



Recent Advances in Optimal Income Taxation*

ROBIN BOADWAY

Queen's University, Canada and CESifo

Abstract

This paper reviews recent contributions to the theory of optimal taxation, particularly those that are policy-relevant. These include refinements of the Mirrlees optimal income tax model, optimal redistribution when workers make labour decisions along the extensive margin, generalizations of the Atkinson-Stiglitz and Deaton Theorems on uniform commodity taxation, the implications of involuntary unemployment for redistribution and unemployment insurance, dynamic optimal taxation and the taxation of capital income, non-tax instruments for redistribution, and problems arising if preferences are heterogeneous.

Keywords: Optimal Taxation, Intensive Margin, Extensive Margin.

JEL classification: H21, H24, H25.

1. Introduction

There has been a resurgence of interest in optimal taxation, an area that was very active in 1970s and early 1980s following the key contributions of Diamond and Mirrlees (1971) and Mirrlees (1971). Some of the impetus has come from fundamental changes in tax systems around the world, including the widespread adoption of value-added taxation, the flattening of progressive income tax rate schedules, the fall in capital income tax rates, the adoption of dual income tax systems in some countries, the increasing sheltering of asset income, the demise of wealth transfer taxation, and the change in the form of business income tax systems to emphasize rent taxation as opposed to the taxation of corporate equity income. In some cases, these changes have been informed by the evolving literature on optimal taxation, and in others research has been stimulated by reforms taking place around the world.

* *Acknowledgments:* The ideas in this paper have benefited from discussions and collaborations with Katherine Cuff, Laurence Jacquet, Michael Keen, Nicolas Marceau, Pierre Pestieau, Motohiro Sato, Jean-François Tremblay, and the late Maurice Marchand. I have drawn on material in Boadway (2012), which treats some of the themes of this paper in much greater length. I appreciate very much the opportunity that the Editor, Javier Suárez-Pandiello, has given me to participate in this special 200th issue of *Hacienda Pública Española/Review of Public Economics*, honouring its contribution to public economics research in Spain.

There have been occasional major tax policy reports that have drawn on the optimal tax literature. In the Anglo-Saxon world, these include President's Advisory Panel on Federal Tax Reform (2005) in the USA, the Henry Report (Australian Treasury 2010) in Australia, and the Mirrlees Review (2011) in the UK. These documents illustrate the policy relevance of the main results in the optimal tax literature.

The classical results that have stood the test of time include the Corlett and Hague (1953) Theorem on differential commodity taxation; the Mirrlees (1971) formulation of the optimal progressivity of the income tax; the Atkinson-Stiglitz (1978) Theorem on uniform commodity taxation; the Deaton (1979) conditions, which extended the Atkinson-Stiglitz Theorem to linear income tax systems; the Diamond and Mirrlees (1971) production efficiency theorem, which forms one rationale for the value-added tax; and the case for cash-flow business taxation, which was proposed by the Meade Report (1978).

The purpose of this paper is to recount some of the major advances in optimal tax theory in recent years, especially those that are policy-relevant. Some of these include a refinement of the Mirrlees approach to optimal income taxation to take account of observed responses of taxable income to changes in the marginal tax rate and to other information available to the government, such as age, gender and location of residence. A major innovation that has been widely exploited has been the application of optimal income tax theory to the case where workers make discrete labour market choices, such as participation and job choice. The case for uniform commodity taxation has been given additional support by tax reform results that extend the Atkinson-Stiglitz Theorem and the Deaton conditions to situations in which income taxes are not optimally chosen. Endogenous involuntary unemployment introduces new considerations that affect both optimal redistribution and the design of unemployment insurance. The recent literature on dynamic optimal income taxation (the so-called new dynamic public finance) has shed new light on the case for capital income taxation, especially when there are capital market imperfections, such as liquidity constraints and absence of wage rate insurance. Finally, there is an emerging literature that seeks to cope with the normative problems that arise when households differ in preferences, so interpersonal comparisons are particularly difficult. We discuss all of these problems in the following sections.

2. Optimal Income Taxation and the Intensive Margin

Our benchmark is the classic statement of the optimal income tax problem in Mirrlees (1971). There is a population of workers who differ in their skills, denoted w , where w is distributed according to $F(w)$ for $0 \leq \underline{w} \leq w \leq \bar{w} \leq \infty$. The total population is normalized to unity so the density $f(w)$ can be thought of as the proportion of the population with skill w . Production is linear, with w representing the output produced by a type- w worker per unit of time. Because of linearity, there are no pure profits. All workers have the same concave utility function in consumption and labour, $U(c, \ell)$ or equivalently, $v(c, y, w)$ where $y = w\ell$ is labour income before taxes.

The benevolent government, or social planner, imposes a nonlinear income tax function $T(y)$ to maximize an additive social welfare function $W = \int_{\underline{w}}^{\bar{w}} \omega(v(\cdot)) dF(w)$, where social utility $\omega(v)$ is increasing and concave. The degree of concavity, indicated by the elasticity of social utility, reflects the aversion to inequality of the planner, and can range from zero (utilitarianism) to infinity (maximin). The government must raise some exogenous amount of revenue, $R \geq 0$, so faces a budget constraint $\int_{\underline{w}}^{\bar{w}} T(w) F(w) \geq R$, which must always be binding. Equivalently, letting $(c(w), y(w))$ be the consumption-income bundle of a type- w worker, the budget constraint can be written $\int_{\underline{w}}^{\bar{w}} (y(w) - c(w)) dF(w) = R$, since $T(w) = y(w) - c(w)$.

If the government could observe workers' types, it could set a tax function based on skills, $T(w)$, to maximize social welfare subject to the budget constraint. However, the government is assumed to be able to observe only incomes $y(w)$, and neither w or labour supply (effort) ℓ . Redistributive taxes based on income face the problem that attempts to redistribute from type w to, say, type w' may induce the former to pretend to be the latter by mimicking their income, $y = y'$. This would cause redistribution to break down, so to preclude that, incentive constraints must be added. Let $u(w)$ be the utility achieved by a type- w worker. The incentive constraint stipulates that $u(w) = v(c(w), y(w), w) \geq v(c(w'), y(w'), w)$ for all w, w' , which is equivalent to requiring $u(w) = \text{Max}_{w'} v(c(w'), y(w'), w)$. The envelope theorem evaluated at $w = w'$ at yields the first-order incentive constraint, $\dot{u}(w) = v_w(c(w), y(w), w)$. This is only a necessary condition for incentive compatibility. Sufficiency is ensured by a second-order incentive constraint that takes the form $\dot{y}(w) \geq 0$ (e.g., Myles 1995). We assume in what follows that the second-order constraint is not binding. Otherwise there will be bunching of workers of different types at common income levels (Ebert 1992). When the incentive constraint is satisfied, the utility of a type- w worker can be written $u(w) = v(c(w), y(w), w)$. This means $u(w)$, $y(w)$ and $c(w)$ are not independent, and in what follows we invert $v(\cdot)$ to obtain $c(w) = c(u(w), y(w))$ where by comparative statics, $c(\cdot)/u(w) = 1/v_c(\cdot)$ and $c(\cdot)/y(w) = -v_y(\cdot)/v_y(\cdot)$.

