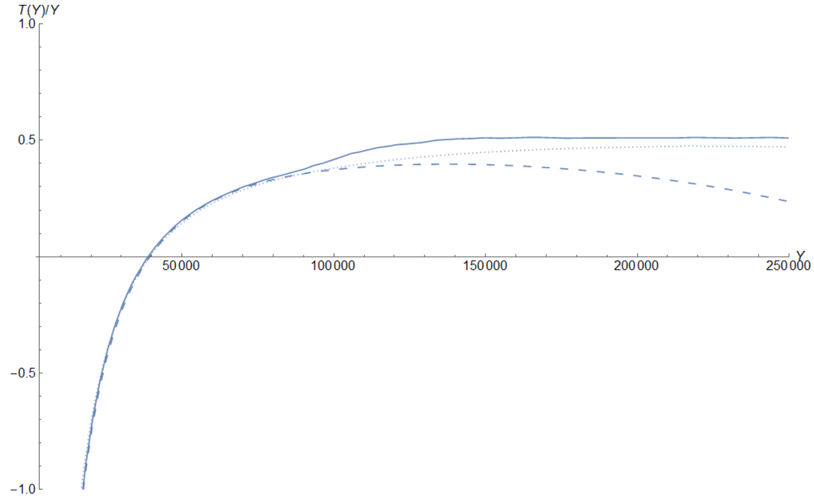

Figure 8: The figure depicts average tax schedules. The dotted curve refers to the the uniform-taxation benchmark. The full (resp., dashed) curves refer to the services (resp., manufacturing) sector in the optimal tax system under Scenario 4.

increases efficiency while generating high tax collection (as high-income agents in the services sector are taxed mostly in a "lump-sum" fashion).

In turn, the marginal taxes on the manufacturing sector are smaller than those in both the services sector and the uniform-taxation benchmark. At low or intermediate income levels (let us say, less than \$120,000) this results from the migration of low-ability agents from the services to the manufacturing sector, which magnifies the welfare costs of distorting the provision of labor in the manufacturing sector. At high-income levels (e.g., higher than \$120,000), low marginal taxes in the manufacturing sector are the welfare-maximizing way of implementing the optimal occupational choice rule, which exhibits a highly steep threshold function. Such low marginal tax rates help discouraging high-ability agents from migrating to the services sector (as their degree of skill transferability is higher than that of similar-skilled agents in services).

Interestingly, the marginal taxes on the services sector increase after \$200,000, converging relatively fast to its asymptotic level (which coincides with the asymptotic level of the uniform taxation benchmark). By contrast, the marginal taxes on the manufacturing sector keep decreasing after \$200,000, and are constant at zero for very high incomes. These results follow from the fact that almost all very-high-income earners (above \$200,000) are in the services sector. In turn, the income distribution in the manufacturing sector, which contains few very high earners, is akin to a bounded distribution. This explains why marginal taxes are zero at the very top of the manufacturing sector when income tax schedules are sector-specific.

Finally, Figure 8 displays the optimal average tax schedules in both sectors, as well as the optimal average taxes when taxation is uniform across sectors. Relative to the uniform benchmark, sector-specific income taxes enable the government to increase the average tax rate by approximately 10% at income levels around \$120,000 in the services sector (containing a large fraction of finance and legal professionals). Differential taxation also leads to a significant reduction in the average tax rate paid by individuals in the manufacturing sector. The ability to significantly change the average tax incidence across sectors and income levels underscores the potential of sectorial taxation to generate welfare improvements.

Needless to say, the policy prescriptions derived above should not be taken at face value. We view this numerical exercise primarily as an illustration of the methods of this paper. Incorporating in the simulations other margins that are important in reality (such as tax evasion and avoidance, to name a few) is likely to bring further insights.

Relatedly, the optimal sector-specific tax system derived above takes as given the joint distribution of skills. In the long-run, it is likely that agents, facing different tax burdens across sectors, adjust their human capital investments in response to the tax system (e.g., by reducing investments to acquire skills that are little transferable across sectors). Incorporating this "long-run" margin in the analysis is an interesting direction of future research.

6 Discussion and Conclusions

Summary. This paper studies differential taxation in a setting where workers' skills are not perfectly transferable across sectors/occupations. 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 agents across sectors) is held fixed, and where the government chooses a tax system to maximize welfare *subject to* implementing that occupational choice rule. Next, we consider 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.

The primal-dual approach described above generates a number of insights. First, it delivers a generalized Mirrlees formula that determines how the government optimally balances intensive-margin distortions in labor supply *across sectors*, for any desired occupational choice rule. Second, it yields an Euler equation that summarizes all the trade-offs faced by the planner in choosing the optimal occupational choice rule. The formula shows how, at the optimum, the marginal revenue losses from further distorting the allocation of workers across sectors (through an increase in marginal taxes is *some* sectors) come from the dissipation of tax revenues due to the migration of workers across sectors. The marginal gains, in turn, come from better tailoring the tax schedule in each sector to its endogenous distribution of talent (tagging). These results imply failure of the Diamond-Mirrlees theorem: Social welfare can be increased by inducing certain agents to work in a sector in which their productivity is not the highest (i.e., production efficiency is violated at the second-best).

We next quantify the welfare gains of differential taxation. First, we show that these gains can be substantive under the empirically plausible assumption that the degree of skill transferability of agents decreases with income. Second, we show that "simpler" tax systems (involving uniform income taxes but differential sales or payroll taxes) leave a sizable portion of these gains unrealized. Third, we document the incidence of production inefficiencies and the tax collection gains that differential taxation generates from high earners in occupations with large wage premia.

k-Sector Extension. In the Supplementary Material, we extend the analysis to economies with k sectors. We show how the primal-dual approach described above can be extended to yield useful formulas also in these richer settings. First, we show how the occupational choice rule can be conveniently described by a collection of absolutely continuous and nondecreasing threshold functions $c \equiv (c_{jl})_{j,l \in K, l \neq j}$, one for each pair of sectors j, l, $l \neq j$, such that an agent with productivity type $\mathbf{n} = (n_1, n_2, ..., n_k) \in [\underline{n}, \bar{n}]^k$ is induced to join sector j if, and only if, for all $l \neq j$, $n_l < c_{jl}(n_j)$.

