# Redistributive Taxation in the Roy Model*

**Casey Rothschild**
Wellesley College

**Florian Scheuer**
Stanford University and NBER

Revised Version: January 2013

**Abstract**

We consider optimal redistribution in a model where individuals can self-select into one of several possible sectors based on heterogeneity in a multidimensional skill vector. We first show that when the government does not observe the sectoral choice or underlying skills of its citizens, the constrained Pareto frontier can be implemented with a single non-linear income tax. We then characterize this optimal tax schedule. If sectoral inputs are complements, a many-sector model with self-selection leads to optimal income taxes that are less progressive than the corresponding taxes in a standard single-sector model under natural conditions. However, they are more progressive than in canonical multi-sector economies with discrete types and without occupational choice or overlapping sectoral wage distributions.

Keywords: Multi-dimensional Screening, Optimal Taxation, Occupational Choice

JEL-codes: H2, D5, D8, E2, E6, J3, J6

*Email addresses: crothsch@wellesley.edu, scheuer@stanford.edu. We are grateful to the editor, Robert Barro, three anonymous referees, as well as Craig Brett, Philippe Choné, Emmanuel Farhi, John Weymark, and seminar participants at the 2012 Taxation Theory Conference (Vanderbilt), SED Annual Meeting (Limassol), NTA Annual Meeting (Providence), 2013 ASSA Meeting (San Diego), University of Michigan and LMU Munich for helpful comments. All remaining errors are our own. This paper is a significantly revised version of the earlier working paper that circulated under the title "Entrepreneurial Taxation and Occupational Choice."

# 1 Introduction

The Roy (1951) model of self-selection is one of the workhorse models in labor economics. It has been used to study immigration and locational choice (Borjas, 1987, Dahl, 2002), schooling (Willis and Rosen, 1979), choice of occupation or industry (Heckman and Sedlacek, 1985, 1990), employment in union versus non-union (Lee, 1978) and private versus public sectors (Borjas, 2002), female labor force participation (Heckman, 1974), training program participation (Ham and LaLonde, 1996), and the growth-retarding impact of racial and gender discrimination in labor markets (Hsieh et al., 2011), for example. Its essential feature is that individuals optimally self-select into one of several sectors based on which one affords them the highest returns. One would expect this sort of self-selection to have important implications for the design of redistributive income taxes. It is surprising, then, that these implications have not been studied formally heretofore. This paper takes a step towards understanding them by analyzing optimal Mirrleesian income taxation in a two-sector Roy model.

Incorporating self-selection à la Roy in an optimal taxation framework à la Mirrlees raises some challenges. In the Mirrleesian approach, the government effectively uses income taxes to screen individuals based on their unobserved skill (or wage). When individuals can choose among multiple sectors, the underlying skill is naturally multi-dimensional: each individual has a skill in each possible sector. It is well-known that multi-dimensional screening problems are typically challenging (Rochet and Choné, 1998). We show that the particular screening problem that arises in a many-sector model of optimal income taxation is tractable despite the underlying multi-dimensional heterogeneity.

We do this by demonstrating that a single non-linear income tax schedule is sufficient for implementing any incentive compatible allocation if the government cannot or does not want to condition taxes on sectoral choice. With a single tax schedule, allocations only depend on the realized wage, not on sectoral choice per se, and hence the multi-dimensional screening problem "collapses" into an almost standard single-dimensional "screening on wages" problem. The tools developed by Mirrlees (1971) and others therefore apply, with one important difference: the wage distribution will typically be *endogenous* when there are many sectors to choose from. This is because the productivity of effort in any given sector will, in general, depend on the aggregate effort expended in this and in other sectors. We characterize optimal taxes accounting for this endogeneity and show that, under natural and general assumptions, it implies a force for *less* progressive taxation relative to a world with a single sector and an exogenous wage distribution.

The basic intuition for this result can be understood as follows. Suppose that there are

two sectors, a "blue collar" and a "white collar" sector, and individuals are free to choose to work in either. The government aims at redistributing from high- to low-income individuals and designs an income tax system accordingly. For administrative, informational, or political reasons, it may not distinguish between the two sectors: a white collar and blue collar worker earning the same income $y$ pay the same tax $T(y)$. Even if sectoral choice is observable, such a restriction could result from the fact that individuals can rather easily relabel their type of occupation and would have strong incentives to do so in the presence of differential taxation, leading to distortions. Unless the government collects detailed information about the kind of tasks that individuals do, such a reclassification may be hard or costly to prevent, which motivates our focus on uniform income taxation. It also corresponds to the actual tax systems in many countries, where the income tax is typically not sector- or occupation-specific.[1]

With a linear production technology (i.e., when equivalent units of white and blue collar efforts are perfect substitutes), the fact that there are two sectors would be irrelevant (unless the government has an intrinsic sectoral preference). Individuals would choose to work in the sector in which they are more productive, as reflected in their wage, and this choice and the resulting wage distribution would be independent of tax policy. The optimal tax would therefore be exactly the same as it would be in a single-sector Mirrlees model with the same wage distribution.

Contrast this with the case in which the two sectors are gross complements. In this case, sectoral choices and wages will be endogenous to tax policy. Lowering taxes at income levels that are dominated by white collar workers, for example, will differentially encourage white collar effort. This will reduce the marginal productivity of white collar effort (assuming diminishing marginal products within a given sector) and raise the marginal productivity of blue collar effort (by complementarity across sectors). It therefore *indirectly redistributes* from white to blue collar workers.

Suppose now that the blue collar sector is the low-income sector, i.e., that there are disproportionately more white collar workers at higher incomes. Then this indirect redistribution channel will lead the government to choose a tax system which is less progressive than in a Mirrleesian world with exogenous wages: Lowering taxes on high earners will disproportionately spur effort in the white collar sector, which will indirectly redistribute from the relatively high-income white collar workers to the lower-income blue collar workers by raising blue collar wages and lowering white collar wages. Similarly,

[1]There are exceptions to this pattern, e.g. the tax treatment of the self-employed versus employed in some countries. See Scheuer (2012a,b) for an analysis of differential taxation of entrepreneurs versus workers.

raising taxes on lower earners will differentially discourage effort in the blue collar sector, again increasing their wage. This indirect redistribution channel (sometimes referred to as "trickle down" effects, since lower earners can benefit from tax cuts on higher earners) therefore leads to less progressive taxes. For example, it implies an optimal top income tax rate that will generally be negative when the skill distribution is bounded—i.e., strictly below the well-known zero top rate result.

This result does not say taxes should be regressive *per se*. Rather, it says that optimal taxes will be *less* progressive than they would be in the alternative allocation that would obtain if the endogeneity of wages implied by a multi-sector Roy model were neglected. Making such a comparison requires formalizing this alternative allocation. We use the notion of a self-confirming policy equilibrium (SCPE), developed for a different context by Rothschild and Scheuer (2011), for this. An SCPE describes the tax system that would emerge in the same economy if the government assumed that it was operating in a standard exogenous-wage world. In such a world, a government would, following the standard approach in public finance pioneered by Saez (2001), infer an underlying skill distribution from the income distribution that it observes given an existing tax system. Taking this skill distribution as given, it would then design the optimal income tax system. In an SCPE, this newly computed optimal income tax system would coincide with the existing tax system, thus "confirming" its optimality. Our results show that taxes in such an SCPE are not, in fact, optimal in a multi-sector economy, since the wage-cum-skill distribution is not, in fact, exogenous. In particular, the optimal taxes would be less progressive.

**Related Literature.** Most closely related to our analysis is Stiglitz (1982), who considers optimal nonlinear taxation in a two-type model with endogenous wages but without occupational choice. He also shows that progressive redistributional motives will lead the optimal top marginal tax rate to be negative when the two types' efforts are complements. The indirect redistribution channel driving his results are similar to those driving ours. Our model differs in two significant ways, however. First, our continuous type model allows us to study the progressivity of the entire tax schedule, rather than just the top marginal tax rate. Second, we identify several *extra* effects that arise in a general Roy model, effects which result from (i) endogenous occupational choice and (ii) the fact that the sectoral wage distributions will typically overlap in a general model with continuous types, whereas Stiglitz's discrete type model generically—and somewhat unrealistically—rules out workers in different sectors earning the same wage.

We show that these extra effects mitigate the general equilibrium effects of taxation found in Stiglitz (1982) and therefore make optimal taxes *more* progressive than in a dis-

crete type model without occupational choice. To understand why, suppose we reduce taxes at the top to increase the effort of the top earners. This is desirable insofar as it indirectly redistributes from high to low incomes by raising the wages of workers in the low wage sector and lowering the wages of workers in the high wage sector. When there is endogenous occupational choice, however, there is an additional effect: the change in wages leads some individuals to shift out of the high wage into the low wage sector. This undoes some of the original increase in aggregate effort in the high wage sector and blunts the desirable effects of the original reduction in taxes. As a result, the optimal progressivity of the tax schedule in our general Roy model is bounded between a standard Mirrlees model at the progressive end and Stiglitz's (1982) model at the regressive end.

Naito (1999) has studied the role of sector-specific income taxes in Stiglitz's (1982) two-type model and observed that introducing additional commodity taxes or production inefficiencies can be desirable when it relaxes incentive constraints by manipulating relative wages, thus contradicting the results by Atkinson and Stiglitz (1976) and Diamond and Mirrlees (1971). We do not consider sector-specific or commodity taxation in this paper. Saez (2004) has contrasted Stiglitz's model (with fixed occupations but flexible hours) to a model with fixed hours but flexible occupational choice and shown that the standard Diamond and Mirrlees (1971) optimal tax formulas apply and that Naito's results cease to hold in the latter setting. Saez's results depend on the assumption that each wage corresponds to a unique occupation and there is no intensive margin. In contrast, we allow the wage distributions of different occupations to overlap and consider both endogenous occupational choice and labor supply.

Our paper also relates to earlier research on optimal income taxation in models with endogenous wages and occupational choice, such as Feldstein (1973), Zeckhauser (1977), Allen (1982), Boadway et al. (1991), and Parker (1999). This literature has largely restricted attention to linear taxation. An exception is the work by Moresi (1997), who considers non-linear taxation of profits in a model of occupational choice between workers and entrepreneurs. The occupational choice margin in his model is considerably simplified, however, and heterogeneity is confined to affect one occupation only, not the other.

Restricting heterogeneity to affect one occupation only, or tax schedules to be linear, sidesteps the complexities of multidimensional screening, which emerge naturally in the present model. Few studies in the optimal taxation literature have attempted to deal with multidimensional screening problems until recently. Kleven, Kreiner, and Saez (2009) have made progress along these lines in a study of the optimal income taxation of couples and Scheuer (2012b) has considered the taxation of entrepreneurs and workers with occupational choice, but the second dimension of heterogeneity in their models takes a

specialized additive form.[2] Choné and Laroque (2010) and Brett and Weymark (2003) use an exogenous "type aggregator" akin to our technique, which can be interpreted as an *endogenous* type aggregator. Moreover, the types of heterogeneity they study (tastes for labor and education, respectively) are again quite distinct from the multi-dimensional skill types that arise naturally in our Roy model.

The methodological approach that we pursue here to characterize redistributive taxation in the Roy model builds on Rothschild and Scheuer (2011), who consider optimal income taxation in an economy with a productive and a rent-seeking sector. While that paper shares the overall structure of two-dimensional skill heterogeneity and occupational choice between two sectors, the emphasis is on the *corrective* role of income taxation in a setting where wages deviate from the social marginal product of effort due to rent-seeking externalities, issues that are absent from the present framework.

More generally, this paper follows the large optimal income taxation literature building on the seminal contributions of Mirrlees (1971) and Diamond (1998). Until recently, the theoretical literature focused on deriving results for a given skill distribution and social welfare function. Saez (2001) instead inferred skills and optimal taxes from the observed income distribution, and Laroque (2005), Werning (2007), and Choné and Laroque (2010) explored the conditions under which an existing set of taxes is potentially Pareto efficient. In the same spirit, we characterize Pareto efficient tax policies rather than focusing on a particular social welfare function. With multiple complementary sectors, however, the wage distribution is endogenous to the tax code, so existing tests for optimality—e.g., Werning (2007), who infers wage-*cum*-skill distributions from income distributions as a test of optimality—are potentially misleading. One might conclude that the tax code is indeed Pareto efficient given the inferred skill distribution under the (implicit and incorrect) assumption that the skill distribution is independent of taxes. Our concept of an SCPE, described above, captures this situation. It is related to the literature on self-confirming equilibria in learning models (e.g. Sargent, 2009, and Fudenberg and Levine, 2009).

The paper proceeds as follows. Section 2 describes the basic model and shows that a single non-linear income tax is a fully general policy tool for a government who observes income but not effort, wage, or sectoral choice. Section 3 characterizes and compares optimal and SCPE non-linear taxes. Section 4 compares our results to Stiglitz (1982) and points out the novel role of occupational choice and overlapping wage distributions in our model. Section 5 illustrates our theoretical results with a simple empirical calibration. Section 6 extends our results to allow for unbounded skill distributions and additional across-occupation cost heterogeneity. Section 7 concludes. All proofs are in the Appendix.

[2] See also Brett (2007) for a four type model of family taxation with two-dimensional heterogeneity.

# 2 The Model

## 2.1 Setup

We consider an economy where a unit mass of individuals can choose between working in either of two sectors. Accordingly, individuals have a two-dimensional skill type $(\theta, \varphi) \in \Theta \times \Phi$ with $\Theta = [\underline{\theta}, \overline{\theta}]$ and $\Phi = [\underline{\varphi}, \overline{\varphi}]$. $\theta$ captures an individual's productivity when working in the $\Theta$-sector and $\varphi$ captures her $\Phi$-sector skill. These skills are distributed in the population according to a continuous two-dimensional cdf $F(\theta, \varphi)$ with density $f(\theta, \varphi)$. In section 6, we will demonstrate how our results extend to the case where individuals also differ in their cost of working in one of the sectors, capturing, e.g., differential education requirements or tastes for different occupations.

Individuals have preferences over consumption $c$ and effort $e$ captured by the strictly concave utility function $U(c, e)$ with $U_c > 0$, $U_e < 0$. We denote the consumption, effort, utility, and sector assigned to an individual of type $(\theta, \varphi)$ by $c(\theta, \varphi)$, $e(\theta, \varphi)$, $V(\theta, \varphi) \equiv U(c(\theta, \varphi), e(\theta, \varphi))$, and $S(\theta, \varphi) \in \{\Theta, \Phi\}$, respectively.