The government maximizes $W = \int_{\underline{w}}^{\bar{w}} \omega(u(w)) dF(w)$ subject to $\int_{\underline{w}}^{\bar{w}} (y(w) - c(u(w), y(w))) dF(w) = R$, and $u(w) = v_w(y(w), c(u(w), y(w)), w)$. The control variable is $y(w)$, and $u(w)$ is a state variable. The first-order conditions for this problem constitute a planning solution for the variables $y(w)$, $u(w)$ and therefore $c(w)$. This solution can be implemented in a decentralized setting by a nonlinear tax function $T(w)$ where worker utility-maximizing behaviour satisfies $c(\cdot)/y(w) = 1 - T'(y(w))$. Suppose utility is additively separable in c and ℓ or y , so $v(c(w), y(w), w) = v(c(w)) - h(y(w)/w)$. Then, using above properties of $c(\cdot)$ and the behavioural relationship involving $T'(y(w))$, the first-order conditions from the planner's problem can be simplified to the following expression for marginal tax rates, where we have suppressed function arguments:

$$\frac{T'}{1 - T'} = v' \int_{\underline{w}}^{\bar{w}} \left(\frac{1}{v'} - \frac{\omega'}{\lambda} \right) dF(w) \cdot \frac{1 + \frac{\ell h''}{h'}}{w f} \quad (1)$$

From (1), we can infer that, except in the maximin case, $T'(y(\bar{w})) = T'(y(\underline{w})) = 0$ (assuming there is no bunching at the bottom), and $0 < T'(y(w)) < 1$ in the interior. Otherwise, the

pattern of marginal tax rates is ambiguous. For example, while the integral term on the righthand side increases in w for equity reasons, the factor wf may be increasing. Simulations suggest that the pattern of marginal tax rates tend to rise only modestly in the interior, falling rapidly to zero only at the ends (Tuomala 1990). However, different patterns emerge under different assumptions. In the maximin case, under reasonable assumptions marginal tax rates will be positive at the bottom and declining throughout the skill distribution, leading to an increasing but strictly concave tax function and hump-shaped average tax rates $T(y(w))/y(w)$ (Boadway and Jacquet 2008). In the case where the skill distribution is unbounded, the pattern of marginal tax rates can be U-shaped above the mode if the skill distribution is Pareto, so $wf(w)$ is constant, and utility is quasilinear in consumption (Diamond 1998).

The special case where preferences are quasilinear in consumption, so $v(c(w), y(w), w) = c(w) - h(y(w)/w)$, yields further insights. Since $v'(c) = 1$ and $\ell h''/h' = 1/\epsilon$, where ϵ is the elasticity of labor supply, (1) can be rewritten:

$$\frac{T'}{1-T'} = \frac{1+\epsilon}{\epsilon} \cdot \frac{\int_w^{\bar{w}} \left(1 - \frac{\omega'}{\lambda}\right) dF(w)}{1-F(w)} \cdot \frac{1-F(w)}{wf(w)} = \left(\frac{1+\epsilon}{\epsilon}\right) (1+g(w)) \cdot \frac{1-F(w)}{wf(w)} \quad (2)$$

where $g(w) = \int_w^{\bar{w}} (1 - \omega'/\lambda) dF/(1-F(w))$ is the average social value in terms of government revenue of giving one euro to all persons with skills above w . The first term in (2) is an efficiency effect: a higher elasticity of labour supply will reduce T' . The last two terms reflect equity considerations. An increase in the marginal tax rate on type- w workers, holding all other T' constant, will cause tax liabilities for all workers with wages higher than w to go up, of whom there are $1-F(w)$. Because preferences are quasilinear, labour supply will not change. The additional taxes paid will have a net social value of $1-g(w)$ for each of these higher skilled workers. The term $1-g(w)$ will increase with w , while $1-F(w)$ will fall, leading to ambiguity in the pattern of marginal tax rates that will be moderated depending on how ϵ varies with w and on $wf(w)$ which depends on the skill distribution.

Saez (2001) obtained further policy-relevant insight into the shape of the optimal tax function by transforming the skill distribution into an income distribution, and deriving the optimal tax structure using a perturbation method. Let $H(y)$ be the distribution of income, which is endogenous, with density $h(y)$. The labour supply elasticity, ϵ , is also the elasticity of income y with respect to $1-T'(y)$ ¹. Suppose the optimal income tax is in place. A small perturbation in $T'(y)$ at any y will have zero net effect on social welfare. An increase in T' by dT' over the interval $y+dy$ will increase taxes by all those with $y' > y$ by $dT' dy$ leading to a change in government revenue of $(1-H(y))dT'dy$. Net social welfare will increase on this account by $ds = (1-g(y))(1-H(y))dT'dy$, where, as above, $g(y)$ is the average change in social welfare per person from giving an additional euro to all persons with $y' > y$. In addition, for those in the interval $y+dy$ whose marginal tax rate has increased, labour supply and income will fall, and given ϵ , their tax payments will go down by $dR = -\epsilon y dT'/(1-T')$. (Their

utility is unchanged since they are on the margin.) Since the tax structure is optimal to begin with, $dS + dR = 0$, which leads to:

$$\frac{T'(y)}{1-T'(y)} = \frac{1}{\epsilon} \cdot (1+g(y)) \cdot \frac{1-H(y)}{yh(y)} \quad (3)$$

This has a similar interpretation to (2), but now the components of the righthand side are based on income, so are observable. Following Feldstein (1999), the elasticity of earnings, ϵ , can be given a broader and more realistic interpretation. Earnings might vary not just because of changes in labour supply but also because of tax avoidance—e.g., converting labour income into capital income or moving to less risky, but lower paying jobs— or evasion (mis-reporting earnings)². To the extent that evasion and avoidance increase the elasticity of earnings, ϵ , optimal marginal tax rates will fall (except to the extent that the avoidance or evasion represents simply a transfer of tax liabilities to another agent, as Chetty (2009) points out). Empirical estimates suggest that the elasticity of reported earnings is relatively high at higher income levels (Gruber and Saez 2002). This argues in favour of lower marginal tax rates at higher skill levels, something that recent policy has accommodated, except to the extent that avoidance can be reduced by eliminating special tax preferences or evasion can be deterred.

