

Optimal Income Taxation and Risk: The Extensive-Margin Case

by

Robin Boadway, Queen's University, Canada
Motohiro Sato, Hitotsubashi University, Japan

December, 2012

Abstract

The optimal income tax structure is studied in a setting in which workers make discrete labor market decisions and earnings are uncertain. Workers differ continuously along a single dimension that reflects their skills as well as their disutility of work in different jobs. A discrete number of skill-types of jobs are available in perfectly elastic supply. Each job yields a stochastic distribution of wages, where the distribution differs among skill-types. The amount of work in each job is fixed, so there is no intensive labor-supply decision and wages reflect earnings. Expected wages for a given skill-type of job are higher for higher-skilled workers. Workers first choose a job based on the distribution of wages they expect to earn in different jobs. Once jobs are chosen, wages are revealed and workers decide whether to participate in the job or to become voluntarily unemployed. Each job will be associated with a distribution of wages, and the same wage will be paid by more than one type of job. Under reasonable conditions, workers segment themselves by skill levels into jobs. We analyze the optimal income tax structure given these two margins of decision-making, job choice and participation. The optimal tax will reflect insurance (since earnings are uncertain when jobs are chosen), redistribution (since persons of higher skills earn more), and efficiency (since taxes affect both job choice and participation). The form of the tax structure is comparable to that obtained when labor supply can be varied along the intensive margin.

Key Words: optimal income taxation, risk, extensive margin

JEL Classification: H21, H24

Acknowledgments: Financial support of the Social Sciences and Humanities Research Council of Canada (Boadway) and of Grants-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (Sato) are gratefully acknowledged.

Email: <boadwayr@econ.queensu.ca>, <satom@econ.hit-u.ac.jp>

1 Introduction

The optimal income tax literature has developed considerably in recent years. In the seminal approach of Mirrlees (1971), workers who differ in ability or productive skills decide how much labor to supply given their wage rate and a nonlinear tax on their earnings. The government observes earnings but not wage rates, and chooses a nonlinear income tax to maximize a concave social welfare function subject to a budget constraint and some incentive constraints, taking into account workers' labor market behavior. The resulting income tax structure reflects various factors, including the form of the social welfare function, worker preferences, and the distribution of skills. The intuitive effects of these various factors, as well as their implications for the progressivity of the tax structure, are well-understood (see, for example, Saez 2001 and Kaplow 2008).

The Mirrlees approach has been extended in a number of directions. Multiple consumption goods have been allowed, opening up the possibility of differential commodity taxation alongside the nonlinear income tax (Atkinson and Stiglitz 1976; Christiansen 1984; and Guesnerie 1995). Worker choices have been constrained to the extensive margin by assuming that labor time on the job is not variable, so that workers can choose whether to work or not (Diamond 1980), and if they do work which job to choose (Saez 2002, Christiansen 2012). Extensive-margin models open up the possibility of involuntary unemployment, and optimal income taxation has been explored in this context in models where search frictions are responsible for the unemployment (Hungerbühler et al 2006). More recently, the dynamic public finance literature has extended the Mirrlees approach to an intertemporal setting (Golosov et al 2007; Kocherlakota 2010). In a dynamic setting, capital income taxes become relevant along with lifetime tax schedules (Conesa, Kitao and Krueger 2009; Bastani, Blomquist and Micheletto 2010; Weinzierl 2011).

The dynamic public finance literature has also emphasized an element that has been studied only sparingly in the optimal income tax literature, and that is the effect of uncertainty. In this literature, the usual approach is to assume that wages evolve stochastically over time, and there is limited, if any, insurance available. Labor supply in each period is chosen after wages are revealed, so earnings are certain. However, saving and possibly purchases of consumer durables (Cremer and Gahvari 1995) and human capital (Anderberg 2009) are chosen before future states are revealed, so they incorporate precautionary elements.

Our interest is in labor choices that are made before future wages or earnings are known. The literature on optimal income taxation under wage uncertainty is relatively limited. Much of the focus has been on the case of ex ante identical individuals who choose labor supply with uncertain wage rates (Tuomala 1984; Cremer and Gahvari 1999; Low and Maldoom 2004; Henriët, Pintus and Trannoy 2012). Although workers are heterogeneous ex post, so redistributive taxation is desirable, the motive is essentially one of providing insurance. Analyses of workers with different skills ex

ante, which combine insurance with redistribution, are less common. A well-known early example is Eaton and Rosen (1980), who investigate the use of a linear progressive tax in a model with two skill-types of workers and wage uncertainty. The complexity of the problem forces them to resort to simulations to get a sense of the optimal form of policy. Diamond, Helms and Mirrlees (1980) consider the case of optimal linear progressive taxation when workers of different skills face uncertainty about whether they are able to work or not. The general case of optimal nonlinear taxation in a Mirrlees setting where labor supply must be chosen before wage rates are realized was studied by Boadway and Sato (2011). The structure of the optimal nonlinear marginal income tax rates can be specified as a composite of an equity effect, an insurance effect, and an incentive effect. In all these papers, labor supply is along the intensive margin.

In this paper, we explore optimal nonlinear income taxation when worker choices are along extensive margins. We distinguish between decisions taken before wage rates are revealed and those taken after. In particular, there is a continuous distribution of workers by skill type who must choose among a discrete number of jobs, each of which caters to a different skill level. Employment in a job involves a fixed labor supply and pays a given, but uncertain, wage. Jobs requiring higher skills pay higher average wages. At the time that a job is chosen, there is uncertainty about wages. Once a job is chosen, the wage rate is revealed and workers decide whether to participate. This depends both on the wage rate and on their utility of leisure, which is correlated with their skill parameter. The government observes wages, but not the workers' types. Policies include a nonlinear tax function in terms of wages and a payment to those who do not participate. We characterize the structure of optimal taxes.

2 Basic Setting

There is a continuum of workers who differ in a single dimension denoted by θ , which is distributed over $[\underline{\theta}, \bar{\theta}]$ by the distribution function $F(\theta)$, with density $f(\theta) = F'(\theta)$. The population size of workers is normalized to unity. We refer to θ as a skill parameter, although it affects both the wage distribution and the disutility of work. There are J types of jobs, denoted by the subscripts $j = 1, 2, \dots, J$, that are available in perfectly elastic supply. Job $j + 1$ can be interpreted as being suited to higher skilled workers than job j . Each worker chooses a job that yields an uncertain wage \tilde{w} , whose true value w is revealed after the job is chosen. The revealed wage w can be the same for workers of different skills θ in the same job or in different jobs, and it can differ for workers of the same skill level. Job choice is assumed to be irreversible.

After w is revealed, the worker decides whether to work. If a worker of type θ who has chosen job j decides to work, the level of earnings obtained is fixed and equals the revealed value of w . Thus, there is no intensive variation in labor supply. Those who do not work have no earnings and rely on a government transfer to finance consumption. This ex post participation decision, taken

after uncertainty is resolved, is identical to the pure extensive-margin models of Diamond (1980) and Saez (2002).

A fixed disutility of work $\phi_j(\theta)$ is incurred in job j . We assume $\phi_j(\theta)$ is independent of w and satisfies the following properties:

Assumption 1: The disutility of work function satisfies:

$$(i) \phi_{j+1}(\theta) > \phi_j(\theta), \quad (ii) \phi'_j(\theta) < 0, \quad (iii) \phi'_{j+1}(\theta) < \phi'_j(\theta), \quad \forall j, \theta$$

Assumption 1(i) says that higher-skilled jobs are more costly for any given type of worker. Assumption 1(ii) says that in any given job, the disutility of work is decreasing in skill. Assumption 1(iii) says that the disutility of work decreases more rapidly with skills in higher skilled jobs. The latter assumption plays a role in leading to a single-crossing property that results in matching workers to jobs by skill levels.

The government can observe earnings w and imposes a nonlinear income tax function $t(w)$. After-tax earnings associated with the wage w are $z(w) = w - t(w)$. In what follows, we sometimes require $z(w)$ to be non-decreasing in w for incentive purposes. The utility of a type- θ worker earning a wage w in job j is given by $u(z(w)) - \phi_j(\theta)$, where $u(\cdot)$ is increasing and strictly concave, so workers are risk-averse. For all non-participants, utility is simply $u(b)$, where b is a transfer that is the same for all non-participants since the government cannot observe their types.

The distribution of wages depends on both the worker's type and the job. Let $q_j(w|\theta)$ be the probability of a type- θ worker obtaining wage w in job j . The conditional distribution of w for a type- θ worker in job j is then $Q_j(w|\theta) = \int_{\underline{w}}^w q_j(v|\theta)dv$, where \underline{w} and \bar{w} are the bounds of wages and $\int_{\underline{w}}^{\bar{w}} q_j(w|\theta)dw = 1$. Thus, $Q_j(w|\theta)$ is the proportion of type- θ worker earning no more than wage w in job j . We assume for notational simplicity that \underline{w} and \bar{w} are the same for all jobs, although for some purposes we may assume that only high-skilled workers can earn \bar{w} . Assume the conditional distribution function satisfies the following properties.

