Pareto Efficient Income Taxation*

Iván Werning, MIT and NBER

April, 2007

Abstract

I study the set of Pareto efficient tax schedules in Mirrlees' optimal tax model and provide a simple test for the efficiency of a given tax schedule. The efficiency condition generalizes the the well-known zero-tax-at-the-top result: taxes should be low in regions where the density of income falls rapidly. Both the set of efficient and inefficient tax schedules is large. I use the framework to explore the optimality of a flat tax, to bound the top tax rate of a nonlinear schedule, and to evaluate the efficiency of a tax system that does not condition on observable traits.

1. Introduction

This paper derives the restrictions on the income-tax schedule implied by efficiency. I work with Mirrlees's (1971) original framework, but, instead of maximizing a Utilitarian objective or any other social welfare function, I derive the conditions required by constrained Pareto efficiency (e.g., (★) and (★★) below)—I adopt the positive aspect of Mirrlees's model, but not the normative one. In other words, I characterize tax schedules that cannot be reformed without harming some individuals. Conversely, I identify tax schedules that can be reformed to everyone's benefit.

The main results of the analysis are as follows:

1. For any increasing income-tax schedule, there exists a set of skill distributions for which the allocation is Pareto efficient and another for which it is not. Thus, theory alone, without empirical knowledge on the distribution of skills, provides no guidance: *anything goes*.

2. Any income-tax schedule in place induces a distribution of income which identifies the underlying skill distribution (Saez, 2001). One can then characterize the entire set of Pareto efficient tax schedules consistent with it. Both the set of efficient and inefficient tax schedules is "large". In particular, this shows that Pareto efficiency is not a vacuous requirement.

3. I derive a simple test for the efficiency of any current tax schedule. A graphical version of this test compares the actual distribution of income against an hypothetical one, defined as that which would have made the tax schedule optimal for a Rawlsian (see Figure 2).

4. The restrictions I derive on the shape of the tax schedule can be seen as a quantitative generalization of the well-known qualitative "zero-tax-at-the-top" result.

5. Simple formulas are obtained for the following: (i) to test the Pareto efficiency of a flat tax; (ii) to bound the top tax rate of a nonlinear schedule; and (iii) to evaluate the efficiency of a tax system that does not condition on observable traits.

Looking forward, it might be interesting to see how far one can go verifying whether the conditions for efficiency hold in the United States and other countries. I provide a tentative analysis along these lines for the United States. The rest of the Introduction discusses related literature.

Diamond (1998) and Saez (2001) reinvigorated the interest in Mirrlees' model of optimal taxation. Both papers explored the importance of the underlying skill distribution, as well as other parameters. I adopt the perspective of Saez (2001) that the distribution of income should be used to identify the underlying distribution of skills through the lens of the model. I express my tests for efficiency directly in terms of the distribution of income.

Applied work in this area typically adopts particular welfare functions and explores the solutions to the optimal tax schedule obtained numerically. Since the choice of a welfare function

cannot be determined empirically, sometimes several welfare functions are employed and the effects of this choice are highlighted (Tuomala, 1990; Diamond, 1998; Saez, 2001). In contrast, my paper inverts the direction of these exercises and asks whether an existing or hypothetical tax schedule is consistent with Pareto efficiency.[1] More generally, I emphasize studying the *set of efficient tax schedules*, as opposed to *the optimal tax schedule* for a particular welfare criterion. I show that such an approach can yield useful new insights and remains tractable enough for practical quantitative explorations, of the kind pursued by Saez (2001) and Gruber and Saez (2002).

A more theoretical line of research advocates and explores a Pareto efficient approach to optimal taxation (see Stiglitz, 1987, and the references therein). In particular, Brito, Hamilton, Slutsky, and Stiglitz (1990) study Pareto efficient taxation and derive some qualitative implications, such as a "zero-tax-at-the-top" results. However, these results do not greatly restrict the income tax schedule. Less explored are the implications of Pareto efficiency the overall shape of the income tax schedule. In a recent paper, Chone and Laroque (2006) study optimal taxation with quasi-linear preferences. They focus on the optimality of negative marginal tax rates with discrete labor supply choices, but their approach and methodology is close to this paper in that Pareto efficiency is used as the criterion for optimality.[2]

In addition to these formal contributions, economists have been ardent advocates of various tax reforms. In particular, flat taxes have been proposed on efficiency, simplicity and even fairness considerations (e.g. Friedman, 1962; Hall and Rabushka, 1995). Yet, these proposals are seemingly at odds with the output from Mirrlees' (1971) model, which consistently produces nonlinear tax schedules. In this paper, I show that this is misleading: flat taxes, as well as more progressive tax systems, may well be Pareto efficient within Mirrlees' economy.

2. Preliminaries

The model economy is populated by a continuum of workers. For simplicity, I assume additively separable preferences of the following form:[3]

$$U(c, Y, \theta) = u(c) - \theta h(Y),$$

where θ indexes the heterogeneous disutility from producing output Y.[4] It is worth remarking that, given the focus on Pareto efficiency, no interpersonal comparisons of utility will be needed. Thus,

---

[1] Saez (2001, pg.) suggests, but does not pursue, the possibilities of such an approach.

[2] Chone and Laroque (2005) study the Rawlsian optimal tax schedule and compare it to the optimal tax schedule for a Utilitarian planner.

[3] However, some of the notation I employ in what follows is more general, in preparation for a possible extension to the more general case

[4] This specification obtains from a common utility function with heterogeneity in productivity if the disutility of effort is a power function.

the cardinality of preferences is completely irrelevant and only the ordinal features of preferences matter. Let the expenditure function $e(v, Y, \theta)$ represent the inverse of $U(\cdot, Y, \theta)$. Let $F(\theta)$ be the distribution of θ in the population. I assume there are no mass points, so that the distribution can be represented by its density $f(\theta)$.

Now take any tax function $T(Y)$. Workers maximize their utility and obtain

$$v(\theta) \equiv \max_Y U(Y - T(Y), Y, \theta).$$

For a worker of type θ, let $c(\theta)$ and $Y(\theta)$ be the resulting allocation (solving this maximization) for consumption, $Y - T(Y)$, and output, Y, respectively. Note that another expression for consumption is $c(\theta) = e(v(\theta), Y(\theta), \theta)$.

An allocation $(c(\theta), Y(\theta))$ is *resource feasible* if

$$\int \big(Y(\theta) - c(\theta)\big) \, dF(\theta) + e \geq 0$$

where e is the net endowment (non-labor output net of government consumption). The allocation generated by some tax schedule is (constrained) *Pareto efficient* if there is no other tax schedule that induces a resource feasible allocation where nobody is worse off and some workers are strictly better off.

On first pass, I simplify by assuming that $T(Y)$ is differentiable and induces a continuous, differentiable and strictly monotone allocation (i.e., no "bunching"). It will be useful to define the marginal tax rate

$$\tau(\theta) \equiv T'(Y(\theta)) = 1 + \frac{U_Y(c(\theta), Y(\theta), \theta)}{U_c(c(\theta), Y(\theta), \theta)} = 1 - \frac{\theta h'(Y(\theta))}{u'(c(\theta))} = 1 - e_Y(v(\theta), Y(\theta), \theta).$$

3. Conditions for Pareto Efficiency

In this section I introduce the Pareto planning problem and derive the necessary and sufficient conditions for optimality.