A refinement of the standard optimal income tax model concerns the use of some observable signal, or tag, that is correlated with ability. Following Akerlof (1978) and Parsons (1996), this tag allows one to divide the population between two groups, the tagged and the untagged. As long as the distribution of skills between the tagged and untagged groups differs, tagging can generally improve social welfare. A separate nonlinear income tax system can be applied to each group, and a lump-sum transfer can be made between groups (from the group with the highest average skill level to the other group). Immonen, Kanbur, Keen and Tuomala (1998) studied tagging in the Mirrlees (1971) model using simulation techniques. They found that the marginal income tax rate declined with skills in the group with higher average skills, and increased in the other group. The gain in welfare from tagging was relatively high. Boadway and Pestieau (2006) obtained analytical results for the case where preferences were quasilinear in consumption and the number of skills was discrete. Using a maximin social welfare function, they found that the tax would be more progressive in the group with higher average skill levels. Some authors have suggested tagging according to demographic characteristics, such as gender (Alesina, Ichino and Karabarbounis 2007; Cremer, Gahvari and Lozachmeur 2010) or height (Mankiw and Weinzierl 2010). These kinds of variables raise issues of social acceptability of tagging, and others have raised concerns about stigmatization (Jacquet and Van der Linden 2006). A final application of tagging involves the decentralization of nonlinear income taxation to subnational governments in a federation. Gordon and Berry Cullen (2012) argue that such decentralization is desired not only because national social welfare can be improved by tagging accompanied by interregional transfers as above, but also because vertical fiscal externalities are eliminated when there is no federal income tax. They study the effects of decentralizing nonlinear income taxes to the state level of government in the USA, and allow for costly migration among states.

3. Differential Commodity Taxation

The nonlinear income tax seems to be a blunt instrument for redistribution since it implicitly taxes all spending the same whether it is for necessity goods or luxuries. It is natural to wonder whether redistribution, and therefore social welfare, might not be improved by imposing higher taxes on the consumption of goods that have higher income elasticities of demand. Atkinson and Stiglitz (1976) proved a remarkable theorem. They showed that, even if the government could impose separate nonlinear taxes on the consumption of different goods, it would be optimal not to use them as long as preferences are weakly separable in goods and leisure or labour (i.e., utility in goods x_i and labour can be written $u(f(x_1, \dots, x_n), \ell)$) and the optimal nonlinear income tax is in place. This is the case even though weak separability does not rule out goods having very different income elasticities of demand. The intuition is rather subtle. The factors that prevent nonlinear income taxation from achieving a first-best outcome are the incentive constraints that preclude higher-skilled persons from mimicking the income of lower-skilled ones. If preferences are weakly separable, differential taxes on consumption cannot relax the incentive constraints but will introduce distortions in good purchases. That is because workers of different w who earn the same income, obtain the same disposable income but supply different amount of labour. Because of weak separability, they will consume the same bundles of goods, so imposing differential taxes cannot penalize the mimickers. The Atkinson-Stiglitz Theorem was extended by Deaton (1979) to the case where the government can only use linear progressive taxes. In this case, as long as the linear progressive tax is set optimally, differential commodity taxes cannot improve social welfare if preferences are weakly separable and if the demand for goods is quasi-homothetic. Again, this does not preclude the possibility of goods having very different income elasticities of demand (as the case of Stone-Geary preferences confirms).

Recently, both the Atkinson-Stiglitz Theorem and the Deaton Theorem have been generalized in important ways. Laroque (2005a) and Kaplow (2006), following an approach suggested by Konishi (1995), showed the following. Start with a set of differential indirect commodity taxes and any arbitrary nonlinear income tax, not necessary optimal, and assume that preferences are weakly separable. There exists a Pareto-improving tax reform that eliminates differential commodity taxes and adjusts the income tax such that both the government revenue and incentive constraints are satisfied. So, as long as there are no restrictions on changing the nonlinear income tax, there is a case for eliminating differential commodity taxes (or adopting a uniform VAT)³. Similarly, the Deaton Theorem has been generalized by Hellwig (2009). Suppose preferences satisfy weak separability and quasi-homotheticity in goods, and start with a differential commodity tax system and an arbitrary linear progressive tax. He proves that there is a Pareto-improving tax reform that eliminates differential commodity taxes and adjusts the linear progressive income tax such that the government budget constraint is satisfied. (Incentive constraints are always satisfied for linear progressive taxes.)

These results cast doubt on the preferential taxation of necessity goods, which is a common feature of VAT systems. If something precludes the government from making the in-

come tax system as progressive as desired, such as evasion or avoidance, a case can be made for treating necessities preferentially, as Boadway and Pestieau (2011) argue. However, they also show paradoxically that if low-income workers are constrained from purchasing some luxury goods, necessities should be taxed at a higher rate than luxuries even if the income tax is optimized.

If preferences are not weakly separable, differential commodity taxes can improve social welfare in principle, although differentiation introduces administrative complications that detract significantly. Even here, the preferential taxation of necessities is not prescribed. Theory suggests that higher tax rates should be imposed on goods that are relatively more complementary with leisure. This result is reminiscent of the famous Corlett-Hague (1953) Theorem, but based on different reasoning. Imposing a higher tax on leisure-complements relaxes the incentive constraints and allows the nonlinear income tax to be more progressive (Edwards, Keen and Tuomala 1994). A worker who is mimicking the income of a lower-wage worker will obtain the same disposable income but have more leisure so will consume more of those goods that complement leisure. Imposing higher taxes on such goods makes it more difficult to mimic so relaxes the incentive constraint. Goods that are more complementary with leisure are not necessarily luxury goods, as borne out by empirical analysis performed for the Mirrlees Review (2011), reported in Blundell (2011).

Other considerations also support differential goods taxation, though again not necessarily preferential treatment of necessities. Examples include differences in unobserved endowments of some goods (Cremer, Pestieau and Rochet 2001); differences in preferences (Saez 2002a; Marchand, Pestieau and Racionero 2003; Blomquist and Christiansen 2008); differences in the need for particular goods differ (Boadway and Pestieau 2003). Boadway and Gahvari (2006) show that a higher tax should be levied on goods whose consumption is more time-intensive as long as consumption time is a substitute for labor. Finally, Cremer and Gahvari (1995) show that, if consumer durables must be purchased before wage rates are known, a case can be made for preferential tax treatment of durables. This makes it more difficult for those who turn out to have high wages *ex post* to mimic lower-wage persons. We return to this case below.

4. Extensive-Margin Labour Supply

So far we have assumed that any changes in labour supply involved hours of work. There may be rigidities that constrain hours of work. Diamond (1980) considered the extreme case where hours of work and earnings in a given job were fixed, and the only choice was to participate in work or not. Participation was then determined by heterogeneity in the disutility of work, or the benefit of leisure or home production. The result had important consequences for the structure of the income tax system. Saez (2002b) generalized Diamond's analysis to allow workers to choose not only whether to participate, but also in which job. Consider each in turn.

