NBER WORKING PAPER SERIES

REDISTRIBUTIVE TAXATION IN THE ROY MODEL

Casey Rothschild
Florian Scheuer

Working Paper 18228
http://www.nber.org/papers/w18228

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2012

We are grateful to Phillipe Choné, Craig Brett, John Weymark, and seminar participants at the 2012 Taxation Theory Conference (Vanderbilt), SED Annual Meeting (Limassol) and LMU Munich. All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Redistributive Taxation in the Roy Model
Casey Rothschild and Florian Scheuer
NBER Working Paper No. 18228
July 2012
JEL No. D5,D80,E2,E6,H2,J3,J6

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.

Casey Rothschild
Wellesley College
106 Central Street
Wellesley, MA 02481
crothsch@wellesley.edu

Florian Scheuer
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
and NBER
scheuer@stanford.edu

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, naturally the underlying skill is 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), but 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 showing that a single non-linear income tax schedule is sufficient for implementing any incentive compatible allocation that is feasible for a government that cannot condition on sectoral choice. Since incomes depend on the realized wage, not on sectoral choice per se, 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, however: 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, the income tax system may not distinguish between the two sectors: a white collar and blue collar worker earning the same income y pay the same tax $T(y)$.

With a linear production technology (i.e., when equivalent units of white and blue collar effort 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, 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. A SCPE describes the tax system that would emerge in the same economy if the government *naively* believed that it was operating in a standard exogenous-wage world. In such a world, a government would, following Saez (2001), infer an underlying skill distribution from the income distribution it observes given an existing tax system. Taking this skill distribution as exogenous, it would then compute the optimal income tax system. In a 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 a 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.

Most closely related to our analysis is Stiglitz (1982), who considers optimal non-linear 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.[1] 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 discrete 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

[1] Naito (1999) studies the role of sector-specific taxes in a two-type model and observes that introducing production inefficiencies (manipulating wages) can be desirable when it relaxes incentive constraints. We do not consider sector-specific taxation in this paper.

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.

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, but the second dimension of heterogeneity in their model takes a specialized additive form. Choné and Laroque (2010) and Brett and Weymark (2003) use a "type aggregator" which is closer in spirit to the techniques we use, but the types of heterogenity they consider (a taste for labor, and an aptitude for education, respectively) are quite distinct from the multi-dimensional skill types that arise naturally in our Roy model.

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 the tax code. Our concept of a self-confirming policy equilibrium, described above, is meant to capture this situation. It is closely related to the recent literature on self-confirming equilibria in learning models (e.g., Sargent, 2009, and Fudenberg and Levine, 2009).

The paper proceeds as follows. In section 2, we describe the basic model and show 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. In section 3, we characterize and compare optimal and SCPE non-linear taxes. In Section 4, we explicitly compare our results to Stiglitz (1982) and point out the novel role of occupational choice and overlapping wage distributions in our model. Section 5 empirically calibrates a simple Roy model of the U.S. economy to explore the potential quantitative importance of sectoral choice effects for optimal taxes. Section 6 extends our results to allow for unbounded skill distributions and additional heterogeneity in tastes or costs for different occupations. Section 7 concludes. All proofs are in the Appendix.

2 The Model

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 $\Phi = [\underline{\varphi}, \overline{\varphi}]$. θ captures an individual's productivity when working in the Θ-sector; φ captures her Φ-sector skill. These skills are distributed in the population according to a continuous two-dimensional cdf $F(\theta, \varphi)$ with density $f(\theta, \varphi)$.

Individuals have preferences over consumption c and effort e captured by the 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 (θ, φ) 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 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

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

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

2.1 A Direct Implementation Approach

We start by characterizing a general, direct implementation, where individuals announce their privately known type (θ, φ) 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. we want to send type (θ, φ) to the Θ-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 (θ, φ) to imitate another type (θ', φ'), namely by earning (θ', φ')'s income either in the Θ- or the Φ-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 following lemma shows that an incentive compatible allocation can be implemented by offering a single (non-linear) income tax schedule.

Lemma 1. *Suppose that only income y and consumption c are observable, whereas an individual's skill type (θ, φ), effort e and sectoral choice S are private information. Then an incentive compatible allocation can be implemented by offering a schedule of $c(w), y(w)$-bundles with $w \equiv \max\{Y_\theta(x)\theta, Y_\varphi(x)\varphi\}$ and*

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

which is equivalent to offering a single (non-linear) income tax schedule.

Proof. In Appendix A. □

In other words, if an individual's sectoral choice is not observable, there is no loss in considering the set of allocations implemented by offering a single non-linear income tax schedule and letting individuals choose their preferred income and sector given this.

In particular, any such allocation will treat any two individuals who earn the same wage (albeit in different sectors) the same, and individuals will work in the sector in which they can achieve a higher wage. We study these allocations below.

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 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, and analogously 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

$$\tilde{x}(x) \equiv \frac{\int_{\underline{w}_x}^{\overline{w}_x} \frac{w}{Y_\theta(x)} e(w) f_x^\theta(w) dw}{\int_{\underline{w}_x}^{\overline{w}_x} \frac{w}{Y_\varphi(x)} e(w) f_x^\varphi(w) dw} = x, \quad \text{or} \tag{2}$$

$$(1-\alpha(x))\int_{\underline{w}_x}^{\overline{w}_x} we(w)f_x^{\theta}(w)dw - \alpha(x)\int_{\underline{w}_x}^{\overline{w}_x} we(w)f_x^{\varphi}(w)dw = 0, \tag{3}$$

where $\alpha(x)$ denotes the aggregate income share of the Θ-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 can therefore write the Pareto problem for income taxation in the Roy model as

$$\max_x W(x) \equiv \max_{e(w),V(w)} \int_{\underline{w}_x}^{\overline{w}_x} V(w)dG_x(w) \tag{4}$$

subject to (3),

$$V'(w) + U_e(c(V(w),e(w)),e(w))\frac{e(w)}{w} = 0 \;\; \forall w \in [\underline{w}_x, \overline{w}_x], \tag{5}$$

and

$$\int_{\underline{w}_x}^{\overline{w}_x} (we(w) - c(V(w),e(w)))\, f_x(w)dw \geq 0. \tag{6}$$

We employ the standard Mirrleesian approach of optimizing directly over consumption-income allocations or, equivalently, $e(w), V(w)$-bundles, where $V(w) \equiv U(c(w),e(w))$ and $c(V,e)$ denotes the inverse of $U(.)$ w.r.t. its first argument. We refer to the three constraints (3), (5) and (6) as the consistency condition, the incentive constraints and the resource constraint, respectively.[2]

As suggested by the previous discussion and the notation in (4), it is useful to decompose the Pareto problem (3) to (6) 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 (4) when maximizing over $V(w), e(w)$ subject to (3), (5) and (6) (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 the subset of allocations on the Pareto frontier that result from attaching the same Pareto weight to any two individuals who end up earning the same wage (and thus income), regardless of their sectoral choice. These allocations are obtained by solving the Pareto problem using relative welfare weights $\Psi(F(\theta,\varphi))$ rather than the more general Pareto weights $G(\theta,\varphi)$.

[2]Note that we have made use of the local version of the incentive constraints in (5). 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 (5) together with the monotonicity constraint that income $y(w)$ must be non-decreasing in w are equivalent to the global incentive constraints $V(w) \equiv U(c(w),e(w)) = \max_{w'} U(c(w'),(e(w')w')/w)$ 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.

Intuitively, $\Psi(.)$ determines how much social welfare weight is attached to different quantiles of the type distribution, and hence the resulting wage distribution $F_x(w)$.

To compare Pareto optimal tax schedules to those that would be optimal in a standard Mirrlees model with an exogenous skill and thus wage distribution, we consider a naive social planner who incorrectly presumes the wage distribution is exogenous. We define a self-confirming equilibrium (SCPE) as a mutually consistent tax schedule and wage distribution pair: in a SCPE, the planner designs an optimal tax schedule taking the wage distribution as given, and this tax policy happens to induce exactly the original wage distribution, so that the planner finds herself confirmed in her view that the wage distribution is exogenous.[3] This notion captures the standard approach to optimal taxation in our environment. For example, Saez (2001) suggests computing optimal taxes via a two step process: first, identify the wage distribution from the observed income distribution and for a given tax schedule; second, design an optimal tax schedule taking this identified wage distribution as given. This process is self-consistent at, and only at, a SCPE.