3.1. The Planning Problem

An allocation $(c(\theta), Y(\theta))$, that delivers utility $v(\theta)$, is Pareto efficient *if and only if* $(Y(\theta), v(\theta))$ solves the following planning problem:[5]

---

[5]If some feasible allocation does not solve this problem then there exists an alternative allocation where the resource constraint is slack that provides the same or more utility. A Pareto improvement would always be possible: if another allocation provided the same utility but increased net resources, then these resources can be used to construct another allocation that increases utility for some workers and is resource feasible. Conversely, if an allocation solves this problem, then all alternative allocations that provide the same or more utility cannot satisfy the resource constraint,

$$\max_{\tilde{Y},\tilde{v}} \int \left(\tilde{Y}(\theta) - e(\tilde{v}(\theta), \tilde{Y}(\theta), \theta)\right) dF(\theta)$$

subject to,

$$\tilde{v}(\theta) = \tilde{v}(\bar{\theta}) - \int_{\theta}^{\bar{\theta}} U_\theta\big(e(\tilde{v}(z), \tilde{Y}(z), z), \tilde{Y}(z), z\big)dz$$

$$\tilde{Y}(\theta) \text{ nonincreasing}$$

$$\tilde{v}(\theta) \geq v(\theta)$$

where $v(\theta)$ represents the our original utility profile. The objective is to maximize aggregate net resources, output minus consumption. The first constraint is simply the familiar condition that $\dot{v} = U_\theta$, but in integral form. The second constraint imposes that output be monotone decreasing in θ, so that more skilled workers produce more. Together these two constraints ensure incentive compatibility. The last constraint requires that workers are not made worse off.

To verify whether the original allocation solves this planning problem, I evaluate first-order conditions at the original allocation $(Y(\theta), v(\theta))$. Since first-order conditions are necessary, if they are violated it indicates that the original allocation cannot be Pareto efficient. Conversely, because a transformed version of this problem (choosing h instead of Y) is convex, the first-order conditions are sufficient. Hence, if they are verified the original allocation solves the planning problem.

To derive the first-order conditions, define the Lagrangian

$$\mathcal{L} \equiv \int \left(\tilde{Y}(\theta) - e(\tilde{v}(\theta), \tilde{Y}(\theta), \theta)\right) dF(\theta)$$

$$+ \int \left(\tilde{v}(\theta) - \tilde{v}(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} U_\theta\big(e(\tilde{v}(z), \tilde{Y}(z), z), \tilde{Y}(z), z\big)dz\right) d\mu(\theta).$$

Integrating the second term by parts,

$$\mathcal{L} = \int \left(\tilde{Y}(\theta) - e(\tilde{v}(\theta), \tilde{Y}(\theta), \theta)\right) dF(\theta) - \tilde{v}(\bar{\theta})\mu(\bar{\theta}) + \mu(\underline{\theta})\tilde{v}(\underline{\theta})$$

$$+ \int \tilde{v}(\theta)d\mu + \int \mu(\theta)U_\theta\big(e(\tilde{v}(\theta), \tilde{Y}(\theta), \theta), \tilde{Y}(\theta), \theta\big)d\theta.$$

---

since if there were another allocation with higher utility for some workers that yields the same net resources, then one can find another one that increases net resources that still satisfies $\tilde{v}(\theta) \geq v(\theta)$.

3.2. Efficiency Conditions

The first-order condition for $\tilde{Y}(\theta)$ evaluated at $(Y(\theta), v(\theta))$ gives

$$\big(1 - e_Y(v(\theta), Y(\theta), \theta)\big) f(\theta) = -\mu(\theta) \Big(U_{\theta c}\big(e(v(\theta), Y(\theta), \theta), Y(\theta), \theta\big) e_v(v(\theta), Y(\theta), \theta)$$
$$+ U_{\theta Y}\big(e(v(\theta), Y(\theta), \theta), Y(\theta), \theta\big)\Big),$$

implying

$$\mu(\theta) = \tau(\theta) \frac{f(\theta)}{h'(Y(\theta))}. \tag{1}$$

The first-order condition for $v(\bar{\theta})$ gives $\mu(\bar{\theta}) \geq 0$. Likewise, if θ is bounded away from zero, the first-order condition for $v(\underline{\theta})$ gives $\mu(\underline{\theta}) \leq 0$. Thus,

$$\tau(\bar{\theta}) \geq 0 \quad \text{and} \quad \tau(\underline{\theta}) \leq 0. \tag{2}$$

For interior θ, the first-order condition with respect to $\tilde{v}(\theta)$ evaluated at $(Y(\theta), v(\theta))$ gives

$$\dot{\mu}(\theta) \leq e_v\big(v(\theta), Y(\theta), \theta\big) f(\theta). \tag{3}$$

Differentiating equation (1) gives

$$\dot{\mu}(\theta) = \mu(\theta) \left(\frac{\tau'(\theta)}{\tau(\theta)} + \frac{f'(\theta)}{f(\theta)} - \frac{h''(Y(\theta))}{h'(Y(\theta))} Y'(\theta) \right). \tag{4}$$

Substituting equations (1) and (4) into the first-order condition (3) and rearranging gives

$$\tau(\theta) \left(\frac{d \log \tau(\theta)}{d \log \theta} + \frac{d \log f(\theta)}{d \log \theta} - \frac{d \log h'(Y(\theta))}{d \log \theta} \right) \leq 1 - \tau(\theta), \qquad (\bigstar)$$

a restrictions on the shape of the tax function and the allocation it generates.[6] The integral form of this efficiency condition is

$$\frac{\tau(\theta) f(\theta)}{h'(Y(\theta))} + \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\tilde{\theta}))} f(\tilde{\theta}) \, d\tilde{\theta} \quad \text{is nonincreasing}, \qquad (\textstyle\int\bigstar)$$

which is derived formally in Appendix A, requires weaker differentiability requirements on the tax schedule and the resulting allocation.

---

[6]This expression resulted in an effort to find a simple expression, but unnecessarily assumes $\tau(\theta) > 0$ so that $\log \tau(\theta)$ is well defined. This will be rewritten in future versions.

Proposition 1. *Given the utility function $U(c, Y, \theta)$ and a density of skills $f(\theta)$, a differentiable tax function $T(Y)$ inducing an allocation $(c(\theta), Y(\theta))$ is Pareto efficient if and only if condition $(\int \bigstar)$ holds, where $\tau(\theta) = T'(Y(\theta))$.*

Notably, the distribution of skills, through its effect on $d \log f(\theta)/d \log \theta$, is key. Indeed, for any initial tax schedule and allocation one can violates or satisfy the inequality by an appropriate choice of the distribution of skills. The next result follows immediately from this observation.

Proposition 2. *For any tax schedule $T(Y)$ and its resulting allocation there is a set of skill distributions $F(\theta)$ and net endowments e for which the outcome is Pareto efficient and another set of skill distributions $F(\theta)$ and net endowments e for which it is Pareto inefficient.*

Suppose one is given the utility functions u and the disutility of labor function h. Condition $(\bigstar)$ can then be implemented after identifying the distribution of skills and the implied allocation as a function of θ. This can be easily done. First, for any smooth tax schedule $T(Y)$, the workers' maximization problem gives an allocation for output $Y(\theta) \in \arg \max U(Y - T(Y), Y, \theta)$. Second, any worker's choice of output uniquely determines the θ from the first-order condition

$$\theta(Y) = (1 - T'(Y)) \frac{u'(Y - T(Y))}{h'(Y)}.$$

Now suppose one observes the distribution $G(Y)$ of output Y across workers. This implies the relation $F(\theta(Y)) = 1 - G(Y)$. If we assume the tax schedule is sufficiently smooth, so that it induces a distribution of output representable by a density then $f(\theta(Y)) = -g(Y)/\theta'(Y)$. Thus, in this way, one identifies the skill distribution for skills from the distribution of observed output and the tax function (see Saez, 2001).