4.1. Participation choice only

Suppose there are n_i workers of skill type i , $i = 1, \dots, s$, as well as $n_0 = 0$ workers unable to work. There are also s types of jobs, one for each skill level. Working in job i involves a fixed number of hours and generates fixed earnings of y_i such that $y_i > y_{i-1}$ for all i . The utility if working is simply $c_i = y_i - t_i$, where t_i is the tax on earnings y_i , which the government can observe. Those not working obtain utility of $c_0 + \tilde{m}_i = -t_0 + \tilde{m}_i$, where $-t_0$ is the transfer to all non-participants (since the government cannot distinguish their type) and $\tilde{m}_i$ is the value of leisure. The latter is distributed by $\Gamma_i(m_i)$ which can vary by types. For the marginal type $-i$ participant, $y_i - t_i = t_0 + \hat{m}_i$, so the number of type participants is $n_i \Gamma_i(\hat{m}_i) = n_i \Gamma_i(y_i - t_i + t_0) \equiv h_i(y_i - t_i + t_0)$, with $h'_i(\cdot) > 0$, and the number of non-participants is $1 - \sum_{i=1}^s n_i \Gamma_i(y_i - t_i + t_0) \equiv h_0$.

Social welfare is given by $\sum_{i=1}^s n_i h_i(y_i - t_i + t_0) u(y_i - t_i) + \sum_{i=0}^s n_i \hat{m}_i u(-t_0 + m_i) d\Gamma_i(i)$, where $u(\cdot)$ is the concave social utility function. The government chooses taxes t_i to maximize social welfare subject to the budget constraint $\sum_{i>0} h_i(y_i - t_i + t_0) t_i + (1 - \sum_{i>0} h_i(y_i - t_i + t_0)) t_0 = 0$. Letting the Lagrange multiplier be λ , the first-order conditions with respect to t_i and t_0 are $-h_i u'_i + \lambda(h_i - (t_i - t_0) u'_i) = 0$, for $i > 0$ and $-\sum_{i=0}^s n_i \hat{m}_i u'_{i0} d\Gamma_i(m_i) + \lambda(h_0 + \sum_{i>0} (t_i - t_0) h'_i) = 0$.

To interpret these, define the social utility of a euro in terms of government revenue to participants and non-participants as $g_i \equiv u'_i / \lambda$ and $g_0 \equiv \sum_{i=0}^s n_i \hat{m}_i u'_{i0} d\Gamma_i(m_i)$. Then, the first-order conditions reduce to

$$\frac{t_i - t_0}{c_i - c_0} = \frac{1 - g_i}{\eta_i}, \quad \forall i > 0, \quad \text{and} \quad \sum_{i \geq 0} h_i g_i = 1 \quad (4)$$

where $\eta_i = (c_i - c_0) h'_i(\cdot) / h_i(\cdot)$ is the elasticity of participation for type $-i$'s, and $t_i - t_0$ can be interpreted as the participation tax.

This has the following implications. Suppose $g_0 > g_1 > g_2 \dots$, so social welfare weights decrease with skills in the optimum (that is, c_i is increasing). Then, there will be some skill level i^* such $g_i > 1$ that if and only if $i < i^*$. In this case, the participation tax is negative for low skill levels, and becomes positive for $i < i^*$. This corresponds with what is referred to as an earned income tax credit scheme (based on the US terminology), and might be contrasted with negative income tax formats where the tax paid on increments of earnings is always positive.

Two cautionary notes should be mentioned about this argument. First, if the social welfare function is maximin, $g_i = 0$ for all $i > 0$, so $i^* = 0$. In this case, the participation tax is always positive. Second, the social weights $g_i = u'_i / \lambda$ are endogenous, and there is no guarantee that they will be declining with skills. On the contrary, they should be heavily influenced by the elasticity of participation, η_i . For groups with a low elasticity of participation, (4) indicates that they should have a high tax rate. There is no reason why their tax rate could not be such that g_i is induced to be very high even if they are high skilled. Thus, the pattern

of participation tax rates can be very non-monotonic, leading to non-monotonic patterns for average and marginal tax rates (possibly even more than 100 percent) defined in the conventional way.

4.2. Participation and job choice

Suppose, following Saez (2002b), that in addition to deciding whether to participate, workers can choose to work in the job most suited for them or the next-least skilled one for simplicity. Analogous to above, let the number of workers who opt for job i be given by $h_i(c_i - c_0, c_{i+1} - c_i, c_i - c_{i-1})$, where $c_i = y_i - t_i$ and $c_0 = -t_0$ as above. The first argument, $c_i - c_0$, represents the reward from participating (not including the loss in value of leisure). The elasticity of labour supply $h_i(\cdot)$ with respect $c_i - c_0$ to is the elasticity of participation for a type $-i$, η_i . The second argument, $c_{i+1} - c_i$, is the reward to a type $-(i+1)$ worker from opting for job i . The third argument is the reward to a type $-i$ worker for choosing a less-skilled job. The elasticity of job choice for a type $-i$ worker is $\epsilon_i = (c_i - c_{i-1})/h_i \cdot h_i(\cdot)/(c_i - c_{i-1})$.

The structure of optimal earnings taxes can be obtained by following a similar perturbation procedure as above. Suppose that all taxes $t_0, \dots, t_s$ are set optimally, and consider a small perturbation $dt = dt_j$ for $j \neq i$. This induces a change dt in reward for type $-i$ workers choosing between jobs i and $i-1$, as well as an increase dt in the participation tax for all workers of types $j \neq i$. The latter will all transfer euros to the government, which is valued at $1 - g_j$ for each type j . In addition, there will be a loss in net tax revenue resulting from the reduction in participation of all types $j \neq i$, based on the participation elasticities, η_j , and a move of some type $-i$'s to job $i-1$, determined by ϵ_i . Adding all these changes up and summing them to zero, we obtain:

$$\frac{t_i - t_{i-1}}{c_i - c_{i-1}} = \frac{1}{h_i \epsilon_i} \sum_{j=1}^s h_j \left(1 - g_j - \eta_j \frac{t_j - t_0}{c_j - c_0} \right), \quad \forall i > 0 \quad (5)$$

Suppose that there is no participation choice, so $\eta_j = 0$. Then, since $h_i g_i = 1$ as above and $h_1 = 1$ the righthand side will be positive for $i > 0$ as long as the social weights, g_i , are decreasing in skills. In this case, participation tax rates are all positive, unlike in the pure participation choice model. When the participation elasticity η_j is positive, the participation tax rate is reduced at the bottom. Whether it is negative or positive depends on the relative weight η_j of and ϵ_i , something which empirical analysis can help resolve (Blundell 2011).