Optimal taxes in an exogenous-wage environment, and hence SCPE tax rates, can be computed using relative welfare weights $\Psi(F_x(w))$.[4] The inner SCPE problem for a given x is then

$$\max_{e(w),V(w)} \int_{\underline{w}_x}^{\overline{w}_x} V(w) d\Psi(F_x(w)) \tag{7}$$

subject to (5) and (6). The consistency constraint (3) disappears in this problem because such a planner is not aware of the fact that the wage distribution is endogenous and the allocation has to be consistent with x. To ensure that the allocation in aggregate is consistent, however, we have to ensure that the value of x implied by the solution to the inner problem in fact equals x. Formally, the outer problem requires finding a fixed point of the mapping $x \to \tilde{x}^*(x)$, where $\tilde{x}^*(x)$ is defined by (2), evaluated at the allocation $e(w), V(w)$ that solves the inner SCPE problem. We summarize as follows:

Definition 1. *A self-confirming policy equilibrium (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.*

This formulation suggests a natural comparison between SCPE and Pareto optimal allocations, namely between a SCPE for given quantile weights $\Psi(F_x(w))$ and the Pareto optimum that results with two-dimensional relative welfare weights $\Psi(F(\theta, \varphi))$, using the same $\Psi(.)$. We use this comparison in our quantitative explorations in section 5.

[3]See Rothschild and Scheuer (2011) for an analogous concept in the context of rent-seeking.

[4]Since the planner in a SCPE views wages as exogenous economic fundamentals and is unaware of the multi-sector character of the economy, the notion of SCPE is not consistent with general Pareto weights $G(\theta, \varphi)$ over sector-specific skill types, but only with wage or wage-distribution $F_x(w)$ specific weights.

3.2 Example: Entrepreneurs and Workers

Before characterizing optimal income tax schedules in Pareto optima and SCPE, we briefly demonstrate that our model can readily capture the situation where agents in one sector are hired by agents in the other sector, as in Scheuer's (2011) model of entrepreneurship and endogenous firm formation. In particular, suppose Θ is the entrepreneurial sector and Φ the workers' sector, and individuals have skills θ and φ in the respective activities. Let workers supply effective labor $l \equiv \varphi e$ at (endogenous) wage ω. Entrepreneurs hire effective labor in a competitive labor market taking the wage ω as given, and combine this labor L with their own effective entrepreneurial effort $E \equiv \theta e$ using a constant returns to scale technology $Y(E, L)$ (with diminishing marginal products of E and L) to produce a consumption good.

Entrepreneurs will choose E and L to maximize utility

$$\max_{E,L} U\left(Y(E,L) - \omega L - T(Y(E,L) - \omega L), E/\theta\right),$$

where $T(\pi)$ (with $T' \leq 1$) is the nonlinear income (profit) tax. This immediately implies that, for any given effective effort E and wage ω, entrepreneurs set their labor demand by solving the subproblem $\max_L Y(E, L) - \omega L$ with the necessary condition $Y_L(E, L) = \omega$. By constant returns to scale, Y_L depends only (and monotonically) on the ratio E/L. This necessary condition therefore implicitly defines the input mix of all firms via $E/L \equiv \chi(\omega)$, where $\chi(\cdot)$ is the inverse of Y_L.

Moreover, by constant returns to scale, profits are $\pi = Y - \omega L = Y_E E + Y_L L - \omega L = Y_E E$. We can therefore view entrepreneurs as facing a wage $\tilde{\omega} \equiv Y_E$ on their effective entrepreneurial effort, which, by $\tilde{\omega} = Y_E(\chi(\omega))$, is a decreasing function of ω. Consequently, ω fully determines the occupational choice and wage for each individual: Just as in the Roy model, individuals will select the sector which affords them a higher return on their effort, so that $w \equiv \max\{\tilde{\omega}(\omega)\theta, \omega\varphi\}$. We can therefore again write allocations $c(w), e(w)$ as a function of the wage w and obtain a wage distribution $F_\omega(w)$ and sectoral densities $f_\omega^\varphi(w)$, $f_\omega^\theta(w)$, and so forth, conditional on ω.

In equilibrium, ω has to be such that the labor market clears, i.e., such that total effective labor hired by all individuals who decide to enter entrepreneurship is equal to total effective labor supplied by all workers. Effective labor demand per firm is $L = E/\chi(w)$ and effective entrepreneurial effort is $E = e\theta = ew/\tilde{\omega}$. Analogously, effective labor supply per worker can be written as $l = e\varphi = ew/\omega$. Hence, labor market clearing requires

$$\int_{\underline{w}_\omega}^{\overline{w}_\omega} e(w)\frac{w}{\tilde{\omega}\chi(\omega)} f_\omega^\theta(w)dw = \int_{\underline{w}_\omega}^{\overline{w}_\omega} e(w)\frac{w}{\omega} f_\omega^\varphi(w)dw. \tag{8}$$

This shows that, in this model, the consistency condition (2) can be interpreted as a labor market clearing condition: Equation (8) is completely equivalent to (2) since there is a one-to-one relationship between ω and χ, $\omega = Y_L(\chi)$, $\tilde{\omega} = Y_E(\chi)$, and $\chi = x$.

Finally, note that using a firm-level production function is equivalent to the aggregate production function $Y(E_\theta, E_\varphi)$ that we consider for the Roy model, since total output produced by all entrepreneurs is

$$\begin{aligned}
&\int_{\underline{w}_\omega}^{\overline{w}_\omega} Y\left(e(w)\frac{w}{\tilde{\omega}}, e(w)\frac{w}{\tilde{\omega}\chi(\omega)}\right) f_\omega^\theta(w)dw = Y\left(\frac{1}{\tilde{\omega}}, \frac{1}{\tilde{\omega}\chi(\omega)}\right)\int_{\underline{w}_\omega}^{\overline{w}_\omega} e(w)wf_\omega^\theta(w)dw \\
= \;& Y\left(\int_{\underline{w}_\omega}^{\overline{w}_\omega} e(w)\frac{w}{\tilde{\omega}} f_\omega^\theta(w)dw, \int_{\underline{w}_\omega}^{\overline{w}_\omega} e(w)\frac{w}{\tilde{\omega}\chi(\omega)} f_\omega^\theta(w)dw\right) \\
= \;& Y\left(\int_{\underline{w}_\omega}^{\overline{w}_\omega} e(w)\frac{w}{\tilde{\omega}} f_\omega^\theta(w)dw, \int_{\underline{w}_\omega}^{\overline{w}_\omega} e(w)\frac{w}{\omega} f_\omega^\varphi(w)dw\right),
\end{aligned}$$

where we repeatedly used the fact that Y has constant returns to scale and the last step used condition (8).

3.3 Inner Problem

We now return to the Roy model and use the inner problem, for fixed x, to derive formulas for marginal tax rates. We start with Pareto optimal tax schedules using general Pareto weights $G(\theta, \varphi)$.

Proposition 1. *Let μ denote the multiplier on the resource constraint (6), let $\mu\xi$ denote the multiplier on the consistency condition (3), and use $\varepsilon^u(w)$ and $\varepsilon^c(w)$ to denote the uncompensated and compensated labor supply elasticities, respectively. For given x and Pareto weights G, marginal tax rates satisfy*

$$1 - T'(y(w)) = \left(1 + \xi\left(\frac{f_x^\theta(w)}{f_x(w)} - \alpha(x)\right)\right)\left(1 + \frac{\eta(w)}{wf_x(w)}\frac{1+\varepsilon^u(w)}{\varepsilon^c(w)}\right)^{-1} \tag{9}$$

where

$$\eta(w) = \int_w^{\overline{w}_x}\left(1 - \frac{g_x(z)}{f_x(z)}\frac{U_c(z)}{\mu}\right)\exp\left(\int_w^z\left(1 - \frac{\varepsilon^u(s)}{\varepsilon^c(s)}\right)\frac{dy(s)}{y(s)}\right) f_x(z)dz. \tag{10}$$

Proof. See Appendix B. □

Since the inner SCPE problem differs from the inner Pareto problem by the absence of the consistency condition (3), the marginal tax rates for the SCPE can be found by using relative Pareto weights Ψ and setting $\xi = 0$ in Proposition 1, as in the following Corollary.

Corollary 1. *For given x and relative Pareto weights Ψ, marginal tax rates in any SCPE satisfy*

$$1 - T'(y(w)) = \left(1 + \frac{\eta(w)}{wf_x(w)} \frac{1 + \varepsilon^u(w)}{\varepsilon^c(w)}\right)^{-1}, \tag{11}$$

where

$$\eta(w) = \int_w^{\overline{w}_x} \left(1 - \Psi'(F_x(z))\frac{U_c(z)}{\mu}\right) \exp\left(\int_w^z \left(1 - \frac{\varepsilon^u(s)}{\varepsilon^c(s)}\right) \frac{dy(s)}{y(s)}\right) f_x(z)dz. \tag{12}$$

The formulas for the 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 ξ and a comparison between the aggregate income share of the Θ-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 $\xi > 0$ (we will show in section 3.5 that this corresponds to the case where Θ 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 Θ-sector exceeds its aggregate income share. This disproportionately encourages Θ-sector effort and therefore raises wages in the Φ-sector relative to the Θ-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 problem relative to the SCPE: If Θ 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

[5]Note that the formulas in Corollary 1 are evaluated at endogenously determined x-values. Since, in general, the level of x in the SCPE and the solution to the Pareto problem will differ for a given economy, even when based on the same weights Ψ, 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 skill distributions.