Assumption 2: The conditional distribution of wages for a given skill-type in jobs j and $j + 1$ satisfies the following for all j and $w \in (\underline{w}, \bar{w})$:

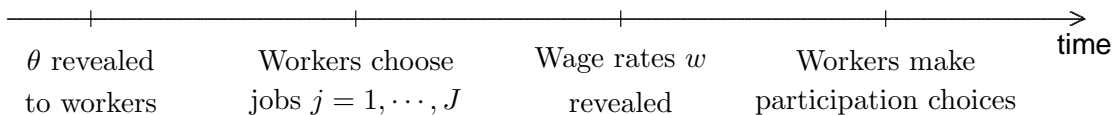
$$(i) \frac{\partial Q_j(w|\theta)}{\partial \theta} < 0, \quad (ii) Q_j(w|\theta) > Q_{j+1}(w|\theta), \quad (iii) -\frac{\partial Q_{j+1}(w|\theta)}{\partial \theta} > -\frac{\partial Q_j(w|\theta)}{\partial \theta}$$

Assumption 2(i) indicates that the wage distribution for skill level θ' in a given job j first-order stochastically dominates that for θ when $\theta' > \theta$. In other words, the expected wage is increasing in θ in a given job. That is the sense in which θ reflects a worker's skills. Assumption 2(ii) says that jobs requiring higher skills tend to be more productive, so pay higher expected wages for a given skill level. Assumption 2(iii) says that for any given skill level θ , an increase in skills will

increase expected wages more in jobs requiring higher skills. This will also favor the single-crossing property mentioned above and discussed further below.

Individuals of each skill-type choose their jobs based on their expected utility, given the skill-specific distribution of wages in each job. We give a more precise specification of that below. Note for now that, since there are a discrete number of jobs and a continuum of worker-types, workers of different θ 's will choose the same job. As well, the same wage can be earned by workers in different jobs. This leads to a matching problem and differs from the standard extensive-margin approaches where all workers in a given job, regardless of skill type, obtain the same wages (e.g., Saez 2002).

The timeline of events is as follows.



3 Worker Behavior

Workers thus make two decisions in sequence. They choose a job knowing only the distribution of wages they can expect, and once the wage is revealed they decide whether to participate in the workforce. We consider these in reverse order.

Participation Decision

Once the job has been chosen and earnings revealed, the participation choice is made. Workers of type θ who have chosen job j will participate if and only if $u(z(w)) - \phi_j(\theta) \geq u(b)$. Let the skill-type of the marginal participant earning w in job j be given by θ_j^* . Then, θ_j^* satisfies:

$$\phi_j(\theta_j^*) = u(z(w)) - u(b), \quad \text{where} \quad \frac{\partial \theta_j^*}{\partial z(w)} = \frac{u'(z(w))}{\phi_j'(\theta_j^*)} < 0, \quad \frac{\partial \theta_j^*}{\partial b} = -\frac{u'(b)}{\phi_j'(\theta_j^*)} > 0 \quad (1)$$

The inequalities follow from $\phi_j'(\theta) < 0$ by Assumption 1(ii). Solving (1), we obtain the skill levels of marginal participants earning different wage rates in job j , $\theta_j^*(z(w), b)$, which is decreasing in $z(w)$ and increasing in b . All workers of type $\theta \geq \theta_j^*(z(w), b)$ who have chosen job j and been offered wage w will participate, while those with lower θ will not. Note that for any given job, there will be many marginal participants, each one associated with a different revealed wage w .

Given Assumption 1(i) that $\phi_{j+1}(\theta) > \phi_j(\theta)$, the following lemma is apparent.

Lemma 1: For any given $z(w)$ and b , $\theta_{j+1}^*(z(w), b) > \theta_j^*(z(w), b)$.

That is, cutoff skill levels for given wage rates and policies will be higher in higher-skilled jobs.

Job Choice

Workers must choose their job-type before uncertainty is resolved, anticipating the ex post participation decision. The expected utility for a type- θ worker in job j is given by:

$$V_j(\theta) = \int_{\underline{w}}^{\bar{w}} \max \left[u(z(w)) - \phi_j(\theta), u(b) \right] q_j(w|\theta) dw$$

which can be rewritten as:

$$V_j(\theta) = u(b) \int_{\underline{w}}^{\bar{w}} \left(1 - \delta_j(\theta_j^*(\cdot), \theta) \right) q_j(w|\theta) dw + \int_{\underline{w}}^{\bar{w}} \delta_j(\theta_j^*(\cdot), \theta) \left(u(z(w)) - \phi_j(\theta) \right) q_j(w|\theta) dw \quad (2)$$

where $\delta_j(\theta_j^*(z(w), b), \theta) = 1$ iff $\theta \geq \theta_j^*(z(w), b)$, so the worker participates.

Define $w_j^*(\theta)$ as the value of w that satisfies $\theta_j^*(z(w_j^*), b) = \theta$. Thus, $w_j^*(\theta)$ is the cutoff wage level in job j for a type- θ worker, and varies with θ . Of course, $w_j^*(\theta)$ depends on government policies, although we have suppressed that in the notation for simplicity.

In what follows, we make the following further assumption unless otherwise indicated, where we follow the convention of using a ‘dot’ to indicate a derivative with respect to w .

Assumption 3: $\dot{z}(w) \geq 0$

This plays a similar role to the assumption in the standard optimal nonlinear income tax problem that the second-order incentive constraint is satisfied (Ebert 1992). Given that $\theta_j^*(\cdot)$ is decreasing in $z(w)$, Assumption 3 implies that $w_j^*(\theta)$ will be uniquely defined for given j , θ , and government policies $t(w)$ and b . Note that workers will participate if and only if $w_j(\theta) \geq w_j^*(\theta)$.

Using this definition of $w_j^*(\theta)$, expected utility in (2) can be rewritten as:

$$V_j(\theta) = u(b)Q_j(w_j^*|\theta) + \int_{w_j^*(\theta)}^{\bar{w}} u(z(w))q_j(w|\theta)dw - \phi_j(\theta) \left(1 - Q_j(w_j^*|\theta) \right) \quad (3)$$

Applying partial integration to the integral on the right-hand side of (3) and using $Q_j(\bar{w}|\theta) = 1$, this becomes:

$$\begin{aligned} V_j(\theta) &= u(b)Q_j(w_j^*|\theta) + \int_{w_j^*(\theta)}^{\bar{w}} u'(z(w))\dot{z}(w) \left(1 - Q_j(w_j^*|\theta) \right) dw + \left(u(z(w_j^*)) - \phi_j(\theta) \right) \left(1 - Q_j(w_j^*|\theta) \right) \\ &= \int_{w_j^*(\theta)}^{\bar{w}} u'(z(w))\dot{z}(w) \left(1 - Q_j(w_j^*|\theta) \right) dw + u(b) \end{aligned} \quad (4)$$

where the last step uses the fact that $u(b) - u(z(w_j^*)) + \phi_j(\theta) = 0$ for the marginal participant.

The way in which $V_j(\theta)$ varies by jobs determines which job a worker of a given skill chooses. More generally, how $V_j(\theta)$ varies by both jobs and skills determines how workers of different types sort themselves among jobs. Differentiating (4) with respect to θ yields:

$$V_j'(\theta) = - \int_{w_j^*(\theta)}^{\bar{w}} u'(z(w))\dot{z}(w) \frac{\partial Q_j(w|\theta)}{\partial \theta} dw - u'(z(w_j^*))\dot{z}(w_j^*) \left(1 - Q_j(w_j^*|\theta) \right) \frac{\partial w_j^*}{\partial \theta} \quad (5)$$

Differentiating $u(z(w_j^*)) - \phi_j(\theta) = u(b)$ with respect to θ gives $u'(z(w_j^*))\dot{z}(w_j^*)\partial w_j^*/\partial\theta = \phi_j'(\theta)$. Substituting this in (5), we obtain:

$$V_j'(\theta) = - \int_{w_j^*(\theta)}^{\bar{w}} u'(z(w))\dot{z}(w) \frac{\partial Q_j(w|\theta)}{\partial\theta} dw - \phi_j'(\theta) \left(1 - Q_j(w_j^*|\theta)\right) \quad (6)$$

Since $Q_j(\cdot)$ is decreasing in θ by Assumption 2(i), $\dot{z}(w) > 0$ by Assumption 3, and $\phi_j'(\theta) < 0$ by Assumption 1(ii), we have the following lemma.

Lemma 2: $V_j(\theta)$ is increasing in θ for given j .