It is also possible to rewrite condition $(\bigstar)$ in a number of ways. Appendix B shows how it can be expressed directly in terms of the density over output $g(Y)$ and the tax schedule $T(Y)$ as follows:

$$T'(Y) \left(-\frac{d \log(1 - T'(Y))}{d \log Y} + \frac{d \log h'(Y)}{d \log Y} - \frac{d \log g(Y)}{d \log Y} - 1 + \frac{d \log \varepsilon_{\theta, Y}}{d \log Y} \right)$$
$$\leq -2 \frac{d \log(1 - T'(Y))}{d \log Y} + \frac{d \log h'(Y)}{d \log Y} - \frac{d \log u'(Y - T(Y))}{d \log Y} \quad (\bigstar\bigstar)$$

where $\varepsilon_{\theta, Y}$ is the elasticity of $\theta(Y)$ with respect to Y,

$$\varepsilon_{\theta, Y} \equiv \left| \frac{Y \theta'(Y)}{\theta(Y)} \right| = -\frac{d \log(1 - T'(Y))}{d \log Y} - \frac{d \log u'(Y - T(Y))}{d \log Y} + \frac{d \log h'(Y)}{d \log Y}.$$

Although more involved than $(\bigstar)$, this condition can be evaluated directly with information on output and taxes, without the need to transition through the identification of θ and its distribution.

It is useful to think of this result as a test for efficiency of any triplet (U, T, g). The formal proof of this result is in Appendix A.

Proposition 3. *Given the utility function $U(c, Y, \theta)$, a tax function $T(Y)$ inducing a density of output $g(Y)$ is Pareto efficient* if and only if *condition* (★★) *holds.*

A few remarks on tax schedules that are not differentiable are in order. In the model, if the tax schedule $T(Y)$ has concave kinks then some mass of workers will accumulate at these points (convex kinks are irrelevant). Even in such bunching cases, conditions (★) and (★★) continue to be sufficient for optimality; these inequalities are also necessary over intervals where agents are not bunched. Note that the skill density $f(\theta)$ over an interval of workers that are bunched at a kink in the tax schedule cannot be identified by observing their output choices. Then, without independent information on the skill distribution, it follows that conditions (★) and (★★) are also necessary for optimality.

In the real world, however, statutory tax schedules with kinks do not seem to induce workers to accumulate at these points. There are several possible explanations for this. For example, workers may face uncertainty, for any given work effort, in their output. In addition to mechanically avoiding bunching in terms of output, this smooths the rewards to their work effort choices which prevents bunching in terms of work effort. Similar effects may result without uncertainty in a dynamic model where skills are accumulated over the life cycle. Although richer models could incorporate such elements, it seems more sensible within the context of the present static Mirrlees (1971) to interpret the relevant tax schedule as smooth.

3.3. A Simple Reform and Laffer Effects

I now provide a variational result that provides intuition for the form that the efficiency condition (★) (and by association (★★)), and especially for the crucial role played by the distribution of skills. The result states that, if $d \log f(\theta)/d \log \theta$ is high enough, a particularly simple tax reform can create a Pareto improvement.

Proposition 4. *Suppose an increasing and differentiable tax schedule $T(Y)$ induces a feasible allocation such that the inequality*

$$\tau(\theta) \left(\frac{d \log \tau(\theta)}{d \log \theta} + 2 \frac{d \log f(\theta)}{d \log \theta} - \frac{d \log h'(Y(\theta))}{d \log \theta} \right) \le 3(1 - \tau(\theta)) \tag{5}$$

is violated around some interior point $\hat{\theta}$. Then, for some $\varepsilon > 0$ the alternative tax schedule

$$\hat{T}(Y) = \begin{cases} T(Y(\hat{\theta})) - \varepsilon & Y = Y(\hat{\theta}) \\ T(Y) & Y \ne Y(\hat{\theta}) \end{cases}$$
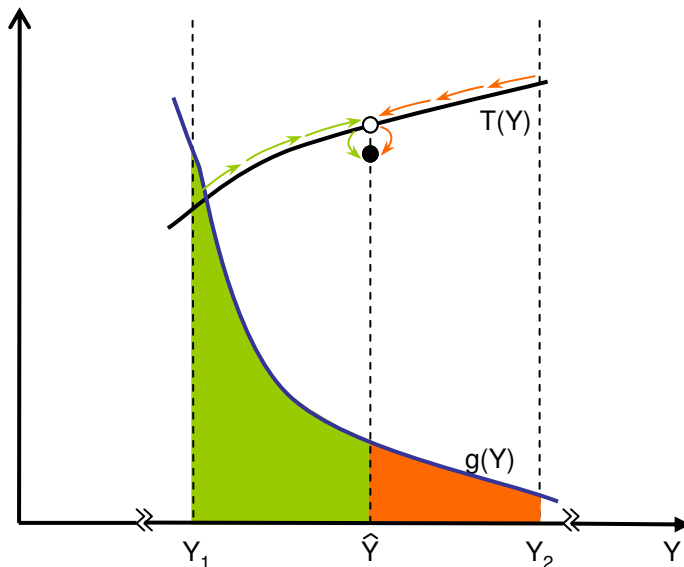
7

Figure 1: A local tax reform that reduces taxes discontinuously at one point $\hat{Y}$. Neighboring worker to the right of $\hat{Y}$ decrease output and pay less taxes. Neighboring worker to the left increase output and may pay more taxes. For a small reform, the net effect on tax receipts depends critically on the rate of growth of the density $g(Y)$ (and, hence, of $f(\theta)$).

induces a feasible allocation that Pareto improves the original one.

The proof is in Appendix C, but the main argument is as follows. That workers are better off facing a lower tax schedule is immediate. What is less obvious is whether tax collections rise or fall, which is equivalent to asking whether the resulting allocation is resource feasible or not. The reform induces some neighboring workers to produce less and others to produce more. Tax receipts are definitely lost from the former group, but can be gained from the latter. This situation is illustrated in Figure 1. The relative fraction of workers moving up and down is then critical. A low enough value of $g'(\theta)/g(\theta)$ or a high enough value of $f'(\theta)/f(\theta)$ makes the relative fraction of workers that increase their output favorable enough, so that tax total revenues rise.

A test for efficiency based on condition ($\star$) is more powerful than one based on inequality (5).[7] That is, the simple tax reform used in Proposition 4 is powerful enough to detect some possible Pareto improvements, but it misses others requiring alternative tax reforms.

The simple variation provided by Proposition 4 illustrates the general point that a Pareto improvement requires a reduction in the tax schedule. Otherwise, if taxes are higher in some region, then workers producing there would be better off with the original schedule.