[6]The same would be true if Φ were the high-income sector. Then ξ will be negative and hence, per (9), marginal tax rates will again be higher (lower) in the Pareto optimum for low (high) income levels.

disappear and preferences are of the quasilinear form $U(c,e) = c - h(e)$. Then $U_c(w) = \mu = 1$ and $\varepsilon^c(w) = \varepsilon^u(w)$, 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 Θ-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*

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

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 ξ, which will generally be such that this top marginal tax rate is negative.

3.4 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.

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

$$W'(x) = -\frac{1}{x}\left[\xi\mu\alpha(x)Y_\varphi(x)E_\varphi + \frac{1}{\sigma(x)}\left(I + R + \xi\mu(S + C)\right)\right], \tag{13}$$

where

$$S \equiv \frac{1}{Y_\theta(x)Y_\varphi(x)}\int_{\underline{w}_x}^{\overline{w}_x} w^2 e(w) f\left(\frac{w}{Y_\theta(x)}, \frac{w}{Y_\varphi(x)}\right) dw > 0, \tag{14}$$

$$I \equiv \mu \int_{\underline{w}_x}^{\overline{w}_x} \eta(w) w \frac{V'(w)}{U_c(w)} \frac{d}{dw} \left(\frac{f_x^{\varphi}(w)}{f_x(w)} \right) dw \tag{15}$$

$$R \equiv \int_{\underline{w}_x}^{\overline{w}_x} w V'(w) \frac{f_x^{\theta}(w) f_x^{\varphi}(w)}{f_x(w)} \left(\frac{g_x^{\theta}(w)}{f_x^{\theta}(w)} - \frac{g_x^{\varphi}(w)}{f_x^{\varphi}(w)} \right) dw, \tag{16}$$

and

$$C \equiv \int_{\underline{w}_x}^{\overline{w}_x} w^2 e'(w) \frac{f_x^{\varphi}(w) f_x^{\theta}(w)}{f_x(w)} dw. \tag{17}$$

Proof. In Appendix B.2. □

We provide here a heuristic derivation that reveals the intuition behind this result. To that end, we break the welfare effects of a small change in x into four effects:

1. A *direct* effect: the effect of changing x on $\alpha(x)$ in (3), holding the wage, sector and allocation of each individual constant.

2. A *direct wage shift* effect: the effect arising from the change in wages induced by the change in x, holding each individual's allocation and sector constant.

3. An *indirect wage shift* effect: the effect arising from the change in allocations caused by the change in wages induced by the change in x, holding the individual's sector constant.

4. A *sectoral shift* effect: the effect that arises from individuals changing sectors, holding their wage and allocation constant.

Notice first that if technology is linear, so that $\sigma(x) = \infty$, then setting $W'(x) = 0$ in (13) 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 wage shift and sectoral shift effects driven by the endogeneity of wages in the finite $\sigma(x)$ case that provide scope for using additional tools for accomplishing redistributive objectives.

Because individual allocations are held constant, the direct wage shift has no effect on the objective. Because of constant returns to scale, it also has no effect on the resource constraint. However, it affects both the incentive and consistency constraints. The effect on the latter can be combined with the direct effect (of changing x on $\alpha(x)$) to yield $-\alpha(x)Y_{\varphi}(x)E_{\varphi}/x$. To wit: an alternative formulation of the consistency constraint is $E_{\theta}/E_{\varphi} - x = 0$. Since these effects hold sector and allocation $e(w)$ constant for each individual, the net effect would be the negative of the Lagrange multiplier on this constraint in this formulation; and (3) can be written as $\left(\alpha(x)Y_{\varphi}(x)E_{\varphi}/x\right) \left[E_{\theta}/E_{\varphi} - x\right] = 0$.

The term in expression (13) containing I arises from the direct effect of the wage shift on the incentive constraints. To understand it, consider the effects of a small increase in x for a portion of the wage distribution centered at wage w. Such a increase will raise Φ-sector wages and lower Θ-sector wages. If the share of Θ-sector workers is *increasing* locally, this leads to a local compression of wage distribution. Such a compression eases the incentive compatibility constraints if they are binding in the downward direction (i.e., higher wage individuals are tempted to imitate lower wage individuals), in which case $\eta(w) > 0$. An 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.

The sectoral shift and indirect wage-shift effects are effects that are *not* present in Stiglitz's (1982) framework. It is therefore worth elaborating on why they take the form they do, and in particular, why they reinforce each other whenever $e'(w) > 0$. We consider the sectoral shift effect first.

3.4.1 The sectoral shift

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 Θ-sector, and similarly for $Z_\varphi(x)$. Consider a small increase Δx in x, holding efforts and wages constant. This will lead some individuals to shift from the Θ- to the Φ-sector, as illustrated in figure 1. Let $\Delta Z_\theta(x)$ denote the resulting change in Θ-sector income. Since there is an equal and opposite change in Φ-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 Θ-sector skills between θ and $\theta + d\theta$ who are in the Θ-sector at x but in the Φ-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

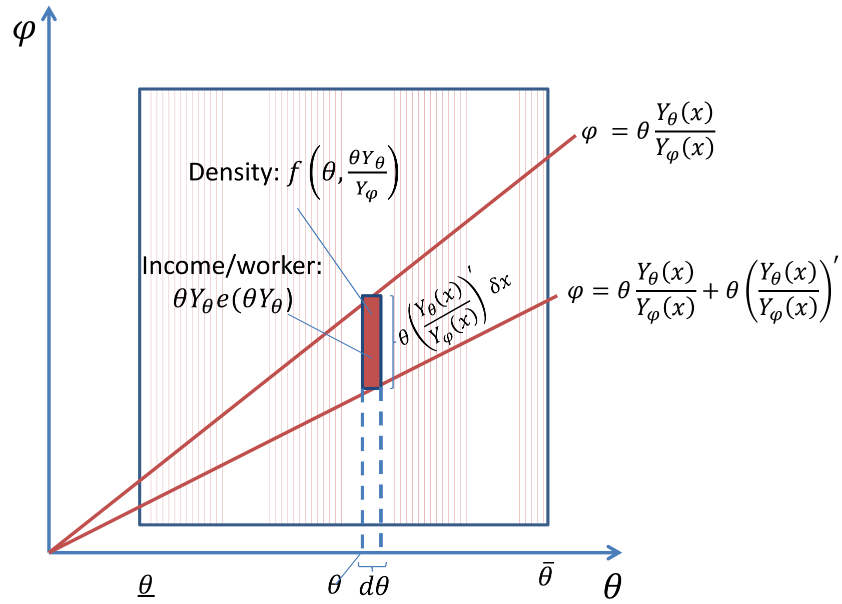

Figure 1: Illustrating the computation of the sectoral shift effect

density, and the per individual income, and then integrating over θ gives:

$$\Delta Z_\theta(x) = \Delta x \int_{\underline{\theta}}^{\bar{\theta}} \left(\frac{Y_\theta'(x) Y_\varphi(x) - Y_\theta(x) Y_\varphi'(x)}{Y_\varphi(x)^2} \right) Y_\theta(x) \theta^2 e\left(\theta Y_\theta(x)\right) f\left(\theta, \frac{\theta Y_\theta(x)}{Y_\varphi(x)}\right) d\theta$$

$$= \Delta x \frac{-1/(x\sigma(x))}{Y_\theta(x) Y_\varphi(x)} \int_{\underline{w}}^{\overline{w}} w^2 e\left(w\right) f\left(\frac{w}{Y_\theta(x)}, \frac{w}{Y_\varphi(x)}\right) dw,$$

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 (14) follows directly.

3.4.2 The indirect wage shift effect

There is no simple graphical representation of the indirect wage shift effect, but a similar heuristic could be used to derive the terms C and R. We omit the algebraic details here, and instead provide the basic intuition behind those effects. (See Appendix B.2 for a formal treatment.)

Imagine increasing x by a small amount while holding the tax code constant so that allocations $e(w)$ and $V(w)$ are an unchanged function of wages. The two types of individuals at original wage w^* are affected differently by the change in x: individuals in the

Φ-sector find their wage increases; Θ-sector individuals see their wage decrease. The Θ-sector individuals move down along the (fixed) schedules $e(w)$ and $V(w)$, and Φ-sector individuals move up. Depending on the schedules and the proportions of the two types at w^*, the net effort and utility effect on wage w^*-individuals may be positive or negative. The algebraic manipulations in Appendix B.2 are motivated by thinking of these shifts in two steps: a level shift of all wage w^*-individuals that absorbs the *net* shift of effort and utility, and a *re-allocation* of effort and utility across the two types. The former involves a particular shift in the $e(w)$ and $V(w)$ schedules, and has a zero net welfare effect by an envelope argument (the original schedules were optimal). The re-allocation of effort and utility in the latter give rise to the terms C and R in Lemma 2.

Because it arises from a re-allocation of utility across individuals at the same w^* induced by the change in x, the term R 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. It is straightforward to show that this will be the case whenever $G(\theta, \varphi)$ takes the form $G(\theta, \varphi) = \Psi(F(\theta, \varphi))$. In contrast, when the social planner has an intrinsic preference for the Θ-sector individuals at wage w^*, the re-allocation of utility from the Θ- to the Φ-sector is welfare reducing.