The technology in the economy is described by a constant returns to scale aggregate production function $Y(E_\theta, E_\varphi)$ that combines the skill-weighted aggregate effort in the two sectors to produce the consumption good. Formally, aggregate efforts are given by

$$E_\theta \equiv \int_{P \subset \Theta \times \Phi} \theta e(\theta, \varphi) dF(\theta, \varphi) \quad \text{and} \quad E_\varphi \equiv \int_{\Theta \times \Phi \setminus P} \varphi e(\theta, \varphi) dF(\theta, \varphi)$$

with $P \equiv \{(\theta, \varphi) | S(\theta, \varphi) = \Theta\}$. Since the technology is linear homogeneous, the marginal products only depend on the ratio of aggregate effort in the two sectors $x \equiv E_\theta / E_\varphi$ and are therefore denoted by $Y_\theta(x)$ and $Y_\varphi(x)$. We define an individual's wage $w$ as the marginal return to effort, so that

$$w(\theta, \varphi) = \begin{cases} Y_\theta(x)\theta & \text{if} \quad S(\theta, \varphi) = \Theta, \\ Y_\varphi(x)\varphi & \text{if} \quad S(\theta, \varphi) = \Phi. \end{cases} \tag{1}$$

An individual's income is then given by $y(\theta, \varphi) \equiv w(\theta, \varphi) e(\theta, \varphi)$. As is standard, we assume $U(c, e)$ to satisfy the single-crossing property, i.e. the marginal rate of substitution between income and consumption, $-U_e(c, y/w)/(wU_c(c, y/w))$, is decreasing in $w$.

## 2.2 Implementation

We start by characterizing a general, direct implementation, where individuals announce their privately known type $(\theta, \varphi)$ and then get assigned consumption $c(\theta, \varphi)$, income

$y(\theta,\varphi)$ and a sector to work in $S(\theta,\varphi)$. Assuming that income and consumption are observable but not sectoral choice, wage, or effort, the resulting incentive constraints that guarantee truth-telling of the agents are as follows. First, suppose $S(\theta,\varphi)=\Theta$, i.e. type $(\theta,\varphi)$ is assigned to the $\Theta$-sector. Then incentive compatibility requires that

$$U\left(c(\theta,\varphi),\frac{y(\theta,\varphi)}{Y_\theta(x)\theta}\right)\geq\max\left\{U\left(c(\theta',\varphi'),\frac{y(\theta',\varphi')}{Y_\theta(x)\theta}\right),U\left(c(\theta',\varphi'),\frac{y(\theta',\varphi')}{Y_\varphi(x)\varphi}\right)\right\}\quad\forall\theta',\varphi'$$

since there are two ways for type $(\theta,\varphi)$ to imitate another type $(\theta',\varphi')$, namely by earning $(\theta',\varphi')$'s income either in the $\Theta$- or the $\Phi$-sector. Analogously, if $S(\theta,\varphi)=\Phi$, we need

$$U\left(c(\theta,\varphi),\frac{y(\theta,\varphi)}{Y_\varphi(x)\varphi}\right)\geq\max\left\{U\left(c(\theta',\varphi'),\frac{y(\theta',\varphi')}{Y_\theta(x)\theta}\right),U\left(c(\theta',\varphi'),\frac{y(\theta',\varphi')}{Y_\varphi(x)\varphi}\right)\right\}\quad\forall\theta',\varphi'.$$

The assumption that only income, but not the wage or effort is observable is standard to optimal income tax models since Mirrlees (1971). Even though sectoral choice may be observable in some applications, it is worth emphasizing again that conditioning allocations on occupations may be hard to enforce in practice. This motivates our approach of treating occupational choice as de facto non-contractible for the government. The following lemma shows that, under these assumptions, any incentive compatible allocation can be decentralized by offering a single non-linear income tax $T(y)$.

**Lemma 1.** *Suppose that only income y and consumption c are observable, whereas an individual's skill type $(\theta,\varphi)$, effort e and sectoral choice S are private information. Then for any incentive compatible allocation $\{c(\theta,\varphi),y(\theta,\varphi),S(\theta,\varphi),x\}$, the following properties hold:*
*(i)* $w(\theta,\varphi)=\max\{Y_\theta(x)\theta,Y_\varphi(x)\varphi\}$ *and*

$$S(\theta,\varphi)=\begin{cases}\Theta & \text{if}\quad Y_\theta(x)\theta>Y_\varphi(x)\varphi,\\ \Phi & \text{if}\quad Y_\varphi(x)\varphi>Y_\theta(x)\theta;\end{cases}\tag{2}$$

*(ii)* $U(c(\theta,\varphi),y(\theta,\varphi)/w)=U(c(\theta',\varphi'),y(\theta',\varphi')/w)$ *for all* $(\theta,\varphi),(\theta',\varphi')$ *such that* $w(\theta,\varphi)=w(\theta',\varphi')=w$;
*(iii)* $\{c(\theta,\varphi),y(\theta,\varphi),S(\theta,\varphi),x\}$ *can be implemented by offering the single non-linear income tax schedule* $T^*(y)$ *corresponding to the retention function* $R^*(y)=y-T^*(y)$ *defined by*

$$R^*(y)\equiv\max_c\left\{c\,\middle|\,U\left(c(\theta,\varphi),\frac{y(\theta,\varphi)}{w(\theta,\varphi)}\right)\geq U\left(c,\frac{y}{w(\theta,\varphi)}\right)\forall(\theta,\varphi)\in\Theta\times\Phi\right\}\tag{3}$$

*and letting all agents choose one of their most preferred* $(c,y)$*-bundles from the resulting budget set* $B^*=\{(c,y)|c\leq y-T^*(y)\}$.

*Proof.* In Appendix A. □

In other words, individuals work in the sector in which they earn a higher wage, any two individuals with the same wage achieve the same utility, and the so-called *principle of taxation* extends to our setting. The first two properties follow directly from incentive compatibility. For part (iii), the retention function $R^*(y)$ constructed in (3) to implement a given incentive compatible allocation is the lower envelope of the indifference curves of all individuals through the consumption-income bundle assigned to them. Given these properties, the direct mechanism boils down to allocating $(c, y)$-pairs to a set of individuals with different wages, as in a standard optimal taxation problem.

Lemma 1 does not rule out the possibility that two individuals with the same wage (e.g. in different sectors) choose different consumption-income bundles, even though, by property (ii), these bundles must be located on the same indifference curve in the $(c, y)$-space. Since the type distribution $F(\theta, \varphi)$ is continuous, we can nevertheless restrict attention to allocations that pool all same-wage individuals at the same allocation. To see why, observe that for any $\hat{w}$ such that two types who both earn wage $\hat{w}$ receive distinct allocations $(c_1, y_1)$ and $(c_2, y_2)$, the retention function $R^*(y)$ described in (3) and the wage-$\hat{w}$ indifference curve must coincide between $y_1$ and $y_2$. The number of such wages $\hat{w}$ must therefore be countable and the measure of types at such wages must be zero. Pooling all individuals at each such $\hat{w}$ at a bundle $(R^*(y), y)$ with $y$ between $y_1$ and $y_2$ is therefore incentive compatible and resource feasible, and, moreover, affects neither the aggregate effort in either sector nor, a fortiori, $x$. Since pooling in this manner does not alter any type's utility, the Pareto frontier can be traced out by using only allocations $\{c(w), y(w), x\}$ that pool all same-wage individuals and where wages $w$ and sectoral choice $S(\theta, \varphi)$ are as described in Lemma 1. We focus on such allocations henceforth.

# 3 Characterizing Optimal Income Taxes

## 3.1 Pareto Optima and Self-Confirming Equilibria

For any given $x$, the marginal productivities in the two sectors, $Y_\theta(x)$ and $Y_\varphi(x)$, the two-dimensional skill distribution, $F(\theta, \varphi)$, and the implied sectoral choice described in (2) together induce a one-dimensional wage distribution characterized by the cdf and sectoral densities

$$F_x(w) \equiv F\left(\frac{w}{Y_\theta(x)}, \frac{w}{Y_\varphi(x)}\right) \quad \text{and}$$

$$f_x^\theta(w) = \frac{1}{Y_\theta(x)} \int_{\underline{\varphi}}^{w/Y_\varphi(x)} f\left(\frac{w}{Y_\theta(x)}, \varphi\right) d\varphi, \;\; f_x^\varphi(w) = \frac{1}{Y_\varphi(x)} \int_{\underline{\theta}}^{w/Y_\theta(x)} f\left(\theta, \frac{w}{Y_\varphi(x)}\right) d\theta,$$

with associated cdfs $F_x^\theta(w)$ and $F_x^\varphi(w)$ and with $f_x(w) = f_x^\theta(w) + f_x^\varphi(w)$. We also define the bottom and top wages given $x$ as

$$\underline{w}_x = \max\left\{Y_\theta(x)\underline{\theta}, Y_\varphi(x)\underline{\varphi}\right\}, \;\; \overline{w}_x = \max\left\{Y_\theta(x)\overline{\theta}, Y_\varphi(x)\overline{\varphi}\right\}.$$

To trace out the Pareto frontier, we assign Pareto weights $G(\theta, \varphi)$ in the two-dimensional skill space. As with the type distribution $F(\theta, \varphi)$, we can integrate the Pareto weights $G(\theta, \varphi)$ to obtain a distribution of Pareto weights over wages given $x$, denoted by $G_x(w)$, with the corresponding densities $g_x(w) = g_x^\theta(w) + g_x^\varphi(w)$ (and cdfs $G_x^\theta(w)$ and $G_x^\varphi(w)$).

Fixing $x$, the optimal income tax problem in a Roy model thus collapses to a one-dimensional screening problem despite the underlying two-dimensional heterogeneity in the economy. In particular, the Pareto problem is an almost standard Mirrlees problem with the additional constraint that the individuals' efforts and sectoral choices must be consistent with $x$. Formally, we must require that

$$\tilde{x}(x) \equiv \frac{\int_{\underline{w}_x}^{\overline{w}_x} \frac{w}{Y_\theta(x)} e(w) f_x^\theta(w) dw}{\int_{\underline{w}_x}^{\overline{w}_x} \frac{w}{Y_\varphi(x)} e(w) f_x^\varphi(w) dw} = x, \quad \text{or equivalently} \tag{4}$$

$$(1 - \alpha(x)) \int_{\underline{w}_x}^{\overline{w}_x} we(w) f_x^\theta(w) dw - \alpha(x) \int_{\underline{w}_x}^{\overline{w}_x} we(w) f_x^\varphi(w) dw = 0, \tag{5}$$

where $\alpha(x)$ denotes the aggregate income share of the $\Theta$-sector:

$$\alpha(x) \equiv \frac{xY_\theta(x)}{xY_\theta(x) + Y_\varphi(x)} = \frac{Y_\theta(x)E_\theta}{Y(E_\theta, E_\varphi)}.$$

We employ the standard approach of optimizing directly over $c(w), y(w)$-allocations. In fact, it is equivalent but formally more useful to optimize over $e(w), V(w)$-bundles, where $V(w) = U(c(w), e(w))$ and $e(w) = y(w)/w$. With this, we can write the Pareto problem for income taxation in the Roy model as

$$\max_x W(x) \equiv \max_{e(w), V(w)} \int_{\underline{w}_x}^{\overline{w}_x} V(w) dG_x(w) \tag{6}$$

subject to (5),

$$V'(w) + U_e(c(V(w), e(w)), e(w))\frac{e(w)}{w} = 0 \quad \forall w \in [\underline{w}_x, \overline{w}_x], \tag{7}$$

and

$$\int_{\underline{w}_x}^{\overline{w}_x} (we(w) - c(V(w), e(w)))\, f_x(w) dw \geq 0, \tag{8}$$

where $c(V, e)$ denotes the inverse of $U(.)$ w.r.t. its first argument. We refer to the three constraints (5), (7) and (8) as the consistency condition, the incentive constraints and the resource constraint, respectively.[3]

As suggested by the notation in (6), it is useful to decompose the Pareto problem (5) to (8) into an "inner" problem, given $x$, and an "outer" problem, which maximizes over $x$. Formally, fix some $x$ and let $W(x)$ denote the value of the objective (6) when maximizing over $V(w), e(w)$ subject to (5), (7) and (8) (the inner problem). Then the outer problem is simply $\max_x W(x)$.

For some of the subsequent analysis, it will be useful to restrict attention to solutions of the Pareto problem when the Pareto weights have the special form $G(\theta, \varphi) = \Psi(F(\theta, \varphi))$. It is straightforward to show that $g_x^\theta(w)/f_x^\theta(w) = g_x^\varphi(w)/f_x^\varphi(w)$ and $G_x(w) = \Psi(F_x(w))$ for such weights and any given $x$. The same welfare weight is thus assigned to any two individuals with the same (endogenously determined) wage and we can think of $\Psi$ as a measure of how much social welfare weight is attached to different quantiles of the (endogenous) wage distribution. We refer to such weights as *relative welfare weights* and use $\Psi$ instead of $G$ to distinguish this special case from the general one.

In order to compare the Pareto optimal tax schedules described so far with the corresponding optimal tax schedule in a standard Mirrlees model, we solve the following benchmark inner-outer problem decomposition using relative welfare weights $\Psi$. For a given $x$, we consider the relaxed inner problem:

$$\max_{e(w), V(w)} \int_{\underline{w}_x}^{\overline{w}_x} V(w) d\Psi(F_x(w)) \tag{9}$$

subject to (7) and (8)—i.e., we drop the consistency constraint (5) and solve a standard Mirrleesian problem taking the wage distribution $F_x(w)$ as given. To ensure that the re-

[3]Note that we have made use of the local version of the incentive constraints in (7). It is a standard result (see e.g. Fudenberg and Tirole (1991), Theorems 7.2 and 7.3) that, under the single-crossing condition, the local incentive constraints (7) together with the monotonicity constraint that income $y(w)$ must be non-decreasing in $w$ are equivalent to the global incentive constraints $V(w) = U(c(w), e(w)) = \max_{w'} U(c(w'), (e(w')w')/w)$ for all $w, w'$. We follow the usual approach of dropping the monotonicity constraint and verifying ex post that the solution satisfies it, abstracting from issues of bunching.

sulting allocation is consistent with the original economy, however, we need to require that the $\tilde{x}$ induced by the solution to the inner problem is equal to the original $x$ that was taken as given to solve the standard Mirrlees problem. The "outer" problem thus requires that $x$ be a fixed point of the mapping $x \to \tilde{x}^*(x)$, where $\tilde{x}^*(x)$ is defined by (4), evaluated at the effort allocation $e(w)$ that solves the relaxed inner problem (7) to (9).