Next, we consider a primal problem analogous to the one in the two-sector model where the occupational choice rule is held fixed and where the government maximizes welfare subject to a budget constraint, and to the requirement that the tax system implement the given occupational choice rule. The solution to this primal problem delivers a multi-sector generalization of the formula in Proposition 1 above.

We then proceed by solving a dual problem in which the labor supply in one sector is held fixed and where the government maximizes welfare subject to the budget constraint and the requirement that the resulting tax equilibrium be consistent with the labor supply schedule in that sector. The solution to this dual problem yields an Euler equation analogous to the one in Proposition 2. Interestingly, despite the complexity of the environment, these formulas are remarkably similar to those in the two-sector model. In particular, the Euler equation describing the optimal occupational choice rule balances the same revenue collection, welfare, migration, and continuity correction effects as the simple formula in Proposition 2. It can be derived heuristically from the same type of α-payroll tax reforms considered above. However, migration in these richer economies is more intriguing, for it involves migration from the sector in which the reform takes place to each other sector in the economy. The key trick in solving such complex problems is a change in measure that permits one to express the endogenous talent distribution in all but one sector of the economy as a function of the talent distribution in the remaining sector. As we explain below, we expect these techniques will be useful also in other multi-dimensional screening problems. The k-sector model can also be fruitfully applied in quantitative exercises that extend the analysis of Section 5 to more than two sectors.

Other Applications. Our analysis is conducted in the context of a multi-sector economy where workers choose their occupation by comparing their productivity differential across sectors with the corresponding differential in 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.[35] The planner's problem studied in this paper (be it Ralwsian or concave Utilitarian) 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.[36]

Another application of the techniques 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 taxation), which are typically easier to enforce than income taxes.[37] The techniques in Sections 4 and 4.4 above can be adapted to study the optimal combination of income and sale taxes in economies with an informal sector.[38]

The design of dual tax systems, where corporate and wage incomes are taxed differently, constitute yet another natural application of our results. Similarly to the problem examined in the present paper, the choice by many workers between employment and self-employment is often based on the comparison between the productivity differential and the differential in the tax burden across the two regimes. More broadly, the study of the interplay between tax enforceability and occupational choice, and its implications for the design of optimal tax systems, is a fascinating topic for future research.

Finally, the techniques developed in this paper can be applied to other multi-dimensional screening problems in which agents face rival choices. Consider first the application to nonlinear pricing by a multi-product monopolist with rival product lines (for instance, a car dealer designing price-quality schedules for various car categories, such as sport and family cars). Given their preferences for each product line and the price-quality schedules offered by the seller, buyers choose which type of car to buy, and then, select, within the chosen category, the desired model (identified by a

[35]See Kleven et al. (2013, 2014) and the references therein for empirical studies of how workers respond to differential taxation by migrating from one state to another.

[36]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.

[37]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 economies.

[38]See also Best et al. (2015) and the references therein for other examples of economies in which governments optimally sacrifice production efficiency to boost tax revenues when their ability to enforce tax collection is limited.

combination of price and quality).[39]

Another application is the design of managerial compensation schemes with rival career paths. In this multi-dimensional extension of Laffont and Tirole (1986), each manager chooses between different career paths as a function of his productivity in each path and its respective performance-pay schedule. In designing the remuneration in each career, the firm has to trade-off the profit losses from misallocating managers across tasks with the reductions in informational rents that differential compensation schemes make possible.

Both the nonlinear pricing and the managerial compensation applications can be studied with the primal-dual approach of this paper. The solution to the primal and dual problems deliver analogs of the generalized Mirrlees formula and the Euler equation, which can be used to derive properties of optimal contracts when agents face exclusive choices and have multi-dimensional private information.

[39]See Calzolari and Denicolò (2013, 2015) and Choné and Linnemer (2015) for recent contributions to the literature on price discrimination under adverse selection and exclusive contracting.

7 Appendix: Omitted Proofs

Proof of Lemma 1. *Necessity.* We first establish that, for the allocation $(\mathcal{C}, h_a, h_b)$ to be implemented at the wage rates $\mathbf{w}$ by the tax system $\mathcal{T}$, the properties in parts 1-3 must jointly hold. That wages must satisfy the condition in part (1) for the labor market to clear follows directly from the fact that the production function exhibits constant returns to scale.

Next, consider the properties in part 2. For any $j = a, b$, 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 sector-j productivity n_j obtains from generating income y in sector j. Because the function satisfies the strict increasing difference property, the correspondence $\hat{y}_j(n_j) = \arg\max_y \left\{y - T_j(y) - \psi\left(y/(w_j n_j)\right)\right\}$ must be nondecreasing over $\bar{N}$ in the strong set order sense. Because $\bar{N}$ is compact, this means that $\hat{y}_j(n_j)$ is bounded over $\bar{N}$. Next note that $\hat{u}_j(y; n_j)$ is differentiable over $\bar{N}$. Because $\hat{y}_j(n_j)$ is bounded over $\bar{N}$, $\hat{u}_j(y; n_j)$ is equi-Lipschitz continuous over $\hat{y}_j(\bar{N}) \times \bar{N}$, with $\hat{y}_j(\bar{N})$ denoting the range of $\hat{y}_j(\cdot)$. Standard envelope theorems (e.g., Milgrom and Segal (2002)) then imply that the value function

$$\tilde{u}_j(n_j) \equiv \max_h\{w_j h n_j - T_j(w_j h n_j) - \psi(h)\} = \max_y\{\hat{u}_j(y; n_j)\} = \max_{y \in \hat{y}_j(N)}\{\hat{u}_j(y; n_j)\}$$

must be Lipschitz continuous over N with derivative equal to

$$\tilde{u}_j'(n_j) = \psi'\left(\frac{y_j(n_j)}{w_j n_j}\right)\frac{y_j(n_j)}{w_j n_j^2}$$

for almost every $n_j \in N$, where $y_j : N \to \hat{y}_j(N)$ is an arbitrary selection from the correspondence $\hat{y}_j(\cdot)$. Using the fact that, for any $n_j \in N_j$,

$$h_j(n_j) = \frac{y_j(n_j)}{w_j n_j}$$

along with the fact that $u_j(n_j) = \tilde{u}_j(n_j)$ for all $n_j \in N_j$, we then have that the properties in Part 2 are necessary.