The term C arises from an analogous re-allocation of effort. If the effort schedule is increasing (i.e. $e'(w) \geq 0$), then an increase in x effectively re-allocates effort from the Θ- to the Φ-sector, because Θ-workers move down and Φ-workers up along the $e(w)$ schedule. This effect therefore reinforces the sectoral shift effect.

3.5 Marginal Tax Rate Results

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

$$\xi = -\frac{\dfrac{I+R}{\mu\sigma(x)}}{\alpha(x)Y_\varphi(x)E_\varphi + \dfrac{C+S}{\sigma(x)}}. \tag{18}$$

We summarize the resulting conditions for the sign of ξ in the following corollary:

Corollary 4. *With 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 indirect wage shift term C reinforces the sectoral shift effect. The latter holds whenever the (average) marginal social value of consumption, given by $U_c(w)g_x(w)/f_x(w)$, is decreasing. This is guaranteed with quasilinear-in-consumption preferences and weakly progressive welfare weights (such that $g_x(w)/f_x(w)$ is increasing), 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 Θ is the high-skill sector, and an increase in x, by raising wages in the Φ-sector and lowering them in the Θ-sector, has desirable wage compression effects, as in Stiglitz (1982). This desirable effect is reinforced by 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 Φ-sector workers than on Θ-sector workers at any given wage, and the wage changes induced by an increase in x also have direct 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 Θ is the high-skilled sector, i.e. $f_x^{\varphi}(w)/f_x(w)$ is decreasing so that $\xi > 0$ by Corollary 4. Then the marginal keep share in the Pareto optimum is scaled down relative to the SCPE wherever the local income share in the Φ-sector is higher than in aggregate. This disproportionately reduces Φ-sector effort and therefore indirectly increases wages in the Φ-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. I.e., 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 $G_x(w) = \Psi(F_x(w))$. Then $g_x(w) = \Psi'(F_x(w))f_x(w)$ and hence

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

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 ξ on the consistency constraint accounts for the overall redistributive motive across sectors, i.e. whether we want to redistribute from Θ to Φ or vice versa. 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 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 a 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 θ and φ and with fractions f_θ and $f_\varphi = 1 - f_\theta$ in the population. We put (relative) Pareto weights ψ_θ and ψ_φ on them such that $f_\theta\psi_\theta + f_\varphi\psi_\varphi = 1$. Without loss of generality, we will think of θ as the high wage sector and φ 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 θ-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

$$W(x) = \max_{e_\theta, e_\varphi, V_\theta, V_\varphi} f_\theta \psi_\theta V_\theta + f_\varphi \psi_\varphi V_\varphi \tag{19}$$

subject to

$$V_\theta \geq U\left(c(V_\varphi, e_\varphi), e_\varphi \frac{w_\varphi}{w_\theta}\right), \tag{20}$$

$$(1 - \alpha(x)) f_\theta w_\theta e_\theta - \alpha(x) f_\varphi w_\varphi e_\varphi = 0, \tag{21}$$

$$f_\theta w_\theta e_\theta + f_\varphi w_\varphi e_\varphi \geq f_\theta c(V_\theta, e_\theta) + f_\varphi c\left(V_\varphi, e_\varphi\right). \tag{22}$$

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

We focus on the top marginal tax rate (i.e., the optimal θ-type allocation). Denoting by μ and $\xi\mu$ the multipliers on (22) and (21), the first order condition w.r.t. e_θ 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 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, Θ-sector individual, is $-(1 - \alpha(x))\xi$ as in Corollary 3.

4.2 Outer Problem

We next turn to the outer problem to determine ξ. We can decompose the welfare effect of a marginal change in x into a direct effect (of changing $\alpha(x)$ in (21)) and a direct wage shift effect (via changes in w_θ and w_φ induced by the change in x). By the envelope theorem, we can compute the wage shift effect by holding V_i and e_i constant, for $i = \theta, \varphi$. Since this (trivially) implies aggregate effort and utility at any given wage are held constant, there is no indirect wage shift effect. That is, sectors do not overlap at any wage, so the re-allocation across sectors that led to R and C in the continuous model in section 3.4 is absent here. There is also no sectoral shift effect S since sectors are exogenous.

Exactly as in section 3.4, the direct effect and the direct wage shift on the consistency constraint combine to yield $-\xi\mu\alpha(x)Y_\varphi(x)E_\varphi/x$. Similarly, the (direct) effect of the wage shift on the objective is again identically zero, and the effect on the resource constraint (22) is zero by constant returns to scale. Finally, putting a multiplier $\hat{\eta}\mu$ on (20) (and using the same algebraic tricks employed in the proof of Lemma 2) we find that the effect of the wage shift on the incentive constraint is:

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

The effect $\hat{I} \equiv \mu\hat{\eta}U_e\left(c_\varphi, e_\varphi\frac{w_\varphi}{w_\theta}\right)e_\varphi\frac{w_\varphi}{w_\theta}$ is the discrete incentive constraint analog of I from the general Roy model.[7]

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

$$\xi = -\frac{\hat{I}/(\mu\sigma(x))}{\alpha(x)Y_\varphi(x)E_\varphi}, \tag{23}$$

which coincides with the general formula (18) if $S = C = R = 0$ and when we replace I with $\hat{I}$. Moreover, the sign of $\hat{I}$ is opposite of the sign of $\hat{\eta}$. This means that $\xi > 0$ (and hence top marginal taxes are negative) precisely when we want to redistribute from the θ- to the φ-types, so that the downward incentive constraint binds.

This is analogous to our results in the general Roy model, but the addition of S and C will make ξ, 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

[7] To see this, observe that in the limit where $f_x^\theta(w)/f_x(w)$ is 0 up until w_θ, and 1 thereafter, $(d/dw)[f_x^\theta(w)/f_x(w)]$ is a Dirac δ-function and the integral in (15) 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 (5). The only difference from $\hat{I}$ is that it has $e_\varphi w_\varphi/w_\theta$ instead of e_θ and c_φ rather than c_θ (and $\hat{\eta} = \eta/U_c$ is discrete rather than continuous). In the density limit, the θ-type would be imitating an infinitesimally close individual. If we let w_φ be arbitrarily close to w_θ, then we would get e_θ and c_θ, as in the limit case of I.

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 discrete 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. We show in Appendix C that ξ is *also* bounded between Mirrlees (i.e. 0) and Stiglitz (i.e. formula (23)) in a general model with a discrete, two-dimensional type-distribution. This means that the Stiglitz formula (23) is a special case even *within* the class of discrete-type models.

5 A Numerical Example

In this section, 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 income taxes. 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, $Y = E_\theta^\alpha E_\varphi^{1-\alpha}$.[8] 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 (θ_i, φ_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 μ_θ, μ_φ, variances σ_θ^2, σ_φ^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[9]

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

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and cumulative distribution of the standard nor-

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

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

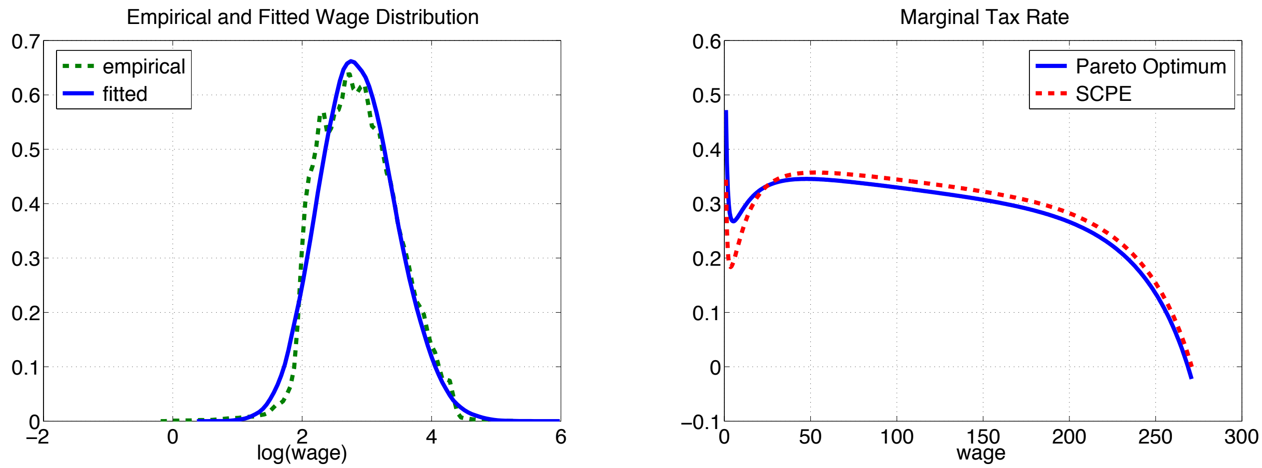

Figure 2: Empirical/estimated wage distributions and optimal/SCPE tax schedules

mal 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 Θ-sector, given by α, can be inferred from this estimated bivariate wage distribution by using the estimated parameters μ_θ, μ_φ, σ_θ^2, σ_φ^2, and $\sigma_{\theta\varphi}$ to draw a (large) sample of (w_θ, w_φ). 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 τ is the marginal tax rate, which we take to be 25% for our calibration), we can compute the sectoral incomes Y^θ and Y^φ and hence $\alpha = Y^\theta/(Y^\theta + Y^\varphi)$.