The solution $(\hat{T}(y), x^*)$ to this benchmark problem is interpretable as a self-confirming policy equilibrium (SCPE) as developed in a different context in Rothschild and Scheuer (2011). Suppose that taxes happen to be set at $\hat{T}(y)$ (thereby inducing $x = x^*$) and consider a social planner with Pareto weights $\Psi(F_{x^*}(w))$ who takes the wage distribution to be exogenous to the tax code, as in a standard Mirrlees model. This social planner would, following Saez (2001), observe the income distribution induced by $\hat{T}(y)$ and infer the underlying wage distribution $F_{x^*}(w)$ from it. It would then compute the optimal Mirrleesian taxes given $\Psi$ and $F_{x^*}(w)$ and "confirm" the optimality of the existing tax code $\hat{T}(y)$. $\hat{T}(y)$ would not, in fact, be optimal in light of the endogeneity of wages to the tax code, but a social planner who takes wages to be exogenous would have no reason to explore other regions of the tax-policy space, given the "confirmed" optimality of the status quo.[4]

**Definition 1.** *An SCPE is a value of $x$ and an allocation $V(w), e(w)$ such that (i) $x$ is a fixed point of $\tilde{x}^*(x)$ and (ii) $V(w)$ and $e(w)$ solve the inner SCPE problem given $x$.*

Hence, our definition of an SCPE is the most straightforward way of capturing standard Mirrleesian taxation in our framework in a consistent way. We will compare SCPE and Pareto optima for given weights $\Psi$ in our numerical explorations in section 5.

## 3.2 Inner Problem

We start by using the inner problem, taking $x$ as given, to derive formulas for marginal tax rates that have to be satisfied in any Pareto optimum.

**Proposition 1.** *Let $\mu$ denote the multiplier on the resource constraint (8), let $\mu\xi$ denote the multiplier on the consistency condition (5), and use $\varepsilon^u(w)$ and $\varepsilon^c(w)$ to denote the uncompensated and compensated labor supply elasticities, respectively. For any given Pareto weights $G$, optimal marginal tax rates satisfy*

$$1 - T'(y(w)) = \left(1 + \xi\left(\frac{f_x^\theta(w)}{f_x(w)} - \alpha(x)\right)\right)\left(1 + \frac{\eta(w)}{w f_x(w)}\frac{1+\varepsilon^u(w)}{\varepsilon^c(w)}\right)^{-1} \tag{10}$$

[4]Note the importance of using relative welfare weights here: the SCPE social planner is unaware of the multi-sector character of the economy, so using general Pareto weights $G(\theta, \varphi)$ would be inconsistent with the planner's incorrect model of the economy.

*where*

$$\eta(w) = \int_w^{\overline{w}_x} \left(1 - \frac{g_x(z)}{f_x(z)} \frac{U_c(z)}{\mu}\right) \exp\left(\int_w^z \left(1 - \frac{\varepsilon^u(s)}{\varepsilon^c(s)}\right) \frac{dy(s)}{y(s)}\right) f_x(z)dz. \tag{11}$$

*Proof.* See Appendix B. □

Since the inner SCPE problem differs from the inner Pareto problem only by the absence of the consistency condition (5), the marginal tax rates for an SCPE can be found by using relative Pareto weights $\Psi$ and setting $\xi = 0$ in Proposition 1, as in the following Corollary.

**Corollary 1.** *For any given relative Pareto weights $\Psi$, marginal tax rates in the resulting SCPE satisfy*

$$1 - T'(y(w)) = \left(1 + \frac{\eta(w)}{w f_x(w)} \frac{1 + \varepsilon^u(w)}{\varepsilon^c(w)}\right)^{-1}, \tag{12}$$

*where*

$$\eta(w) = \int_w^{\overline{w}_x} \left(1 - \Psi'(F_x(z)) \frac{U_c(z)}{\mu}\right) \exp\left(\int_w^z \left(1 - \frac{\varepsilon^u(s)}{\varepsilon^c(s)}\right) \frac{dy(s)}{y(s)}\right) f_x(z)dz. \tag{13}$$

The formulas for an SCPE are the same as for a standard Mirrlees model (see e.g. Saez, 2001). In contrast, in any Pareto optimum, the formula for marginal keep shares $1 - T'$ is adjusted compared to the SCPE by a correction factor that depends on $\xi$ and a comparison between the aggregate income share of the $\Theta$-sector, given by $\alpha(x)$, with its local income share $y(w)f_x^\theta(w)/(y(w)f_x(w)) = f_x^\theta(w)/f_x(w)$.[5]

This is intuitive. For instance, suppose $\xi > 0$—which we will show in section 3.4 corresponds to the case where $\Theta$ is the high-income sector. Then the marginal keep share is scaled up in the Pareto problem relative to the SCPE whenever, at the given wage (or equivalently income) level, the local income share of the $\Theta$-sector exceeds its aggregate income share. This disproportionately encourages $\Theta$-sector effort and therefore raises wages in the $\Phi$-sector relative to the $\Theta$-sector. Hence, the solution to the Pareto problem uses this "trickle down" channel through wages in order to redistribute to the low-income sector, which is desirable for relative Pareto weights with $\Psi(F) \geq F$ for all $F \in [0,1]$. Note that this implies a force towards less progressivity in the Pareto optimum relative to the

[5]The formulas in Proposition 1 and Corollary 1 will eventually be evaluated at the optimal $x$-values from solving the respective outer problems. Since, in general, the level of $x$ in the SCPE and the Pareto optimum will differ for a given economy, even when based on the same Pareto weights $\Psi$, the formulas do not permit a direct comparison of tax *rates* at the two solutions. One interpretation, however, is as a comparison of the tax rates in two *different* economies with the same endogenous wage distributions.

SCPE: If $\Theta$ is the high-income sector, marginal tax rates will be scaled up (down) in the Pareto problem compared to the SCPE for low (high) income levels.[6]

As usual, $\eta(w)$ captures the redistributive motives of the social planner as well as income effects. The optimal tax formula therefore simplifies considerably if income effects disappear and preferences are of the quasilinear form $U(c,e) = c - h(e)$. Then $U_c(w) = \mu = 1$ and $\varepsilon^c(w) = \varepsilon^u(w) \equiv \varepsilon(w)$, which leads to the following corollary:

**Corollary 2.** *With quasilinear preferences, the marginal tax rate in any Pareto optimum satisfies*

$$1 - T'(y(w)) = \left(1 + \xi\left(\frac{f_x^\theta(w)}{f_x(w)} - \alpha(x)\right)\right)\left(1 + \frac{G_x(w) - F_x(w)}{w f_x(w)}\left(1 + \frac{1}{\varepsilon(w)}\right)\right)^{-1}.$$

Without income effects, the redistributive effect of taxation is simply given by $\eta(w) = G_x(w) - F_x(w)$, which is increasing in the degree to which $G_x(w)$ puts weight on low-wage earners over and above their population share $F_x(w)$. The marginal tax rate is also decreasing in the elasticity $\varepsilon(w)$, which captures the distortionary effects of taxation as in Diamond (1998). Again, the correction factor $\xi(f_x^\theta(w)/f_x(w) - \alpha(x))$ comparing aggregate and local income shares of the $\Theta$-sector is applied to marginal keep shares.

Independent of whether preferences exhibit income effects or not, the top marginal tax rate is generally not zero in a Pareto optimum, as the following result demonstrates.

**Corollary 3.** *The top marginal tax rate is zero in any SCPE and given by*

$$T'(y(\overline{w}_x)) \;=\; \xi\left(\alpha(x) - \frac{f_x^\theta(\overline{w}_x)}{f_x(\overline{w}_x)}\right)$$

*in any Pareto optimum. In particular, if* $f_x^\theta(\overline{w}_x)/f_x(\overline{w}_x) = 0$ (respectively $f_x^\theta(\overline{w}_x)/f_x(\overline{w}_x) = 1$), *then* $T'(y(\overline{w}_x)) = \xi\alpha(x)$ (respectively $T'(y(\overline{w}_x)) = -\xi(1-\alpha(x))$).

In the next subsections, we will use the outer problem to determine the sign of $\xi$, which will generally be such that this top marginal tax rate is negative.

## 3.3 Outer Problem

Denoting the substitution elasticity of the production function $Y(E_\theta, E_\varphi)$ by $\sigma(x)$, we can derive the following decomposition of the welfare effect of a marginal change in $x$.

[6]The same would be true if $\Phi$ were the high-income sector. We will show below that, in this case, $\xi$ is negative and hence, per (10), marginal tax rates will again be higher (lower) in the Pareto optimum for low (high) income levels.

**Lemma 2.** *For any Pareto weights $G$, the welfare effect of a marginal change in $x$ can be decomposed as follows:*

$$W'(x) = -\frac{1}{x}\left[\xi\mu\alpha(x)Y_\varphi(x)E_\varphi + \frac{1}{\sigma(x)}\left(I + R + \xi\mu(S+C)\right)\right], \tag{14}$$

*where*

$$S \equiv \frac{1}{Y_\theta(x)Y_\varphi(x)}\int_{\underline{w}_x}^{\overline{w}_x} w^2 e(w) f\left(\frac{w}{Y_\theta(x)}, \frac{w}{Y_\varphi(x)}\right) dw > 0, \tag{15}$$

$$I \equiv \mu\int_{\underline{w}_x}^{\overline{w}_x} \eta(w) w \frac{V'(w)}{U_c(w)} \frac{d}{dw}\left(\frac{f_x^\varphi(w)}{f_x(w)}\right) dw, \tag{16}$$

$$R \equiv \int_{\underline{w}_x}^{\overline{w}_x} wV'(w)\frac{f_x^\theta(w)f_x^\varphi(w)}{f_x(w)}\left(\frac{g_x^\theta(w)}{f_x^\theta(w)} - \frac{g_x^\varphi(w)}{f_x^\varphi(w)}\right) dw, \tag{17}$$

*and*

$$C \equiv \int_{\underline{w}_x}^{\overline{w}_x} w^2 e'(w)\frac{f_x^\varphi(w)f_x^\theta(w)}{f_x(w)} dw. \tag{18}$$

*Proof.* In Appendix B.2. □

We provide a heuristic derivation to illustrate the intuition behind this lemma and the terms in (15) to (18). Notice first that if technology is linear, so that $\sigma(x) = \infty$, then setting $W'(x) = 0$ in (14) immediately implies $\xi = 0$. By Proposition 1, the marginal tax formulas for the SCPE and Pareto problems coincide. This is intuitive: in this case, wages are exogenous to the tax code, so the fact that there are two sectors is irrelevant. It is only the additional effects driven by the endogeneity of wages in the finite $\sigma(x)$ case that provide scope for using additional tools for accomplishing redistributive objectives.

In particular, observe that when $\sigma(x) < \infty$, a small increase in $x$ has several effects. First, it has a *direct effect* on the consistency condition (5), or equivalently,

$$\frac{E_\theta}{E_\varphi} - x = 0. \tag{19}$$

Second, it has *wage effects*, since type $(\theta, \varphi)$'s wage is given by $w = \max\left\{\theta Y_\theta(x), \varphi Y_\varphi(x)\right\}$ and hence depends on $x$. Finally, it has a *sectoral shift* effect since it will induce some individuals to migrate from the $\Theta$- to the $\Phi$-sector due to the decreased wages in the $\Theta$-sector and increased wages in the $\Phi$-sector. The term $S$ in equation (15) in Lemma 2 captures this sectoral shift effect, which we derive first. We then return to a discussion of the other terms in equations (16) to (18), which are attributable to the direct and wage effects.

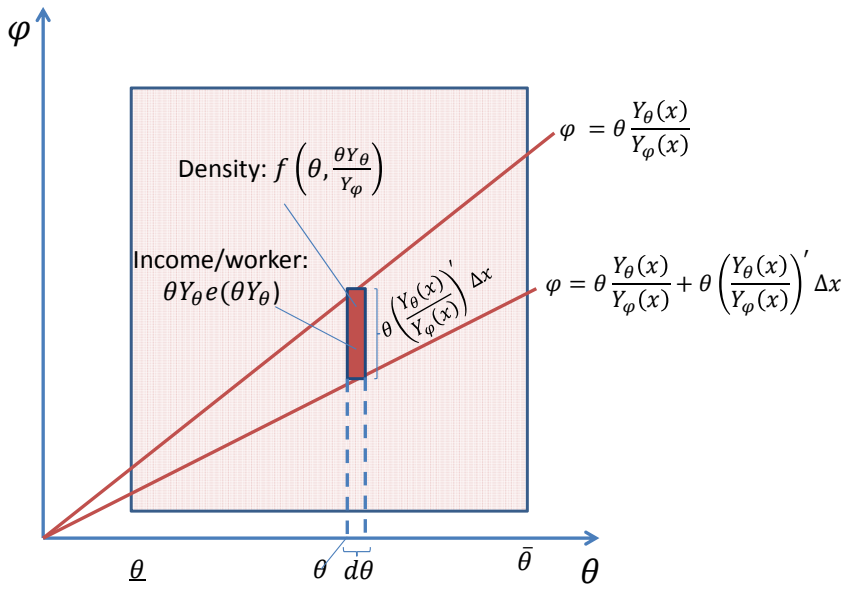



Figure 1: Illustrating the computation of the sectoral shift effect

### 3.3.1 The Sectoral Shift Effect

To derive the sectoral shift effect, it is useful to write the consistency condition as $(1-\alpha(x))Z_\theta(x) - \alpha(x)Z_\varphi(x) = 0$, where $Z_\theta(x) \equiv \int_{\underline{w}}^{\overline{w}} we(w) f_x^\theta(w) dw$ is the income earned in the $\Theta$-sector, and similarly for $Z_\varphi(x)$. Consider a small increase $\Delta x$ in $x$, holding efforts and wages constant. This will lead some individuals to shift from the $\Theta$- to the $\Phi$-sector, as illustrated in figure 1. Let $\Delta Z_\theta(x)$ denote the resulting change in $\Theta$-sector income. Since there is an equal and opposite change in $\Phi$-sector income, the sectoral shift effect can be written as $S = \Delta Z_\theta(x)$. Figure 1 illustrates the computation of $\Delta Z_\theta(x)$. It considers the mass element of individuals with $\Theta$-sector skills between $\theta$ and $\theta + d\theta$ who are in the $\Theta$-sector at $x$ but in the $\Phi$-sector at $x + \Delta x$. The height of this element is

$$\frac{d\left(Y_\theta(x)/Y_\varphi(x)\right)}{dx}\theta\Delta x = \left(\frac{Y_\theta'(x)Y_\varphi(x) - Y_\theta(x)Y_\varphi'(x)}{Y_\varphi(x)^2}\right)\theta\Delta x.$$

The income earned by each individual in that element is $\theta Y_\theta(x) e\left(\theta Y_\theta(x)\right)$, and the density of individuals is $f\left(\theta, \theta Y_\theta(x)/Y_\varphi(x)\right)$. Multiplying the width $(d\theta)$ by the height, the

density, and the per capita income, and then integrating over $\theta$ gives:

$$\begin{aligned}\Delta Z_\theta(x) &= \Delta x \int_{\underline{\theta}}^{\overline{\theta}} \left( \frac{Y'_\theta(x)Y_\varphi(x) - Y_\theta(x)Y'_\varphi(x)}{Y_\varphi(x)^2} \right) Y_\theta(x)\theta^2 e\left(\theta Y_\theta(x)\right) f\left(\theta, \frac{\theta Y_\theta(x)}{Y_\varphi(x)}\right) d\theta \\ &= \Delta x \frac{-1/(x\sigma(x))}{Y_\theta(x)Y_\varphi(x)} \int_{\underline{w}}^{\overline{w}} w^2 e\left(w\right) f\left(\frac{w}{Y_\theta(x)}, \frac{w}{Y_\varphi(x)}\right) dw,\end{aligned}$$

where the second step involves changing variables to $w = \theta Y_\theta(x)$ and using $x\sigma(x) = Y'_\varphi(x)/Y_\varphi(x) - Y'_\theta(x)/Y_\theta(x)$ (*viz* Lemma 3 in Appendix B.2.) The sectoral shift term $S$ defined in expression (15) follows directly.