4.3. Intensive and extensive margins

Workers may be able to vary their labour supply as well as being able to make discrete choices about participation and/or jobs. Jacquet, Lehmann and Van der Linden (2010) consider a standard Mirrlees model with a participation possibility added. Workers of all skill levels incur a disutility of participation that is drawn from some distribution function. As in

the pure extensive-margin model, non-participants will be drawn from all skill levels, and must be treated alike by the government. Participants face the usual nonlinear income tax schedule. The addition of this further margin of decision-making to the Mirrlees model causes optimal marginal tax rates to be reduced. The larger are the elasticities of participation, the greater are the reductions in marginal income tax rates. It is possible for participation tax rates to be negative at the bottom, but marginal income tax rates will always be nonnegative.

Boadway, Song and Tremblay (2012) instead allow workers to choose both a job and an intensity of labour supply. There is a job most suited for each skill-type of worker, and workers earn their highest wage in the most suited job. Such wage rates rise with skills. A worker who chooses a job requiring less skill earns a lower wage than the worker for whom the job is most suited. Although workers can choose their labour supply, there are upper and lower bounds to the incomes that can be earned in each job. In this setting, the pattern of marginal tax rates differs considerably from the standard model. Redistribution from higher to lower-skilled workers is constrained by precluding the former from imitating the income and job of the latter. This involves an upward binding incentive-constraint since lower-skilled workers are more productive in their most suited job than higher skilled mimickers. The consequence is that marginal tax rates can be negative for lower skilled workers since encouraging them to work discourages mimicking. In addition to marginal income tax rates, there can be ranges of the income tax schedule where average tax rates are declining.

4.4. Involuntary unemployment

One final extension of the extensive-margin model has been to allow for endogenous involuntary unemployment. Typically this is modeled using a variant of search models in the tradition of Mortensen and Pissarides (1999) and Pissarides (2000), surveyed in Rogerson, Shimer and Wright (2005). In these models, firms compete for workers by posting job vacancies at a fixed cost per vacancy, and unemployed workers who choose to search also incur a cost. Only a portion of workers succeed in attaining a match because of some unspecified frictions in the labour market. In the simplest model, the number of vacancies filled is based on a constant-returns-to-scale matching function of the Cobb-Douglas form, $M(U, V) = U^\alpha V^{1-\alpha}$, where U is the number of workers searching and V is the number of vacancies. The probability of a vacancy being filled is given by $\pi(\theta) = M(U, V)/V$, $\pi'(\theta) < 0$ where $\theta = V/U$ is a measure of labour market looseness. The probability of a worker obtaining a job, assumed to be the same for all workers, is $M(V, U)/U = \theta\pi(\theta)$ which is increasing in θ .

Once a match is found, the wage rate w is determined by Nash Bargaining over the surplus. (Alternatively, wages could be set ex ante and competitively by firms, which yields qualitatively similar results.) If ρ is the bargaining power of workers, assumed to be given, w maximizes the Nash product $(w - t(w) - b)^\rho (a - w)^{1-\rho}$, where $t(w)$ is the tax function on earnings as in the extensive-margin model, b is the transfer to the unemployed, and a is the skill level of the worker. In an equilibrium, firm entry generates zero expected profits, and

the number of workers searching is based on the marginal worker getting as much expected utility from searching as from remaining unemployed and obtaining the transfer b . It is usually assumed that all unemployed workers obtain the same benefit b regardless of their type, or whether they have chosen to search. In practice, transfers to the unemployed differ between the voluntary and involuntary unemployed. This is enforced by monitoring workers to verify if they are truly involuntarily unemployed and are actively searching for a job. Though monitoring is an important feature of transfer programs for the unemployed, there are relatively few models that incorporate it. Exceptions are Boadway and Cuff (1999), Boadway, Marceau and Sato (1999) and Boone, Fredriksson, Holmlund and van Ours (2007).

Searching and posting vacancies yields potential benefits to the searching worker and posting firm, but causes potentially offsetting externalities. When a firm creates a vacancy, this increases the probability of workers finding a job, but decreases the probability that other firms find a match. And, when a worker chooses to search, that increases the probability that firms will find a match, but decreases the probability of other workers finding a job. As Hosios (1990) showed, these externalities are offsetting and search is efficient if $\rho = \alpha$. The share of the worker's surplus from bargaining then equals the worker's relative productivity at generating matches, and similarly for firms, so workers' search effort and firms' choice in vacancies will be efficient. In this case, the equilibrium bargaining wage maximizes the expected surplus of workers, $\theta\pi(\theta)(a - t(w) - b)$.

These matching models have been deployed in two different contexts. In one, static models are used and involuntary unemployment is treated as permanent. Then, the relevant policy response involves optimal income redistribution. In the other, frictional unemployment is temporary, and the policy response takes the form of insurance.

4.4.1 *I* a a d d b

The rigorous treatment of involuntary unemployment as a redistribution problem was initiated by Hungerbühler, Lehmann, Parmentier and Van der Linden (2006). Workers choose whether to participate in the labour market in a job corresponding to their skill level, denoted a . Participants search in a skill-specific job market, and a matching mechanism of the above sort applies at each skill level. Workers have a common value of leisure, which implies there is a cutoff skill level $\tilde{a}$ such that workers participate if and only if $a \geq \tilde{a}$. Wage bargaining at each skill level is assumed to be efficient (i.e., $\rho = \alpha$). The government observes negotiated wages w , and imposes a nonlinear income tax $t(w)$ and a transfer to the unemployed b . However, worker skill a is private information to the worker and employer, so firms employing workers of one skill level can mimic the bargaining outcome of any other. An incentive constraint thus applies to wage bargains, which restricts the ability of government to redistribute among skill levels. The nonlinear tax affects wage bargains and therefore employment at each participating skill level ($a \geq \tilde{a}$), but it does so in contradictory ways. A higher tax $t(w)$ reduces worker surplus, so increases w , while a higher $t'(w)$ reduces the incentive for workers to bargain leading to a lower w . Naturally, employment varies inversely

with w . In the social optimum, the participation tax applying at the marginal skill level $\tilde{a}$ is positive, unlike in the Diamond-Saez extensive-margin model, while the marginal tax rate $t'(w) > 0$ for all w , and the average tax rate $t(w)/w$ is increasing. Marginal tax rates tend to be much higher than in the standard intensive-margin model.