Finally, with Cobb-Douglas technology, we can, w.l.o.g., take the underlying skills (θ, φ) to coincide with (w_θ, w_φ)—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 θ-skills by $k > 0$ scales E_θ by k, this implies that Y_θ scales by k^{-1}. This re-scaling leaves wages $w = \theta Y_\theta$, efforts, and incomes unchanged. In other 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 Φ-sector, which has an income share $1-\alpha$ of only 5.8%. The left panel of figure 2 compares

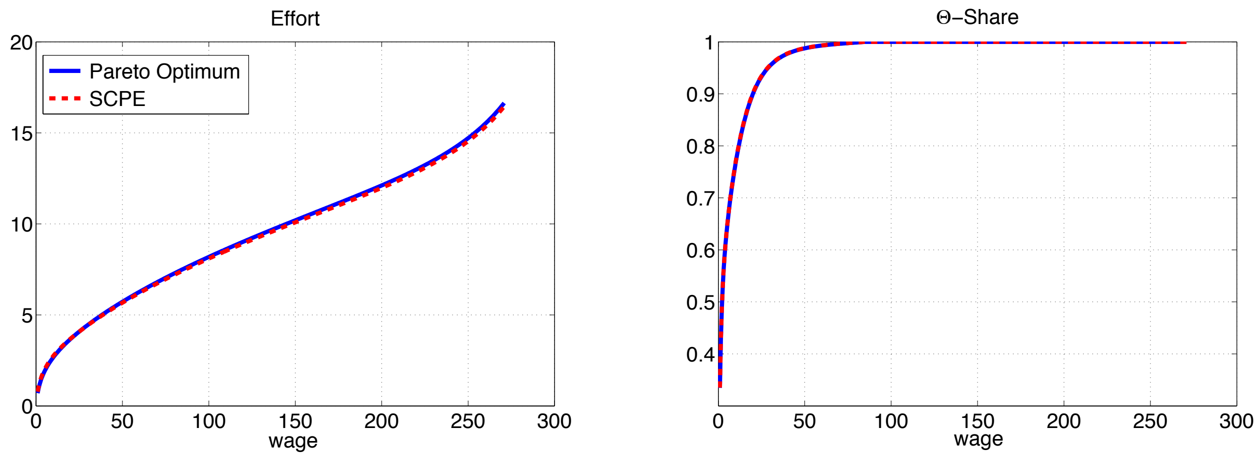

Figure 3: Effort and share of workers in the Θ-sector for Pareto optimum/SCPE

the estimated to the empirical wage distribution; it shows a reasonably good fit.

The Φ-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 our tax computations.

To compute the Pareto optimal and SCPE taxes for this economy, 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 w both for the Pareto optimum and the SCPE resulting from our parametrization.[10] 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 Φ (recall that, by Corollary 3, the top marginal tax rate is given

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

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

We consider two extensions of the baseline model here: an unbounded skill distribution, and allowing individuals to have different costs or tastes for effort across the two sectors.

6.1 Unbounded Skill Distribution

We focused attention on bounded skill distributions for ease of interpretation and for 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 3.4 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:

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 Θ-sector: $\lim_{w\to\infty} f_x^{\theta}(w)/f_x(w) = 1$
(iii) the Θ-sector skill distribution has a Pareto tail with parameter κ: $\lim_{w\to\infty} \dfrac{1-F_x(w)}{wf_x(w)} = \kappa$
(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 $\xi > 0$ *and the asymptotic marginal tax rate is*

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

Proof. In Appendix D. □

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 explore the potential quantitative importance of this downscaling, we have to replace the thinner-tailed lognormal distributions from the calibration in section

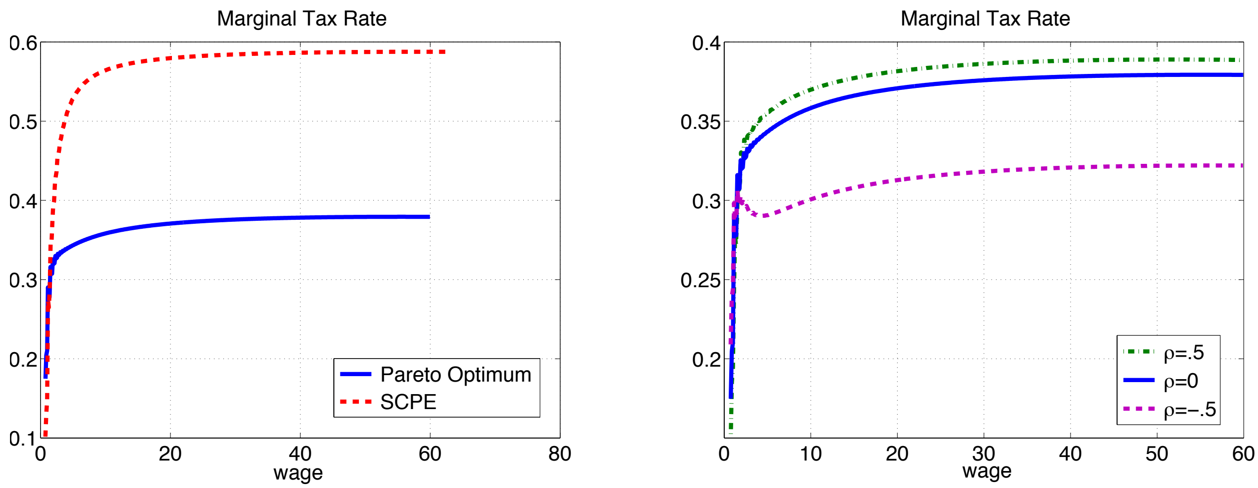

Figure 4: Pareto optimal/SCPE tax rates, and Pareto optimal tax rates for varying σ

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 φ-dimension compared to θ, and Φ 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 Θ 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 far to a CES production function $Y(E_\theta, E_\varphi) = \left[\alpha E_\theta^\rho + (1-\alpha)E_\varphi^\rho\right]^{1/\rho}$, where 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 Θ- or Φ- sector afforded them a higher wage. It is straightforward to extend the model to applications in which individuals have different costs of working in the two sectors.

Let the types be described by a triple $t = (\theta, \varphi, \beta)$, where β parameterizes the cost of Θ-sector effort relative to Φ-sector effort and, as above, θ and φ measure the Θ- and Φ-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 Θ-sector. (With CRRA, a transformed β is interpretable as a *proportional* consumption cost.)

Within the Θ-sector, there will generally be two dimensions of heterogeneity—θ and β. 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 $\tilde{\theta}_1 = \tilde{\theta}_2$ 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

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

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

By interpreting w as an *effective* wage, $\max\{\tilde{\theta} Y_\theta(x), \varphi Y_\varphi(x)\}$, the Pareto optimal and SCPE tax rates are characterized exactly as in Section 3. In particular, marginal taxes are given by Proposition 1, with ξ as in equation (18), 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 can be quantitatively significant. More detailed empirical calibrations would be 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.

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.

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., 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.

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*.

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.

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

SCHEUER, F. (2011): "Entrepreneurial Taxation and Occupational Choice," *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

We prove the result in the following four steps:

Step 1. It is an immediate consequence of the incentive constraints that the utility of a type sent to a given sector can only depend on his wage in that sector. Formally, consider two types (θ_0, φ_0) and (θ_1, φ_1) with

$$S(\theta_0, \varphi_0) = S(\theta_1, \varphi_1) = \Theta \text{ and } Y_\theta(x)\theta_0 = Y_\theta(x)\theta_1 = w.$$

Whenever

$$U\left(c(w/Y_\theta(x), \varphi_0), \frac{y(w/Y_\theta(x), \varphi_0)}{w}\right) \neq U\left(c(w/Y_\theta(x), \varphi_1), \frac{y(w/Y_\theta(x), \varphi_1)}{w}\right),$$

either type $(w/Y_\theta(x), \varphi_0)$'s or type $(w/Y_\theta(x), \varphi_1)$'s incentive constraint is violated. An analogous argument applies to types sent to the Φ-sector. Since w is the wage of the agent in the sector that he is sent to, we can thus write utilities as $V_\theta(w)$ for all types for which $S(\theta, \varphi) = \Theta$ and $V_\varphi(w)$ for all types with $S(\theta, \varphi) = \Phi$.