Proposition 5. *If the tax schedule $T_1(Y)$ induces an allocation that Pareto dominates that implied*

---

[7]If condition ($\star$) holds then it is possible to show, using the relation $1 - \tau(\theta) = \theta h'(Y(\theta))/u'(c(\theta))$, that $\tau(\theta)d \log f(\theta)/d \log \theta \leq 2(1 - \tau(\theta))$. This then implies the inequality (5).

by $T_0(Y)$, then $T_1(Y) \leq T_0(Y)$ for any $Y \in \arg\max_{\tilde{Y}} (\tilde{Y} - T(\tilde{Y}), \tilde{Y}, \theta)$.

If the resource constraint originally held with equality, then the alternative tax schedule must collect the same or more revenue. In this sense, Pareto improvements require a Laffer-like effect: lower taxes that do not lower tax revenue, that "pay for themselves". However, the point is more subtle than a lowering of taxes across the board yielding an increase in revenue or for the revenue increase to be uniform across workers. In particular, when a tax reform targets a particular point on the schedule, lowering it, total revenues may rise even if the revenue collected from some workers falls. Indeed, except possibly at the top, revenues collected from some workers always decrease.

3.4. A Graphical Test with a Rawlsian Connection

I now provide a simple reinterpretation of the test for optimality (★★). If the marginal tax rate $T'(Y)$ is everywhere positive then this condition is equivalent to

$$\frac{d \log g(Y)}{d \log Y} \geq a(Y), \qquad (\bigstar\bigstar')$$

where $a(Y)$ is a function of Y that can be computed given any tax schedule $T(Y)$ and a specification for $u(c)$ and $h(Y)$. Let $\alpha(Y)$ be the unique density that satisfies this condition with equality $d \log \alpha(Y)/d \log Y = a(Y)$:

$$\alpha(Y) = \frac{\exp\left(\int_0^Y a(z)\, dz\right)}{\int_0^\infty \exp\left(\int_0^Y a(z)\, dz\right)}.$$

This density has the following interpretation. If the tax schedule $T(Y)$ generates an income density $\alpha(Y)$ then this schedule is optimal for a Rawlsian social welfare function. That is it solves $\min_\theta v(\theta)$ subject to incentive compatibility and resource feasibility. Equivalently, one can set up the problem as the maximization of net resources, but imposes a constant minimum $v(\theta) = \underline{v}$. I call $\alpha(Y)$ the Rawlsian density.

Since $d \log \alpha(Y)/d \log Y = a(Y)$, condition (★★') is equivalent to

$$\frac{g(Y)}{\alpha(Y)} \quad \textbf{is nondecreasing.} \qquad (\bigstar\bigstar'')$$

This admits a simple graphical interpretation. To satisfy this condition, the density $g(Y)$ should cross the Rawlsian density at most once from below. Indeed, it should cross any multiple of the density $A\alpha(Y)$, for any $A > 0$, from below at most once. In other words, if one plots the Rawlsian density and its family of multiples, then the actual density of output should cross all of these from below. Figure 2 illustrates a density that passes the test.
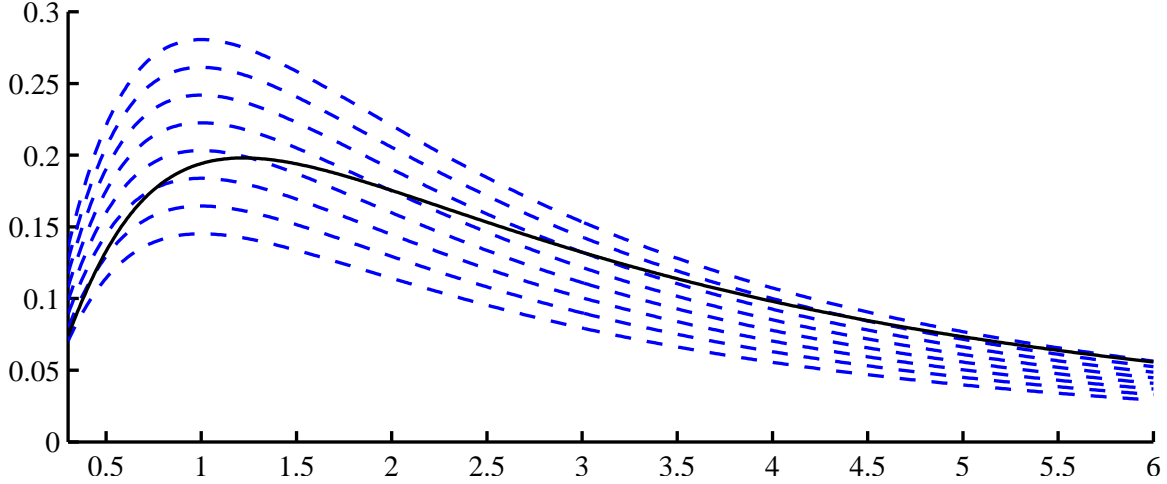
9

Figure 2: An example of the graphical test. The dashed lines show the Rawlsian density and scaled versions of it. The solid line represents a density that passes the test for efficiency.

By implication, to pass the test for efficiency, a distribution of income must first-order stochastically dominate the Rawlsian distribution of income. Thus, if the distribution of income is stochastically dominated by the Rawlsian distribution, then the tax schedule in question cannot be efficient. The Rawlsian distribution is the lowest distribution (in the sense of first order dominance) that passes the efficiency test.

3.5. Quantifying Potential Inefficiencies

Suppose a triplet (U, T, g) does not pass the test for Pareto efficiency ($\bigstar\bigstar$). This is a qualitative conclusion, but what is a natural quantitative measure for the importance of the inefficiency?

The analysis suggests using the difference between the maximized net resources, obtained the planning problem, and the original value of net resources

$$\Delta \equiv \int \left(\tilde{Y}^*(\theta) - \tilde{c}^*(\theta) \right) dF(\theta) - \int \left(Y(\theta) - c(\theta) \right) dF(\theta).$$

The welfare gain measure requires computing the solution to the planning problem, after identifying the density of skills $f(\theta)$ (as discussed in Section 3). This measure is expressed in consumption units and represents the additional resources that could be saved by implementing the planning problem's solution. Of course, such resource savings could then be used, at another step, to increase the utility of some workers still further.

In addition to saving resources, the planning problem's solution makes some workers strictly better off. The measure Δ does not account for these gains. However, this omission is not accidental: while information on these utility improvements may be of complementary interest, any attempt to summarize them into a single number runs counter to the Pareto-efficiency spirit of the

10

exercise. One cannot do so without making interpersonal comparisons, or using some particular social welfare function.[8]

4. Some Explorations

In this section I explore implications of efficiency condition derived in the previous section.

4.1. The Tax Rate at the Top

If θ is not bounded away from zero (i.e., productivity is unbounded above), the support for the distribution of output is potentially unbounded above. What restrictions does Pareto efficiency put on this top tax rate? In this subsection I show that it imposes an upper bound.[9]

Suppose one is asked to verify the efficiency of a tax schedule for which the asymptotic tax rate $\bar{\tau} \equiv \lim_{\theta \to 0} \tau(\theta) = \lim_{Y \to \infty} T'(Y)$ exists and is positive and less than 1. In addition, suppose the tax schedule is such that the limits of all the terms in ($\bigstar\bigstar$) exist. Assume power utility $u(c) = c^{1-\sigma}/(1-\sigma)$ and $h(Y) = \kappa Y^{\eta}$ for $\sigma > 0$, $\eta > 1$ and $\kappa > 0$.