We assume the following single-crossing property to facilitate our analysis.

Assumption 4: $V_{j+1}'(\theta) > V_j'(\theta)$

This says that expected utility increases in skills at a higher rate in jobs requiring higher skills.

The plausibility of Assumption 4 can be checked by using (6) for jobs j and $j + 1$ to give:

$$\begin{aligned} V_{j+1}'(\theta) - V_j'(\theta) = & - \int_{w_{j+1}^*(\theta)}^{\bar{w}} u'(z(w))\dot{z}(w) \frac{\partial Q_{j+1}(w|\theta)}{\partial\theta} dw + \int_{w_j^*(\theta)}^{\bar{w}} u'(z(w))\dot{z}(w) \frac{\partial Q_j(w|\theta)}{\partial\theta} dw \\ & - \phi_{j+1}'(\theta) \left(1 - Q_{j+1}(w_{j+1}^*|\theta)\right) + \phi_j'(\theta) \left(1 - Q_j(w_j^*|\theta)\right) \end{aligned} \quad (7)$$

where $-\partial Q_{j+1}/\partial\theta > -\partial Q_j/\partial\theta$ by Assumption 2(iii), and $w_{j+1}^* > w_j^*$ follows from Lemma 1. The two integrals on the right-hand side of (7) can be written as:

$$- \int_{w_{j+1}^*(\theta)}^{\bar{w}} u'(z(w))\dot{z}(w) \left(\frac{\partial Q_{j+1}(w|\theta)}{\partial\theta} - \frac{\partial Q_j(w|\theta)}{\partial\theta} \right) dw + \int_{w_j^*(\theta)}^{w_{j+1}^*(\theta)} u'(z(w))\dot{z}(w) \frac{\partial Q_j(w|\theta)}{\partial\theta} dw$$

The first term is positive and the second is negative, but it is reasonable to suppose that the first outweighs the second for jobs below the most-skilled. The last two terms in (7) can be written:

$$\phi_j'(\theta) \left(Q_{j+1}(w_{j+1}^*|\theta) - Q_j(w_j^*|\theta) \right) + \left(\phi_j'(\theta) - \phi_{j+1}'(\theta) \right) \left(1 - Q_{j+1}(w_{j+1}^*|\theta) \right)$$

The first term is ambiguous, while the second one is positive since $\phi_j'(\theta) - \phi_{j+1}'(\theta) > 0$ by Assumption 1(iii). The two terms together will be positive if $\phi_j'(\theta) - \phi_{j+1}'(\theta)$ is large enough. Overall, Assumption 4 is reasonable, and implies the following lemma.¹

¹ The reasonableness of Assumption 4 can further be illustrated by considering the extreme case where there is no wage uncertainty. Each job offers a given wage such that $w_1 < w_2 < \dots < w_J$. Workers select jobs according to their skill level θ , with the lowest- θ workers choosing not to participate. In this case, (7) reduces to $V_{j+1}' - V_j' = \phi_j' - \phi_{j+1}' > 0$ by Assumption 1. Therefore, $V_{j+1}' > V_j'$, so the single-crossing property is satisfied.

Lemma 3: Given Assumption 4, there will be unique cutoff skill levels $\hat{\theta}_j$ such that $V_{j+1}(\hat{\theta}_j) = V_j(\hat{\theta}_j)$ for all $j = 1, \dots, J - 1$, with $\hat{\theta}_1 < \hat{\theta}_2 < \dots < \hat{\theta}_{J-1}$.

This implies that workers will sort themselves out among jobs such that those with skill $\theta \in (\hat{\theta}_j, \hat{\theta}_{j+1})$ will choose job $j + 1$, for all j . Figure 1 depicts this. Workers with $\underline{\theta} \leq \theta < \hat{\theta}_1$ will choose job 1, while those with $\bar{\theta}_{J-1} < \theta \leq \bar{\theta}$ will choose job J . The allocation of all workers to jobs is shown in Figure 2.²

4 The Overall Allocation of Workers

We are now in a position to characterize the number of workers at each wage level, and the number who choose to participate. Assuming that workers sort themselves among jobs as implied by Lemma 3, the total number of workers whose wage has been revealed to be w , denoted $N(w)$, can be expressed as follows:

$$N(w) = \sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} q_j(w|\theta) f(\theta) d\theta$$

where, recall, $q_j(w|\theta)$ is the proportion of type- θ workers who earn wage rate w in job j . The total number of workers then satisfies:

$$\int_{\underline{w}}^{\bar{w}} N(w) dw = \sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \left(\int_{\underline{w}}^{\bar{w}} q_j(w|\theta) dw \right) f(\theta) d\theta = 1 \quad (8)$$

As above, let $\delta_j(\theta_j^*, \theta) = 1$ if $\theta > \theta_j^*(z(w), b)$, so the worker is a participant. For non-participants, $\delta_j(\theta_j^*, \theta) = 0$. The number of participants who obtain wage rate w can then be written:

$$L(w) = \sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \delta_j(\theta_j^*, \theta) q_j(w|\theta) f(\theta) d\theta \quad (9)$$

where the marginal skill levels are $\theta_j^* = \theta_j^*(z(w), b)$ as earlier. Differentiating (9) with respect to the government's policy variables $z(w)$ and b , we obtain:³

$$\frac{\partial L(w)}{\partial z(w)} = \sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \frac{\partial \delta_j}{\partial z(w)} q_j(w|\theta) f(\theta) d\theta > 0, \quad \frac{\partial L(w)}{\partial b} = \sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \frac{\partial \delta_j}{\partial b} q_j(w|\theta) f(\theta) d\theta < 0 \quad (10)$$

² We have assumed that all workers choose some job, and then decide whether to participate when wages are revealed. Alternatively, it might be the case that for the lowest- θ workers, they choose not to participate ex ante. To account for this, we could designate $j = 0$ as the selection of these workers. The cutoff skill level between $j = 0$ and $j - 1$ would be $\hat{\theta}_1$. Moving up to jobs of higher skills, $\hat{\theta}_j$ would be the cutoff skill level between jobs $j - 1$ and j . This would involve a slight reinterpretation of Figures 1 and 2, but the optimal tax results derived and their interpretation would not change.

³ Note that the marginal skill values, $\hat{\theta}_j$ and $\hat{\theta}_{j-1}$, are being held constant in these derivatives. That is because later we use these bounds as control variables for the government's optimal tax problem. The derivative of $L(w)$ with respect to $\hat{\theta}_j$ is given below.

Recall that $z(w) = w - t(w)$. The inequalities in (10) imply that labor participation $L(w)$ is decreasing in $t(w) + b$, which can be interpreted as the participation tax for those earning w in any job. The following lemma characterizing a property of (10) will prove useful in what follows.

Lemma 4: The number of participants of wage rate w satisfies:⁴

$$\frac{1}{u'(b)} \frac{\partial L(w)}{\partial b} + \frac{1}{u'(z(w))} \frac{\partial L(w)}{\partial z(w)} = 0$$

Differentiating $L(w)$ in (9) with respect to the skill boundary $\hat{\theta}_j$ yields:

$$\frac{\partial L(w)}{\partial \hat{\theta}_j} = f(\hat{\theta}_j) \left(\delta_j(\theta_j^*, \hat{\theta}_j) q_j(w|\hat{\theta}_j) - \delta_{j+1}(\theta_{j+1}^*, \hat{\theta}_{j+1}) q_{j+1}(w|\hat{\theta}_j) \right) \quad (11)$$

It will be useful to rewrite (11) as follows, using $w_j^*(\theta) < w_{j+1}^*(\theta)$:

$$\frac{\partial L(w)}{\partial \hat{\theta}_j} = \begin{cases} 0 & \text{if } w < w_j^*(\theta) \\ f(\hat{\theta}_j) q_j(w|\hat{\theta}_j) & \text{if } w_j^*(\theta) \leq w < w_{j+1}^*(\theta) \\ f(\hat{\theta}_j) (q_j(w|\hat{\theta}_j) - q_{j+1}(w|\hat{\theta}_j)) & \text{if } w \geq w_{j+1}^*(\theta) \end{cases} \quad (12)$$

We turn next to the government problem and characterize the properties of the optimal non-linear tax for this economy.

5 The Government Problem

The government can observe the wages workers earn regardless of the job they have chosen, and can condition taxes on wages. We assume that it does not observe the quality of jobs, so does not condition taxes on jobs. If it could, it would enhance the ability to redistribute, since job choice is correlated with skills. Recall that the tax imposed on wage level w is $t(w) = w - z(w)$. Given the transfer b made to all non-participants, government revenue can be written, using $\int_{\underline{w}}^{\bar{w}} N(w) dw = 1$ by (8):

$$R = \int_{\underline{w}}^{\bar{w}} L(w) t(w) dw - \int_{\underline{w}}^{\bar{w}} (N(w) - L(w)) b dw = \int_{\underline{w}}^{\bar{w}} L(w) (t(w) + b) dw - b \quad (13)$$

As mentioned, the term $t(w) + b$ is the participation tax for workers earning wage w , regardless of the job they have chosen.