Step 2. We now show that the consumption and income allocated to a type who is sent to a given sector can also only depend on his skill in that sector. To see this, consider again two types (θ_0, φ_0) and (θ_1, φ_1) with

$$S(\theta_0, \varphi_0) = S(\theta_1, \varphi_1) = \Theta \text{ and } \theta_0 = \theta_1 = w/Y_\theta(x).$$

Consider the expression

$$\begin{aligned} H(w, w') \quad = \quad & \left[U\left(c(w'/Y_\theta(x), \varphi_0), \frac{y(w'/Y_\theta(x), \varphi_0)}{w'}\right) - U\left(c(w/Y_\theta(x), \varphi_0), \frac{y(w/Y_\theta(x), \varphi_0)}{w'}\right)\right] \\ & - \left[U\left(c(w'/Y_\theta(x), \varphi_1), \frac{y(w'/Y_\theta(x), \varphi_1)}{w'}\right) - U\left(c(w/Y_\theta(x), \varphi_1), \frac{y(w/Y_\theta(x), \varphi_1)}{w'}\right)\right]. \end{aligned}$$

Mechanically, we have

$$\left.\frac{\partial H(w, w')}{\partial w'}\right|_{w'=w} = 0,$$

since the partial derivative of each bracketed term, evaluated at $w' = w$, is individually zero. On the other hand, we showed in 1. that

$$U\left(c(w'/Y_\theta(x),\varphi_0),\frac{y(w'/Y_\theta(x),\varphi_0)}{w'}\right) = U\left(c(w'/Y_\theta(x),\varphi_1),\frac{y(w'/Y_\theta(x),\varphi_1)}{w'}\right) = V_\theta(w'),$$

so that $H(w,w')$ reduces to

$$H(w,w') = U\left(c(w/Y_\theta(x),\varphi_1),\frac{y(w/Y_\theta(x),\varphi_1)}{w'}\right) - U\left(c(w/Y_\theta(x),\varphi_0),\frac{y(w/Y_\theta(x),\varphi_0)}{w'}\right).$$

Single-crossing implies that

$$U_w\left(c(w/Y_\theta(x),\varphi_1),\frac{y(w/Y_\theta(x),\varphi_1)}{w}\right) \gtrless U_w\left(c(w/Y_\theta(x),\varphi_0),\frac{y(w/Y_\theta(x),\varphi_0)}{w}\right)$$

whenever

$$U\left(c(w/Y_\theta(x),\varphi_1),\frac{y(w/Y_\theta(x),\varphi_1)}{w}\right) = U\left(c(w/Y_\theta(x),\varphi_0),\frac{y(w/Y_\theta(x),\varphi_0)}{w}\right)$$

and $c(w/Y_\theta(x),\varphi_1) \gtrless c(w/Y_\theta(x),\varphi_0)$.[11] Hence, it must be that

$$c(w/Y_\theta(x),\varphi_1) = c(w/Y_\theta(x),\varphi_0) \text{ and } y(w/Y_\theta(x),\varphi_1) = y(w/Y_\theta(x),\varphi_0).$$

The same argument applies to types sent to the Φ-sector. We can thus write allocations as $c_\theta(w), y_\theta(w)$ for all types with $S(\theta,\varphi) = \Theta$ and $c_\varphi(w), y_\varphi(w)$ for all types with $S(\theta,\varphi) = \Phi$.

Step 3. It is also straightforward to see that two types who earn the same wage in different sectors must get the same utility. Consider two types (θ_0,φ_0) and (θ_1,φ_1) with

$$S(\theta_0,\varphi_0) = \Theta,\ S(\theta_1,\varphi_1) = \Phi \text{ and } Y_\theta(x)\theta_0 = Y_\varphi(x)\varphi_1 = w.$$

Next, assume w.l.o.g.

$$U\left(c_\theta(w),\frac{y_\theta(w)}{w}\right) > U\left(c_\varphi(w),\frac{y_\varphi(w)}{w}\right).$$

Then the $(\theta_1, w/Y_\varphi(x))$-type could imitate the $(w/Y_\theta(x),\varphi_0)$-type by producing income $y_\theta(w)$ but doing so in the Φ-sector, i.e. using his skill $w/Y_\varphi(x)$. His utility from this deviation would be exactly the LHS of the above inequality, contradicting incentive compatibility. This shows

$$V_\theta(w) = V_\varphi(w) \equiv V(w)\ \forall w.$$

Step 4. We finally show that the incentive constraints also imply that allocations have to be such that

$$c_\theta(w) = c_\varphi(w) \equiv c(w) \text{ and } y_\theta(w) = y_\varphi(w) \equiv y(w)\ \forall w,$$

[11] Recall that single crossing implies that $-U_e/(wU_c)$ is decreasing in w, or equivalently that U_y/U_c is increasing in w and hence $U_{yw}U_c - U_yU_{cw} > 0$. Therefore, the change in U_w from a marginal increase in c and y along w's indifference curve (i.e. such that $dc = -(U_y/U_c)dy$) is $dU_w = U_{wy} - (U_y/U_c)U_{wc} > 0$.

i.e. two types who earn the same wage in different sectors must get the same consumption and income. For this purpose, consider again the expression

$$\begin{aligned} H(w,w') &= \left[U\left(c_\theta(w'),\frac{y_\theta(w')}{w'}\right)-U\left(c_\theta(w),\frac{y_\theta(w)}{w'}\right)\right] \\ &\quad -\left[U\left(c_\varphi(w'),\frac{y_\varphi(w')}{w'}\right)-U\left(c_\varphi(w),\frac{y_\varphi(w)}{w'}\right)\right]. \end{aligned}$$

Now consider again $\partial H/\partial w'|_{w'=w}$. On the one hand, this is mechanically zero since the partial derivative of each bracketed term is individually zero. On the other hand,

$$U\left(c_\theta(w'),\frac{y_\theta(w')}{w'}\right)=U\left(c_\varphi(w'),\frac{y_\varphi(w')}{w'}\right)$$

by step 3., so

$$H(w,w')=U\left(c_\varphi(w),\frac{y_\varphi(w)}{w'}\right)-U\left(c_\theta(w),\frac{y_\theta(w)}{w'}\right),$$

and, by single crossing

$$U_w\left(c_\varphi(w),\frac{y_\varphi(w)}{w}\right)\gtreqless U_w\left(c_\theta(w),\frac{y_\theta(w)}{w}\right)$$

whenever $U(c_\varphi(w),y_\varphi(w)/w)=U(c_\theta(w),y_\theta(w)/w)$ and $c_\varphi(w)\gtreqless c_\theta(w)$. Hence, it must be that $c_\varphi(w)=c_\theta(w)$ and $y_\varphi(w)=y_\theta(w)$ for all w.

These four steps prove that any incentive compatible allocation can be implemented by offering a schedule of $c(w),y(w)$ bundles with $w\equiv\max\{Y_\theta(x)\theta,Y_\varphi(x)\varphi\}$, which is equivalent to a single non-linear tax schedule.

B Proofs for Section 3

B.1 Proof of Proposition 1

Putting multipliers μ on (6), $\xi\lambda$ on (3) and $\hat\eta(w)\mu$ on (5), the Lagrangian corresponding to (4)-(6) is, after integrating by parts (5),

$$\begin{aligned} \mathcal{L} &= \int_{\underline{w}_x}^{\overline{w}_x}V(w)g_x(w)dw-\int_{\underline{w}_x}^{\overline{w}_x}V(w)\hat\eta'(w)\mu dw+\int_{\underline{w}_x}^{\overline{w}_x}U_e(c(V(w),e(w)),e(w))\tfrac{e(w)}{w}\hat\eta(w)\mu dw \\ &\quad +\xi\mu(1-\alpha(x))\int_{\underline{w}_x}^{\overline{w}_x}we(w)f_x^\theta(w)dw-\xi\mu\alpha(x)\int_{\underline{w}_x}^{\overline{w}_x}we(w)f_x^\varphi(w)dw \\ &\quad +\mu\int_{\underline{w}_x}^{\overline{w}_x}we(w)f_x(w)dw-\mu\int_{\underline{w}_x}^{\overline{w}_x}c(V(w),e(w))f_x(w)dw. \end{aligned} \tag{24}$$

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

$$\hat\eta'(w)\mu=g_x(w)-\mu f_x(w)\frac{1}{U_c(w)}+\hat\eta(w)\mu\frac{U_{ec}(w)}{U_c(w)}\frac{e(w)}{w}. \tag{25}$$

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

$$\eta'(w)=g_x(w)\frac{U_c(w)}{\mu}-f_x(w)+\eta(w)\frac{U_{cc}(w)c'(w)+U_{ce}(w)e'(w)+U_{ce}(w)e(w)/w}{U_c(w)}. \tag{26}$$

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 (26) 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 (26) and rearranging yields

$$-\frac{\partial MRS(w)}{\partial c}e(w)\frac{y'(w)}{y(w)}\eta(w) = f_E(w) - \psi_E(w)\frac{U_c(w)}{\mu} + \eta'(w). \tag{27}$$

Integrating this ODE gives

$$\begin{aligned}\eta(w) &= \int_w^{\overline{w}_x} \left(f_x(w) - \psi_x(z)\frac{U_c(z)}{\mu}\right) \exp\left(\int_w^z \frac{\partial MRS(s)}{\partial c}e(s)\frac{y'(s)}{y(s)}ds\right) dz \\ &= \int_w^{\overline{w}_x} \left(1 - \frac{\psi_x(z)}{f_x(z)}\frac{U_c(z)}{\mu}\right) \exp\left(\int_w^z \left(1 - \frac{\varepsilon^u(s)}{\varepsilon^c(s)}\right)\frac{dy(s)}{y(s)}\right) f_x(z)dz,\end{aligned} \tag{28}$$

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

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

which after some algebra can be rewritten as

$$wf_x(w)\left(1 - \frac{MRS(w)}{w}\right) + \xi w\left((1-\alpha(x))f_x^\theta(w) - \alpha(x)f_x^\varphi(w)\right) = \eta(w)\left(\frac{\partial MRS(w)}{\partial e}\frac{e}{w} + \frac{MRS(w)}{w}\right). \tag{29}$$

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

$$1 + \xi\frac{(1-\alpha(x))f_x^\theta(w) - \alpha(x)f_x^\varphi(w)}{f_x(w)} = (1 - T'(y(w)))\left[1 + \frac{\eta(w)}{wf_x(w)}\left(1 + \frac{\partial MRS(w)}{\partial e}\frac{e}{MRS(w)}\right)\right]. \tag{30}$$

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 (28) and (30).