Finally, consider the properties in part 3. That the occupational choice rule $\mathcal{C}$ must satisfy these properties follows directly from the fact that each $\tilde{u}_j(\cdot)$ is Lipschitz continuous and strictly increasing over N along with the fact that the payoff $\tilde{u}_j(n_j)$ that each agent obtains from joining each sector j and then choosing his labor supply optimally is independent of his productivity in any other sector.

Sufficiency. Now suppose that $(\mathcal{C}, h_a, h_b)$, along with the wage rates $\mathbf{w}$ and the tax system $\mathcal{T}$, satisfies all the properties in parts 1-3 in the lemma. Then let $\hat{\mathcal{T}}$ be any tax system such that, for all $j = a, b$, (a) $\hat{\tau}_j = \tau_j$, (b) $\hat{T}_j(y) = T_j(y)$ for all y such that $y = w_j n_j h_j(n_j)$ for some $n_j \in N_j$, and (c) $\hat{T}_j(y) = Q \in \mathbb{R}_{++}$ for all y such that there exists no $n_j \in N_j$ such that $y = w_j n_j h_j(n_j)$. It is easy to see that there exists $Q \in \mathbb{R}_{++}$ large enough such that, for any $j = a, b$, any $n_j \in N$,

$$\arg\max_y \left\{y - \hat{T}_j(y) - \psi\left(\frac{y}{w_j n_j}\right)\right\} \subset \left\{y : y = w_j n_j' h_j(n_j'),\ n_j' \in N_j\right\}.$$

That the allocation $(\mathcal{C}, h_a, h_b)$ is implemented at the wage rates $\mathbf{w}$ by the tax system $\hat{\mathcal{T}}$ then follows from the following facts. Employing the effective labor

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

is profit maximizing for each firm $j = a, b$, given that the wages $\mathbf{w}$ satisfy the condition in part (1) and that the production function exhibits constant returns to scale. For each type $\mathbf{n}$ joining sector $j \in \{a, b\}$ (i.e., for which $\mathcal{C}(\mathbf{n}) = j$), the labor supply $h_j(n_j)$ satisfies the incentive compatibility condition on the intensive margin: $u_j(n_j) = \tilde{u}_j(n_j)$. This follows from the fact that $y_j(n_j) = w_j h_j(n_j) n_j$ is nondecreasing over N_j along with the fact that the equilibrium payoff satisfies the envelope condition

$$u_j(n_j) = u_j(n_j') + \int_{n_j'}^{n_j} \psi'\left(h_j(x)\right) \frac{h_j(x)}{x} dx$$

for any pair $n_j, n_j' \in N_j$ and the fact that, for any y for which there exists no $n_j \in N_j$ such that $y = w_j n_j h_j(n_j)$, $\hat{T}_j(y) = Q$.

Lastly, that the occupational choice rule $\mathcal{C}$ satisfies the incentive compatibility condition on the extensive margin follows from the fact that, for any $\mathbf{n}$ any $j \in \{a, b\}$ $\mathcal{C}(\mathbf{n}) = j$ only if $\tilde{u}_j(n_j) \geq \tilde{u}_l(n_l)$ all $l \neq j$. This follows directly from the threshold structure of $\mathcal{C}$ along with the fact that each $\tilde{u}_j(\cdot)$ is strictly increasing, any $j = a, b$. Q.E.D.

Proof of Remark 1. 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. First, note that, 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 Conditions (7) and (9), we then have that, for almost every $n_a \in N$ such that (i) $c(n_a) \in N$ and (ii) $T_a(y_a(n_a))$ and $T_b(y_b(c(n_a)))$ are differentiable, Condition (11) must hold.

Let us now consider the government's problem. It 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:[40]

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

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:

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

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

$$u'_a(n_a) = \psi'\left(h_a(n_a)\right) \frac{h_a(n_a)}{n_a} \text{ for almost every } n_a \in N_a, \tag{25}$$

$$u'_b(n_b) = \psi'\left(h_b(n_b)\right) \frac{h_b(n_b)}{n_b} \text{ for almost every } n_b \in N_b, \tag{26}$$

(iii) the occupational-choice incentive-compatibility constraints:[41]

$$h_b(c(n_a)) = J_c[n_a]h_a(n_a), \text{ for all } n_a \in (n'_a, n''_a), \tag{27}$$

where

$$J_c[n_a] \equiv \left(c(n_a)/\left(n_a \cdot c'(n_a)\right)\right)^{\xi},$$

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 (a) for any $n_a \in (n'_a, \bar{n})$,

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

along with the fact that (b) 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

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

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

[41]To see that the occupational-choice incentive-compatibility constraint takes the form in (27), recall that, for almost every $n_a \in N$ such that $c(n_a) \in N$, $u'_a(n_a) = u'_b(c(n_a))c'(n_a)$. Next, use (25) and (26), along with the fact that $\psi'(h) = \xi h^{\xi-1}$, to arrive at (27).

Using the definition of the density

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

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

$$\int_{N_a} \left\{ \mathbf{1}_x^{CU} \phi(u_a(n_a))g(n_a|c) + [1 - \mathbf{1}_x^{CU}]f_a(n_a)u_a(n_a') \right\} dn_a \tag{28}$$

subject to the budget constraint

$$\begin{aligned}&\int_{N_a} \left\{ [h_a(n_a)n_a - \psi\,(h_a(n_a)) - u_a(n_a)]\, f_a(n_a)F_{b|a}(c(n_a)|n_a) \right\} dn_a \\ &+ \int_{N_a} \left\{ [J_c[n_a]h_a(n_a)c(n_a) - \psi\,(J_c[n_a]h_a(n_a)) - u_a(n_a)]\, c'(n_a)f_b(c(n_a))F_{a|b}(n_a|c(n_a)) \right\} dn_a, \\ &\geq B\end{aligned} \tag{29}$$

and the IC constraints

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

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

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

where λ is the Lagrange multiplier associated to the common budget constraint (29) and where μ is the co-state variable associated with the law of motion of u_a. The transversality conditions are:

$$\mu(n_a') = \mu(\bar{n}) = 0. \tag{30}$$

From the Pontryagin Maximum Principle,

$$\mu'(n_a) = -\frac{\partial H}{\partial u_a} = \left[\lambda - \mathbf{1}_x^{CU} \phi'(u_a(n_a))\right] g(n_a|c). \tag{31}$$