⁴ Proof: This will be the case if

$$\frac{1}{u'(b)} \frac{\partial \delta_j}{\partial b} + \frac{1}{u'(z(w))} \frac{\partial \delta_j}{\partial z(w)} = 0 \quad \text{or, equivalently} \quad \frac{1}{u'(b)} \frac{\partial \hat{\theta}_j^*}{\partial b} + \frac{1}{u'(z(w))} \frac{\partial \hat{\theta}_j^*}{\partial z(w)} = 0$$

which is satisfied using the comparative static properties of $\hat{\theta}_j^*$ in (1).

The objective function of government is taken to be a weighted utilitarian social welfare function in expected utilities of the different skill-types of workers:⁵

$$SW = \sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \beta(\theta) V_j(\theta) f(\theta) d\theta$$

where $\beta(\theta)$ are weights associated with individual skill levels, with $\int_{\underline{\theta}}^{\bar{\theta}} \beta(\theta) f(\theta) d\theta = 1$. We assume that these weights are non-increasing in θ . Note that this formulation for social welfare puts the same weight on all workers of a given skill level regardless of whether they choose to participate or not. This may be controversial since skill levels reflect both earning ability and the disutility of work, and some may prefer not to give social weight to differences in the disutility of work. The treatment of differences in preferences affects the case for social weights $\beta(\theta)$ that are non-increasing in θ .⁶

Using (2) for $V_j(\theta)$, the expected utility of a type- θ worker in job j , social welfare may be rewritten:

$$\begin{aligned} SW = u(b) \int_{\underline{w}}^{\bar{w}} \left(\sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \beta(\theta) (1 - \delta_j(\theta_j^*, \theta)) q_j(w|\theta) f(\theta) d\theta \right) dw + \int_{\underline{w}}^{\bar{w}} \bar{\beta}(w) L(w) u(z(w)) dw \\ - \int_{\underline{w}}^{\bar{w}} \left(\sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \beta(\theta) \delta_j(\theta_j^*, \theta) \phi_j(\theta) q_j(w|\theta) f(\theta) d\theta \right) dw \end{aligned} \quad (14)$$

where

$$\bar{\beta}(w) \equiv \frac{1}{L(w)} \sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \beta(\theta) \delta_j(\theta_j^*, \theta) q_j(w|\theta) f(\theta) d\theta$$

and

$$\int_{\underline{w}}^{\bar{w}} \left(\sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \beta(\theta) q_j(w|\theta) f(\theta) d\theta \right) dw = \sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \beta(\theta) \left(\int_{\underline{w}}^{\bar{w}} q_j(w|\theta) dw \right) f(\theta) d\theta = 1$$

Thus, $\bar{\beta}(w)$ is the average social weight given to participating workers who earn w , which includes workers of various θ -types. Given that $\beta(\theta)$ is decreasing except for the utilitarian case, $\bar{\beta}(w)$ will

⁵ Alternatively, one could assume that the social welfare function is concave in utilities. This complicates the analysis and interpretation without leading to different insights. The weights $\beta(\theta)$ can also be thought of as Pareto weights in the government's Pareto-optimizing problem.

⁶ Boadway et al (2000) and Choné and Laroque (2010) study the consequences for optimal income taxation of alternative social weights given to preferences for leisure. Fleurbaey and Maniquet (2011) investigate the difficulties involved in constructing social preferences that neither penalize nor reward differences in preferences, and analyze some consequences for optimal income taxation in a setting of full certainty.

be decreasing as well.⁷ In the utilitarian case, $\bar{\beta}(w) = 1$ for all w . Let L_0 be the total number of non-participants, and denote their social weight by β_0 . Then, we can rewrite (14) as:

$$SW = \beta_0 L_0 u(b) + \int_{\underline{w}}^{\bar{w}} \bar{\beta}(w) L(w) u(z(w)) dw - \int_{\underline{w}}^{\bar{w}} \left(\sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \beta(\theta) \delta_j(\theta_j^*, \theta) \phi_j(\theta) q_j(w|\theta) f(\theta) d\theta \right) dw \quad (15)$$

where $\beta_0 L_0 = 1 - \int_{\underline{w}}^{\bar{w}} \bar{\beta}(w) L(w) dw$, since the social welfare weights sum to unity. Note that by the envelope theorem, small changes in $\hat{\theta}_j$ and δ_j or θ_j^* do not affect social welfare. In the following, we can therefore ignore changes in β_0 , $\bar{\beta}(w)$ and $L(w)$ when evaluating the effect of differential policy changes on social welfare (though not the effect on worker behavior in the constraints).

The government anticipates both the job and participation choices of workers. The latter is incorporated into the participation cutoff functions $\theta_j^*(z(w), b)$ which solve (1), or equivalently $\delta_j(\theta_j^*, \theta)$. Job choice is captured by $\hat{\theta}_j$, the cutoff skill level between job j and job $j + 1$, which satisfies Lemma 3, which is endogenous and affected by government policy. Rather than using $V_{j+1}(\hat{\theta}_j) = V_j(\hat{\theta}_j)$ in Lemma 3 to obtain a solution for $\hat{\theta}_j$, we treat it as a constraint and allow the government to choose $\hat{\theta}_j$ as an artificial choice variable for all $j = 1, \dots, J - 1$. Given Assumption 3 that after-tax earnings $z(w)$ are non-decreasing, workers segment into jobs according to their skills as in Figure 2. The condition $V_{j+1}(\hat{\theta}_j) = V_j(\hat{\theta}_j)$ serves as an incentive constraint in the government's optimal tax problem, along with the participation constraint embedded in $\theta_j^*(z(w), b)$.

The Lagrangian expression for the government problem can then be written as follows, using the social welfare expression in (15) and the government budget constraint in (13):

$$\begin{aligned} \mathcal{L} = & \beta_0 L_0 u(b) + \int_{\underline{w}}^{\bar{w}} \bar{\beta}(w) L(w) u(z(w)) dw - \int_{\underline{w}}^{\bar{w}} \left(\sum_{j=1}^J \int_{\hat{\theta}_{j-1}}^{\hat{\theta}_j} \beta(\theta) \delta_j(\theta_j^*, \theta) \phi_j(\theta) q_j(w|\theta) f(\theta) d\theta \right) dw \\ & + \lambda \left(\int_{\underline{w}}^{\bar{w}} L(w) (t(w) + b) dw - b - R \right) + \sum_{j=1}^{J-1} \mu_j \left(V_{j+1}(\hat{\theta}_j) - V_j(\hat{\theta}_j) \right) \end{aligned}$$

⁷ Proof: Define the proportion of workers earning wage w who are type- θ as:

$$g(\theta|w) = \frac{\delta_j(\theta_j^*, \hat{\theta}_j) q_j(w|\theta) f(\theta)}{L(w)}, \quad \text{with } \int_{\theta} g(\theta|w) d\theta = 1$$

Then, $\bar{\beta}(w) = \int_{\theta} \beta(\theta) g(\theta|w) d\theta$. Using partial integration, we obtain:

$$\bar{\beta}(w) = \int_{\theta} \beta'(\theta) (1 - G(\theta|w)) d\theta - [\beta(\theta) (1 - G(\theta|w))]_{\underline{\theta}}^{\bar{\theta}} = \int_{\theta} \beta'(\theta) (1 - G(\theta|w)) d\theta + \beta(\underline{\theta})$$

Differentiation yields

$$\dot{\bar{\beta}}(w) = - \int_{\theta} \beta'(\theta) \frac{\partial}{\partial w} G(\theta|w) d\theta < 0$$

since $\beta'(\theta) < 0$ and $\partial G(\theta|w) / \partial w < 0$.

where λ is the shadow price of government revenues, and μ_j is the multiplier on the incentive constraint for $\hat{\theta}_j$.