Subsequent contributions relaxed some of the stronger assumptions of the Hungerbiller et al approach. Lehmann, Parmentier and Van der Linden (2011) assume that the value of leisure is heterogeneous among all workers, so there is a marginal participant at each skill level, as in the Diamond-Saez pure participation model. Otherwise, the assumptions of Hungerbiller et al apply. They emphasize the maximin case, and find that $t'(w) > 0$ throughout and higher than in the competitive labour market setting. As in the Diamond-Saez maximin case, participation tax rates are positive at the bottom. Assuming that the participation elasticity fall with skills, average tax rates are rising. These results continue to apply with more general social welfare functions, although the participation tax rate at the bottom can then be negative. Jacquet, Lehmann and Van der Linden (2011) revert to the pure extensive-margin case by assuming the government can observe wage bargains by skill level, so mimicking is not possible. There is still a participation margin at each skill level since costs of search differ among workers. They simplify wage bargaining so that the shares of the surplus going to workers and firms are fixed. If ρ is the bargaining power of workers, the wage at skill level a becomes $w(a) = \rho a + (1 - \rho)(t(w(a)) - b)$, which is increasing in $t(w)$. An increase $t(w)$ then reduces labour demand, and also reduces labour supply via the participation effect. The optimal participation tax for a type- a worker is

$$\frac{t(w(a)) + b}{c(a) - c_0} = \frac{1 - g(a)(1 + \eta^D(a))\rho}{(\eta^P(a) + \eta^D(a) + \eta^P(a)\eta^D(a))\rho}$$

where $\eta^P(a)$ is the elasticity of participation of type $-a$ workers and $\eta^D(a)$ is the elasticity of demand with respect to surplus $a - w(a)$. This is a generalization of the pure extensive margin participation tax rate in (4) to take account of involuntary as well as voluntary unemployment. The addition of labour demand elasticities reduces optimal participation taxes.

4.4.2 *I* a a d a c

Arguably, search models are better suited for explaining temporary rather than permanent unemployment, in which case unemployment insurance becomes a relevant policy instrument. There is a large literature on the optimal design on unemployment insurance, and we cannot do justice to it here. A survey of search-based theories of optimal unemployment insurance may be found in Coles (2008). Earlier models of unemployment considered sources other than matching frictions, such as efficiency wage or turnover cost models (Phelps 1968, 2003), temporary layoffs and implicit contracts (Baily 1974; Azariadis 1975; Feldstein 1978) and displaced workers (LaLonde 2007). Instead, we focus on some innovations based on search models, particularly those that might be relevant for optimal redistribution policy.

Recent models highlight one important trade-off in designing efficient unemployment insurance systems in search environments (Chetty 2008). Unemployment insurance improves efficiency by undoing the consequences of liquidity constraints. These preclude households from self-insuring to smooth their consumption across time when faced with the shock of unemployment. At the same time, unemployment insurance induces inefficiency by reducing search effort by households, the moral hazard effect. The moral hazard effect arises because the government cannot observe either the behaviour of workers when unemployed or whether unemployment is voluntary versus involuntary. Therefore, unemployment insurance is necessarily incomplete, even though the problem can be mitigated by monitoring the unemployed. Optimal unemployment insurance—including eligibility requirements, wage replacement rates, and duration—trades off these liquidity and moral hazard effects.

This view of unemployment insurance has been extended in various fruitful ways. Chetty and Saez (2010) allow for the fact that some private unemployment insurance exists in practice, if only implicitly, an example being negotiated severance pay. Publicly provided unemployment insurance crowds out private insurance, so reduces the optimal amount of unemployment insurance, that is the earnings replacement rate. Spinnewijn (2010) supposes that the unemployed may mistakenly both overestimate the chances of finding an employment match and underestimate the return to search. These effects tend to make the optimal level of unemployment benefits higher. Acemoglu and Shimer (1999) emphasize the consequences of risk aversion. Introducing risk aversion into search models reduces wages and therefore increases employment and output. Unemployment insurance reduces the effect of risk aversion and induces workers to direct their search to higher risk jobs. As a result, firms invest in such jobs and this causes aggregate output to rise as well as improving risk-sharing (although moral hazard would mitigate that). Crossley and Low (2011) extend unemployment insurance to a life-cycle setting. They argue that the relevance of liquidity constraints falls with age, so the benefit of unemployment insurance does as well. Landais, Michaillat and Saez (2010) analyze optimal unemployment insurance over the business cycle where both cyclical and frictional (search) unemployment exist. In a recession, unemployment is caused by wage rigidity so there is job rationing. This implies that the adverse effects of moral hazard are mitigated since there is a shortage of jobs. Moreover, search intensity causes a negative externality on other searchers since the number of available jobs is fixed. Then, optimal unemployment insurance should provide higher earnings replacement during recessions. In an earlier contribution, Cremer, Marchand and Pestieau (1996) study optimal duration in a matching model. Workers have abilities in non-market activities and can reject a job offer to continue searching for a better job match. Higher unemployment insurance benefits increase the duration of unemployment, which leads to better matching. It induces some workers who are productive in non-market activities to remain unemployed. Optimal unemployment benefits will decrease with unemployment duration.

These results emphasize efficiency arguments, particularly those arising from insurance market failure due to moral hazard and inadequate self-insurance due to liquidity constraints. There are at the same time important equity issues in designing unemployment insurance schemes. Liquidity constraints are systematically tighter for persons with low family wealth

and for younger, lower-income workers who have not been able to accumulate assets. Self-insurance is much less costly for workers from high-wealth families and for workers who have better access to capital markets. The cost of adjusting to employment shocks is therefore much higher for low-income workers (Chetty and Looney 2006). In addition, low-skilled workers are exposed to relatively higher risk of unemployment, and to a longer duration. These factors suggest that unemployment insurance should be more generous for low-skilled workers, especially those from low-wealth families. In addition, supplementary active labour market policies, such as training, that help low-skilled workers get into the labour force, possibly with higher skills, can improve outcomes.

In fact, there has been relatively little literature that has studied the redistributive role of unemployment insurance. One exception in a search context is Mortensen and Pissarides (2003). In their model, workers of different skills search in skill-specific job markets, and those successful receive wages determined by Nash bargaining. The costs of job posting and job creation are increasing in skills, while the value of leisure is independent of skills. In equilibrium, the wage and employment rates are higher for high-skilled workers. They argue, on the basis of simulations, that wage and employment subsidies financed by a payroll tax increases low-skilled wages and employment. Phelps (1997) and Hoon and Phelps (2003), deploying a turnover-cost efficiency-wage model of unemployment, also find that low-skilled workers have higher unemployment rates and longer duration. They propose an employment subsidy financed by a payroll tax. See also Snower (1994). Unlike participation subsidies, which work through increasing participation and can reduce wage rates through supply-side effects, employment subsidies both stimulate employment and increase low-skilled wage rates.

5. Dynamic Optimal Taxation

In an intertemporal setting, capital income taxation becomes relevant. The recent literature on the so-called ‘new dynamic public finance’ (Golosov, Tsyvinski and Werning 2007; Kocherlakota 2010) has shed considerable light on the rationale for capital taxation, or more accurately on the rationale for imposing tax wedges on intertemporal consumption margins. If it is optimal for there to be an intertemporal tax wedge on consumption, that can be implemented by various time-specific taxes, of which capital income taxes are only one form (Kocherlakota 2004). In what follows, we focus on the case for intertemporal tax wedges.