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* $\sigma(x) = -\frac{1}{x}\left(\frac{Y'_\theta(x)}{Y_\theta(x)} - \frac{Y'_\varphi(x)}{Y_\varphi(x)}\right)^{-1}$

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 (24) to compute

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

with

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

Integrating by parts the five integral terms yields

$$\begin{aligned} W'(x) = &B_1 + B_2 - \int_{\underline{w}_x}^{\overline{w}_x} V'(w)\frac{dG_x(w)}{dx}dw + \mu\int_{\underline{w}_x}^{\overline{w}_x}\left(\frac{V'(w)}{U_c(w)} + MRS(w)e'(w)\right)\frac{dF_x(w)}{dx}dw - \xi\mu\alpha'(x)Y(E_\theta,E_\varphi) \\ &+ \mu\xi\left(\int_{\underline{w}_x}^{\overline{w}_x}(we'(w)+e(w))\left(\alpha(x)\frac{dF_x^\varphi(w)}{dx} - (1-\alpha(x))\frac{dF_x^\theta(w)}{dx}\right)dw\right) - \mu\int_{\underline{w}_x}^{\overline{w}_x}(we'(w)+e(w))\frac{dF_x(w)}{dx}dw \end{aligned} \tag{31}$$

with

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

By the first order conditions (27) and (29) 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

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

are both equal to zero. Adding them to (31), using (5) and re-arranging yields

$$\begin{aligned}
W'(x) &= B_1 + B_2 - \xi\mu\alpha'(x) Y(E_\theta, E_\varphi) + \int_{\underline{w}_x}^{\overline{w}_x} V'(w) \left(\frac{g_x(w)}{f_x(w)} \frac{dF_x(w)}{dx} - \frac{dG_x(w)}{dx} \right) dw \\
&- \mu \int_{\underline{w}_x}^{\overline{w}_x} e(w) \frac{dF_x(w)}{dx} dw - \mu \int_{\underline{w}_x}^{\overline{w}_x} \left(\frac{\eta(w)}{w} \frac{d\left[MRS(w)e(w)\right]}{dw} + \eta'(w) \frac{V'(w)}{U_c(w)} \right) \frac{1}{f_x(w)} \frac{dF_x(w)}{dx} dw \\
&+ \xi\mu \int_{\underline{w}_x}^{\overline{w}_x} \left((e(w) + we'(w)) \left(\alpha(x) \frac{dF_x^\varphi(w)}{dx} - (1-\alpha(x)) \frac{dF_x^\theta(w)}{dx} \right) + we'(w) \left((1-\alpha(x)) \frac{f_x^\theta(w)}{f_x(w)} - \alpha(x) \frac{f_x^\varphi(w)}{f_x(w)} \right) \frac{dF_x(w)}{dx} \right) dw.
\end{aligned} \tag{32}$$

From Lemma 4,

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

The first integral in (32) is therefore

$$\int_{\underline{w}_x}^{\overline{w}_x} V'(w) \left(\frac{g_x(w)}{f_x(w)} \frac{dF_x(w)}{dx} - \frac{dG_x(w)}{dx} \right) dw = \lambda(x) R \tag{33}$$

Combining the terms with $e(w)$ on the second and third line of (32) and using lemma 4 and the identity $Y_\theta'(x) E_\theta + Y_\varphi'(x) E_\varphi \equiv 0$ gives:

$$\begin{aligned}
&-\mu \int_{\underline{w}_x}^{\overline{w}_x} e(w) \frac{dF_x(w)}{dx} dw + \xi\mu \int_{\underline{w}_x}^{\overline{w}_x} e(w) \left(\alpha(x) \frac{dF_x^\varphi(w)}{dx} - (1-\alpha(x)) \frac{dF_x^\theta(w)}{dx} \right) dw \\
&= \mu \left[\frac{Y_\theta'(x)}{Y_\theta(x)} \int_{\underline{w}_x}^{\overline{w}_x} we(w) f_x^\theta(w) dw + \frac{Y_\varphi'(x)}{Y_\varphi(x)} \int_{\underline{w}_x}^{\overline{w}_x} we(w) f_x^\varphi(w) dw \right] \\
&\quad + \xi\mu \int_{\underline{w}_x}^{\overline{w}_x} we(w) \left((1-\alpha(x)) \frac{Y_\theta'(x)}{Y_\theta(x)} f_x^\theta(w) - \alpha(x) \frac{Y_\varphi'(x)}{Y_\varphi(x)} f_x^\varphi(w) \right) dw - \xi\mu \int_{\underline{w}_x}^{\overline{w}_x} e(w) \Omega_x(w) dw \\
&= 0 + \xi\mu Y_\theta' E_\theta - \xi\mu \int_{\underline{w}_x}^{\overline{w}_x} e(w) \Omega_x(w) dw.
\end{aligned} \tag{34}$$

The terms with $we'(w)$ in the last line (32) can be written, using Lemma 4 again, as

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

$$=\xi\mu\int_{\underline{w}_x}^{\overline{w}_x} w^2 e'(w)\left[(1-\alpha(x))\frac{Y'_\theta(x)}{Y_\theta(x)}f_x^\theta(w)-\alpha(x)\frac{Y'_\varphi(x)}{Y_\varphi(x)}f_x^\varphi(x)-\left((1-\alpha(x))\frac{f_x^\theta(w)}{f_x(w)}-\alpha(x)\frac{f_x^\varphi(w)}{f_x(w)}\right)\left(\frac{Y'_\theta(x)}{Y_\theta(x)}f_x^\theta(w)+\frac{Y'_\varphi(x)}{Y_\varphi(x)}f_x^\varphi(x)\right)\right]dw$$

$$-\xi\mu\int_{\underline{w}_x}^{\overline{w}_x} we'(w)\Omega_x(w)dw$$

$$=\lambda(x)\xi\mu C-\xi\mu\int_{\underline{w}_x}^{\overline{w}_x} we'(w)\Omega_x(w)dw, \tag{35}$$

where the first term in the last step follows after some tedious algebra. Combining the terms with $\Omega_x(w)$ from (34) and (35) gives $-\xi\mu\int_{\underline{w}_x}^{\overline{w}_x}(we(w))'\Omega_x(w)dw$, which can be integrated by parts to yield:

$$-B_3+\xi\mu\frac{\lambda(x)}{Y_\theta(x)Y_\varphi(x)}\int_{\underline{w}_x}^{\overline{w}_x} w^2 e(w)f\left(\frac{w}{Y_\theta(x)},\frac{w}{Y_\varphi(x)}\right)dw=\xi\mu\lambda(x)S, \tag{36}$$

with $B_3 = \xi\mu\overline{w}_x e(\overline{w}_x)\Omega_x(\overline{w}_x)$ since $\Omega_x(\underline{w}_x) = 0$. Finally, use the incentive constraint (5), rewritten as $V'(w)/U_c(w) = MRS(w)e(w)$, to write the last line of (32) as

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

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,

$$\mu\int_{\underline{w}_x}^{\overline{w}_x}\eta(w)w\frac{V'(w)}{U_c(w)}\frac{d}{dw}\left(-\left(\frac{Y'_\theta(x)}{Y_\theta(x)}\frac{f_x^\theta(w)}{f_x(w)}-\frac{Y'_\varphi(x)}{Y_\varphi(x)}\frac{f_x^\varphi(w)}{f_x(w)}\right)\right)dw=\lambda(x)I \tag{37}$$

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

$$\frac{d\tilde{F}(\overline{w}_x,x)}{dx}=\frac{\partial\tilde{F}(\overline{w}_x,x)}{\partial x}+\frac{\partial\tilde{F}(\overline{w}_x,x)}{\partial w}\frac{d\overline{w}_x}{dx}=\frac{dF_x(\overline{w}_x)}{dx}+f_x(\overline{w}_x)\frac{d\overline{w}_x}{dx}=0. \tag{38}$$

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 (33), (34), (35), (36) and (37) in (32) yields

$$W'(x)=\lambda(x)\left(I+R+\xi\mu\left[\alpha(x)Y_\varphi(x)E_\varphi\sigma(x)+S+C\right]\right), \tag{39}$$

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 A General Discrete-Type Model

Stiglitz's (1982) results can be viewed as a special case of a discrete version of a Roy model, with a unit measure of *two* discrete types $(\theta,0)$ and $(0,\varphi)$, so that sectoral choices are trivial. We generalize the discrete version here by allowing any finite number of discrete types (θ_j,φ_j), $j=(1,...,J)$ with masses f_j; we assume that the types are generic, in the sense that (i) $i\neq j\Rightarrow\theta_i\neq\theta_j$ and $\varphi_i\neq\varphi_j$ and (ii) $i\neq j\Rightarrow\varphi_i/\theta_i\neq\varphi_j/\theta_j$. For each type j, let x_j be the unique solution to $\theta_j Y_\theta(x_j)=\varphi_j Y_\varphi(x_j)$; since types are generic, $x_j=x_i\Leftrightarrow i=j$.