The first-order conditions with respect to $z(w)$, b and $\hat{\theta}_j$ respectively are:

$$\bar{\beta}(w)L(w)u'(z(w)) - \lambda \left(L(w) - (t(w) + b) \frac{\partial L(w)}{\partial z(w)} \right) + \sum_{j=1}^{J-1} \mu_j \left(\frac{\partial V_{j+1}(\hat{\theta}_j)}{\partial z(w)} - \frac{\partial V_j(\hat{\theta}_j)}{\partial z(w)} \right) = 0 \quad (16)$$

$$\beta_0 L_0 u'(b) - \lambda \left(L_0 - \int_{\underline{w}}^{\bar{w}} (t(w) + b) \frac{\partial L(w)}{\partial b} dw \right) + \sum_{j=1}^{J-1} \mu_j \left(\frac{\partial V_{j+1}(\hat{\theta}_j)}{\partial b} - \frac{\partial V_j(\hat{\theta}_j)}{\partial b} \right) = 0 \quad (17)$$

$$\lambda \int_{\underline{w}}^{\bar{w}} (t(w) + b) \frac{\partial L(w)}{\partial \hat{\theta}_j} dw + \mu_j \left(V'_{j+1}(\hat{\theta}_j) - V'_j(\hat{\theta}_j) \right) = 0 \quad (18)$$

Using (2) to evaluate the derivatives of $V_{j+1}(\hat{\theta}_j)$ and $V_j(\hat{\theta}_j)$ with respect to $z(w)$ and b , first-order conditions (16) and (17) can be rewritten as follows:

$$\left(\bar{\beta}(w)L(w) + \sum_{j=1}^{J-1} \delta_j \mu_j \Delta q_j(w|\hat{\theta}_j) \right) u'(z(w)) - \lambda \left(L(w) - (t(w) + b) \frac{\partial L(w)}{\partial z(w)} \right) = 0 \quad (19)$$

$$\left(\beta_0 L_0 + \int_{\underline{w}}^{\bar{w}} \sum_{j=1}^{J-1} (1 - \delta_j) \mu_j \Delta q_j(w|\hat{\theta}_j) dw \right) u'(b) - \lambda \left(L_0 - \int_{\underline{w}}^{\bar{w}} (t(w) + b) \frac{\partial L(w)}{\partial b} dw \right) = 0 \quad (20)$$

where

$$\Delta q_j(w|\hat{\theta}_j) \equiv q_{j+1}(w|\hat{\theta}_j) - q_j(w|\hat{\theta}_j) \quad (21)$$

6 Interpretation

The first-order conditions to the government's problem can be used to characterize the properties of the optimal tax-transfer system, which here include the taxes on uncertain earnings $t(w)$ and the transfer to the non-participants b , and therefore participation tax rates $t(w) + b$. Using first-order conditions (19) and (20), we obtain:⁸

$$\left(\frac{\lambda}{u'(b)} - \beta_0 \right) L_0 + \int_{\underline{w}}^{\bar{w}} \left(\frac{\lambda}{u'(z(w))} - \bar{\beta}(w) \right) L(w) dw = 0 \quad (22)$$

⁸ Proof: Combining (19) and (20), we obtain:

$$\begin{aligned} & \lambda \int_{\underline{w}}^{\bar{w}} \left(\frac{1}{u'(b)} \frac{\partial L(w)}{\partial b} + \frac{1}{u'(z(w))} \frac{\partial L(w)}{\partial z(w)} \right) (t(w) + b) dw \\ &= \left(\frac{\lambda}{u'(b)} - \beta_0 \right) L_0 + \int_{\underline{w}}^{\bar{w}} \left(\frac{\lambda}{u'(z(w))} - \bar{\beta}(w) \right) L(w) dw - \sum_{j=1}^{J-1} \mu_j \int_{\underline{w}}^{\bar{w}} \Delta q_j dw \end{aligned}$$

By Lemma 4, the left-hand side is zero. As well, $\int_{\underline{w}}^{\bar{w}} \Delta q_j dw = \int_{\underline{w}}^{\bar{w}} q_{j+1}(w|\hat{\theta}_j) dw - \int_{\underline{w}}^{\bar{w}} q_j(w|\hat{\theta}_j) dw = 0$. Therefore, (22) follows.

Since $\beta_0 L_0 + \int_{\underline{w}}^{\bar{w}} \bar{\beta}(w) L(w) dw = 1$, (22) implies the following lemma.

Lemma 5:

$$\frac{\lambda}{u'(b)} L_0 + \int_{\underline{w}}^{\bar{w}} \frac{\lambda}{u'(z(w))} L(w) dw = \mathbb{E} \left[\frac{\lambda}{u'(z(w))} \right] = 1 \quad (23)$$

This says that the expected value of the marginal cost of public funds is zero. A similar result applies in the case where uncertain earnings occur in the intensive-margin model of labor supply (eq. (13) in Boadway and Sato 2012). Condition (23) is analogous to a similar expression in Saez (2002) stating that the weighed sum of social values of transfers to workers and non-participants equals unity. It is useful in interpreting the pattern of optimal income tax rates.

First-order condition (19) gives some insight into participation tax rates. Rearranging it, we obtain:

$$\frac{t(w) + b}{z(w) - b} = \frac{\tau(w)}{1 - \tau(w)} = \frac{1}{\eta(w)} \left(1 - \frac{u'(z(w))}{\lambda} (\bar{\beta}(w) + A(w)) \right) \quad (24)$$

where

$$\tau(w) \equiv \frac{t(w) + b}{w}, \quad \eta(w) \equiv \frac{\partial L(w)}{\partial z(w)} \frac{z(w) - b}{L(w)}, \quad \text{and} \quad A(w) \equiv \frac{\sum_{j=1}^{J-1} \delta_j \mu_j \Delta q_j}{L(w)}$$

That is, $\tau(w)$ is the participation tax rate as a proportion of earnings, $\eta(w)$ is the elasticity of participation with respect to the net consumption gain from participating, and $A(w)$ is a term involving μ_j , the shadow price of the job choice constraint. Eq. (24) is comparable to the standard extensive-margin optimal tax result under certainty of Diamond (1980) and Saez (2002), with the additional term $A(w)$ involving the influence of job choice. However, our participation elasticity $\eta(w)$ is somewhat more complicated since workers earning the wage rate w can include those with different skills in different jobs.

Next, we can obtain an expression for marginal earnings tax rates, $dt(w)/dw = \dot{t}(w)$, by differentiating (24) with respect to w . Routine manipulation yields the following:⁹

$$\frac{\dot{t}(w)}{1 - \dot{t}(w)} = \left(\eta(w) + G(w) \right)^{-1} \left(1 - \frac{u'(z(w))}{\lambda} \left(1 - \frac{z(w) - b}{z(w)} \sigma(w) \right) \cdot (\bar{\beta}(w) + A(w)) - G(w) \right) \quad (25)$$

where

$$G(w) = (t(w) + b) \dot{\eta}(w) + \frac{u'(z(w))}{\lambda} (z(w) - b) (\dot{\bar{\beta}}(w) + \dot{A}(w))$$

and $\sigma(w) \equiv -u''(z(w))z(w)/u'(z(w))$ is the coefficient of relative risk aversion.

To obtain some insight into the participation tax and marginal tax rates in (24) and (25), it is useful to consider first the case where there is no job choice, so the only decision workers make concerns participation.

⁹ The left-hand side of (25) can be interpreted as the marginal tax rate based on after-tax income, z . To see this, note that $z = w - t$, so $dz/dt = dw/dt - 1 = 1/\dot{t}(w) - 1 = (1 - \dot{t}(w))/\dot{t}(w)$, implying that $dt/dz = \dot{t}(w)/(1 - \dot{t}(w))$.

Participation Choice Only

If job choice is predetermined, workers still face uncertainty about their earnings. However, they make their participation choices after that uncertainty is resolved. In the absence of job choice, the constraint $V_{j+1}(\hat{\theta}_j) = V_j(\hat{\theta}_j)$ no longer applies, so $\mu_j = 0$ and therefore $A(w) = 0$. Eq. (24) then reduces to:

$$\frac{\tau(w)}{1 - \tau(w)} = \frac{1}{\eta(w)} \left(1 - \frac{\bar{\beta}(w)u'(z(w))}{\lambda} \right) = \frac{1 - g(w)}{\eta(w)} \quad (26)$$

where $g(w) \equiv \bar{\beta}(w)u'(z(w))/\lambda$ is the marginal social value in terms of government revenue of a transfer to workers of different skills earning wage rate w . Recall that $E[\beta] = E[\lambda/u'] = 1$, and both $\bar{\beta}(w)$ and $u'(z(w))$ are decreasing. This implies that the participation tax rate is negative for low values of w ($\tau(w) < 0$) and rises monotonically, becoming positive at some wage rate. This means that in the absence of job choice, the structure of participation tax rates in (26) parallels that in Diamond (1980) and Saez (2002). Not surprisingly, the absolute value of participation tax rates is inversely related to the elasticity of participation.

Additional intuition for participation tax rates can be obtained by rewriting (26) as follows:

$$\frac{\tau(w)}{1 - \tau(w)} = \frac{1}{\eta(w)} \left(1 - \frac{u'(z(w))}{\lambda} + \frac{(1 - \bar{\beta}(w))u'(z(w))}{\lambda} \right)$$

The term $1 - u'(z(w))/\lambda$ corresponds to an insurance, or consumption-smoothing term, and appears even in the case in which all workers are ex ante identical. As above, since $E[\lambda/u'] = 1$ and $u'(z(w))$ is decreasing in $z(w)$, this term is negative at low wage rates and positive for higher ones. The term $(1 - \bar{\beta}(w))u'(z(w))/\lambda$ represents additional equity considerations if the government's aversion to inequality is strictly positive. It reinforces the first term. In the utilitarian case, $\bar{\beta}(w) = 1$ for all w , so this term disappears.