There are some important challenges to applying optimal tax analysis in an intertemporal setting, and not all of these have been resolved. One is that second-best policy outcomes are generally time-inconsistent. Unless one assumes the government can commit to future taxes, one is forced to look for time-consistent (subgame-perfect) policies, which can be significantly inferior to second-best policies. Although some progress has been made in this area (Pereira 2009), we follow the standard approach of assuming that the government can commit. Second, the general optimal nonlinear tax problem requires that we imagine a lifetime tax system with lifetime incentive constraints. In practice, this is difficult to implement,

and the loss from being unable to do so can be significant (Weinzierl 2011). Third, it is natural to think that in an intertemporal setting, there will be uncertainty about the future. We sidestep the first two issues by assuming that the government can commit to future tax rates, and that the tax system is applied on a lifetime basis. Some insight can be obtained from this procedure. We present some of the key insights from the dynamic public finance literature using a simple two-period two-type setting with and without uncertainty (following Diamond 2007).

Consider the simple benchmark case of two skill-types ($i = 1, 2$) who have wage rates w_i^1 and w_i^2 , supply labour ℓ_i^1 and ℓ_i^2 , and consume c_i^1 and c_i^2 in the first and second periods. Type-2's are high-skilled, so $w_2^j > w_1^j$ for $j = 1, 2$. The two-period utility function is $u(c_i^1) - h(y_i^1/w_i^1) + \beta_i(u(c_i^2) - h(y_i^2/w_i^2))$ where $y_i^j = w_i^j \ell_i^j$ and we allow for the possibility that the two skill-types have a different utility discount factor, β_i . The government observes incomes y_i^1 and y_i^2 as well as consumption in each period, or equivalently saving. The second-best optimal nonlinear income tax solves the following utilitarian problem:

$$\text{Max} \sum_{i=1,2} n_i \left(u(c_i^1) - h\left(\frac{y_i^1}{w_i^1}\right) + \beta_i \left(u(c_i^2) - h\left(\frac{y_i^2}{w_i^2}\right) \right) \right)$$

subject to

$$\sum_{i=1,2} n_i \left(y_i^1 - c_i^1 + \frac{y_i^2 - c_i^2}{1+r} \right) = R$$

and

$$u(c_2^1) - h\left(\frac{y_2^1}{w_2^1}\right) + \beta_2 \left(u(c_2^2) - h\left(\frac{y_2^2}{w_2^2}\right) \right) \geq u(c_1^1) - h\left(\frac{y_1^1}{w_2^1}\right) + \beta_2 \left(u(c_1^2) - h\left(\frac{y_1^2}{w_2^2}\right) \right)$$

where the first constraint is the government's intertemporal budget and the second one is the lifetime incentive constraint, which applies to type-2's only.

The first-order conditions with respect to c_i^j give the following intertemporal tax wedges:

$$\frac{u'(c_2^1)}{\beta_2 u'(c_2^2)} = 1 + r = \frac{n_2 - \gamma}{n_2 - \gamma \beta_2 / \beta_1} \cdot \frac{u'(c_1^1)}{\beta_1 u'(c_1^2)}$$

If the two skill-types have the same utility discount factors ($\beta_1 = \beta_2$), there is no intertemporal tax wedge, so no need for capital income taxation. The marginal rate of substitution between c_i^2 and c_i^1 equals $1 + r$ for both types. This is just an intertemporal version of the Atkinson-Stiglitz Theorem. On the other hand, if $\beta_1 > \beta_2$ as Diamond and Spinnewijn (2011) have argued, then $u'(c_1^1)/(\beta_1 u'(c_1^2)) < 1 + r$, so there is a positive intertemporal tax wedge on the type-1's and the Atkinson-Stiglitz Theorem fails. There is then a case for a positive capital income tax, and that will also be true in a model where the government can

only impose a uniform capital income tax on both types. The first-order conditions on y_t^j lead to marginal income tax rates on the two types. These are zero for the high-skilled workers in both periods, and positive for the low-skilled workers. Marginal tax rates generally vary over time for the type-1's, so there is no intertemporal tax smoothing. Diamond (2007) shows that if the intertemporal wage profile is steeper for the type-2's than the type-1's, marginal tax rates tend to rise over time.

Extending the above model to a setting in which future wage rates are not known when first period saving decisions are made leads to one of the main insights of the dynamic public finance literature. Suppose we simplify matters by assuming all workers have an identical wage rate w^1 in the first period, so choose the same values of ℓ^1 and c^1 (or income y^1 and saving). In the second period, n_1 and n_2 turn out to be type-1's and type-2's. Expected utility of all workers at the beginning of the first period is $u(c^1) - h(y^1/w^1) + \beta \sum_{i=1,2} n_i (u(c_i^2) - h(y_i^2/w_i^2))$. The government chooses income-consumption bundles to maximize expected utility subject to a present-value revenue constraint, $y^1 - c^1 + (1 - r)^{-1} \sum_{i=1,2} n_i (y_i^2 - c_i^2) = R$, and a second-period incentive constraint, $u(c_2^2) - h(y_2^2/w_2^2) \geq u(c_1^2) - h(y_1^2/w_1^2)$. The standard nonlinear income tax applies to earnings in the second period. In addition, there is a positive intertemporal tax wedge:

$$\frac{u'(c^1)}{\beta \sum_i n_i u'(c_i^2)} < 1 + r$$

In effect, reducing saving by taxing capital income makes it more difficult for high-skill workers to mimic low-skilled ones in the second period.

Taken together, these results suggest that capital income tax rates should be positive (Banks and Diamond 2010). This argument is reinforced if workers face liquidity constraints early in life that prevent them from borrowing to smooth their lifetime consumption. A capital income tax shifts tax liabilities to the second period and serves to relax the liquidity constraint (Hubbard and Judd 1987). These effects can be substantial. Conesa, Kitao and Krueger (2009) study an overlapping generations model of life-cycle households with stochastic wage rates. They calculate the optimal tax system consisting of a progressive age-dependent labour income tax and a proportional capital income tax in a model calibrated to the US economy, allowing for liquidity constraints. Remarkably, the optimal capital income tax turns out to be of the order of 36 percent.

One final contribution to the dynamic optimal taxation literature worth noting is by Cremer and Gahvari (1995). They extend the above model with uncertain future wages by allowing for multiple commodities that can be taxed indirectly. Suppose utility functions are weakly separable in goods and leisure. When all commodities are purchased after wages are revealed, the Atkinson-Stiglitz Theorem applies directly, so there is no need for differential commodity taxes. Suppose, however, that one good must be purchased before wage rates are known. They interpret this as a durable good. Now, it is optimal to impose a uniform tax on all goods purchased ex post, but a lower tax rate on the durable. Analogous to the case of sav-

ing mentioned above, encouraging greater spending on the durable good makes is more difficult for workers who turn out to be high-wage to mimic low-wage workers. This would support preferential treatment of consumer durables such as housing in the income tax system.