Integrating the right-hand side of (31) and using the transversality condition (30) we have that

$$\mu(n_a) = -\lambda \int_{n_a}^{\overline{n}} \left[1 - m_a(\tilde{n}_a)\mathbf{1}_x^{CU}\right] g(\tilde{n}_a|c)d\tilde{n}_a, \tag{32}$$

where we used the definition of

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

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

$$\frac{\partial H}{\partial h_a} = \lambda \left[n_a - \psi'(h_a(n_a)) \right] f_a(n_a) F_{b|a}(c(n_a)|n_a) +$$
$$\lambda \left\{ J_c[n_a] c(n_a) - \psi'(J_c[n_a] h_a(n_a)) J_c[n_a] \right\} c'(n_a) f_b(c(n_a)) F_{a|b}(n_a|c(n_a)) +$$
$$\mu(n_a) \frac{\psi'(h_a(n_a)) + \psi''(h_a(n_a)) h_a(n_a)}{n_a} = 0 \tag{33}$$

Combining (32) with (33) 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:

$$\left[n_a - \psi'(h_a(n_a)) \right] g_a(n_a|c) + \left\{ J_c[n_a] c(n_a) - \psi'(J_c[n_a] h_a(n_a)) J_c[n_a] \right\} c'(n_a) g_b(c(n_a)|c) \tag{34}$$
$$= \left\{ \frac{\psi'(h_a(n_a)) + \psi''(h_a(n_a)) h_a(n_a)}{n_a} \right\} \int_{n_a}^{\overline{n}} \left[1 - m_a(\tilde{n}_a) \mathbf{1}_x^{CU} \right] g(\tilde{n}_a|c) d\tilde{n}_a$$

Recall the first-order condition

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

and that

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

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 (34) can be rewritten as the usual Mirrlees condition

$$\xi \frac{T_a'(y_a(n_a))}{1 - T_a'(y_a(n_a))} n_a f_a(n_a) = \int_{n_a}^{\overline{n}} \left[1 - \mathbf{1}_x^{CU} m_a(\tilde{n}_a) \right] f_a(\tilde{n}_a) d\tilde{n}_a, \tag{35}$$

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 (34) can be rewritten as

$$\frac{T_a'(y_a(n_a))}{1 - T_a'(y_a(n_a))} n_a g_a(n_a|c) + \frac{T_b'(y_b(c(n_a)))}{1 - T_a'(y_a(n_a))} J_c[n_a] c'(n_a) c(n_a) g_b(c(n_a)|c) \tag{36}$$
$$= \xi^{-1} \int_{n_a}^{\overline{n}} \left[1 - m_a(\tilde{n}_a) \mathbf{1}_x^{CU} \right] g(\tilde{n}_a|c) d\tilde{n}_a.$$

Replacing (11) into (36) 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 $n_a \in (n_a', n_a'')$,

$$\xi \frac{T_a'(y_a(n_a))}{1 - T_a'(y_a(n_a))} n_a g_a(n_a|c) + \xi \frac{T_b'(y_b(c(n_a)))}{1 - T_b'(y_b(c(n_a)))} c(n_a) g_b(c(n_a)|c) \tag{37}$$
$$= \int_{n_a}^{\overline{n}} \left[1 - \mathbf{1}_x^{CU} m_a(\tilde{n}_a) \right] g(\tilde{n}_a|c) d\tilde{n}_a.$$

Combining the results establishes equation (10). 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

$$R^x_j\left(n_j\right) \equiv \mathbf{1}^{CU}_x \phi\left(u_j\left(n_j\right)\right) + \lambda\left\{h_j(n_j)n_j - \psi\left(h_j(n_j)\right) - u_j\left(n_j\right)\right\}, \tag{38}$$

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

$$\textstyle\sum_{j=a,b} \displaystyle\int_{\underline{n}}^{\bar{n}} R^x_j\left(n\right) g_j(n|c)dn + [1 - \mathbf{1}^{CU}_x]u_a(n'_a) \tag{39}$$

subject to the incentive compatibility constraints for labor supply

$$u'_a(n_a) = \psi'\left(h_a(n_a)\right)\frac{h_a(n_a)}{n_a} \text{ for almost every } n_a \in N_a, \tag{40}$$

$$u'_b(n_b) = \psi'\left(h_b(n_b)\right)\frac{h_b(n_b)}{n_b} \text{ for almost every } n_b \in N_b, \tag{41}$$

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 $n_b \in \operatorname{int}(N_b)$,

$$h_b(n_b) = J_c[c^{-1}(n_b)]h_a(c^{-1}(n_b)), \tag{42}$$

and

$$u_b(n_b) = u_a(c^{-1}(n_b)), \tag{43}$$

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

$$\int_{n'_a}^{\bar{n}} \left[R^x_a\left(n_a\right) g_a(n_a|c) + \hat{R}^x_b\left(n_a|c\right) g_b(c(n_a)|c)c'(n_a)\right] dn_a + [1 - \mathbf{1}^{CU}_x]u_a(n'_a) \tag{44}$$

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 $\hat{R}_b^x(n_a|c) = 0$ 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 (44). 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 (40). 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 (42) and (43), respectively. Finally, the tax schedules in the two sectors are given by

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

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 $\tilde{R}_b^x : \mathbb{R}^3 \to \mathbb{R}$ be the function defined by

$$\tilde{R}_b^x(n_a, c, J) \equiv \mathbf{1}_x^{CU}\phi(u_a(n_a)) + \lambda\{Jh_a(n_a)c - \psi(Jh_a(n_a)) - u_a(n_a)\}, \tag{45}$$

and denote by $\partial\tilde{R}_b^x/\partial n_a$, $\partial\tilde{R}_b^x/\partial c$, and $\partial\tilde{R}_b^x/\partial J$ 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

$$J(n, c, c') = \left(\frac{c}{nc'}\right)^{\xi}, \tag{46}$$

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

$$g_a(n_a|c) \equiv f_a(n_a)F_{b|a}(c(n_a)|n_a) = \int_{\underline{n}}^{c(n_a)} f(n_a, x)dx \tag{47}$$

and

$$g_b(c(n_a)|c) = f_b(c(n_a))F_{a|b}(n_a|c(n_a)) = \int_{\underline{n}}^{n_a} f(x, c(n_a))dx \tag{48}$$