### 3.3.2 The Direct and Wage Effects

Towards understanding the other terms in Lemma 2, first consider the thought experiment of changing $x$—and hence the wage distribution—while holding effort and sector constant for each individual. The first term in expression (14) is equal to the effect of this change on the consistency condition (5), which can be seen by observing that the effect on the reformulated constraint (19) is $-1$ and that the original and reformulated constraints differ by a factor of $\alpha(x)Y_\varphi(x)E_\varphi/x$.

This thought experiment is not actually feasible, however, since it requires assigning two individuals with the same wage (but in different sectors) different $(e(w), V(w))$-bundles after the change in $x$. The approach we take in the formal proof in Appendix B.2 is instead motivated by a related but feasible thought experiment: increase $x$ by a small amount while holding the $e(w)$ and $V(w)$ schedules fixed *at each wage* $w$ rather than for each type.[7]

This thought experiment leads to changes in aggregate efforts and utilities, $e_{agg}(w) \equiv e(w)f_x(w)$ and $V_{agg}(w) \equiv V(w)f_x(w)$, at any given $w$. The algebraic manipulations in our proof are driven by thinking of the thought experiment in two steps: first, a change in the schedules $e(w)$ and $V(w)$ at the original $x$ and wage distribution that absorbs the aggregate change in $e_{agg}(w)$ and $V_{agg}(w)$ induced by the change in $x$ at each wage; and second, a change in $x$—and the wage distribution—coupled with a reversion to the original $e(w)$ and $V(w)$ schedules. This second part, by construction, holds $e_{agg}(w)$ and $V_{agg}(w)$ at each $w$ constant, but *re-allocates* that effort and utility across individuals who originally earned the same wage but in different sectors. By an envelope argument, only the second, re-allocating step has welfare effects. Since the resource constraint (8) depends only on

[7] An alternative approach, leading to the same results, is to break the incentive constraint into two sector-specific constraints. See Rothschild and Scheuer (2011) for an example of this approach in a related problem.

$e_{agg}(w)$ and $V_{agg}(w)$ at each wage, it is unaffected in this second step. The terms $C$, $I$ and $R$ arise from the effects on the other constraints (5) and (7) and the objective (6).

The effect of the re-allocating change on the reformulated consistency constraint (19) has two components: the $-1$ term from the (infeasible) thought experiment considered above, and a second term—labeled $C$ in expression (18)—which arises because the across-sector re-allocation of effort at each wage affects $E_\theta / E_\varphi$. Intuitively, an increase in $x$ raises the wage of $\Phi$-sector wage-$w$ types and lowers the wage of $\Theta$-sector wage-$w$ types. If $e'(w) > 0$, this effectively re-allocates effort from the $\Theta$- to the $\Phi$-sector since $\Theta$-workers move down and $\Phi$-workers move up along the $e(w)$ schedule. It therefore reinforces the re-allocation of effort across sectors from the sectoral shift effect $S$.

The term $R$ in (17) arises analogously: the re-allocation of $V$ across sectors at each $w$ affects the objective insofar as the planner assigns different welfare weights to workers in different sectors but at the same wage. It therefore disappears when there is no intrinsic sectoral preference—i.e., when $g_x^\theta(w)/g_x^\varphi(w) = f_x^\theta(w)/f_x^\varphi(w)$ for all $w$, as will be the case with relative welfare weights $G(\theta, \varphi) = \Psi(F(\theta, \varphi))$. In contrast, when the social planner has an intrinsic preference for the $\Theta$-sector individuals at wage $w$, the re-allocation of utility from the $\Theta$- to the $\Phi$-sector is welfare reducing.

Finally, the term $I$ in equation (16) arises from the incentive effect of the re-allocating step of the thought experiment. To understand it, suppose that the share of $\Theta$-sector workers is locally increasing at some $w$. Then an increase in $x$, which raises $\Phi$-sector wages and lowers $\Theta$-sector wages, leads to a local compression of the wage distribution. Such a compression eases the local incentive compatibility constraints if they are downward-binding, in which case $\eta(w) > 0$. The increase in $x$ therefore leads to a welfare improvement insofar as $\eta(w) d\left(f_x^\varphi(w)/f_x(w)\right)/dw < 0$, and the magnitude of this improvement will be related to how steeply increasing the utility distribution is. As we will formalize in the subsequent section, $I$ can be thought of as a (generalized) Stiglitz (1982) effect: with endogenous wages, increasing (decreasing) effort at high (low) wages will raise (lower) wages at low (high) wages. In contrast, the sectoral shift and other wage effects, captured by $S$, $C$ and $R$, are not present in Stiglitz's (1982) framework.

### 3.4 Marginal Tax Rate Results

We can use the decomposition in Lemma 2 to sign the multiplier on the consistency condition $\xi$ at an optimal $x$ by setting $W'(x) = 0$ in equation (14):

$$\xi = -\frac{I+R}{\mu\sigma(x)} \Big/ \left(\alpha(x) Y_\varphi(x) E_\varphi + \frac{C+S}{\sigma(x)}\right). \tag{20}$$

We summarize the resulting conditions for the sign of $\xi$ in the following corollary:

**Corollary 4.** *With a linear technology ($\sigma(x) = \infty$), $\xi = 0$.*
*For $\sigma(x) \in (0, \infty)$, the following holds for any Pareto optimum with (i) increasing effort ($e'(w) \geq 0$) and (ii) downwards-binding incentive constraints ($\eta(w) \geq 0$ for all $w$):*

1. *$\xi \geq 0$ if $f_x^\theta(w)/f_x(w)$ is increasing in $w$ and $g_x^\theta(w)/f_x^\theta(w) \leq g_x^\varphi(w)/f_x^\varphi(w)$ $\forall w$,*
2. *$\xi \leq 0$ if $f_x^\varphi(w)/f_x(w)$ is increasing in $w$ and $g_x^\theta(w)/f_x^\theta(w) \geq g_x^\varphi(w)/f_x^\varphi(w)$ $\forall w$.*

*The inequalities in 1. and 2. are strict if $\eta(w)$ is not identically zero.*

Conditions (i) and (ii) are sufficient, but not necessary. The former ensures that the wage shift term $C$ reinforces the sectoral shift effect. By equation (11), the latter holds whenever the social marginal value of consumption, given by $U_c(w)g_x(w)/f_x(w)$, is decreasing in $w$. This is guaranteed with quasilinear-in-consumption preferences and weakly progressive welfare weights (i.e., decreasing $g_x(w)/f_x(w)$), for example. It ensures that a compression of the wage distribution eases the incentive compatibility constraints. If $f_x^\theta(w)/f_x(w)$ is increasing in $w$, then $\Theta$ is the high-skill sector, and an increase in $x$, by raising wages in the $\Phi$-sector and lowering them in the $\Theta$-sector, has desirable wage compression effects, as in Stiglitz (1982). This desirable effect is reinforced by the redistribution effect $R$ whenever $g_x^\varphi(w)/f_x^\varphi(w) \geq g_x^\theta(w)/f_x^\theta(w)$ $\forall w$. In this case, the social planner puts higher social welfare weight on $\Phi$-sector workers than on $\Theta$-sector workers at any given wage, and the wage changes induced by an increase in $x$ also have direct welfare benefits.

Combining these results from the outer problem with the marginal tax rate results from the inner problem has crisp implications for the comparison between Pareto optimal and SCPE tax schedules. For instance, suppose $\Theta$ is the high-skilled sector, i.e. $f_x^\varphi(w)/f_x(w)$ is decreasing so that $\xi > 0$ under conditions (i) and (ii) in Corollary 4 (and assuming no intrinsic redistributive motives across sectors). Then by equation (10) in Proposition 1, the marginal keep share in the Pareto optimum is scaled down relative to the SCPE wherever the local income share in the $\Phi$-sector is higher than in aggregate. This disproportionately reduces $\Phi$-sector effort and therefore indirectly increases wages in the $\Phi$-sector, achieving redistribution to the low-skilled sector. In particular, since $f_x^\varphi(w)/f_x(w)$ is decreasing, this means that marginal keep shares are scaled down for low wages and scaled up for high wages, and the top marginal tax rate is negative.

On the other hand, suppose $f_x^\varphi(w)/f_x(w)$ is increasing, so that $\xi < 0$. Marginal keep shares will be scaled down whenever $f_x^\varphi(w)/f_x(w)$ is low, i.e., again for high wages. The

top marginal tax rate is also again negative. In *both* of the two cases the general equilibrium effects in the Roy model work in favor of *less* progressive taxation. We summarize these insights in the following Proposition:

**Proposition 2.** *If $\sigma(x) \in (0, \infty)$, then the top marginal tax rate is negative in any Pareto optimum with*
*(i) a decreasing i-sector share of workers $f_x^i(w)/f_x(w)$, $i \in \{\theta, \varphi\}$,*
*(ii) an increasing effort schedule $e(w)$,*
*(iii) a decreasing social marginal utility of consumption schedule $U_c(w)g_x(w)/f_x(w)$ and*
*(iv) a weak intrinsic social preference for the i-sector, i.e. $g_x^i(w)/f_x^i(w) \geq g_x^j(w)/f_x^j(w)$ for all $w$, $j \neq i \in \{\theta, \varphi\}$.*

Notably, consider the special case with relative welfare weights. Then $G_x(w) = \Psi(F_x(w))$, $g_x(w) = \Psi'(F_x(w))f_x(w)$, and hence

$$g_x^\theta(w) = \Psi'(F_x(w))f_x^\theta(w) \quad \text{and} \quad g_x^\varphi(w) = \Psi'(F_x(w))f_x^\varphi(w).$$

This immediately implies $g_x^\theta(w)/f_x^\theta(w) = g_x^\varphi(w)/f_x^\varphi(w)$ $\forall w$ and thus $R = 0$. With relative welfare weights, condition (iv) can therefore be dropped.

Hence, these results reveal the following intuitive separation: Per Corollary 4, the sign of the multiplier $\xi$ on the consistency constraint accounts for the overall redistributive motive across sectors, i.e., whether redistribution from $\Theta$ to $\Phi$ or vice versa is desirable. Then, conditional on this direction, the nonlinear marginal tax rate correction in the Pareto optimum relative to the SCPE is determined by comparing local and aggregate income shares between sectors, per equation (10) in Proposition 1.

# 4 The Role of Occupational Choice and Continuous Types

In this section, we relate our results to those in Stiglitz (1982), who considers optimal nonlinear taxation in a two-type model with endogenous wages but without occupational choice. We demonstrate that the general Roy model, with continuous types and occupational choice, features three extra effects, as captured by $S$, $C$ and $R$ in the previous section, that do not appear in Stiglitz's model. The disappearance of the sectoral shift effect $S$ in a model without occupational choice is obvious. In addition, the Roy model with continuous types generates overlapping wage distributions in the two sectors, which gives rise to the effects $C$ and $R$. In contrast, in a discrete type model, generically—and somewhat unrealistically—there are no workers in different sectors earning the same wage.

The extra Roy effects that emerge in our model do not change the sign of the general equilibrium effects found in Stiglitz (1982), but they mitigate them. In this sense, optimal redistributive taxation in the Roy model involves a less progressive tax schedule than a standard Mirrlees model (as captured by an SCPE) but a *more* progressive tax schedule than a discrete type model without occupational choice.

We start by reformulating Stiglitz's (1982) model in terms of the decomposition into an inner problem (for fixed $x$) and outer problem (optimizing over $x$) as above. Let there be two types with skills $\theta$ and $\varphi$ and with fractions $f_\theta$ and $f_\varphi = 1 - f_\theta$ in the population. We put (relative) Pareto weights $\psi_\theta$ and $\psi_\varphi$ on them such that $f_\theta \psi_\theta + f_\varphi \psi_\varphi = 1$. Without loss of generality, we will think of $\theta$ as the high wage sector and $\varphi$ as the low wage sector, so that regular welfare weights satisfy $\psi_\theta \leq 1$ and $\psi_\varphi \geq 1$. As in Stiglitz (1982), we therefore focus on the case where only the $\theta$-type's incentive constraint binds.

## 4.1 Inner Problem

Individuals are paid their marginal products, $w_\theta = \theta Y_\theta(x)$, and $w_\varphi = \varphi Y_\varphi(x)$. Hence, we can write the inner problem for fixed $x$ as

$$W(x) = \max_{e_\theta, e_\varphi, V_\theta, V_\varphi} f_\theta \psi_\theta V_\theta + f_\varphi \psi_\varphi V_\varphi \tag{21}$$

subject to

$$V_\theta \geq U\left(c(V_\varphi, e_\varphi), e_\varphi \frac{w_\varphi}{w_\theta}\right), \tag{22}$$

$$(1 - \alpha(x)) f_\theta w_\theta e_\theta - \alpha(x) f_\varphi w_\varphi e_\varphi = 0, \tag{23}$$

$$f_\theta w_\theta e_\theta + f_\varphi w_\varphi e_\varphi \geq f_\theta c(V_\theta, e_\theta) + f_\varphi c\left(V_\varphi, e_\varphi\right). \tag{24}$$

As before, the outer problem is just $\max_x W(x)$.[8]

We focus on the top marginal tax rate (i.e., the optimal allocation for type $\theta$). Denoting by $\mu$ and $\xi\mu$ the multipliers on (24) and (23), the first order condition w.r.t. $e_\theta$ is

$$-f_\theta \mu \left(\partial c(V_\theta, e_\theta)/\partial e_\theta - w_\theta\right) + (1 - \alpha(x)) f_\theta \xi \mu w_\theta = 0.$$

Using $\partial c/\partial e = -U_e/U_c = MRS$, this simplifies to $MRS_\theta = w_\theta(1 + (1 - \alpha(x))\xi)$. By

[8] Observe that we again pool all individuals of a given type at the same allocation $(V, e)$. This is without loss of generality, but the reasoning is somewhat different from the continuous type case in section 2.2. In this discrete-type setting, as in Blackorby, Brett, and Cebreiro (2007), any allocation that treats individuals who earn the same wage differently can be replaced by an alternative incentive compatible allocation that keeps each individual's utility unchanged and requires fewer resources. Importantly for our endogenous wage setting, this alternative allocation also does not change aggregate effort in either sector or $x$.

the first order condition for the worker's utility maximization problem, i.e., $MRS/w = 1 - T'(y)$, this implies that the marginal tax rate for the high-wage, $\Theta$-sector individual, is $-(1-\alpha(x))\xi$ as in Corollary 3.