Next, consider marginal income tax rates, $\dot{t}(w)$. With $A(w) = 0$, (25) simplifies to:

$$\frac{\dot{t}(w)}{1 - \dot{t}(w)} = \left(\eta(w) + G(w) \right)^{-1} \left(1 - \frac{u'(z(w))}{\lambda} \left(1 - \frac{z(w) - b}{z(w)} \sigma(w) \right) \bar{\beta}(w) - G(w) \right) \quad (27)$$

where

$$G(w) = (t(w) - b)\eta(w) + \frac{u'(z(w))}{\lambda} (z(w) - b)\dot{\bar{\beta}}(w)$$

From this, we can infer that the marginal tax rate $\dot{t}(w)$ will be higher the more negative is $\dot{\bar{\beta}}(w)$, that is, the greater the aversion to inequality there is in the social welfare function. As well, if the elasticity of participation, $\eta(w)$, is declining, the larger is $\dot{t}(w)$, and vice versa. These effects are intuitive.

Further insight can be obtained into the structure of marginal tax rates by considering the special case where the participation elasticity is constant, so $\eta(w) = \eta$, and the government objective

is utilitarian, so $\bar{\beta}(w) = 1$. Then, (27) becomes:

$$\frac{\dot{t}(w)}{1 - \dot{t}(w)} = \frac{1}{\eta} \left(1 - \frac{u'(z(w))}{\lambda} \left(1 - \frac{z(w) - b}{z(w)} \sigma(w) \right) \right) \quad (28)$$

A number of observations can be made about the pattern of marginal tax rates in (28). First, a higher coefficient of relative risk aversion $\sigma(w)$ leads to higher marginal tax rates, reflecting more consumption smoothing. Second, since $u'(z(w))/\lambda > 1$ for low incomes, $\dot{t}(w) < 0$ at low income levels if $\sigma(w)$ is small enough. Third, given that $u'(z(w))/\lambda < 1$ for high incomes, $\dot{t}(w)/(1 - \dot{t}(w))$ can be substantially higher than unity at high income levels, especially if η is low. Indeed, as shown below for the general case with job choice, given reasonable assumptions about the conditional wage distribution functions, the marginal tax rate at the top of the wage distribution will exceed 100 percent, being restricted only if $\dot{z}(w) \geq 0$ is imposed (Assumption 3). Since in this section, job choice has been suppressed, there is no reason to require that $\dot{z}(w) \geq 0$. Overall, the marginal tax rate tends to be increasing in w , possibly being negative at the bottom and exceeding 100 percent at the top.

These results have all suppressed the job choice decision. We next turn to the influence of job choice on optimal tax rates.

Job Choice

If workers are also allowed to choose their jobs, the constraint $V_{j+1}(\hat{\theta}_j) = V_j(\hat{\theta}_j)$ becomes relevant, and participation tax rates in (24) include the extra term $A(w) = \sum_{j=1}^{J-1} \delta_j \mu_j \Delta q_j / L(w)$. The sign of $A(w)$ depends on the signs of both μ_j and Δq_j , where $\Delta q_j = q_{j+1}(w|\hat{\theta}_j) - q_j(w|\hat{\theta}_j)$ by (21).

Consider first the sign of μ_j . It follows from the first-order condition on $\hat{\theta}_j$, (18). Using (12) for $\partial L(w) / \partial \hat{\theta}_j$, (18) can be written as:

$$\begin{aligned} \frac{\mu_j}{\lambda f(\hat{\theta}_j)} \left(V'_{j+1}(\hat{\theta}_j) - V'_j(\hat{\theta}_j) \right) &= \int_{w_{j+1}^*(\hat{\theta}_j)}^{\bar{w}} \left(q_{j+1}(w|\hat{\theta}_j) - q_j(w|\hat{\theta}_j) \right) (t(w) + b) dw \\ &\quad - \int_{w_j^*(\hat{\theta}_j)}^{w_{j+1}^*(\hat{\theta}_j)} q_j(w|\hat{\theta}_j) (t(w) + b) dw \end{aligned}$$

Applying partial integration to the first term on the right-hand side, this can be rewritten as:

$$\begin{aligned} \frac{\mu_j}{\lambda f(\hat{\theta}_j)} \left(V'_{j+1}(\hat{\theta}_j) - V'_j(\hat{\theta}_j) \right) &= \int_{w_{j+1}^*(\hat{\theta}_j)}^{\bar{w}} \left(Q_j(w|\hat{\theta}_j) - Q_{j+1}(w|\hat{\theta}_j) \right) \dot{t}(w) dw \\ &\quad + \left(Q_j(w_{j+1}^*|\hat{\theta}_j) - Q_{j+1}(w_{j+1}^*|\hat{\theta}_j) \right) (t(w_{j+1}^*) + b) - \int_{w_j^*(\hat{\theta}_j)}^{w_{j+1}^*(\hat{\theta}_j)} q_j(w|\hat{\theta}_j) (t(w) + b) dw \quad (29) \end{aligned}$$

From this, we can deduce the following lemma, using Assumption 2(ii), $Q_j(w|\theta) > Q_{j+1}(w|\theta)$, and Assumption 4, $V'_{j+1}(\theta) > V'_j(\theta)$.¹⁰

Lemma 6: Given Assumptions 2(ii) and 4, $\mu_j > 0$ if:

- (i) $\dot{t}(w) > 0$,
- (ii) $t(w_{j+1}^*) + b \geq 0$, and
- (iii) $Q_j(w_j^*|\hat{\theta}_j) > Q_{j+1}(w_{j+1}^*|\hat{\theta}_j)$

The property $\mu_j > 0$ can be given the following interpretation. It is the shadow price associated with the constraint $V_{j+1}(\hat{\theta}_j) = V_j(\hat{\theta}_j)$. If $\mu_j > 0$, social welfare would be increased if $V_j(\hat{\theta}_j)$ could be increased relative to $V_{j+1}(\hat{\theta}_j)$, that is, if there could be more redistribution between higher and lower skilled jobs. The fact that this constraint is binding implies that job choice is a constraint on redistribution.¹¹

An alternative interpretation of Lemma 6 is as follows. Since $V'_{j+1}(\hat{\theta}_j) > V'_j(\hat{\theta}_j)$ by Assumption 4, an increase in $\hat{\theta}_j$ from its optimal value when constrained by job choice would increase social welfare if $\mu_j > 0$. That is, starting in the constrained optimum, shifting workers from job j to job $j + 1$ is welfare-enhancing. Such a shift would enhance net tax revenue. Thus, there are too few workers in more skilled jobs if $\mu_j > 0$.

Note that positive marginal tax rates and participation tax rates are sufficient conditions for $\mu_j > 0$. Since these may well not apply at low income levels, it is possible that $\mu_j < 0$ for low-skilled jobs, so that workers should be shifted from higher to lower skilled jobs at the bottom end. To see this, consider (29) again. If $t(w) + b < 0$ at low wage levels, the last two terms on the righthand side take opposite values. The first term can in principle take either sign if $\dot{t}(w) < 0$ at the bottom.

¹⁰ Proof: If $\dot{t}(w) > 0$, from the last term in (29), we have:

$$\int_{w_j^*(\hat{\theta}_j)}^{w_{j+1}^*(\hat{\theta}_j)} q_j(w|\hat{\theta}_j)(t(w) + b)dw < \left(Q_j(w_{j+1}^*|\hat{\theta}_j) - Q_{j+1}(w_{j+1}^*|\hat{\theta}_j) \right) (t(w_{j+1}^*) + b)$$

Therefore, the last two terms in (29) satisfy

$$\left(Q_j(w_{j+1}^*|\hat{\theta}_j) - Q_{j+1}(w_{j+1}^*|\hat{\theta}_j) \right) (t(w_{j+1}^*) + b) - \int_{w_j^*(\hat{\theta}_j)}^{w_{j+1}^*(\hat{\theta}_j)} q_j(w|\hat{\theta}_j)(t(w) + b)dw >$$

$$\left(Q_j(w_j^*|\hat{\theta}_j) - Q_{j+1}(w_{j+1}^*|\hat{\theta}_j) \right) (t(w_{j+1}^*) + b) > 0$$

if $t(w_{j+1}^*) + b \geq 0$ and $Q_j(w_j^*|\hat{\theta}_j) > Q_{j+1}(w_{j+1}^*|\hat{\theta}_j)$

¹¹ To see this, let the constraint on job choice be written $V_{j+1}(\hat{\theta}_j) \geq V_j(\hat{\theta}_j) - \Delta_j$, so the relevant term in the Lagrangian expression becomes $\sum_j \mu_j (V_{j+1}(\hat{\theta}_j) - V_j(\hat{\theta}_j) + \Delta_j)$. Then, by the envelope theorem, social welfare will increase with Δ_j if $\mu_j > 0$.