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

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

Dropping the arguments from $\tilde{R}_b^x$ 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

$$R_a^x(n_a)\, f(n_a, c(n_a)) + \frac{\partial \tilde{R}_b^x}{\partial c} c'(n_a) g_b(c(n_a)|c) + \frac{\partial \tilde{R}_b^x}{\partial J} J_c c'(n_a) g_b(c(n_a)|c) + \tilde{R}_b^x c'(n_a) \frac{\partial}{\partial c}\left[g_b(c(n_a)|c)\right]$$
$$= \frac{d}{dn_a}\left[\frac{\partial \tilde{R}_b^x}{\partial J} J_{c'} c'(n_a) g_b(c(n_a)|c) + \tilde{R}_b^x g_b(c(n_a)|c)\right]. \tag{49}$$

Use (48) to note that the fourth term of the left-hand-side of (49) can be developed as follows:

$$\begin{aligned}
\tilde{R}_b^x c'(n_a) \frac{\partial}{\partial c}\left[g_b(c(n_a)|c)\right] &= \tilde{R}_b^x c'(n_a) \left(\int_{\underline{n}}^{n_a} \frac{\partial}{\partial n_b} f(x, c(n_a)) dx\right) \\
&= \tilde{R}_b^x \frac{d}{dn_a}\left[g_b(c(n_a)|c)\right] - \tilde{R}_b^x f(n_a, c(n_a)) \\
&= \frac{d}{dn_a}\left[\tilde{R}_b^x g_b(c(n_a)|c)\right] - \frac{d\tilde{R}_b^x}{dn_a} g_b(c(n_a)|c) - \tilde{R}_b^x f(n_a, c(n_a)).
\end{aligned} \tag{50}$$

Substituting (50) into (49) and simplifying, we can rewrite the point-wise Euler equation (49) as follows

$$\left[R_a^x(n_a) - \tilde{R}_b^x\right] f(n_a, c(n_a)) + \frac{\partial \tilde{R}_b^x}{\partial c} c'(n_a) g_b(c(n_a)|c) + \frac{\partial \tilde{R}_b^x}{\partial J} J_c g_b(c(n_a)|c) c'(n_a) - \frac{d\tilde{R}_b^x}{dn_a} g_b(c(n_a)|c)$$
$$= \frac{d}{dn_a}\left[\frac{\partial \tilde{R}_b^x}{\partial J} J_{c'} c'(n_a) g_b(c(n_a)|c)\right]. \tag{51}$$

Multiplying both sides of (51) by $c(n_a)$ and rearranging terms, we obtain that

$$\left[R_a^x(n_a) - \tilde{R}_b^x\right] f(n_a, c(n_a)) c(n_a) + \frac{\partial \tilde{R}_b^x}{\partial c} c'(n_a) c(n_a) g_b(c(n_a)|c) - \frac{d\tilde{R}_b^x}{dn_a} c(n_a) g_b(c(n_a)|c)$$
$$= \frac{d}{dn_a}\left[\frac{\partial \tilde{R}_b^x}{\partial J} J_{c'} c'(n_a) g_b(c(n_a)|c)\right] c(n_a) - \frac{\partial \tilde{R}_b^x}{\partial J} J_c g_b(c(n_a)|c) c(n_a) c'(n_a). \tag{52}$$

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

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

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 (11)).

Replacing these expressions into the right-hand side of (52), we obtain that for any $n_a \in (n_a', n_a'')$, the right-hand side of the Euler equation becomes

$$
\begin{aligned}
&\frac{d}{dn_a}\left[\frac{\partial \tilde{R}_b^x}{\partial J} J_{c'} c'(n_a) g_b(c(n_a)|c)\right] c(n_a) - \frac{\partial \tilde{R}_b^x}{\partial J} J_c g_b(c(n_a)|c) c(n_a) c'(n_a) \\
&= -\xi \left\{ \frac{d}{dn_a}\left[J \frac{\partial \tilde{R}_b^x}{\partial J} g_b(c(n_a)|c)\right] c(n_a) + J \frac{\partial \tilde{R}_b^x}{\partial J} g_b(c(n_a)|c) c'(n_a) \right\} \\
&= -\xi \frac{d}{dn_a}\left[J \frac{\partial \tilde{R}_b^x}{\partial J} g_b(c(n_a)|c) c(n_a)\right]. \qquad (53)
\end{aligned}
$$

Substituting (53) into (52), we then have that, for any $n_a \in (n_a', n_a'')$, the Euler equation becomes

$$
\begin{aligned}
&\left[R_a^x(n_a) - \tilde{R}_b^x\right] f(n_a, c(n_a)) c(n_a) + \frac{\partial \tilde{R}_b^x}{\partial c} c'(n_a) c(n_a) g_b(c(n_a)|c) - \frac{d\tilde{R}_b^x}{dn_a} c(n_a) g_b(c(n_a)|c) \\
&= -\xi \frac{d}{dn_a}\left[J \frac{\partial \tilde{R}_b^x}{\partial J} g_b(c(n_a)|c) c(n_a)\right] \qquad (54)
\end{aligned}
$$

Integrating (54) from n_a' to $n_a \in (n_a', n_a'')$ we then obtain that

$$
\begin{aligned}
&\int_{n_a'}^{n_a} \left[R_a^x(\tilde{n}_a) - \tilde{R}_b^x(\tilde{n}_a)\right] c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)) d\tilde{n}_a \qquad (55) \\
&+ \int_{n_a'}^{n_a} \frac{\partial \tilde{R}_b^x(\tilde{n}_a)}{\partial c} c(\tilde{n}_a) c'(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a - \int_{n_a'}^{n_a} \frac{d\tilde{R}_b^x(\tilde{n}_a)}{dn_a} c(\tilde{n}_a) g_b(c(\tilde{n}_a)|c) d\tilde{n}_a \\
&= -\xi J(n_a) \frac{\partial \tilde{R}_b^x(n_a)}{\partial J} g_b(c(n_a)|c) c(n_a) + \lim_{n_a \to n_a'} \xi J(n_a) \frac{\partial \tilde{R}_b^x(n_a)}{\partial J} g_b(c(n_a)|c) c(n_a)
\end{aligned}
$$

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 (55). 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 (55).