## 4.2 Outer Problem

We next turn to the outer problem to determine $\xi$. By the envelope theorem, we can compute the wage shift effect by holding $V_i$ and $e_i$ constant for $i \in \{\theta, \varphi\}$, which, in contrast to section 3.3, is a feasible thought experiment here as each wage corresponds to a single sector. Since sectors do not overlap at any wage, the re-allocation across sectors that led to the terms $R$ and $C$ in the continuous model in section 3.3 is absent here. There is also no sectoral shift effect $S$ since occupational choice is fixed. As in section 3.3, the effect of the wage shift induced by the change in $x$ on the objective is identically zero, and the effect on the resource constraint (24) is zero by constant returns to scale. Only the direct effect $-\xi\mu\alpha(x)Y_\varphi(x)E_\varphi/x$ and the incentive effects remain.

Specifically, putting a multiplier $\hat{\eta}\mu$ on (22) (and using the same algebraic steps employed in the proof of Lemma 2) we find that the effect of the wage shift on the incentive constraint is:

$$-\mu\hat{\eta}U_e\left(c_\varphi, e_\varphi\frac{w_\varphi}{w_\theta}\right)e_\varphi\frac{\varphi}{\theta}\left[\frac{Y'_\varphi(x)}{Y_\theta(x)} - \frac{Y_\varphi(x)Y'_\theta(x)}{Y_\theta(x)Y_\theta(x)}\right] = -\frac{1}{x\sigma(x)}\hat{I},$$

where $\hat{I} \equiv \mu\hat{\eta}U_e\left(c_\varphi, e_\varphi\frac{w_\varphi}{w_\theta}\right)e_\varphi\frac{w_\varphi}{w_\theta}$ is the discrete incentive constraint analog of $I$ from the general Roy model.[9]

Combining the effects yields $W'(x) = -\frac{1}{x}\left[\xi\mu\alpha(x)Y_\varphi(x)E_\varphi(x) + \frac{1}{\sigma(x)}\hat{I}\right]$.

## 4.3 Marginal Tax Rates

At an optimum, $W'(x) = 0$, so

$$\xi = -\frac{\hat{I}/(\mu\sigma(x))}{\alpha(x)Y_\varphi(x)E_\varphi}, \tag{25}$$

[9]To see this, observe that in the limit where $f_x^\theta(w)/f_x(w)$ is 0 up until $w_\theta$, and 1 thereafter, $(d/dw)[f_x^\theta(w)/f_x(w)]$ is a Dirac $\delta$-function and the integral in (16) evaluates to $-V'(w_\theta)w_\theta\eta(w_\theta) = U_e(c_\theta, e_\theta)e_\theta\eta(w_\theta)$ by the incentive constraint (7). The only difference from $\hat{I}$ is that it has $e_\varphi w_\varphi/w_\theta$ instead of $e_\theta$ and $c_\varphi$ rather than $c_\theta$ (and $\hat{\eta} = \eta/U_c$ is discrete rather than continuous). In the density limit, the $\theta$-type would be imitating an infinitesimally close individual. If we let $w_\varphi$ be arbitrarily close to $w_\theta$, then we would get $e_\theta$ and $c_\theta$, as in the limit case of $I$.

which coincides with the general formula (20) if $S = C = R = 0$ and when we replace $I$ with $\hat{I}$. Moreover, $\hat{I}$ and $\hat{\eta}$ have opposite signs. This means that $\xi > 0$ (and hence top marginal taxes are negative) precisely when redistribution occurs from the $\theta$- to the $\varphi$-types, so that the downward incentive constraint binds.

This is analogous to our results in the general Roy model, but comparing equations (20) and (25) reveals that the addition of the effects $S$ and $C$ will make $\xi$, and hence top marginal taxes, smaller in absolute value. To understand the intuition behind this, suppose we lower taxes at the top to increase the effort of the top earners. This is welfare enhancing because it raises the wages of the low wage sector workers and lowers the wages of high wage sector workers and thus relaxes the downward incentive constraint. However, when there is endogenous occupational choice, the sectoral shift effect works against this, since this change in wages leads some individuals to shift out of the high wage into the low wage sector, undoing some of the original increase in aggregate effort in the high wage sector. The indirect wage shift effect $C$ reinforces the sectoral shift effect (when effort is increasing in wage), since, at any given wage where the sectors overlap, it involves a re-allocation of effort (among individuals who do *not* shift sectors) from workers in the high wage sector to workers in the low wage sector. Optimal taxes in the general Roy model with continuous types will therefore be less progressive than in a Mirrlees model but more progressive than in a Stiglitz model with two types and exogenous occupations.

In fact, the result that the optimal progressivity of taxes in a Roy model is "sandwiched" between Mirrlees at the progressive end and Stiglitz at the regressive end is not particular to our environment with continuous types. As we show in Rothschild and Scheuer (2012), $\xi$ is also bounded between Mirrlees (i.e. 0) and Stiglitz (i.e. formula (25)) in a general model with a discrete, two-dimensional type-distribution and with allocations that condition on wages only. This means that the Stiglitz formula (25) is a special case even *within* the class of discrete-type models.

# 5 A Numerical Example

The purpose of this section is to illustrate the role of sectoral and wage shift effects for the progressivity of optimal income taxes and verify the consistency of the conditions in Corollary 4 and Proposition 2. We use data from the 2011 Current Population Survey (CPS) rotating March sample to calibrate a simple 2-sector Roy model of the U.S. economy and to compute optimal tax schedules. We assume quasilinear preferences $U(c,e) = c - h(e)$ with isoelastic disutility $h(e) = e^{1+1/\varepsilon}/(1+1/\varepsilon)$. We use a labor elasticity $\varepsilon = 0.5$

and Cobb-Douglas technology, $Y = E_\theta^\alpha E_\varphi^{1-\alpha}$.[10] We remain deliberately agnostic about the nature of the two latent sectors. Instead, we build on Basu and Ghosh (1978) and Heckman and Honoré (1990), who show that the parameters of an underlying bivariate normal distribution over $(x_1, x_2)$ can be identified by observing the single-dimensional distribution of the *maximum* of $x_1$ and $x_2$ (up to a permutation of indices).

Specifically, following Mankiw et al. (2009), we use the CPS data on weekly earnings and hours to generate a sample of hourly wages $w_i$ for the U.S. population. We assume that this wage distribution is generated from a two-sector Roy model with individuals whose skills $(\theta_i, \varphi_i)$ are drawn from a bivariate lognormal distribution so that, for a given $x = E_\theta / E_\varphi$, the distribution across individuals of possible wages $(\theta_i Y_\theta(x), \varphi_i Y_\varphi(x))$ is also bivariate lognormal. We estimate the means $\mu_\theta$, $\mu_\varphi$, variances $\sigma_\theta^2$, $\sigma_\varphi^2$, and covariance $\sigma_{\theta\varphi}$ of this bivariate wage distribution by maximum likelihood. The likelihood of an observation $\tilde{w}_i \equiv \log(w_i) = \max\{\log(\theta_i Y_\theta(x)), \log(\varphi_i Y_\varphi(x))\}$ is given by[11]

$$\ell_i = \phi\left(\frac{\mu_\theta - \tilde{w}_i}{\sigma_\theta}\right)\left[1 - \Phi\left(\frac{\tilde{\mu}_\varphi - \tilde{w}_i}{\tilde{\sigma}_\varphi}\right)\right] + \phi\left(\frac{\mu_\varphi - \tilde{w}_i}{\sigma_\varphi}\right)\left[1 - \Phi\left(\frac{\tilde{\mu}_\theta - \tilde{w}_i}{\tilde{\sigma}_\theta}\right)\right],$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and cumulative distribution of the standard normal distribution, respectively, and where

$$\tilde{\mu}_\theta \equiv \left(\mu_\theta - \frac{\sigma_{\theta\varphi}}{\sigma_\varphi^2}\mu_\varphi\right)\frac{1}{1 - \sigma_{\theta\varphi}/\sigma_\varphi^2} \quad \text{and} \quad \tilde{\sigma}_\theta \equiv \frac{\sigma_\theta\sqrt{1 - \sigma_{\theta\varphi}\sigma_\theta^2\sigma_\varphi^2}}{\left|1 - \sigma_{\theta\varphi}/\sigma_\varphi^2\right|}$$

if $\sigma_{\theta\varphi}/\sigma_\varphi^2 \neq 1$ (and $\tilde{\mu}_\theta \equiv \mu_\theta - \mu_\varphi$, $\tilde{\sigma}_\theta \equiv \sqrt{1 - \sigma_{\theta\varphi}\sigma_\theta^2\sigma_\varphi^2}$ otherwise), with symmetric expressions for $\tilde{\mu}_\varphi$ and $\tilde{\sigma}_\varphi$.

The income share of the $\Theta$-sector, given by $\alpha$, can be inferred from this estimated bivariate wage distribution by using the estimated parameters $\mu_\theta$, $\mu_\varphi$, $\sigma_\theta^2$, $\sigma_\varphi^2$, and $\sigma_{\theta\varphi}$ to draw a (large) sample of $(w_\theta, w_\varphi)$. We can infer from this sample both sectoral choices and wages $w = \max\{w_\theta, w_\varphi\}$. From these and the individual optimization condition $e^{1/\varepsilon} = (1 - \tau)w$ (where $\tau$ is the marginal tax rate, which we take to be 25% for our calibration), we can compute the sectoral incomes $Y^\theta$ and $Y^\varphi$ and hence $\alpha = Y^\theta/(Y^\theta + Y^\varphi)$.

Finally, with Cobb-Douglas technology, we can, w.l.o.g., take the underlying skills $(\theta, \varphi)$ to coincide with $(w_\theta, w_\varphi)$—i.e., we can take $Y_\theta(x) = Y_\varphi(x) = 1$. To wit: note that $Y_\theta = \alpha Y / E_\theta$. Since scaling all $\theta$-skills by $k > 0$ scales $E_\theta$ by $k$, this implies that $Y_\theta$ scales by $k^{-1}$. This re-scaling leaves wages $w = \theta Y_\theta$, efforts, and incomes unchanged. In other

[10] This implies a constant substitution elasticity $\sigma = 1$. We explore the effect of varying $\sigma$ in section 6.

[11] See Basu and Ghosh (1978), expressions (2.5) and (6.2).

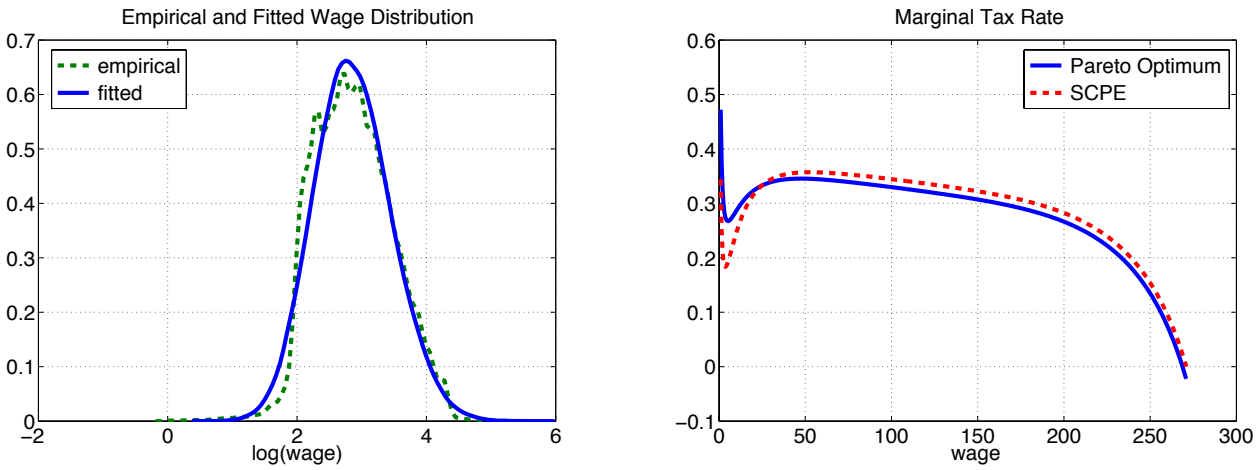



Figure 2: Empirical/estimated wage distributions and optimal/SCPE tax schedules

words, for a given economy, the underlying skills are only defined up to such a re-scaling.

Our baseline estimates (standard errors) are $\mu_\theta = 2.81$ $(.029)$, $\mu_\varphi = 1.74$ $(.714)$, $\sigma_\theta = 0.647$ $(.015)$, $\sigma_\varphi = 0.637$ $(.369)$, and $\rho_{\theta\varphi} = -.030$ $(.630)$, where $\rho_{\theta\varphi} = \sigma_{\theta\varphi}/\sigma_\theta\sigma_\varphi$ is the correlation between the two dimensions. The corresponding mean wages are approximately 7 and 20 for the two sectors, and 12.3% of the workers are estimated to work in the $\Phi$-sector, which has an income share $1-\alpha$ of only 5.8%. The left panel of figure 2 compares the estimated to the empirical wage distribution; it shows a reasonably good fit.

The $\Phi$-sector is the low-income sector here, and it is quite small in quantitative terms. We therefore use a likelihood ratio test to see whether we can reject the two-sector model in favor of a simpler one-sector model. This likelihood ratio test is complicated by the fact that a single-sector model is a "singularity" in parameter space ($\rho_{\theta\varphi}$ ceases to be well-defined under the restriction). We perform a two-step version of the test: first, we observe that $\rho_{\theta\varphi} = 0$ is not rejected. Re-estimating the model with $\rho_{\theta\varphi} \equiv 0$ leads to imperceptible changes in the remaining coefficients. A standard likelihood ratio test for the restriction of *this* model to a single sector yields $\chi^2(2) = 244.2$, easily rejecting the single-sector restriction. We employ the $\rho_{\theta\varphi} = 0$ estimates in the following tax computations.

To compute the Pareto optimal and SCPE taxes for this economy, only the Pareto weights remain to be specified. We use relative weights $\Psi = 1-(1-F)^r$, where $r$ characterizes the magnitude of the government's desire for redistribution from high to low wages. With the quasilinear preferences that we use here, $r = 1$ implies no redistributive motives, and $r \to \infty$ for a Rawlsian social planner. We take $r = 1.3$, so that there is some intermediate desire for redistribution from high- to low-wage earners.