Asymptotically, many of the terms in ($\bigstar\bigstar$) vanish. Since $T'(Y)$ and $\varepsilon_{\theta,Y}$ converge to a constant and the limits are assumed to exist,

$$\lim_{Y \to \infty} \frac{d \log(1 - T'(Y))}{d \log Y} = 0 \quad \text{and} \quad \frac{d \log \varepsilon_{\theta,Y}}{d \log Y} = 0.$$

Furthermore, for high income, consumption becomes proportional to income, so that

$$\lim_{Y \to \infty} \frac{d \log u'(Y - T(Y))}{d \log Y} = -\sigma \quad \text{and} \quad \lim_{Y \to \infty} \frac{d \log h'(Y)}{d \log Y} = \eta - 1.$$

Substituting these expressions into condition ($\bigstar\bigstar$) gives the following upper bound on the top tax rate:

$$\bar{\tau} \leq \frac{\sigma + \eta - 1}{\varphi + \eta - 2}.$$

where $\varphi \equiv -\lim_{Y \to \infty} \frac{d \log g(Y)}{d \log Y}$. The value of $\varphi - 1$, which must be positive to ensure that income has finite mean, is called the asymptotic Pareto parameter.[10]

Note that the bound is less than 1 if and only if $\varphi - 1 > \sigma$. Saez (2001) argues that for the United States $\varphi \approx 3$, which implies that this condition has bite for $\sigma < 2$. Recall that our specification of preferences is free of cardinality. Thus, σ should not be interpreted as risk-aversion, but rather

---

[8]In particular, some welfare function may be completely unaffected, or trivially so, by increases in utility for some particular set of workers, while others may be particularly sensitive to these increases. This ambiguity cannot be resolved without taking a stand on the social welfare function.

[9]See Diamond (1998), Saez (2001) and Roberts (2000) for optimal tax rate formulas derived for specific welfare functions.

[10]The Pareto distribution had a density that is a power function, $g(Y) = AY^{-\varphi}$, so that $d \log g(Y)/d \log Y = -\varphi$.

11

as controlling the income and substitution effects for labor. One natural benchmark is logarithmic utility $\sigma = 1$, where income and substitution effects for labor cancel out, i.e., the balanced-growth specification. Another is the quasi-linear case with $\sigma = 0$, which may be sensible given the low estimates of income elasticities for labor. The quasi-linear specification was adopted by Diamond (1998).

The upper bound accords well with intuition. As $\varphi \to \infty$, for fixed σ and η, the income distribution has a very thin upper tail and the upper bound on the tax rate converges to 0. Thus, the case with thin tails behaves as the case with bounded skills, where the tax rate must be zero at the top. Also, for fixed σ and φ, as $\eta \to \infty$ the bound converges to 1. Thus, any tax rate may be justified on efficiency grounds when labor supply is infinitely inelastic.

4.2. A Flat Tax

Suppose a proportional tax is in place, so that $T(Y) = \bar{\tau} Y$. Then the calculations are very similar to those from the previous subsection and yield

$$\bar{\tau} \leq \frac{\sigma + \eta - 1}{-\frac{d \log g(Y)}{d \log Y} + \eta - 2}. \tag{6}$$

Now suppose that taxes are linear, but not necessarily proportional. Specifically, assume $T(Y) = \bar{\tau} Y - T_0$ with a positive transfer $T_0 > 0$. Then

$$-\frac{d \log u'(Y - T(Y))}{d \log Y} = -\sigma \frac{1 - T'(Y)}{1 - T(Y)/Y} = \sigma \frac{1 - \bar{\tau}}{1 - \bar{\tau} + T_0/Y} \leq \sigma,$$

which starts at 0 for $Y = 0$ and rises monotonically to σ for $Y \to \infty$, so that

$$\frac{d \log \varepsilon_{\theta,Y}}{d \log Y} \geq 0.$$

It follows that (6) remains a necessary condition for ($\bigstar\bigstar$) and, thus, for the efficiency of a flat tax.

Additionally,

$$\frac{d \log}{d \log Y} \left(-\frac{d \log u'(Y - T(Y))}{d \log Y} \right) = \frac{d \log}{d \log Y} \left(\frac{1 - \bar{\tau}}{1 - \bar{\tau} + T_0/Y} \right) = \frac{T_0/Y}{1 - \bar{\tau} + T_0/Y} \leq 1,$$

implying

$$\frac{d \log \varepsilon_{\theta,Y}}{d \log Y} \leq \frac{\sigma}{\sigma + \eta - 1}$$

12

Hence a sufficient condition for (★★) to hold is that

$$\bar{\tau} \leq \frac{\eta - 1}{-\frac{d \log g(Y)}{d \log Y} + \eta - 2 + \frac{\sigma}{\sigma + \eta - 1}} < \frac{\eta - 1}{-\frac{d \log g(Y)}{d \log Y} + \eta - 1}. \tag{7}$$

4.3. Observable and Unobservable Heterogeneity

Suppose that, in addition to skill heterogeneity θ, workers are sorted into groups which affect their preferences and the distribution of skills within their group. For example, one group may supply labor more elastically than another, or one group may have higher average skills than another. In general, a worker's group membership may be observed or not by the tax authority. We treat both cases in turn; Combining the two is straightforward.

Specifically, suppose there are $i = 1, \ldots, I$ groups, with fractions $\pi^1, \ldots, \pi^I$ of the population. Workers in group i have utility $U^i(c, Y, \theta)$. The skill θ within group i is distributed with density $f^i(\theta^i)$. Each worker knows their own group and skill type.

When the group membership of a worker is observed by the tax authority, then it has the *potential* to condition taxes on it. That is, it can tailor a tax schedule $T^i(Y)$ for each group i. Within each group, the analysis is then exactly as before: the condition for efficiency (★) should hold for each worker group $i = 1, \ldots, I$, evaluated separately using each group's tax schedule $T^i(Y)$ and resulting allocation.

It then follows that, although group-contingent tax schedules are possible, it may be Pareto efficient to offer the same schedule to all groups, i.e., to set $T^i(Y) = T^j(Y)$. That is, although taxing an observable trait is feasible it may be Pareto efficient not to do so. This is the case as long as condition (★) holds for each group, when each group faces the same schedule. This is more likely whenever workers different across groups, but not by too much—since then if condition (★) holds for one group it will hold for similar another.

This conclusion contrast with the implications of the typical normative analysis conducted within the Mirrlees (1971) model and and other optimal-taxation studies that adopt a Utilitarian social welfare function. In such settings, conditioning on all available information is strictly optimal. For example, if a certain trait is associated with higher skills, then the optimal solution will generally tax this trait to redistribute towards other workers.