As in the continuous model, type j's wage $w_j = \max\{\theta_j Y_\theta(x), \varphi_j Y_\varphi(x)\}$ is fully determined by x. For any $x \neq x_j$, j's sectoral choice is fully determined. If $x = x_j$, on the other hand, j types are indifferent between the two sectors, and individual optimization is consistent with any fraction $\nu \in [0,1]$ of j-types choosing to work in the Φ-sector. We can therefore view the "outer" problem as a maximization over x and ν of $W(x;\nu)$ (where the ν-dependence is trivial unless $x = x_j$ for some j).

There are three possibilities for the solution (x^*, ν^*) to the outer problem, which we describe in turn below. In all three cases, we assume (purely for expositional ease) that the optimal x^* is such that $w_j \neq w_k$ for any distinct j and k. For notational ease, we define ν_j via $\nu_j = 0$ if $\theta_j Y_\theta(x) > \varphi_j Y_\varphi(x)$, $\nu_j = 1$ if $\theta_j Y_\theta(x) < \varphi_j Y_\varphi(x)$, and $\nu_j = \nu$ if $\theta_j Y_\theta(x) = \varphi_j Y_\varphi(x)$ for any given (x, ν). Then we can write the consistency condition (with multiplier $\xi\mu$) as

$$(1-\alpha(x))\sum_{j=1}^{J} f_j w_j e_j (1-\nu_j) - \alpha(x)\sum_{j=1}^{J} f_j w_j e_j \nu_j = 0.$$

The Stiglitz case: $x^* \neq x_j$ for all j. Since $x^* \neq x_j$, there is no sectoral shift (at the margin). By assumption (true for a generic x), the wages in the two-sectors are non-overlapping. This case is therefore exactly like the two-type Stiglitz case but with additional incentive constraints. In particular, the same logic leads to the first order condition

$$W'(x;\nu) = -\frac{1}{x}\left[\xi\mu\alpha(x)Y_\varphi(x)E_\varphi(x) + \frac{1}{\sigma(x)}\tilde{I}\right] = 0,$$

where $\tilde{I}$ captures the effect of the wage shift induced by x on the incentive constraints. We thus get exactly the same expression for ξ as in the standard Stiglitz framework, but with the more general incentive effect $\tilde{I}$ replacing $\hat{I}$. For example, if we assume only downward-binding adjacent incentive constraints,[12]

$$\tilde{I} = -\mu\sum_{k=2}^{J} d_k \hat{\eta}_k U_e\left(c_{k-1}, e_{k-1}\frac{w_{k-1}}{w_k}\right) e_{k-1}\frac{w_{k-1}}{w_k},$$

where $d_k = 1$ if $S_k = \Theta$ and $S_{k-1} = \Phi$, $d_k = -1$ if $S_k = \Phi$ and $S_{k-1} = \Theta$, and $d_k = 0$ if $S_k = S_{k-1}$.

The Mirrlees case: $x^* = x_k$ for some k and $\nu^* \in (0,1)$. A change in ν at $x^* = x_k$ affects the sectoral composition at wage $\theta_k Y_\theta(x)$ without affecting the aggregate wage distribution. The only effect on the inner problem is therefore via the consistency condition (40). If $\xi > 0$, then lowering ν eases this constraint and is welfare improving. Similarly, if $\xi < 0$, raising ν is welfare improving. Optimality therefore requires $\xi = 0$ in this $\nu^* \in (0,1)$ case. The inner problem is therefore a standard Mirrlees problem; the Roy features can effectively be ignored (locally) in the formulas for optimal taxes from the inner problem.

For some simple intuition for why these features can be ignored in this case, consider the effects of a small change in taxes. The direct effect of such a change will be to change efforts and hence x, for example increasing it by a small amount. This would induce migration of the k-types towards the Φ-sector, leading x to decrease. Since the original tax change was small, this decrease in x would overwhelm the original increase in x if the *entire* discrete mass $(1-\nu)f_k$ of the k-types in the Θ-sector were to migrate (which would in turn lead to migration *back* to the Θ-sector). In equilibrium, net migration by k-types to the Φ-sector must be such that the original x^* is restored at a new, lower ν. The wage distribution will be

[12] A more general treatment of incentive constraints would lead to an intuitively similar but more involved expression for $\tilde{I}$.

unchanged, and the consistency condition will remain satisfied. In short, the endogenous response of k-types to marginal changes in taxes will ensure the consistency condition remains satisfied and the aggregate wage distribution remains unchanged. So, locally, it is as if the wage distribution is exogenous to the tax code and the consistency condition can be ignored.

The intermediate case(s): $x^* = x_j$ **for some** j **and** $\nu^* \in \{0, 1\}$. Consider the $\nu^* = 0$ case first. Then $\xi \geq 0$ (or raising ν would raise W, as in the Mirrlees case). Since lowering x raises the relative wages in the Θ-sector, sectoral choices are constant on $x \in [x^* - \varepsilon, x^*)$, and identical to the choices at x^* for $\nu = 0$. The left-hand derivative $W'_-(x^*; 0) = -\frac{1}{x^*}\left[\xi\mu\alpha(x^*)Y_\varphi(x^*)E_\varphi + \frac{1}{\sigma(x^*)}\tilde{I}\right]$ is therefore well-defined. Since $(x^*, 0)$ is optimal, $W'_-(x^*; 0) \geq 0$. Hence,

$$0 \leq \xi \leq -\frac{\tilde{I}/\sigma(x^*)}{\mu\alpha(x^*)Y_\varphi(x^*)E_\varphi}.$$

This implies that $\tilde{I} \leq 0$ (intuitively: the indirect redistribution caused by an increase in x is desirable), and that ξ is intermediate between the Mirrlees ($\xi = 0$) and Stiglitz ($\xi = -\frac{\tilde{I}/\sigma(x^*)}{\mu\alpha(x^*)Y_\varphi(x^*)E_\varphi}$) cases. It will, in fact, be strictly below the Stiglitz case if $W'_-(x^*; 0) > 0$. The $\nu^* = 1$ case is analogous: it yields a $-\frac{\tilde{I}/\sigma(x^*)}{\mu\alpha(x^*)Y_\varphi(x^*)E_\varphi} \leq \xi \leq 0$ for a situation in which a *decrease* in x is redistributionally beneficial.

Despite the fact that it involves $\nu^* = 0$ or $\nu^* = 1$, there is reason to believe that this intermediate case will actually be typical. If $\xi > 0$ ($\xi < 0$ is symmetric), then $W(x; \nu)$ jumps down when it crosses a mass point x_j, so that $(x_j, \nu = 0)$ will be a *local maximum* whenever W was increasing to the left of x_j.

D Proof of Proposition 3

$$\lim_{w\to\infty} 1 - T'(y(w)) = \lim_{w\to\infty} \frac{1 + \xi\left(1 - \alpha(x) - \dfrac{f_x^\varphi(w)}{f_x(w)}\right)}{1 + \dfrac{\eta(w)}{wf_x(w)}\dfrac{1+\varepsilon^u(w)}{\varepsilon^c(w)}}$$

$$= \frac{1 + \xi\left(1 - \alpha(x) - \lim\limits_{w\to\infty}\left(\dfrac{f_x^\varphi(w)}{f_x(w)}\right)\right)}{1 + \left(1 + \dfrac{1}{\varepsilon}\right)\lim\limits_{w\to\infty}\left(\dfrac{1 - F_x(w)}{wf_x(w)}\right)\left(\dfrac{\Psi(F_x(w)) - F_x(w)}{1 - F_x(w)}\right)} = \frac{1 + \xi(1 - \alpha(x))}{1 + \left(1 + \dfrac{1}{\varepsilon}\right)\kappa}.$$

The first equality is from (9). 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.