The right panel of figure 2 shows the marginal tax schedule $T'(y(w))$ as a function of

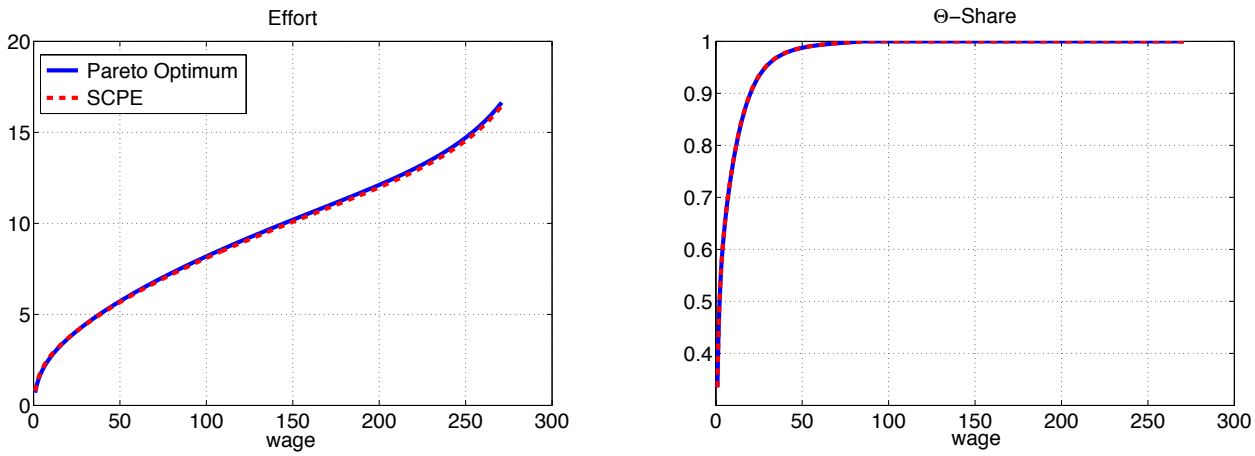



Figure 3: Effort and share of workers in the Θ-sector for Pareto optimum/SCPE

$w$ both for the Pareto optimum and the SCPE resulting from our parametrization.[12] The two tax schedules are similar: Both display U-shaped marginal rates at modest wages, and then falling rates in the upper tail of the distribution. The optimal tax schedule is modestly less progressive than the SCPE, in accord with the theory above, and the top marginal tax rate is negative, albeit small in magnitude at about $-2\%$ due to the small size of the low-wage sector $\Phi$ (recall that, by Corollary 3, the top marginal tax rate is given by $-(1-\alpha)\xi$). Finally, figure 3 demonstrates that the assumptions in Proposition 2 are satisfied in our calibrated example: individual effort $e(w)$ is increasing in the wage, and the shares of Θ- and Φ-sector workers are monotone. A fortiori, income $y(w) = we(w)$ is increasing in $w$, so that bunching does not need to be considered.

# 6 Extensions

## 6.1 Unbounded Skill Distribution

We have focused attention on bounded skill distributions for simplicity and to facilitate a comparison with Stiglitz (1982), but the preceding analysis does not rely on this assumption. In particular, recent studies have emphasized that the top end of the empirical wage distribution is better described by an unbounded Pareto distribution (Saez, 2001). The analysis of the outer problem in section 3.3 can be extended in a straightforward way to such unbounded distributions, and the methods developed in Diamond (1998) and Saez (2001) can be used to compute asymptotic marginal tax rates $T'(y(w))$ for $w \to \infty$. These are particularly transparent in the following case:

[12] We truncate the distribution at the $99.99^{\text{th}}$ percentile so that the top rate is well defined.

**Proposition 3.** *Consider any Pareto optimum (respectively, SCPE) such that*
*(i) preferences are quasilinear and isoelastic:* $U(c,e) = c - e^{1+1/\varepsilon}/(1+1/\varepsilon)$
*(ii) the top earners are all in the* $\Theta$*-sector:* $\lim_{w\to\infty} f_x^\theta(w)/f_x(w) = 1$
*(iii) the* $\Theta$*-sector skill distribution has a Pareto tail with parameter* $\kappa$*, so* $\lim_{w\to\infty} \frac{1-F_x(w)}{wf_x(w)} = \kappa$
*(iv) Pareto weights are relative and progressive:* $G_x(w) = \Psi(F_x(w))$ *with* $\Psi''(x) < 0$ *and*
*(v) zero welfare weight is put on the top earners:* $\Psi'(1) = 0$.
*Then the asymptotic marginal tax rate is*

$$\frac{\kappa\,(1+1/\varepsilon) - \xi(1-\alpha(x))}{\kappa\,(1+1/\varepsilon)+1} \quad \left(\text{respectively,} \quad \frac{\kappa\,(1+1/\varepsilon)}{\kappa\,(1+1/\varepsilon)+1}\right).$$

*Moreover,* $\xi > 0$ *whenever* $e(w)$ *and* $f_x^\theta(w)/f_x(w)$ *are increasing in* $w$.

*Proof.* In Appendix C. □

This implies that asymptotic marginal tax rates are scaled down in the Pareto optimum relative to the SCPE, just like top marginal tax rates in the case of a bounded skill distribution. To illustrate this downscaling numerically, we have to replace the thinner-tailed lognormal distributions from the calibration in section 5 with Pareto tails. The primary advantage of employing a bivariate lognormal distribution was that it could be identified by observing only the empirical wage distribution. This allowed us to study the role of Roy effects without taking a stand on the nature of the underlying sectors. Unfortunately, a bivariate Pareto distribution is *not* identified without additional *sectoral* information (Heckman and Honoré, 1990). We therefore use a simple numerical example that is not explicitly calibrated to the U.S. economy to get a sense for the implications of these thicker tails.

In particular, we consider a skill distribution given by two independent Pareto distributions with support $[1,\infty)$ and parameters $\kappa_\theta = 2$ and $\kappa_\varphi = 4$, respectively. As a result, there is more mass on lower skills in the $\varphi$-dimension compared to $\theta$, and $\Phi$ is again the low skill sector with $\lim_{w\to\infty} f_x^\varphi(w)/f_x(w) = 0$. We again start with the Cobb-Douglas case and set the aggregate income share of the high skill sector $\Theta$ to $\alpha = .2$. All other parameters are as in section 5. The left panel in figure 4 shows the resulting marginal tax schedules in the Pareto optimum and SCPE. It illustrates Proposition 3 and shows that, in principle, the optimal asymptotic marginal tax rate can be considerably lower than in the SCPE, indicating strong general equilibrium effects in the Roy model.

In the right panel, we show how the importance of these effects varies with the elasticity of substitution of the production function. We generalize the technology considered so in section 5 to a CES production function $Y(E_\theta, E_\varphi) = \left[\alpha E_\theta^\rho + (1-\alpha)E_\varphi^\rho\right]^{1/\rho}$, where

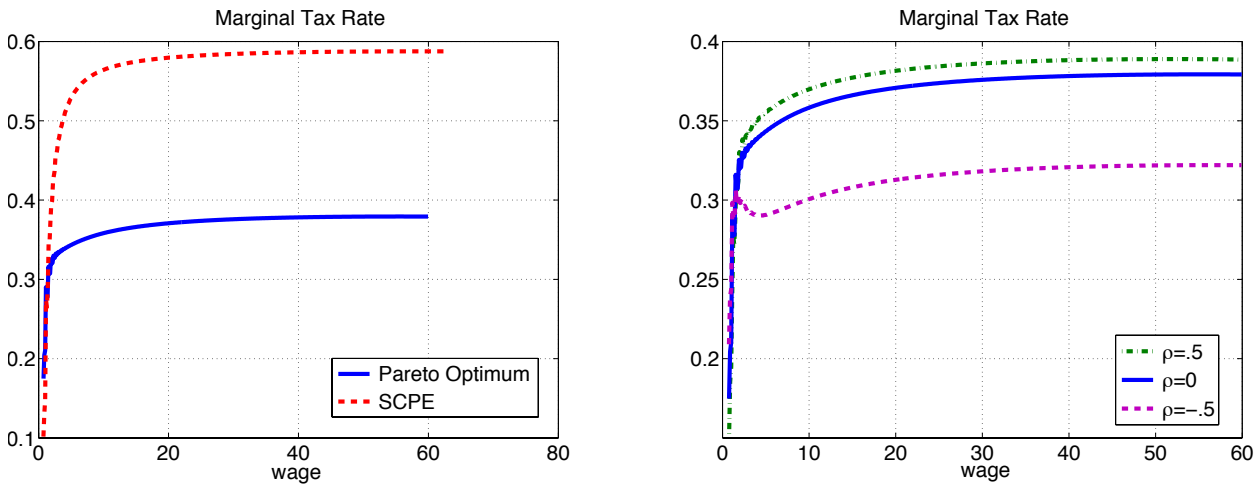



Figure 4: Pareto optimal/SCPE tax rates, and Pareto optimal tax rates for varying $\sigma$

the elasticity of substitution is constant with $\sigma = 1/(1-\rho)$ (Cobb-Douglas obtains as a special case for $\rho = 0$). The figure shows that the optimal asymptotic marginal tax rates fall as we move to lower substitution elasticities. This is because the general equilibrium effects from the endogeneity of wages become more pronounced as we move away from linear technology, with $\sigma = \infty$ ($\rho = 1$) and fixed wages.

## 6.2 Differential Sectoral Costs

In the preceding analysis, individuals based their sectoral choice exclusively on whether the $\Theta$- or $\Phi$-sector afforded them a higher wage. In many applications, however, it is reasonable to assume that occupational choice is also affected by direct costs or tastes for entering specific sectors. For instance, some occupations may require higher levels of education, so that individuals who have a high cost of achieving such education may not select into them even if they could earn a higher wage there. It is straightforward to extend the model to allow for such differential costs of working in the two sectors.

Let the types be described by a triple $t = (\theta, \varphi, \beta)$, where $\beta$ parameterizes the cost of $\Theta$-sector effort relative to $\Phi$-sector effort and, as above, $\theta$ and $\varphi$ measure the $\Theta$- and $\Phi$-sector skills. Let the general cdfs $\hat{F}(\theta, \varphi, \beta)$ and $\hat{G}(\theta, \varphi, \beta)$, with supports $[\underline{\theta}, \overline{\theta}] \times [\underline{\varphi}, \overline{\varphi}] \times [\underline{\beta}, \overline{\beta}]$, denote the type distribution and cumulative welfare weights, respectively, and take preferences to be separable, isoelastic in effort, and dependent on sector $S$ as follows:

$$U(c, e; t, S) = \begin{cases} \beta u(c) - e^{1+/\varepsilon}/(1+1/\varepsilon) & \text{if } S = \Theta \\ u(c) - e^{1+/\varepsilon}/(1+1/\varepsilon) & \text{if } S = \Phi. \end{cases}$$

In particular, with CARA utility of consumption $u(c) = -\exp(-rc)$, we can interpret $\tilde{\beta} \equiv -\log(\beta)/r$ as the consumption cost of working in the $\Theta$-sector. (With CRRA, a similarly transformed $\beta$ is interpretable as a *proportional* consumption cost.)

Within the $\Theta$-sector, there will generally be two dimensions of heterogeneity—$\theta$ and $\beta$. For any $x$, however,

$$U\left(c, \frac{y}{\theta Y_\theta(x)}; (\theta, \varphi, \beta), \Theta\right) = \beta\left[u(c) - \frac{1}{1+1/\varepsilon}\left(\frac{y}{\tilde{\theta}(\theta,\beta)Y_\theta(x)}\right)^{1+1/\varepsilon}\right],$$

where $\tilde{\theta}(\theta,\beta) \equiv \theta\beta^{\varepsilon/(1+\varepsilon)}$. Conditional on $S = \Theta$, $\tilde{\theta}$ is thus a sufficient statistic for preferences over $(c,y)$-bundles (as in Choné and Laroque, 2010). Moreover, any two types $(\theta_1, \varphi_1, \beta_1)$ and $(\theta_2, \varphi_2, \beta_2)$ with $\tilde{\theta}_1(\theta_1,\beta_1) = \tilde{\theta}_2(\theta_2,\beta_2)$ and $\varphi_1 = \varphi_2$ make the same sectoral choice. This means that we can "collapse" the policy relevant type distribution into the two-dimensional distribution of $(\tilde{\theta}, \varphi)$-types, with cdf

$$F(\tilde{\theta}, \varphi) \equiv \int_{\underline{\theta}}^{\theta}\int_{\underline{\varphi}}^{\varphi}\int_{\underline{\beta}}^{(\theta/\tilde{\theta})^{\frac{1+\varepsilon}{\varepsilon}}} d\hat{F}(\theta', \varphi', \beta'),$$

and we can collapse the welfare weights to $G(\tilde{\theta}, \varphi)$ analogously.

By interpreting $w$ as an *effective* wage, given by $\max\{\tilde{\theta}Y_\theta(x), \varphi Y_\varphi(x)\}$, the Pareto optimal and SCPE tax rates are characterized exactly as in Section 3. In particular, marginal taxes are given by Proposition 1, with $\xi$ as in equation (20), and, as in Proposition 2, top marginal tax rates will be negative whenever (i) the $i$-sector is the low $w$ sector, (ii) effort is increasing in $w$, (iii) marginal social utility of consumption is decreasing in $w$, and (iv) there is a weak preference for $i$-sector workers at any given $w$.

# 7 Conclusion

We view this paper as making a two-fold contribution. The first is methodological: we provide a technique for solving a multi-dimensional screening problem in an important class of contexts. Specifically, we show that the multi-dimensional screening problem that arises in designing optimal taxation in a multiple-sector economy can be reduced to a single dimensional optimal income tax problem à la Mirrlees. This basic technique is likely to be applicable more broadly.

Our second contribution is to derive some of the implications that self-selection into occupational sectors can have for optimal income taxation. In particular, we show that

the presence of several complementary sectors in an economy provides a force pushing towards less progressive taxation. This force is a natural extension of Stiglitz's (1982) results to a more general framework with a continuous distribution of types and sectoral mobility, and we show that the extra effects arising in this more general setting mitigate the regressive forces in the basic Stiglitz model.

We also demonstrated through a simple empirical calibration that computing the practical implications of occupational endogeneity for taxation is tractable. Through a theoretical simulation with unbounded skill distributions and fat tails, we also demonstrated that, in principle, these implications could be quantitatively significant. More detailed empirical calibrations are an important next step.

# References


ALLEN, F. (1982): "Optimal Linear Income Taxation with General Equilibrium Effects on Wages," *Journal of Public Economics*, 17, 135–143.

ATKINSON, A., AND J. STIGLITZ (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6, 55–75.

BASU, A., AND J. GHOSH (1978): "Identifiability of the Multinormal and other Distributions under the Competing Risks Model," *Journal of Multivariate Analysis*, 8, 413–429.

BLACKORBY, C., C. BRETT, AND A. CEBREIRO (2007): "Nonlinear Taxes for Spatially Mobile Workers," *International Journal of Economic Theory*, 3, 57–74.

BOADWAY, R., N. MARCEAU, AND P. PESTIEAU (1991): "Optimal Linear Income Taxation in Models with Occupational Choice," *Journal of Public Economics*, 46, 133–162.

BORJAS, G. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77, 531–553.

——— (2002): "The Wage Structure and the Sorting of Workers into the Public Sector," *NBER Working Paper 9313*.

BRETT, C. (2007): "Optimal Nonlinear Taxes for Families," *International Tax and Public Finance*, 14, 225–261.