When a worker's group membership is not observed by the tax authority, then it cannot do better than by imposing a single tax schedule $T(Y)$ for all worker groups $i = 1, \ldots, I$. We now ask, what is the necessary condition for Pareto efficiency in this case? It turns out that it can be
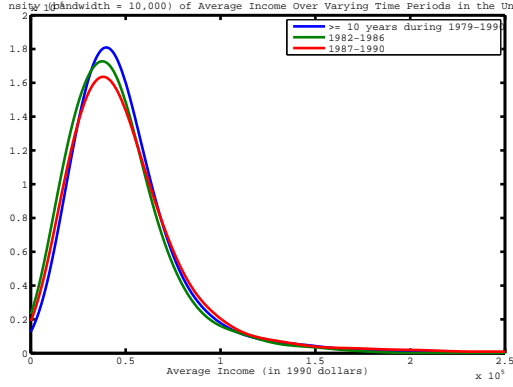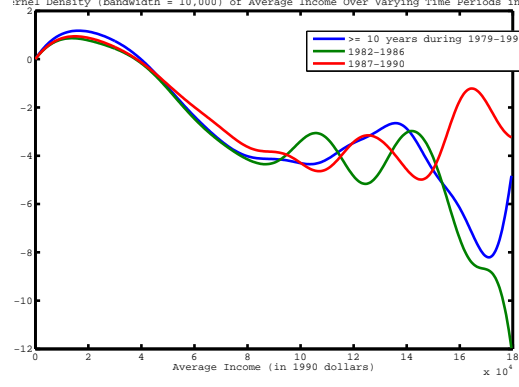
Figure 3: Density of income $g(Y)$.

Figure 4: Implied elasticity $\frac{Y g'(Y)}{g(Y)}$.

expressed as an appropriate average of condition ($\bigstar$):

$$\sum_{i=1}^{N} \pi^i g^i(Y^i(\theta)) \left(\frac{\tau^i(\theta)}{1 - \tau^i(\theta)} \left(\frac{d \log \tau^i(\theta)}{d \log \theta} + \frac{d \log f^i(\theta)}{d \log \theta} - \frac{d \log h^{i\prime}(Y^i(\theta))}{d \log \theta} \right) - 1 \right) \leq 0. \quad (8)$$

The proof of this condition is omitted, but will be included in latter versions of the paper. It is worth briefly remarking that the analysis behind this derivation does not require heterogeneity across worker groups to satisfy single-crossing conditions.

The same analysis applies when a worker's group membership is observable to the government, but we limit the tax authority to offering a single tax schedule for any other reason. This allows us to entertain the possibility that some notion of fairness or horizontal equity may compel societies not to tax workers of different traits differently. That is, conditioning taxes on traits, such as looks, may be viewed as morally unacceptable, even if would allow some Pareto gains. The efficiency condition (8) is then the relevant one. It ensures that no Pareto improvements are possible using a single tax schedule.

4.4. Quantitative Exploration using US Data

In this subsection I provide a preliminary examination of the test for efficiency for the United States. For this purpose, I use the 1979-1990 panel of US Federal Income Tax returns from the IRS's SOI Public use files. The idea is to produce the empirical analog of the theoretical Figure 2.

Since the Mirrlees model is static, it is best interpreted as capturing lifetime decisions over income and consumption. Likewise, the tax schedule should also be interpreted in this lifetime context.[11] This suggests using income averages in the panel as a proxy for lifetime income. Only taxpayers that are married and filing jointly are included, and those older that 65 are also dropped. Figure 3 plots the estimated distribution for three samples: (a) using only individuals with at

---

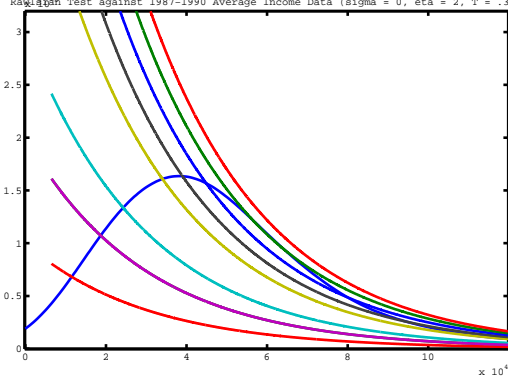[11]This lifetime perspective is likely to smooth out the relevance of any kinks in the yearly tax schedule.

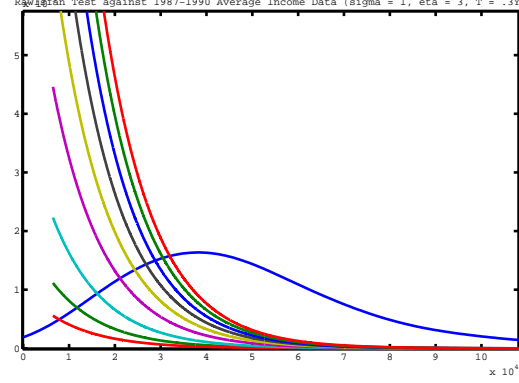Figure 5: Test for $\frac{1}{\eta-1} = \frac{1}{2}$, $\sigma = 0$.

Figure 6: Test for $\frac{1}{\eta-1} = 1$, $\sigma = 1$.

least 10 years of observations; (b) individuals appearing every year from 1982 to 1986; and (c) individuals appearing every year from 1987 to 1990. (Income was adjusted by the CPI to 1990 dollars. The bandwidth was set at 10.000 dollars.) The implied elasticities $Yg'(Y)/g(Y)$ from the kernel estimate of the density are shown in Figure 4. The overall picture confirms the cross-sectional evidence in Saez (2001). Our estimates of $Yg'(Y)/g(Y)$ become quite variable at high income levels.

The two other elements that are needed are required to compute the Rawlsian density. We use specification for utility $U(c, Y; \theta) = c^{1-\sigma}/(1-\sigma) - \theta\kappa Y^\eta$. Following Saez (2001), we simplify the US tax code to the extreme and assume $T(Y) = .3Y$. Figure 5 shows the test for a relatively high Frisch elasticity of labor $\frac{1}{\eta-1} = 1$ and $\sigma = 0$; the test shows a region of inefficiency at intermediate and high levels of income. Figure 6, on the other hand, does the same for a lower Frisch elasticity of labor, equal to $\frac{1}{\eta-1} = \frac{1}{2}$ and $\sigma = 1$; in this case the test is passed.

5. Conclusions

In this paper I have characterized the set of Pareto efficient tax schedules in Mirrlees' (1971) model. The analysis provides versions of the optimality condition that may be useful in testing this condition, or in provide a framework for quantitative work, along the lines of Saez (2001). By avoiding the specification of a normative welfare criterion, the analysis is able to focus on elements of the positive economy. In particular, the optimality conditions shed new light on the importance of the skill distribution and other parameters in shaping efficient tax schedules.

Appendix

A. Proof of Proposition 1

We perform a change in variable and write the planning problem in terms of $\tilde{H}(\theta) = h(\tilde{Y}(\theta))$:

$$\max_{\tilde{v}, \tilde{H}} \int \left(h^{-1}\big(\tilde{H}(\theta)\big) - u^{-1}\left(\tilde{\underline{v}} - \int_{\underline{\theta}}^{\theta} \tilde{H}(z)dz + \theta \tilde{H}(\theta) \right) \right) dF(\theta)$$

subject to,

$$\tilde{\underline{v}} - \int_{\underline{\theta}}^{\theta} \tilde{H}(z)dz - v(\theta) \geq 0 \tag{9}$$

and $\tilde{H}(\theta) \in NI(\Theta)$, where $NI(\Theta)$ represents the set of nonincreasing real-valued functions over Θ. This is now a convex optimization problem: the objective to be maximized is concave and the constraints are (linear) convex.