Overall, we might expect the first term to be positive overall, since $\dot{t}(w) > 0$ except near the bottom. Then, there might still be a presumption that $\mu_j > 0$, although we cannot establish that unambiguously.

Next, consider the term $\Delta q_j = q_{j+1}(w|\hat{\theta}_j) - q_j(w|\hat{\theta}_j)$ in $A(w)$. Its value depends on the properties for the conditional wage distribution functions in jobs of different skills, $Q_j(w|\theta)$, which satisfy Assumption 2. Recall that $q_j(w|\theta)$ is the probability of a type- θ worker obtaining a wage rate w in job j . Since $E[q_j(w|\theta)] = 1$, we have that $E[\Delta q_j] = 0$, implying that Δq_j could be positive or negative depending on the properties of the skill distribution. Suppose that the density function $q_j(w|\theta)$ shifts right when one moves from job j to a higher skilled job $j + 1$. This is reasonable, given assumption 2(ii). Then, for a worker of given skill θ , $q_j(w|\theta) > q_{j+1}(w|\theta)$ for low enough wages, and vice versa for high enough wages. The term Δq_j involves comparisons for a worker whose skills are just marginal in jobs j and $j + 1$. We therefore expect that for worker with skills level $\hat{\theta}_j$, $q_{j+1}(w|\hat{\theta}_j) > q_j(w|\hat{\theta}_j)$ at relatively high wage levels, while $q_{j+1}(w|\hat{\theta}_j) < q_j(w|\hat{\theta}_j)$ when wages are relatively low.

Combining this with the above discussion of μ_j , we expect that $A(w)$ will tend to be positive at higher wage levels where both μ_j and Δq_j are positive. The implication is then that job choice will tend to reduce participation tax rates by (24). This is not surprising since adding a marginal of choice leads to a further source of distortion that causes lower tax rates to be chosen. At low wage levels, Δq_j tends to be negative, but the sign of μ_j is ambiguous. As discussed, there is some presumption that $\mu_j > 0$ even at the bottom, which would suggest that $A(w) < 0$ for low w . This would mitigate the negative participation tax at the bottom. We summarize these effects of job choice on participation tax rates as follows.

Proposition 1: Assuming that the conditions of Lemma 5 apply so $\mu_j > 0$, and that Δq_j is positive at high wage levels and negative at low wage levels, job choice will reduce participation tax rates at high wage levels and increase them at low wage levels.

Turn now to optimal marginal tax rates, $\dot{t}(w)$. These are given for the general case by (25). Given the complexity of this expressions, few qualitative results are apparent. One result is the following. As mentioned above, it is reasonable to expect that the marginal tax rate at the top exceeds 100 percent. The following proposition summarizes the likelihood that $\dot{t}(\bar{w}) > 1$ in the general case.

Proposition 2: Suppose job choice constraints are binding, and assume (i) $q_j(\bar{w}|\theta) = 0$ for all θ and $j < J$, (ii) $q_J(\bar{w}|\theta) = 0$ for all $\theta < \bar{\theta}$, and (iii) $q_J(\bar{w}|\bar{\theta}) > 0$. Then, $\dot{t}(\bar{w}) = 1$ if $z(w)$ is constrained by $\dot{z}(w) \geq 0$; otherwise, $\dot{t}(\bar{w}) > 1$.

The proof of Proposition 2 is as follows. Assumptions (i)–(iii) indicate that workers earning the top wage rate \bar{w} are only the highest skilled workers $\bar{\theta}$ in the highest skilled jobs J . These imply that at the top wage rate, $\eta(\bar{w}) = 0$ and $\dot{A}(\bar{w}) = 0$. Then, condition (25) characterizing the marginal tax rate reduces to the following after routine manipulation:

$$\dot{t}(\bar{w}) = 1 - \frac{z(\bar{w})}{\sigma(\bar{w})(\bar{\beta}(\bar{w}) + A(\bar{w}))} \dot{\bar{\beta}}(\bar{w}) > 1 \text{ if } \dot{\bar{\beta}}(\bar{w}) < 0$$

This says that in the absence of a non-negativity constraint on $\dot{z}(w)$, and assuming that the top wage is earned only by the most-skilled person, the marginal tax rate at the top will actually exceed 100 percent if the social welfare function exhibits strictly positive aversion to inequality ($\dot{\bar{\beta}}(\bar{w}) < 0$). Under a utilitarian social welfare function, the marginal tax rate at the top is 100 percent. The reason for these results, as the proof shows, is that at the top of the wage distribution, the elasticity of participation is zero, so the participation tax rate can increase at the top. Of course, a 100 percent marginal income tax at the top does not imply participation taxes are 100 percent, only that the change in participation taxes exceeds the increase in wages.

At the bottom, we have seen that in the absence of job choice, $\dot{t}(\underline{w}) < 0$ if $\sigma(\underline{w})$ and $-\dot{\bar{\beta}}(\underline{w})$ are not too large. This is reinforced if $A(w) > 0$, and vice versa.

To obtain further results and to focus on the effects of job choice, it is useful to suppress the effects of variable $\eta(w)$ and $\bar{\beta}(w)$ by assuming as above that $\eta(w) = \eta$ and $\bar{\beta}(w) = 1$. Assume also for simplicity that the coefficient of relative risk aversion is constant, so $\sigma(w) = \sigma$. Then, (25) may be written as follows:

$$\begin{aligned} & \frac{\dot{t}(w)}{1 - \dot{t}(w)} \\ &= \left(\eta + \frac{u'(\cdot)}{\lambda} (z(w) - b) \dot{A}(w) \right)^{-1} \left(1 - \frac{u'(\cdot)}{\lambda} \left(1 - \frac{z(w) - b}{z(w)} \sigma \right) (1 + A(w)) - \frac{u'(\cdot)}{\lambda} (z(w) - b) \dot{A}(w) \right) \end{aligned}$$

By the above reasoning, we expect that $A(w) > 0$, which tends to reduce marginal tax rates. However, the sign of $\dot{A}(w)$ is ambiguous. If the job choice constraint becomes tighter for higher skilled jobs, $\dot{A}(w)$ will tend to be positive, which would further reduce marginal tax rates.

Summary of Results

As we have seen, there is some ambiguity about the qualitative results on participation and marginal tax rates. This is largely due to uncertainty about the way in which the job choice constraint affects tax policy. We have argued that under reasonable assumptions, $A(w) > 0$, at least for wages about the lowest ones. Assuming this to be the case, the following tentative results apply.

Participation tax rates

- a) $\tau(w)$ increases in w , with $\tau(\underline{w}) < 0$ and $\tau(\bar{w}) > 0$,

- b) $\tau(w)$ is lower the higher is $\eta(w)$,
- c) $\tau(w)$ increases more with w the greater is the aversion to inequality, $-\dot{\bar{\beta}}(w)$, and
- d) Job choice tends to reduce $\tau(w)$ for high w , and possibly reduce $\tau(w)$ at the bottom.

Marginal tax rates

- a) $\dot{t}(w)$ tends to be increasing in w , with $\dot{t}(\underline{w}) < 0$ and $\dot{t}(\bar{w}) \geq 1$,
- b) $\dot{t}(w)$ is increasing in $\sigma(w)$ (relative risk aversion),
- c) $\dot{t}(w)$ is decreasing in $\eta(w)$,
- d) $\dot{t}(w)$ is increasing in $-\dot{\bar{\beta}}(w)$ (aversion to inequality), and
- e) Job choice tends to reduce $\dot{t}(w)$ if the job constraint is tighter for higher-skill jobs.

7 Concluding Comments

We have explored optimal income taxation in an extensive-margin setting in which earnings in any job are uncertain, and vary according to both worker skills and the skill level of each job. Workers must select a job before they know the wage. Once the wage is revealed, they decide either to participate or to be voluntarily unemployed. By invoking a single-crossing property, workers segment into jobs suited for different skills according to their own skill levels, with higher skilled workers choosing jobs requiring higher skills.

In the absence of wage insurance, the optimal income tax both provides implicit insurance and redistributes imperfectly from higher to lower skilled workers. The progressivity of the tax, characterized by the pattern of participation tax rates, and the structure of marginal tax rates are influenced by both participation and job choice decisions. Given the uncertainty of earnings, the structure of taxes is more complex than in standard extensive-margin models, though it takes a similar form. It is influenced by a number of factors, including aversion to risk, the conditional distribution of wages, and the responsiveness of participation and job choices to taxes. What makes the government's policy choice particularly challenging is the fact that workers of different skills end up earnings the same wage, possibly in different jobs.