BRETT, C., AND J. WEYMARK (2003): "Financing Education using Optimal Redistributive Taxation," *Journal of Public Economics*, 87, 2549–2569.

CHONÉ, P., AND G. LAROQUE (2010): "Negative Marginal Tax Rates and Heterogeneity," *American Economic Review*, 100, 2532–2547.

DAHL, G. B. (2002): "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70, 2367–2420.

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

DIAMOND, P., AND J. MIRRLEES (1971): "Optimal Taxation in Public Production, I: Production Efficiency," *American Economic Review*, 61, 8–27.

FELDSTEIN, M. (1973): "On the Optimal Progressivity of the Income Tax," *Journal of Public Economics*, 2, 356–376.

FUDENBERG, D., AND D. LEVINE (2009): "Self-confirming equilibrium and the Lucas critique," *Journal of Economic Theory*, 144, 2354–2371.

FUDENBERG, D., AND J. TIROLE (1991): *Game Theory*. MIT Press, Cambridge.

HAM, J. C., AND R. J. LALONDE (1996): "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence and Experimental Data on Training," *Econometrica*, 64, 175–205.

HECKMAN, J. (1974): "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679–693.

HECKMAN, J., AND B. HONORE (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–1149.

HECKMAN, J., AND G. SEDLACEK (1985): "Heterogeneity, Aggregation and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93, 1077–1125.

——— (1990): "Self-Selection and the Distribution of Hourly Wages," *Journal of Labor Economics*, 8, S329–S363.

HSIEH, C., E. HURST, C. JONES, AND P. KLENOW (2011): "The Allocation of Talent and U.S. Economic Growth," *Mimeo, Stanford University*.

KLEVEN, H. J., C. T. KREINER, AND E. SAEZ (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77, 537–560.

LAROQUE, G. (2005): "Income maintenance and labor force participation," *Econometrica*, 73, 341–376.

LEE, L. (1978): "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19, 415–433.

MANKIW, G., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxes in Theory and Practice," *Journal of Economic Perspectives*, 23, 147–174.

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

MORESI, S. (1997): "Optimal Taxation and Firm Formation: A Model of Asymmetric Information," *European Economic Review*, 42, 1525–1551.

NAITO, H. (1999): "Re-Examination of Uniform Commodity Taxes Under A Non-Linear Income Tax System And Its Implication for Production Efficiency," *Journal of Public Economics*, 71, 165–188.

PARKER, S. C. (1999): "The Optimal Linear Taxation of Employment and Self-Employment Incomes," *Journal of Public Economics*, 73, 107–123.

ROCHET, J.-C., AND P. CHONÉ (1998): "Ironing, Sweeping and Multidimensional Screening," *Econometrica*, 66, 783–826.

ROTHSCHILD, C., AND F. SCHEUER (2011): "Optimal Taxation with Rent-Seeking," *NBER Working Paper 17035*.

——— (2012): "Redistributive Taxation in the Roy Model," *NBER Working Paper 18228*.

ROY, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135–146.

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

——— (2004): "Direct or Indirect Tax Instruments for Redistributiopn: Short-run versus Long-run," *Journal of Public Economics*, 88, 503–518.

SARGENT, T. (2008): "Evolution and Intelligent Design," *American Economic Review*, 98, 5–37.

SCHEUER, F. (2012a): "Adverse Selection in Credit Markets and Regressive Profit Taxation," *NBER Working Paper No. 18406*.

——— (2012b): "Entrepreneurial Taxation with Endogenous Entry," *Mimeo, Stanford University*.

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

WERNING, I. (2007): "Pareto-Efficient Income Taxation," *Mimeo, MIT*.

WILLIS, R., AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87, S7–S36.

ZECKHAUSER, R. (1977): "Taxes in Fantasy, or Most Any Tax on Labor Can Turn Out to Help the Laborers," *Journal of Public Economics*, 8, 133–150.

# A Proof of Lemma 1

*Part (i).* Consider an incentive compatible allocation $\{c(\theta,\varphi), y(\theta,\varphi), S(\theta,\varphi), x\}$. To obtain a contradiction, suppose there exists a type $(\theta,\varphi)$ such that (e.g.) $\theta Y_\theta(x) < \varphi Y_\varphi(x)$, but $S(\theta,\varphi) = \Theta$ and hence (by equation (1)) $w(\theta,\varphi) = \theta Y_\theta(x)$. Then type $(\theta,\varphi)$'s incentive constraint must be violated, since he could do better by earning the same income $y(\theta,\varphi)$ in the $\Phi$- rather than the $\Theta$-sector, which would result in utility

$$U\left(c(\theta,\varphi), \frac{y(\theta,\varphi)}{\varphi Y_\varphi(x)}\right) > U\left(c(\theta,\varphi), \frac{y(\theta,\varphi)}{\theta Y_\theta(x)}\right).$$

*Part (ii).* Consider two types $(\theta,\varphi)$ and $(\theta',\varphi')$ such that $w(\theta,\varphi) = w(\theta',\varphi') = w$. Again to obtain a contradiction, assume w.l.o.g.

$$U\left(c(\theta,\varphi), \frac{y(\theta,\varphi)}{w}\right) < U\left(c(\theta',\varphi'), \frac{y(\theta',\varphi')}{w}\right).$$

Then type $(\theta,\varphi)$'s incentive constraint must be violated: he could mimic type $(\theta',\varphi')$ by earning $y(\theta',\varphi')$, staying in his original sector. His utility from this deviation would be given by the RHS of the above inequality.

*Part (iii).* We claim that, for each type $(\theta,\varphi)$, the consumption bundle $(c(\theta,\varphi), y(\theta,\varphi))$ is a solution to $\max_{(c,y)\in B^*} U\left(c, y/w(\theta,\varphi)\right)$. First, when faced with the budget set $B^*$, each individual will, taking $x$ as given, choose his sector such that $w(\theta,\varphi) = \max\{\theta Y_\theta(x), \varphi Y_\varphi(x)\}$ and $S(\theta,\varphi)$ as given by equation (2). Second, each bundle $(c(\theta,\varphi), y(\theta,\varphi))$ from the original allocation is included in the budget set $B^*$ by the construction in (3) and incentive compatibility. To see this, suppose there was some $(c(\theta,\varphi), y(\theta,\varphi))$ such that $c(\theta,\varphi) > R^*(y(\theta,\varphi))$. Then by (3) this would imply that there exists some $(\theta',\varphi')$ such that

$$U\left(c(\theta',\varphi'), \frac{y(\theta',\varphi')}{w(\theta',\varphi')}\right) < U\left(c(\theta,\varphi), \frac{y(\theta,\varphi)}{w(\theta',\varphi')}\right),$$

contradicting type $(\theta', \varphi')$'s incentive constraint and thus incentive compatibility of the original allocation. Finally, each type $(\theta, \varphi)$ at least weakly prefers the bundle $(c(\theta,\varphi), y(\theta,\varphi))$ to any other bundle available in $B^*$ by (3).

# B Proofs for Section 3

## B.1 Proof of Proposition 1

Putting multipliers $\mu$ on (8), $\xi\mu$ on (5) and $\hat{\eta}(w)\mu$ on (7), the Lagrangian corresponding to (6)-(8) is, after integrating by parts (7),

$$\begin{aligned}\mathcal{L} \quad &= \int_{\underline{w}_x}^{\overline{w}_x} V(w) g_x(w) dw - \int_{\underline{w}_x}^{\overline{w}_x} V(w)\hat{\eta}'(w)\mu dw + \int_{\underline{w}_x}^{\overline{w}_x} U_e(c(V(w),e(w)),e(w))\tfrac{e(w)}{w}\hat{\eta}(w)\mu dw \\ &\quad + \xi\mu(1-\alpha(x)) \int_{\underline{w}_x}^{\overline{w}_x} we(w) f_x^{\theta}(w) dw - \xi\mu\alpha(x) \int_{\underline{w}_x}^{\overline{w}_x} we(w) f_x^{\varphi}(w) dw \\ &\quad + \mu \int_{\underline{w}_x}^{\overline{w}_x} we(w) f_x(w) dw - \mu \int_{\underline{w}_x}^{\overline{w}_x} c(V(w),e(w)) f_x(w) dw. \end{aligned} \tag{26}$$

Using $\partial c/\partial V = 1/U_c$ and compressing notation, the first order condition for $V(w)$ is

$$\hat{\eta}'(w)\mu = g_x(w) - \mu f_x(w)\frac{1}{U_c(w)} + \hat{\eta}(w)\mu\frac{U_{ec}(w)}{U_c(w)}\frac{e(w)}{w}. \tag{27}$$

Defining $\eta(w) \equiv \hat{\eta}(w)U_c(w)$, this becomes

$$\eta'(w) = g_x(w)\frac{U_c(w)}{\mu} - f_x(w) + \eta(w)\frac{U_{cc}(w)c'(w) + U_{ce}(w)e'(w) + U_{ce}(w)e(w)/w}{U_c(w)}. \tag{28}$$

Using the first order condition corresponding to the optimization problem for an individual worker,

$$U_c(w)c'(w) + U_e(w)e'(w) + U_e(w)\frac{e(w)}{w} = 0,$$

the fraction in (28) can be written as $-(\partial MRS(w)/\partial c)y'(w)/w$ where

$$MRS(w) \equiv -\frac{U_e(c(w),e(w))}{U_c(c(w),e(w))}$$

is the marginal rate of substitution between effort and consumption. Substituting in (28) and rearranging yields

$$-\frac{\partial MRS(w)}{\partial c}e(w)\frac{y'(w)}{y(w)}\eta(w) = f_x(w) - g_x(w)\frac{U_c(w)}{\mu} + \eta'(w). \tag{29}$$

Integrating this ODE gives

$$\begin{aligned}\eta(w) \quad &= \quad \int_w^{\overline{w}_x}\left(f_x(w) - g_x(z)\frac{U_c(z)}{\mu}\right)\exp\left(\int_w^z \frac{\partial MRS(s)}{\partial c}e(s)\frac{y'(s)}{y(s)}ds\right)dz \\ &= \quad \int_w^{\overline{w}_x}\left(1 - \frac{g_x(z)}{f_x(z)}\frac{U_c(z)}{\mu}\right)\exp\left(\int_w^z\left(1-\frac{\varepsilon^u(s)}{\varepsilon^c(s)}\right)\frac{dy(s)}{y(s)}\right)f_x(z)dz, \end{aligned} \tag{30}$$

where the last step follows from $e(w)\partial MRS(w)/\partial c = 1 - \varepsilon^u(w)/\varepsilon^c(w)$ after tedious algebra (e.g. using equations (23) and (24) in Saez, 2001).

Using $\partial c/\partial e = MRS$, the first order condition for $e(w)$ is

$$\begin{aligned} & \mu w f_x(w)\left(1-\frac{MRS(w)}{w}\right)+\xi\mu w\left((1-\alpha(x))f_x^\theta(w)-\alpha(x)f_x^\varphi(w)\right) \\ = \quad & -\hat{\eta}(w)\mu\left[\frac{\left(-U_{ec}(w)U_e(w)/U_c(w)+U_{ee}(w)\right)e(w)}{w}+\frac{U_e(w)}{w}\right], \end{aligned}$$

which after some algebra can be rewritten as

$$w f_x(w)\left(1-\frac{MRS(w)}{w}\right)+\xi w\left((1-\alpha(x))f_x^\theta(w)-\alpha(x)f_x^\varphi(w)\right)=\eta(w)\left(\frac{\partial MRS(w)}{\partial e}\frac{e}{w}+\frac{MRS(w)}{w}\right). \tag{31}$$

Noting that $MRS(w)/w = 1 - T'(y(w))$ from the first order condition of the workers' utility maximization problem and using the definition of $\eta(w)$, this becomes

$$1+\xi\frac{(1-\alpha(x))f_x^\theta(w)-\alpha(x)f_x^\varphi(w)}{f_x(w)}=(1-T'(y(w)))\left[1+\frac{\eta(w)}{wf_x(w)}\left(1+\frac{\partial MRS(w)}{\partial e}\frac{e}{MRS(w)}\right)\right]. \tag{32}$$

Simple algebra again shows that $1+\partial\log MRS(w)/\partial\log e = (1+\varepsilon^u(w))/\varepsilon^c(w)$, and that

$$\frac{(1-\alpha(x))f_x^\theta(w)-\alpha(x)f_x^\varphi(w)}{f_x(w)}=1-\alpha(x)-\frac{f_x^\varphi(w)}{f_x(w)}=\frac{f_x^\theta(w)}{f_x(w)}-\alpha(x).$$

The Proposition follows from (30) and (32).

## B.2 Proof of Lemma 2

We begin with the following two technical lemmas, which will be useful for the proof of Lemma 2.