We now follow the analysis in Luenberger (1969, Chapter 8) to derive the optimality conditions. We shall exploit that $NI(\Theta)$ is a convex closed cone (i.e., closed under multiplication by positive scalars) in the linear space of bounded functions $B(\Theta)$ endowed with the supremum norm. Note that constraint (9) can be expressed as $G(\tilde{H}) \in P$, where the mapping $G : NI(\Theta) \rightarrow C(\Theta)$ is convex, taking nondecreasing functions into continuous functions, and P is the positive cone of $C(\Theta)$, i.e the set of continuous functions. Finally, note that constraint (9) allows for an interior point (e.g., for any $\tilde{\underline{v}} > v(\underline{\theta})$ and $\tilde{H}(\theta) = H(\theta) = h(\tilde{Y}(\theta))$). As a result, all the conditions required in Sections 8.3 and 8.4 in Luenberger (1969) are met, maximizing the Lagrangian is both necessary and sufficient for optimality.

Form the Lagrangian

$$\mathcal{L} \equiv \int \left(h^{-1}\big(\tilde{H}(\theta)\big) - u^{-1}\left(\underline{v} - \int_{\underline{\theta}}^{\theta} H(z)dz + \theta \tilde{H}(\theta) \right) \right) dF(\theta)$$

$$+ \int \left(\underline{v} - \int_{\underline{\theta}}^{\theta} \tilde{H}(z)dz - v(\theta) \right) d\lambda(\theta),$$

for some nondecreasing function $\lambda(\theta)$, the multiplier on the inequality (9), normalized so that $\lambda(\bar{\theta}) = 0$.

Applying the chain rule (Luenberger, 1969, Chapter 7), the Fréchet derivative is given by

$$\partial\mathcal{L}(H;\Delta_{\underline{v}},\Delta_{\tilde{H}}) = \int \left((h^{-1})'\big(H(\theta)\big)\Delta_{\tilde{H}}(\theta) - (u^{-1})'\big(U(\theta)\big)\big(\Delta_{\tilde{v}}(\theta) + \theta\Delta_{\tilde{H}}(\theta)\big) \right) dF(\theta)$$

$$+ \int \Delta_{\tilde{v}}(\theta)\, d\lambda(\theta),$$

where

$$\Delta_{\tilde{v}}(\theta) = \Delta_{\underline{v}} - \int_{\underline{\theta}}^{\theta} \Delta_{\tilde{H}}(z)\, dz.$$

Equivalently, substituting $\Delta_{\tilde{v}}(\theta)$ out and integrating by parts,

$$\partial\mathcal{L}(H;\Delta_{\underline{v}},\Delta_{\tilde{H}}) = \int \left((h^{-1})'\big(H(\theta)\big) - (u^{-1})'\big(U(\theta)\big)\theta \right) f(\theta)\Delta_{\tilde{H}}(\theta)d\theta$$

$$+ \int \left(\int_{\theta}^{\bar{\theta}} (u^{-1})'\big(U(z)\big) f(z)\, dz \right) \Delta_{\tilde{H}}(\theta)\, d\theta + \int \lambda(\theta)\Delta_{\tilde{H}}(\theta)\, d\theta$$

$$- \Delta_{\underline{v}}\left(\lambda(\underline{\theta}) + \int (u^{-1})'\big(U(\theta)\big) f(\theta)d\theta \right)$$

Collecting terms:

$$\partial\mathcal{L}(H;\Delta_{\underline{v}},\Delta_{\tilde{H}}) = \int A(\theta)\Delta_{\tilde{H}}(\theta)\, d\theta = \Delta_{\tilde{H}}(\underline{\theta}) \int_{\underline{\theta}}^{\bar{\theta}} A(z)\, dz + \int \int_{\theta}^{\bar{\theta}} A(z)\, dz\, d\Delta_{\tilde{H}}(\theta), \quad (10)$$

where

$$A(\theta) \equiv \left((h^{-1})'\big(H(\theta)\big) - (u^{-1})'\big(U(\theta)\big)\theta \right) f(\theta) + \int_{\theta}^{\bar{\theta}} (u^{-1})'\big(U(z)\big) f(\theta)\, dz + \lambda(\theta).$$

Since the Lagrangian is convex, the necessary and sufficient conditions for $H \in NI(\Theta)$ to maximize it are:[12]

$$\partial\mathcal{L}(H;\Delta_{\underline{v}},\Delta_{\tilde{H}}) \geq 0 \quad \text{for all } \Delta_{\tilde{H}} \in NI(\Theta), \tag{11}$$

$$\partial\mathcal{L}(H;\underline{v},H) = 0. \tag{12}$$

From (11), since $\Delta_{\underline{v}}$ can be positive or negative, we immediately obtain that

$$\lambda(\underline{\theta}) + \int (u^{-1})'\big(U(\theta)\big) f(\theta)d\theta = 0.$$

---

[12]See Lemma A.2. in Amador, Werning, and Angeletos (2006), which is a simple extension of Lemma 1, pg. 227, in Luenberger (1969) to allow for Gateaux differentials instead of Frechet derivatives.

Since $\Delta_{\tilde{H}}(\underline{\theta})$ can be positive or negative and $\Delta_{\tilde{H}}$ is nondecreasing it follows that we must have that

$$\int_{\underline{\theta}}^{\bar{\theta}} A(z)\,dz = 0 \qquad \int_{\theta}^{\bar{\theta}} A(z)\,dz \le 0.$$

From (12), if the original allocation has $H(\theta) = h(Y(\theta))$ strictly increasing in a neighborhood around θ then it follows that

$$\int_{\theta}^{\bar{\theta}} A(z)\,dz = 0 \quad \Rightarrow \quad A(\theta) = 0.$$

In addition we must have that the resulting $\lambda(\theta)$ is nondecreasing. Using the fact that $(h^{-1})'(H(\theta))-(u^{-1})'(U(\theta))\theta = \tau(\theta)/h'(Y(\theta))$ and that $(u^{-1})'(U(\theta)) = e_v(v(\theta), Y(\theta), \theta)$, we obtain

$$-\lambda(\theta) = \frac{\tau(\theta)f(\theta)}{h'(Y(\theta))} + \int_{\theta}^{\bar{\theta}} e_v(v(z), Y(z), z)f(z)\,dz,$$

must be decreasing. Differentiating this expression and setting $-\lambda'(\theta) \le 0$ gives ($\bigstar$).

In a region where $H(\theta) = h(Y(\theta))$ is constant the optimality condition (12) does not yield any additional constraints.

B. Proof of Proposition 3

I show that condition ($\bigstar$) implies condition ($\bigstar\bigstar$). Proposition 1 then implies the result.

The relation $F(\theta(Y)) = 1 - G(Y)$ implies (using $d\log\theta(Y)/d\log Y = -\varepsilon_{\theta,Y}$ and $\theta'(Y) < 0$)

$$f(\theta(Y)) = -\frac{g(Y)}{\theta'(Y)}$$

$$-\frac{d\log f(\theta(Y))}{d\log\theta}\varepsilon_{\theta,Y} = \frac{d\log g(Y)}{d\log Y} - \frac{d\log -\theta'(Y)}{d\log Y} = \frac{d\log g(Y)}{d\log Y} + 1 - \varepsilon_{\theta,Y} - \frac{d\log\varepsilon_{\theta,Y}}{d\log Y}.$$

Multiplying ($\bigstar$) through by $\varepsilon_{\theta,Y} > 0$ and substituting this last expression gives

$$-\frac{d\log 1 - T'(Y)}{d\log Y} - \frac{T'(Y)}{(1-T'(Y))}\left(\frac{d\log g(Y)}{d\log Y} - \frac{d\log h'(Y)}{d\log Y} + 1 - \varepsilon_{\theta,Y} - \frac{d\log\varepsilon_{\theta,Y}}{d\log Y}\right) \le \varepsilon_{\theta,Y}.$$

After rearranging, this yields ($\bigstar\bigstar$).