The analysis could be further refined by allowing taxes to be conditional not only on earnings but also on job choice, especially since workers segment themselves into jobs according to their skills. As well, we could explore more precise assumptions about the distribution of wages. For example, jobs requiring higher skills might also be riskier in the sense that the variance of wages for a given job are greater. Finally, the use of wage insurance, as proposed by LaLonde (2007), could be explored as an adjunct to nonlinear income taxation. Alternatively, riskiness may differ with the skill of the worker rather than the skill required in the job. More generally, there may be some jobs that are free of risk. These could be chosen at the outset, or they could be options for workers who choose not to participate in risky jobs. These remain open research questions.

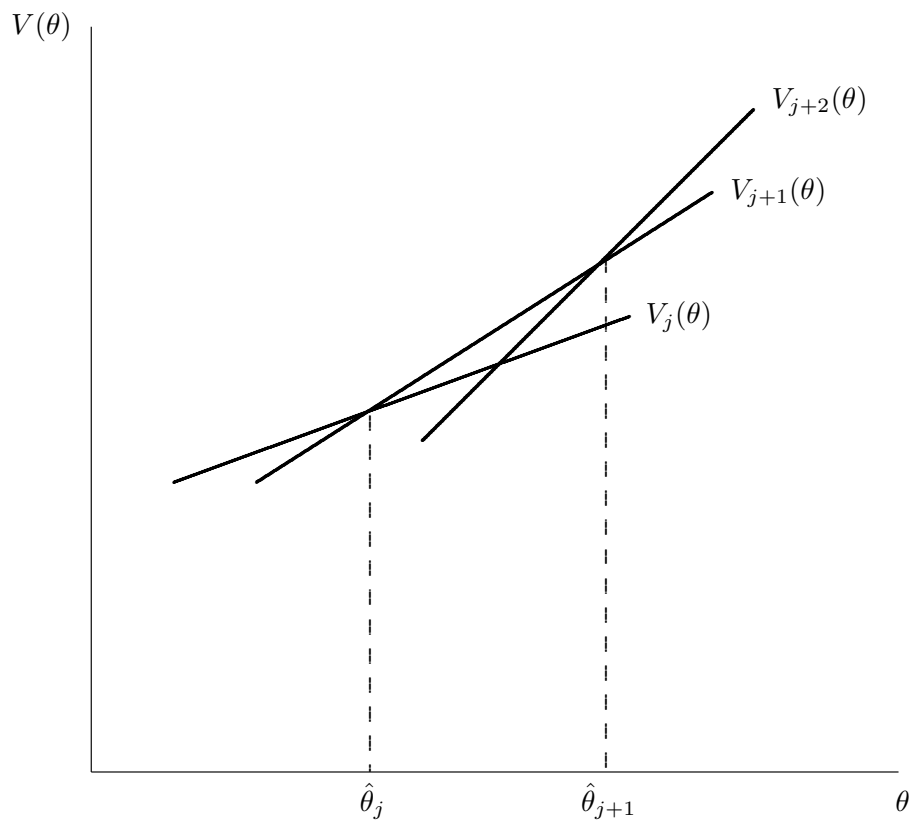


Figure 1

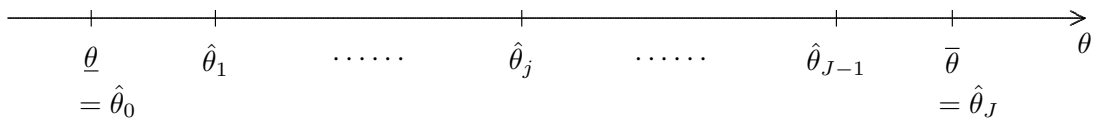


Figure 2

References

- Anderberg, Dan (2009), ‘Optimal Policy and the Risk Properties of Human Capital Reconsidered’, *Journal of Public Economics* **93**, 1017–26.
- Atkinson, Anthony B. and Joseph E. Stiglitz (1976), ‘The Design of Tax Structure: Direct vs. Indirect Taxation’, *Journal of Public Economics* **6**, 55–75.
- Bastani, Spencer, Sören Blomquist and Luca Micheletto (2010), ‘The Welfare Gains of Age Related Optimal Income Taxation’, CESifo Working Paper 3225.
- Boadway, Robin, Maurice Marchand, Pierre Pestieau and Maria del Mar Racionero (2000), ‘Optimal Redistribution with Heterogeneous Preferences for Leisure’, *Journal of Public Economic Theory* **4**, 475–98.
- Boadway, Robin and Motohiro Sato (2011), ‘Optimal Income Taxation with Uncertain Earnings: A Synthesis’, CESifo Working Paper No. 3654.
- Choné, Philippe and Guy Laroque (2010), ‘Negative Marginal Tax Rates and Heterogeneity’, *American Economic Review* **100**, 2532–47.
- Christiansen, Vidar (1984), ‘Which Commodity Taxes Should Supplement the Income Tax?’, *Journal of Public Economics* **24**, 195–220.
- Christiansen, Vidar (2012), ‘Optimal Participation Taxes’, presented at the CESifo Area Conference on Public Sector Economics, April 12–14, 2012.
- Conesa, Juan Carlos, Sagiri Kitao and Dirk Krueger (2009), ‘Taxing Capital? Not a Bad Idea after All’, *American Economic Review* **99**, 25–48.
- Cremer, Helmuth and Firouz Gahvari (1995), ‘Uncertainty, Optimal Taxation and the Direct Versus Indirect Tax Controversy’, *Economic Journal* **105**, 1165–79.
- Cremer, Helmuth and Firouz Gahvari (1999), ‘Uncertainty, Commitment and Optimal Taxation’, *Journal of Public Economic Theory* **1**, 51–70.
- Diamond, Peter A. (1980), ‘Income Taxation with Fixed Hours of Work’, *Journal of Public Economics* **13**, 101–10.
- Diamond, Peter A., L. Jay Helms and James A. Mirrlees (1980), ‘Optimal Taxation in a Stochastic Economy: A Cobb-Douglas Example’, *Journal of Public Economics* **14**, 1–29.
- Ebert, Udo (1992), ‘A Reexamination of the Optimal Nonlinear Income Tax’, *Journal of Public Economics* **49**, 47–73.

- Fleurbaey, Marc and François Maniquet (2011), *A Theory of Fairness and Social Welfare* (New York: Cambridge University Press).
- Golosov, Mikhael, Aleh Tsyvinski and Iván Werning (2007), ‘New Dynamic Public Finance: A Users Guide’, *NBER Macroeconomics Annual 2006*, 317–63.
- Guesnerie, Roger (1995), *A Contribution to the Pure Theory of Taxation* (Cambridge: Cambridge University Press).
- Henriet, Dominique, Patrick A. Pintus and Alain Trannoy (2012), ‘Is the Flat Tax Optimal under Income Risk?’, paper presented at the Taxation Theory Conference, Vanderbilt University, May 18–20, 2012.
- Hungerbühler, Mathias, Etienne Lehmann, Alexis Parmentier and Bruno Van der Linden (2006), ‘Optimal Redistributive Taxation in a Search Equilibrium Model’, *Review of Economic Studies* **73**, 743–67.
- Kaplow, Louis (2008), *The Theory of Taxation and Public Economics*, (Princeton: Princeton University Press).
- Kocherlakota, Narayana R. (2010), *The New Dynamic Public Finance*, Toulouse Lectures in Economics (Princeton: Princeton University Press).
- LaLonde, Robert J. (2007), *The Case for Wage Insurance*, Council on Foreign Relations Special Report No. 30 (New York: Council on Foreign Relations).
- Low, Hamish and Daniel Maldoom (2004), ‘Optimal Taxation, Prudence and Risk-sharing’, *Journal of Public Economics* **88**, 443–64.
- Mirrlees, James A. (1971), ‘An Exploration in the Theory of Optimum Income Taxation’, *Review of Economic Studies* **38**, 175–208.
- Saez, Emmanuel (2001), ‘Using Elasticities to Derive Optimal Income Tax Rates’, *Review of Economic Studies* **68**, 205–29.
- Saez, Emmanuel (2002), ‘Optimal Income Transfer Programs: Intensive vs Extensive Labor Supply Responses’, *Quarterly Journal of Economics* **117**, 1039–73.
- Tuomala, Matti (1984), ‘Optimal Degree of Progressivity under Income Uncertainty’, *Scandinavian Journal of Economics* **86**, 184–93.
- Weinzierl, Matthew C. (2011), ‘The Surprising Power of Age-Dependent Taxes’, *Review of Economic Studies* **78**, 1490–518.