**Lemma 3.** *The substitution elasticity of $Y(E_\theta, E_\varphi)$ is given by $\sigma(x) = -\frac{1}{x\lambda(x)}$ with $\lambda(x) \equiv \frac{Y'_\theta(x)}{Y_\theta(x)} - \frac{Y'_\varphi(x)}{Y_\varphi(x)}$.*

*Proof.* The substitution elasticity is defined as $\sigma(x) \equiv \frac{dx}{x}\frac{Y_\varphi(x)/Y_\theta(x)}{d\left(Y_\varphi(x)/Y_\theta(x)\right)} = \frac{1}{x}\left(\frac{d\log(Y_\varphi(x)/Y_\theta(x))}{dx}\right)^{-1}$, from which the Lemma follows directly. □

**Lemma 4.**

$$\frac{dF_x^\theta(w)}{dx}=-\frac{Y'_\theta(x)}{Y_\theta(x)}wf_x^\theta(w)+\Omega_x(w) \quad \text{and} \quad \frac{dF_x^\varphi(w)}{dx}=-\frac{Y'_\varphi(x)}{Y_\varphi(x)}wf_x^\varphi(w)-\Omega_x(w)$$

*with*

$$\Omega_x(w)\equiv\frac{1}{Y_\theta(x)Y_\varphi(x)}\lambda(x)\int_{\underline{w}_x}^{w}w' f\left(\frac{w'}{Y_\theta(x)},\frac{w'}{Y_\varphi(x)}\right)dw'.$$

*Completely analogous expressions hold for $G_x^\theta(w)$ and $G_x^\varphi(w)$.*

The proof of Lemma 4 involves nothing more than tedious algebra. We now turn to proving Lemma 2

and use (26) to compute

$$W'(x) = \int_{\underline{w}_x}^{\overline{w}_x} V(w)\frac{dg_x(w)}{dx}dw - \mu\int_{\underline{w}_x}^{\overline{w}_x} c(V(w),e(w))\frac{df_x(w)}{dx}dw - \xi\mu\alpha'(x)Y(E_\theta,E_\varphi)$$
$$+\mu\xi\left((1-\alpha(x))\int_{\underline{w}_x}^{\overline{w}_x} we(w)\frac{df_x^\theta(w)}{dx}dw - \alpha(x)\int_{\underline{w}_x}^{\overline{w}_x} we(w)\frac{df_x^\varphi(w)}{dx}dw\right) + \mu\int_{\underline{w}_x}^{\overline{w}_x} we(w)\frac{df_x(w)}{dx}dw + B_1$$

with

$$B_1 \equiv \frac{d\overline{w}_x}{dx}\left[V(\overline{w}_x)g_x(\overline{w}_x) - \mu c(V(\overline{w}_x),e(\overline{w}_x))f_x(\overline{w}_x) + \mu\left(f_x(\overline{w}_x) + \xi\left((1-\alpha(x))f_x^\theta(e(\overline{w}_x)) - \alpha(x)f_x^\varphi(e(\overline{w}_x))\right)\right)\overline{w}_x e(\overline{w}_x)\right]$$
$$-\frac{d\underline{w}_x}{dx}\left[V(\underline{w}_x)g_x(\underline{w}_x) - \mu c(V(\underline{w}_x),e(\underline{w}_x))f_E(\underline{w}_x) + \mu\left(f_x(\overline{w}_x) + \xi\left((1-\alpha(x))f_x^\theta(e(\underline{w}_x)) - \alpha(x)f_x^\varphi(e(\underline{w}_x))\right)\right)\underline{w}_x e(\underline{w}_x)\right].$$

Integrating by parts the five integral terms yields

$$W'(x) = B_1 + B_2 - \int_{\underline{w}_x}^{\overline{w}_x} V'(w)\frac{dG_x(w)}{dx}dw + \mu\int_{\underline{w}_x}^{\overline{w}_x}\left(\frac{V'(w)}{U_c(w)} + MRS(w)e'(w)\right)\frac{dF_x(w)}{dx}dw - \xi\mu\alpha'(x)Y(E_\theta,E_\varphi)$$
$$+\mu\xi\left(\int_{\underline{w}_x}^{\overline{w}_x}(we'(w)+e(w))\left(\alpha(x)\frac{dF_x^\varphi(w)}{dx} - (1-\alpha(x))\frac{dF_x^\theta(w)}{dx}\right)dw\right) - \mu\int_{\underline{w}_x}^{\overline{w}_x}(we'(w)+e(w))\frac{dF_x(w)}{dx}dw \tag{33}$$

with

$$B_2 \equiv \left[V(w)\frac{dG_x(w)}{dx} - \mu c(V(w),e(w))\frac{dF_x(w)}{dx} + \mu\xi we(w)\left((1-\alpha(x))\frac{dF_x^\theta(w)}{dx} - \alpha(x)\frac{dF_x^\varphi(w)}{dx}\right) + \mu we(w)\frac{dF_x(w)}{dx}\right]_{\underline{w}_x}^{\overline{w}_x}.$$

By the first order conditions (29) and (31) with respect to $V(w)$ and $e(w)$ from the inner problem, the terms

$$\mu\int_{\underline{w}_x}^{\overline{w}_x}\frac{e'(w)}{f_x(w)}\left[wf_x(w)\left(1-\frac{MRS(w)}{w}\right)+\xi w\left((1-\alpha(x))f_x^\theta - \alpha(x)f_x^\varphi(w)\right) - \eta(w)\left(\frac{\partial MRS(w)}{\partial e}\frac{e(w)}{w} + \frac{MRS(w)}{w}\right)\right]\frac{dF_x(w)}{dx}dw$$

and

$$\mu\int_{\underline{w}_x}^{\overline{w}_x}\frac{V'(w)}{U_c(w)f_x(w)}\left[g_x(w)\frac{U_c(w)}{\mu} - f_x(w) - \eta'(w) - \eta(w)\frac{\partial MRS(w)}{\partial c}e(w)\frac{y'(w)}{y(w)}\right]\frac{dF_x(w)}{dx}dw$$

are both equal to zero. Adding them to (33), using (7), canceling and re-arranging yields

$$W'(x) = B_1 + B_2 - \xi\mu\alpha'(x)Y(E_\theta,E_\varphi) + \int_{\underline{w}_x}^{\overline{w}_x} V'(w)\left(\frac{g_x(w)}{f_x(w)}\frac{dF_x(w)}{dx} - \frac{dG_x(w)}{dx}\right)dw$$
$$-\mu\int_{\underline{w}_x}^{\overline{w}_x} e(w)\frac{dF_x(w)}{dx}dw - \mu\int_{\underline{w}_x}^{\overline{w}_x}\left(\frac{\eta(w)}{w}\frac{d\left[MRS(w)e(w)\right]}{dw} + \eta'(w)\frac{V'(w)}{U_c(w)}\right)\frac{1}{f_x(w)}\frac{dF_x(w)}{dx}dw$$
$$+\xi\mu\int_{\underline{w}_x}^{\overline{w}_x}\left((e(w)+we'(w))\left(\alpha(x)\frac{dF_x^\varphi(w)}{dx} - (1-\alpha(x))\frac{dF_x^\theta(w)}{dx}\right) + we'(w)\left((1-\alpha(x))\frac{f_x^\theta(w)}{f_x(w)} - \alpha(x)\frac{f_x^\varphi(w)}{f_x(w)}\right)\frac{dF_x(w)}{dx}\right)dw. \tag{34}$$

From Lemma 4,

$$\frac{g_x(w)}{f_x(w)}\frac{dF_x(w)}{dx} - \frac{dG_x(w)}{dx} = -w\frac{Y'_\theta(x)}{Y_\theta(x)}\left[\frac{g_x(w)}{f_x(w)}f_x^\theta(w) - g_x^\theta(w)\right] + w\frac{Y'_\varphi(x)}{Y_\varphi(x)}\left[\frac{g_x(w)}{f_x(w)}f_x^\varphi(w) - g_x^\varphi(w)\right]$$
$$= \left(\frac{Y'_\theta(x)}{Y_\theta(x)} - \frac{Y'_\varphi(x)}{Y_\varphi(x)}\right) w\left[\frac{g_x(w)}{f_x(w)}f_x^\varphi(w) - g_x^\varphi(w)\right] = \lambda(x)w\frac{f_x^\varphi(w)f_x^\theta(w)}{f_x(w)}\left(\frac{g_x^\theta(w)}{f_x^\theta(w)} - \frac{g_x^\varphi(w)}{f_x^\varphi(w)}\right).$$

The first integral in (34) is therefore

$$\int_{\underline{w}_x}^{\overline{w}_x} V'(w)\left(\frac{g_x(w)}{f_x(w)}\frac{dF_x(w)}{dx} - \frac{dG_x(w)}{dx}\right)dw = \lambda(x)R \tag{35}$$

Combining the terms with $e(w)$ on the second and third line of (34) and using Lemma 4 and the identity $Y'_\theta(x)E_\theta + Y'_\varphi(x)E_\varphi \equiv 0$ gives:

$$\begin{aligned}
&-\mu\int_{\underline{w}_x}^{\overline{w}_x} e(w)\frac{dF_x(w)}{dx}dw + \xi\mu\int_{\underline{w}_x}^{\overline{w}_x} e(w)\left(\alpha(x)\frac{dF_x^\varphi(w)}{dx} - (1-\alpha(x))\frac{dF_x^\theta(w)}{dx}\right)dw \\
&= \mu\left[\frac{Y'_\theta(x)}{Y_\theta(x)}\int_{\underline{w}_x}^{\overline{w}_x} we(w)f_x^\theta(w)dw + \frac{Y'_\varphi(x)}{Y_\varphi(x)}\int_{\underline{w}_x}^{\overline{w}_x} we(w)f_x^\varphi(w)dw\right] \\
&\quad + \xi\mu\int_{\underline{w}_x}^{\overline{w}_x} we(w)\left((1-\alpha(x))\frac{Y'_\theta(x)}{Y_\theta(x)}f_x^\theta(w) - \alpha(x)\frac{Y'_\varphi(x)}{Y_\varphi(x)}f_x^\varphi(w)\right)dw - \xi\mu\int_{\underline{w}_x}^{\overline{w}_x} e(w)\Omega_x(w)dw \\
&= 0 + \xi\mu Y'_\theta E_\theta - \xi\mu\int_{\underline{w}_x}^{\overline{w}_x} e(w)\Omega_x(w)dw.
\end{aligned} \tag{36}$$

The terms with $we'(w)$ in the last line (34) can be written, using Lemma 4 again, as

$$\xi\mu\int_{\underline{w}_x}^{\overline{w}_x} we'(w)\left(\alpha(x)\frac{dF_x^\varphi(w)}{dx} - (1-\alpha(x))\frac{dF_x^\theta(w)}{dx} + \left((1-\alpha(x))\frac{f_x^\theta(w)}{f_x(w)} - \alpha(x)\frac{f_x^\varphi(w)}{f_x(w)}\right)\frac{dF_x(w)}{dx}\right)dw$$

$$\begin{aligned}
&= \xi\mu\int_{\underline{w}_x}^{\overline{w}_x} w^2e'(w)\left[(1-\alpha(x))\frac{Y'_\theta(x)}{Y_\theta(x)}f_x^\theta(w) - \alpha(x)\frac{Y'_\varphi(x)}{Y_\varphi(x)}f_x^\varphi(x) - \left((1-\alpha(x))\frac{f_x^\theta(w)}{f_x(w)} - \alpha(x)\frac{f_x^\varphi(w)}{f_x(w)}\right)\left(\frac{Y'_\theta(x)}{Y_\theta(x)}f_x^\theta(w) + \frac{Y'_\varphi(x)}{Y_\varphi(x)}f_x^\varphi(x)\right)\right]dw \\
&\quad - \xi\mu\int_{\underline{w}_x}^{\overline{w}_x} we'(w)\Omega_x(w)dw \\
&= \lambda(x)\xi\mu C - \xi\mu\int_{\underline{w}_x}^{\overline{w}_x} we'(w)\Omega_x(w)dw,
\end{aligned} \tag{37}$$

where the first term in the last step follows after some tedious algebra. Combining the terms with $\Omega_x(w)$ from (36) and (37) gives $-\xi\mu\int_{\underline{w}_x}^{\overline{w}_x}(we(w))'\Omega_x(w)dw$, which can be integrated by parts to yield:

$$-B_3 + \xi\mu\frac{\lambda(x)}{Y_\theta(x)Y_\varphi(x)}\int_{\underline{w}_x}^{\overline{w}_x} w^2e(w)f\left(\frac{w}{Y_\theta(x)}, \frac{w}{Y_\varphi(x)}\right)dw = \xi\mu\lambda(x)S, \tag{38}$$

with $B_3 = \xi\mu\overline{w}_xe(\overline{w}_x)\Omega_x(\overline{w}_x)$ since $\Omega_x(\underline{w}_x) = 0$. Finally, use the incentive constraint (7), rewritten as

$wV'(w)/U_c(w) = MRS(w)e(w)$, to write the second integral in the second line of (34) as

$$-\mu \int_{\underline{w}_x}^{\overline{w}_x} \left( \eta(w) w \frac{d[V'(w)/U_c(w)]}{dw} + \eta'(w) w \frac{V'(w)}{U_c(w)} + \eta(w) \frac{V'(w)}{U_c(w)} \right) \frac{1}{w f_x(w)} \frac{dF_x(w)}{dx} dw$$

or, recognizing the sum of the bracketed terms as $d[\eta(w)wV'(w)/U_c(w)]/dw$, integrating by parts, and using the transversality condition $\eta(\underline{w}_x) = \eta(\overline{w}_x) = 0$ and Lemma 4,

$$\mu \int_{\underline{w}_x}^{\overline{w}_x} \eta(w) w \frac{V'(w)}{U_c(w)} \frac{d}{dw} \left( - \left( \frac{Y'_\theta(x)}{Y_\theta(x)} \frac{f_x^\theta(w)}{f_x(w)} - \frac{Y'_\varphi(x)}{Y_\varphi(x)} \frac{f_x^\varphi(w)}{f_x(w)} \right) \right) dw = \lambda(x) I \tag{39}$$

Define $\tilde{F}(w,x) \equiv F_x(w)$. Since $\tilde{F}(\overline{w}_x, x) \equiv 1$ for all $x$,

$$\frac{d\tilde{F}(\overline{w}_x,x)}{dx} = \frac{\partial \tilde{F}(\overline{w}_x,x)}{\partial x} + \frac{\partial \tilde{F}(\overline{w}_x,x)}{\partial w} \frac{d\overline{w}_x}{dx} = \frac{dF_x(\overline{w}_x)}{dx} + f_x(\overline{w}_x) \frac{d\overline{w}_x}{dx} = 0. \tag{40}$$

Together with an analogous expression at $\underline{w}_x$, the fact that $\Omega_x(\underline{w}_x) = 0$, and Lemma 4, this yields

$$B_1 + B_2 = -\xi \mu \overline{w}_x e(\overline{w}_x) \Omega_x(\overline{w}_x) = B_3.$$

Using (35), (36), (37), (38) and (39) in (34) yields

$$W'(x) = \lambda(x) \left( I + R + \xi \mu \left[ \alpha(x) Y_\varphi(x) E_\varphi \sigma(x) + S + C \right] \right), \tag{41}$$

where we have used $-\alpha'(x)Y(E_\theta, E_\varphi) + Y'_\theta(x)E_\theta = -\alpha(x)Y_\varphi(x)E_\varphi/x$.

# C Proof of Proposition 3

$$\begin{aligned}
\lim_{w\to\infty} 1 - T'(y(w)) &= \lim_{w\to\infty} \frac{1 + \xi \left( 1 - \alpha(x) - \dfrac{f_x^\varphi(w)}{f_x(w)} \right)}{1 + \dfrac{\eta(w)}{w f_x(w)} \dfrac{1+\varepsilon^u(w)}{\varepsilon^c(w)}} \\
&= \frac{1 + \xi \left( 1 - \alpha(x) - \lim\limits_{w\to\infty} \left( \dfrac{f_x^\varphi(w)}{f_x(w)} \right) \right)}{1 + \left( 1 + \dfrac{1}{\varepsilon} \right) \lim\limits_{w\to\infty} \left( \dfrac{1 - F_x(w)}{w f_x(w)} \right) \left( \dfrac{\Psi(F_x(w)) - F_x(w)}{1 - F_x(w)} \right)} = \frac{1 + \xi (1 - \alpha(x))}{1 + \left( 1 + \dfrac{1}{\varepsilon} \right) \kappa}.
\end{aligned}$$

The first equality is from (10). The second uses (iv) and (i), which implies $\eta(w) = \Psi(F_x(w)) - F_x(w)$. The third uses (ii) to simplify the numerator, and (iii) and (v) to take the limits of the two terms in the denominator. The top tax rate result for the Pareto optimum follows with a little re-arranging. Setting $\xi = 0$ yields the result for the SCPE. Corollary 4 and (iv) imply $\xi > 0$, since $\Psi''(x) < 0$ implies that $g_x(w)/f_x(w) = \Psi'(F_x(w))$ is decreasing in $w$.