C. Proof of Proposition 4

For any such change in the tax schedule there is an interval $[\theta_1, \theta_2]$ of agents that now prefer to produce $\hat{Y} = Y(\hat{\theta})$. This induces an allocation that cannot reduce welfare and must strictly increase it for agents in the interval $[\theta_1, \theta_2]$. The rest of this proof shows, that for sufficiently small

ε, the induced allocation is resource feasible.

I shall express $\theta_2 = \Theta_2(\theta_1)$ and $\varepsilon(\theta_1)$ as functions of θ_1. The two indifference conditions

$$v(\theta_1) = u(\hat{c} + \varepsilon(\theta_1)) - \theta_1 \hat{h},$$

$$v\big(\Theta_2(\theta_1)\big) = u(\hat{c} + \varepsilon(\theta_1)) - \Theta_2(\theta_1)\hat{h}$$

determine these functions implicitly. Note that $\Theta_2(\hat{\theta}) = \hat{\theta}$ and $\varepsilon(\hat{\theta}) = 0$. Differentiating these expressions and applying L'Hospital's rule yields the following values for the derivatives of $\Theta_2(\theta)$ and $\varepsilon(\theta)$ which will be needed below:

$$\Theta_2'(\hat{\theta}) = -1 \quad \text{and} \quad \frac{3}{2}\Theta_2''(\hat{\theta}) = -\left(\frac{Y''(\hat{\theta})}{Y'(\hat{\theta})} + \frac{h''(Y(\hat{\theta}))}{h'(Y(\hat{\theta}))}Y'(\hat{\theta})\right)$$

and

$$\varepsilon'(\hat{\theta}) = 0 \quad \text{and} \quad \varepsilon''(\hat{\theta}) = -\frac{1}{\theta}(1 - \tau(\hat{\theta}))Y'(\hat{\theta}),$$

which uses the definition $1 - \tau(\theta) \equiv \theta h'(Y(\theta))/u'(c(\theta))$.

The loss in resources is given by

$$\Delta(\theta_1) \equiv \int_{\theta_1}^{\Theta_2(\theta_1)} \left(Y(\theta) - c(\theta) - (\hat{Y} - \hat{c}) + \varepsilon(\theta_1)\right) f(\theta),$$

with $\Delta(\hat{\theta}) = 0$.

Tedious calculations establish that $\Delta'(\hat{\theta}) = \Delta''(\hat{\theta}) = 0$ and

$$\Delta'''(\hat{\theta}) = -3\Theta_2''\big(Y'(\hat{\theta}) - c'(\hat{\theta})\big)f(\hat{\theta}) - 2\big(Y''(\hat{\theta}) - c''(\hat{\theta}) + \varepsilon''(\hat{\theta})\big)f(\hat{\theta})$$
$$- 4\big(Y'(\hat{\theta}) - c'(\hat{\theta})\big)f'(\hat{\theta}) - 6\varepsilon''(\hat{\theta})f(\hat{\theta}).$$

If $\Delta'''(\hat{\theta}) \geq 0$ then there exists a $\theta_1 < \hat{\theta}$ close enough to $\hat{\theta}$ so that $\Delta(\theta_1) < 0$ and resources rise from the perturbation. The perturbation is then resource feasible and creates a Pareto improvement.

Dividing $\Delta'''(\hat{\theta})$ by $-(Y'(\hat{\theta}) - c'(\hat{\theta}))f(\hat{\theta}) > 0$ implies that $\Delta'''(\hat{\theta}) \geq 0$ if and only if

$$\frac{3}{2}\Theta_2''(\hat{\theta}) + \frac{Y''(\hat{\theta}) - c''(\hat{\theta})}{Y'(\hat{\theta}) - c'(\hat{\theta})} + 2\frac{f'(\hat{\theta})}{f(\hat{\theta})} + 3\frac{\varepsilon''(\hat{\theta})}{Y'(\hat{\theta}) - c'(\hat{\theta})} \leq 0.$$

Note that $Y'(\hat{\theta}) - c'(\hat{\theta}) = \tau(\hat{\theta})Y'(\hat{\theta})$ and $Y''(\hat{\theta}) - c''(\hat{\theta}) = \tau'(\hat{\theta})Y'(\hat{\theta}) + \tau(\hat{\theta})Y''(\hat{\theta})$ so that

$$\frac{Y''(\hat{\theta}) - c''(\hat{\theta})}{Y'(\hat{\theta}) - c'(\hat{\theta})} = \frac{\tau'(\hat{\theta})}{\tau(\hat{\theta})} + \frac{Y''(\hat{\theta})}{Y'(\hat{\theta})} \quad \text{and} \quad \frac{\varepsilon''(\hat{\theta})}{Y'(\hat{\theta}) - c'(\hat{\theta})} = -\frac{1}{\hat{\theta}}\frac{1 - \tau(\hat{\theta})}{\tau(\hat{\theta})}.$$

Substituting these expressions and the one found earlier for $\frac{3}{2}\Theta_2''(\hat{\theta})$ and cancelling yields condition (5).

References

Amador, Manuel, Iván Werning, and George-Marios Angeletos, 2006. "Commitment vs. Flexibility." *Econometrica*. 74 (2): 365–396.

Brito, Dagobert L., Jonathan H. Hamilton, Steven M. Slutsky, and Joseph E. Stiglitz, 1990. "Pareto Efficient Tax Structures." *Oxford Economic Papers*. 42: 61–77.

Chone, Philippe, and Guy Laroque, 2005. "Optimal incentives for labor force participation." *Journal of Public Economics*. 89 (2–3): 395–425.

Chone, Philippe, and Guy Laroque, 2006. "Should low skill work be subsidized?." mimeo.

Diamond, Peter A., 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review*. 88 (1): 83–95.

Friedman, Milton, 1962. *Capitalism and Friedom*, Chicago: University of Chicago Press.

Gruber, Jon, and Emmanuel Saez, 2002. "The elasticity of taxable income: evidence and implications." *Journal of Public Economics*. 84 (1): 1–32.

Hall, Robert, and Alvin Rabushka, 1995. *The Flat Tax*, Hoover Institution Press.

Luenberger, David G., 1969. *Optimization by Vector Space Methods*, John Wiley & Sons, Inc.

Mirrlees, James A., 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies*. 38: 175–208.

Roberts, Kevin, 2000. "A Reconsideration of the Optimal Income Tax." in P.J. Hammond, and H.D. Myles (ed.), *Incentives and Organization: Papers in Honor of Sir james Mirrlees*, Oxford: Oxford University Press.

Saez, Emmanuel, 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*. 68 (1): 205–29.

Stiglitz, Joseph, 1987. "Pareto Efficient and Optimal Taxation and the New New Welfare Economics." in A. Auerbach, and M. Feldstein (ed.), *Handbook on Public Economics* . pp. 991–1042, North Holland: Elsevier Science Publishers.

Tuomala, Matti, 1990. *Optimal Income Taxation and Redistribution*, Oxford University Press, Clarendon Press.