We now express each term in (55) 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 (38) and (45), the first term in (55) is simply:

$$
\begin{aligned}
&\int_{n_a'}^{n_a} \left[R_a^x(\tilde{n}_a) - \tilde{R}_b^x(\tilde{n}_a)\right] c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)) d\tilde{n}_a \\
&= \lambda \int_{n_a'}^{n_a} \left[T_a(y_a(\tilde{n}_a)) - T_b(y_b(c(\tilde{n}_a)))\right] c(\tilde{n}_a) f(\tilde{n}_a, c(\tilde{n}_a)) d\tilde{n}_a. \qquad (56)
\end{aligned}
$$

The second term in (55) is obtained by differentiating (45) with respect to c, which yields

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

$$= \lambda \int_{n_a'}^{n_a} y_b(c(\tilde{n}_a))c'(\tilde{n}_a)g_b(c(\tilde{n}_a)|c)d\tilde{n}_a = \lambda \int_{c(n_a')}^{c(n_a)} y_b(n_b)g_b(n_b|c)dn_b. \tag{57}$$

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 (55) is obtained by totally differentiating (45) with respect to n_a, which gives

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

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

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

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

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

where the last two equalities use (9), and (11). Changing again the variables of integration using the relation $n_b = c(n_a)$, we then obtain that the third term in (55) 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 (55) is obtained by differentiating (45) 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 (55) can be rewritten as

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

$$= \xi \frac{T_b'(y_b\left(c(n_a)\right))}{1 - T_b'(y_b(c(n_a)))} c(n_a)g_b(c(n_a)|c) \cdot \left\{ (1 - T_b'(y_b(c(n_a))))y_b(c(n_a) \right\}. \tag{58}$$

Substituting (56)-(58) into (55) and rearranging yields (12). Q.E.D.

Proof of Proposition 3 The proof has two parts, each establishing the result in the corresponding part of 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 (12), over a positive measure set of types. To see this, first use (11) to observe that, in any equilibrium sustaining production efficiency, $h_a(n) = h_b(n) = h(n)$ for all $n \in N$. Then use (9) and (11) 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:

$$\mathbf{1}_x^{CU} \int_{\underline{n}}^{n} m(\tilde{n}) \left[1 - T'(y(\tilde{n}))\right] y(\tilde{n}) dG_b(\tilde{n}|c) = \int_{\underline{n}}^{n} \left[1 - T'(y(\tilde{n})\varepsilon_y(\tilde{n})\right] y(\tilde{n}) dG_b(\tilde{n}|c) + \xi T'(y(n)) y(n) n g_b(n|c), \tag{59}$$

$$\mathbf{1}_x^{CU} \int_{\underline{n}}^{n} m(\tilde{n}) \left[1 - T'(y(\tilde{n}))\right] y(\tilde{n}) dG_a(\tilde{n}|c) = \int_{\underline{n}}^{n} \left[1 - T'(y(\tilde{n})\varepsilon_y(\tilde{n})\right] y(\tilde{n}) dG_a(\tilde{n}|c) + \xi T'(y(n)) y(n) n g_a(n|c), \tag{60}$$

where we used the 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 (59) and (60) define two linear differential equations in $g_b(n|c)$ and $g_a(n|c)$, respectively. Because the usual Lipschitz conditions hold (as implied by Lemma 1), the Picard-Lindelof theorem implies that each of these equations has a unique solution satisfying some associated boundary condition $g_j(n|c) = \mathring{g}_j$, for $n > \underline{n}$. Because the differential equations defined by (59) and (60) are homogenous, their solutions must satisfy $g_b(n|c) = \delta g_a(n|c)$ for $\delta = \frac{\mathring{g}_b}{\mathring{g}_a}$. 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, we have 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 (44). 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 (22) has to hold for sector b. Finally, that condition (22) holds for sector a follows from the Mirrlees formula (10), after setting the indicator function to zero. Q.E.D.

Proof of Proposition 4. Note that, with uniform labor income taxation and our convention about the labeling of the two sectors (which consists in assuming that $\tau_a \leq \tau_b$), $n_a' = \underline{n}$. 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

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

where

$$u_a(n_a) = u_a(\underline{n}) + \int_{\underline{n}}^{n_a} \psi'\left(h_a(\tilde{n}_a)\right) \frac{h_a(\tilde{n}_a)}{\tilde{n}_a} d\tilde{n}_a \tag{61}$$

is determined by (9) and where the densities under the occupational choice rule corresponding to the sale tax τ_a are given by

$$g_a(n_a|c) = f_a(n_a) F_{b|a}((1-\tau_a)n_a)|n_a) \tag{62}$$

and

$$g_b(n_b|c) = f_b(n_b)F_{a|b}((1-\tau_a)^{-1}n_b)|n_b). \tag{63a}$$

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

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

Then, for any $n_a \in N_a$ let

$$R_a^x\left(n_a\right) \equiv \mathbf{1}_x^{CU}\phi\left(u_a\left(n_a\right)\right) + \lambda\left[h_a\left(n_a\right)n_a - \psi\left(h_a\left(n_a\right)\right) - u_a\left(n_a\right)\right], \tag{64}$$

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

$$\begin{aligned} \hat{R}_b^x\left(n_a|c\right) &\equiv R_b^x\left(c(n_a)\right) \\ &= \mathbf{1}_x^{CU}\phi\left(u_b(c(n_a))\right) + \lambda\left\{h_b(c(n_a))c(n_a) - \psi\left(h_b(c(n_a))\right) - u_b\left(c(n_a)\right)\right\} \\ &= \mathbf{1}_x^{CU}\phi\left(u_a\left(n_a\right)\right) + \lambda\left\{J_c[n_a]h_a(n_a)c(n_a) - \psi\left(J_c[n_a]h_a(n_a)\right) - u_a\left(n_a\right)\right\} \\ &= \mathbf{1}_x^{CU}\phi\left(u_a\left(n_a\right)\right) + \lambda\left[h_a\left(n_a\right)(1-\tau_a)n_a - \psi\left(h_a\left(n_a\right)\right) - u_a\left(n_a\right)\right], \end{aligned} \tag{65}$$

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

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

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

$$\begin{aligned}
&\frac{d}{d\tau_a}\int_{\underline{n}}^{\bar{n}} R_a^x\left(n_a\right) g_a(n_a|c)dn_a \\
&+\frac{d}{d\tau_a}\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a|c\right)(1-\tau_a)g_b((1-\tau_a)n_a|c)dn_a=0.
\end{aligned} \tag{66}$$

The latter condition can be rewritten as

$$\begin{aligned}
&\frac{d}{d\tau_a}\int_{\underline{n}}^{\bar{n}} R_a^x\left(n_a\right) g_a(n_a|c)dn_a+ \\
&+\frac{\bar{n}}{1-\tau_a}\hat{R}_b^x\left(\frac{\bar{n}}{1-\tau_a}|c\right) g_b(\bar{n}|c) \\
&+\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \hat{R}_b^x\left(n_a|c\right)\frac{d}{d\tau_a}\left\{(1-\tau_a)g_b((1-\tau_a)n_a|c)\right\}dn_a \\
&+\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} \frac{d\hat{R}_b^x\left(n_a|c\right)}{d\tau_a}(1-\tau_a)g_b((1-\tau_a)n_a|c)dn_a=0.
\end{aligned} \tag{67}$$

Consider the first term in (67) and note that it is equal to

$$\begin{aligned}
\frac{d}{d\tau_a}\int_{\underline{n}}^{\bar{n}} R_a^x\left(n_a\right) g_a(n_a|c)dn_a &= \frac{d}{d\tau_a}\left\{\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} R_a^x\left(n_a\right) g_a(n_a|c)dn_a+\int_{\frac{\bar{n}}{1-\tau_a}}^{\bar{n}} R_a^x\left(n_a\right) f_a(n_a)dn_a\right\} \\
&= \int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} R_a^x\left(n_a\right)\frac{d}{d\tau_a}[g_a(n_a|c)]dn_a \\
&= -\int_{\underline{n}}^{\frac{\bar{n}}{1-\tau_a}} R_a^x\left(n_a\right) n_a f(n_a,(1-\tau_a)n_a)dn_a
\end{aligned} \tag{68}$$

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 (62).

Next, consider the third term in (67) and use (63a) 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 (67) 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

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

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

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

We conclude that the third term in (67) is thus equal to

$$
\begin{aligned}
&\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\hat{R}_b^x\left(n_a;c\right)\frac{d}{d\tau_a}\left\{(1-\tau_a)g_b((1-\tau_a)n_a|c)\right\}dn_a \qquad (69)\\
&=-\lim_{n_a\to\bar{n}}\frac{n_a}{1-\tau_a}\hat{R}_b^x\left(\frac{n_a}{1-\tau_a}|c\right)g_b(n_a|c) \\
&+\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\left\{\left[\mathbf{1}_x^{CU}\phi'\left(u_a(n_a)\right)-\lambda\right]u_a'\left(n_a\right)+\lambda\left[h_a'\left(n_a\right)n_a\left(1-\tau_a\right)+h_a\left(n_a\right)\left(1-\tau_a\right)-\psi'\left(h_a\left(n_a\right)\right)h_a'\left(n_a\right)\right]\right\}\cdot \\
&\cdot n_a g_b((1-\tau_a)n_a|c)dn_a \\
&+\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\hat{R}_b^x\left(n_a;c\right)n_a f\left(n_a,(1-\tau_a)n_a\right)dn_a.
\end{aligned}
$$

Finally, consider the fourth term in (67). Differentiating (65) with respect to τ_a, we can show that this term is equal to

$$
\begin{aligned}
&\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\frac{d\hat{R}_b^x\left(n_a;c\right)}{d\tau_a}(1-\tau_a)g_b((1-\tau_a)n_a|c)dn_a \qquad (70)\\
&=-\lambda\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}h_a\left(n_a\right)n_a(1-\tau_a)g_b((1-\tau_a)n_a|c)dn_a.
\end{aligned}
$$

Substituting (68), (69) and (70) into (67) and simplifying, we obtain that the optimality condition can be rewritten as

$$\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{ \left[\mathbf{1}_x^{CU} \phi' \left(u_a(n_a) \right) - \lambda \right] u_a' \left(n_a \right) + \lambda \left[h_a' \left(n_a \right) n_a \left(1 - \tau_a \right) - \psi' \left(h_a \left(n_a \right) \right) h_a' \left(n_a \right) \right] \right\} \cdot \tag{71}$$

$$\cdot\, n_a g_b((1-\tau_a)n_a|c) dn_a$$

$$= \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{ R_a^x \left(n_a \right) - \hat{R}_b^x \left(n_a ; c \right) \right\} n_a f(n_a, (1-\tau_a)n_a) dn_a$$

Using the fact that

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

we then have that (71) can be rewritten as

$$\mathbf{1}_x^{CU} \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} m_a \left(n_a \right) u_a' \left(n_a \right) n_a g_b((1-\tau_a)n_a|c) dn_a \tag{72}$$

$$+ \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{ -u_a' \left(n_a \right) + h_a' \left(n_a \right) n_a \left(1 - \tau_a \right) - \psi' \left(h_a \left(n_a \right) \right) h_a' \left(n_a \right) \right\} n_a g_b((1-\tau_a)n_a|c) dn_a$$

$$= \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} [\hat{T}_a(y_a(n_a) - \hat{T}_b(y_b(c(n_a)))] n_a f(n_a, (1-\tau_a)n_a) dn_a$$

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

Using (9), we can then rewrite the first integral in (72) as follows:

$$\begin{aligned} &\mathbf{1}_x^{CU} \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} m_a \left(n_a \right) u_a' \left(n_a \right) n_a g_b((1-\tau_a)n_a|c) dn_a \\ = &\mathbf{1}_x^{CU} \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} m_a \left(n_a \right) \left[1 - T'((1-\tau_a)y_a(n_a)) \right] (1-\tau_a) y_a(n_a) g_b((1-\tau_a)n_a|c) dn_a \end{aligned} \tag{73}$$

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, we can rewrite the second integral in (72) as follows:

$$\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{ -u_a' \left(n_a \right) + h_a' \left(n_a \right) n_a \left(1 - \tau_a \right) - \psi' \left(h_a \left(n_a \right) \right) h_a' \left(n_a \right) \right\} n_a g_b((1-\tau_a)n_a|c) dn_a \tag{74}$$

$$= \int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} \left\{ \begin{array}{c} -[1 - T'((1-\tau_a)y_a(n_a))](1-\tau_a)h_a(n_a) \\ +h_a' \left(n_a \right) n_a \left(1 - \tau_a \right) - \left[1 - T'((1-\tau_a)y_a(n_a)) \right] n_a (1-\tau_a) h_a' \left(n_a \right) \end{array} \right\} \cdot$$

$$\cdot\, n_a g_b((1-\tau_a)n_a|c) dn_a$$

Using the fact that

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

we can rewrite (74) as follows:

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

We conclude that (72) can be rewritten as

$$\mathbf{1}_x^{CU}\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}} m_a\left(n_a\right)\left[1-T'((1-\tau_a)y_a(n_a))\right](1-\tau_a)y_a(n_a)g_b((1-\tau_a)n_a|c)dn_a \tag{75}$$
$$=\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}\left\{1-T'((1-\tau_a)y_a(n_a))\frac{y_a'\left(n_a\right)n_a}{y_a(n_a)}\right\}(1-\tau_a)y_a(n_a)g_b((1-\tau_a)n_a|c)dn_a$$
$$+\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}[\hat{T}_a(y_a(n_a)-\hat{T}_b(y_b(c(n_a)))]n_a f(n_a,(1-\tau_a)n_a)dn_a$$

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 (75) as

$$\mathbf{1}_x^{CU}\int_{\underline{n}(1-\tau_a)}^{\overline{n}} m_b\left(n_b\right)\left[1-T'(y_b(n_b))\right]y_b(n_b)g_b(n_b|c)dn_b \tag{76}$$
$$=\int_{\underline{n}(1-\tau_a)}^{\overline{n}}\left\{1-T'(y_b(n_b))\varepsilon_{y_b}(n_b)\right\}y_b(n_b)g_b(n_b|c)dn_b$$
$$+\int_{\underline{n}}^{\frac{\overline{n}}{1-\tau_a}}[\hat{T}_a(y_a(n_a)-\hat{T}_b(y_b(c(n_a)))](1-\tau_a)n_a f(n_a,(1-\tau_a)n_a)dn_a$$

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 (76) 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, *American Economic Review*, 68, 8-19.

[2] Ales, L., M. Kurnaz and C. Sleet (2015), Technical Change, Wage Inequality, and Taxes, *American Economic Review*, 105, 3061-3101.

[3] Atkinson, A. and J. Stiglitz (1976), The Design of the Tax Structure: Direct Versus Indirect Taxation, *Journal of Public Economics*, 6, 55-75.

[4] Bakija, J., A. Cole and B. T. Heim (2012), Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from US Tax Return Data, *Working Paper.*

[5] Bell, Brian and J. Van Reenen (2014), Bankers and Their Bonuses, *Economic Journal*, 124, F1-F21.

[6] Best, M. A. Brockmeyer, H. Kleven, J. Spinnewijn and M. Waseem (2015), Production vs Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan*, Journal of Political Economy 1*23(6), 1311-1355.

[7] Bloom, N., D. J. Price, F. Guvenen, J. Song, T. von Wachterk (2016), Firming Up Inequality, *Working Paper University of Minnesota.*

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

[9] Calzolari, G. and V. Denicolò (2013), Competition with Exclusive Contracts and Market-share Discounts, *American Economic Review* 103, 2384-2411.

[10] Calzolari, G. and V. Denicolò (2015), Exclusive Contracts and Market Dominance, *American Economic Review*, 3321-3351.

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

[12] Choné, P., and L. Linnemer (2015), Nonlinear Prices and Exclusion: I. Buyer Opportunism, (2015) *Rand Journal of Economics*, 46 (2) 217-40.

[13] 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.

[14] Denk, O. (2015), Financial Sector Pay and Labour Income Inequality: Evidence From Europe, *OECD Economics Department Working Papers*, No. 1225, OECD Publishing, Paris.

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

[16] Diamond, P. and J. Mirrlees (1971), Optimal Taxation and Public Production 1: Production Efficiency, *American Economic Review*, 61, 8-27.

[17] Godechot, O. (2012), Is Finance Responsible for the Rise in Wage Inequality in France?, *Socio-Economic Review,* 10(3), 447-470.

[18] Golosov M., M. Troshkin, A. Tsyvinski, and M. Weinzierl (2013), Preference Heterogeneity and Optimal Capital Income Taxation, *Journal of Public Economics*, 97, 160–175.

[19] Guvenen, F. and B. Kuruscu (2012), Understanding the Evolution of the U.S. Wage Distribution: A Theoretical Analysis, *Journal of European Economic Association*, 10(3), 482-517.

[20] 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.

[21] 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.

[22] Heckman, J. and B. Honoré (1990), The Empirical Content of the Roy Model, *Econometrica*, 42, 1121-1149.

[23] Jacquet, L. and E. Lehmann (2016), Optimal Taxation with Heterogeneous Skills and Elasticities: Structural and Sufficient Statistics Approaches, *Working Paper*.

[24] Kaplan S. N. and J. Rauh (2010), Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?, *Review of Financial Studies*, 23(3), 1004-1050.

[25] Kleven, H, C. Landais, E. Saez (2013) Taxation and International Migration of Superstars: Evidence from the European Football Market, *American Economic Review 10*3, 1892-1924.

[26] Kleven, H, C. Landais, E. Saez and E. Schultz (2014), Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark, *Quarterly Journal of Economics* 129, 333-378.

[27] 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.

[28] Mas-Colell, A., M. Whinston and J. Green (1995), Microeconomic Theory, Oxford University Press.

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

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

[31] Mirrlees, J. (1971), An Exploration in the Theory of Optimum Income Taxation, *Review of Economic Studies*, 38, 175-208.

[32] 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.

[33] Neal, D. (1995), Industry-Specific Human Capital: Evidence from Displaced Workers, *Journal of Labor Economics*, 13(4), 653-677.

[34] Philippon, T. and A. Reshef (2012), Wages and Human Capital in the US Finance Industry: 1909-2006, *Quarterly Journal of Economics*, 127(4), 1551-1609.

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

[36] Roy, A. (1951), Some Thoughts on the Distribution of Earnings, *Oxford Economic Papers*, 3, 135-146.

[37] Saez, E. (2001), Using Elasticities to Derive Optimal Tax Rates, *Review of Economic Studies*, 68, 205-229.

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

[39] Saez, E. (2004), Direct or Indirect Tax Instruments for Redistribution: Short-run Versus Long-run, *Journal of Public Economics*, 88, 503-518.

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

[41] Rothschild, C. and F. Scheuer (2016), Optimal taxation with Rent-Seeking, *Review of Economic Studies*, 83, 1225-1262.

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

[43] Scheuer, F. and I. Werning (2016), Mirrlees meets Diamond-Mirrlees: Simplifying Nonlinear Income Taxation, *Working Paper Stanford University.*