6. Concluding Remarks

The field of optimal income taxation and its use for informing tax reform policy debates has been an active one in recent years. We have recounted some of the highlights of the contributions, but have only scratched the surface of policy-relevant research. In this concluding section, we simply draw attention to a number of other interesting strands of the literature that have policy implications.

One such area involves other policy instruments that can be used to supplement nonlinear income taxation and transfers. In the optimal income tax literature, information or incentive constraints prevent the benevolent government from achieving a first-best outcome. Given that, policy instruments that weaken incentive constraints can be welfare-improving. There are several examples of these in the literature, and they involve either price or quantity controls following Nichols and Zeckhauser (1982) and Guesnerie and Roberts (1984). One is the minimum wage. Assuming that a minimum wage can be enforced by standard legal means, it can provide information to the government that helps target transfers to low-skilled workers, as shown by Marceau and Boadway (1994), Boadway and Cuff (2001), and Hungerbühler and Lehmann (2009).

Another example is workfare, which is the requirement for transfer recipients to supply work to public projects as a condition of receiving transfers (Besley and Coate 1992; Cuff 2000). Workfare can serve various purposes, such as giving work experience to the long-term unemployed, or keeping the latter occupied. In our context, workfare might serve as a device for targeting transfers to the most deserving. It can only do this if it is less costly for intended recipients to engage in workfare than those who are being screened out. It would be counterproductive in the case of recipients who find it relatively costly to work because of a disability. Workfare is more likely to be beneficial, the more productive is the time spent in workfare.

A third category of supplementary policy instruments involves in-kind transfers. There are two versions in the literature. Under the opt-in version, transfers to households are made contingent on them opting to take up the in-kind transfer. These are effective to the extent that the target population put a relatively higher value on the in-kind transfer than others, either because of different preferences (Blackorby and Donaldson 1988) or because the in-kind transfer is a complement with leisure (Blomquist and Christiansen 1995). Restricting the amount that is provided discourages high-skilled persons from opting in and thereby limiting their consumption of the good. The second version is compulsory provision: forcing all households to consume some minimum amount. This can be used to weaken the incentive constraint on high-skilled persons if the good provided is a substitute for leisure. In princi-

ple, excise taxes or subsidies can have the same effect. However, as Boadway, Marchand and Sato (1998) show, compulsory in-kind transfers dominate subsidization.

Another area of the optimal taxation literature involves business taxation. As personal income tax systems have evolved toward progressive consumption taxes or dual income taxes, and as capital has become more mobile internationally, the perceived role of business income taxation has changed from withholding against capital income accumulated within corporations that would otherwise escape immediate personal taxation, to the efficient taxation of economic rents. The Mirrlees Review (2011), following its precursor, the Meade Report (1978), proposed a form of rent taxation to complement progressive consumption taxation. Where the Meade Report had recommended a cash-flow tax, the Mirrlees Review proposed the present-value equivalent of a cash-flow tax. This is referred to as the Allowance for Corporate Equity (ACE) system, and entails firms getting a full deduction for all costs of investment including depreciation and debt and equity costs. The ACE is effectively a cash-flow tax except that not all cash flows are deducted immediately, but can be fully carried forward at a risk-free interest rate⁴. The taxation of rents is particularly important in the natural resource industries. An early version of rent taxation in this context, which is equivalent to the ACE in its effect, is the Resource Rent Tax (RRT) proposed by Garnaut and Clunies Ross (1975), and has recently formed the basis of proposals for mining taxation by the Henry Report in Australia (Australian Treasury 2010). It works by taxing all cash flows above some threshold rate of return, which is typically taken to be the risk-free rate. The case for using the risk-free rate is that there should be no risk associated with the carry-forward of tax credits. Of course, if there is some political risk of the government reneging on this carry-forward, a case can be made for a higher threshold. One problem with both the ACE and the RRT is that for full neutrality to prevail, firms that never make a profit and eventually go out of business must get their negative tax liabilities refunded. In the absence of this, risk is discouraged.

One final area of optimal tax research which is very challenging concerns how to deal with heterogeneous preferences. For example, suppose workers have different preferences for leisure. How should the tax system treat workers who have low incomes because they are high-skilled but have high preferences for leisure compared with those who are low-skilled but hard-working? Depending on the relative social weights one gives to these two types, the pattern of marginal income tax rates can vary widely (Boadway, Marchand, Pestieau and Racionero 2000; Chon and Laroque 2010). One promising approach to this problem involves distinguishing between characteristics over which households have control and those over which they do not. They would only be compensated for the latter characteristics, such as ability in Mirrleesian optimal income tax theory. For characteristics that they are responsible for, of which their preference for leisure might be an example, they should be neither rewarded nor punished. This would imply that two workers with the same skills should pay the same tax regardless of their preferences for leisure (the ‘principle of responsibility’), while those with different skills but the same preferences should have their utilities equalized (the ‘principle of compensation’). It turns out that these two principles cannot be achieved at the same time, so some compromise must be made.

One approach to resolving this, referred to as equality of opportunity by Roemer (1998), is as follows. Within each preference group, the maximin criterion applies. Then, across preference groups, the utilities of the least well-off in each group are averaged, equivalent to utilitarianism. This leads to a social welfare function of the form, $\min_i u(x_{ij}) - \gamma_j g(y_{ij}/w_i)$, where j denotes preference types, i denotes ability types, and individual utility functions are additive with the weight γ_j reflecting the weight put on the disutility of labour by different preference groups. This approach avoids the incompatibility of the principles of compensation and responsibility by arbitrarily averaging the utilities of persons of different skills. Another approach, represented by the work of Fleurbaey and Maniquet (2011), is to eschew trying to trade off the utility of persons with different preferences. Instead, following the approach taken in the theory of fairness, they emphasize equalizing the resources that different persons have available. This is a promising approach but would take us too far afield to discuss it further here.

Notes

1. That is, write utility as

$$v(c, y, w) = c - \frac{y^{1+1/\epsilon}}{1+1/\epsilon} = y - T(y) - \frac{y^{1+1/\epsilon}}{1+1/\epsilon}$$

Then, utility maximization yields $y = (1 - T'(y))^{\epsilon} w^{1+\epsilon}$, which implies that ϵ is the elasticity of earnings y with respect to $1 - T'(y)$.

2. More generally, workers' labour supply responses may be conditioned on whether they agree with the use to which government puts their tax revenue. For example, if the tax system becomes more progressive than workers' social preferences will tolerate, they may begin to respond more vigorously to increases in the marginal tax rate (Boadway, Marceau and Mongrain 2007).
3. The Laroque-Kaplow result might even support a move to uniform commodity taxes even if the income tax reform is not taken out on the principle that the tax reform is potentially Pareto-improving so satisfies a sort of second-best compensation test. For further discussion of second-best compensation tests, see Coate (2000).
4. The ACE was originally proposed by Institute for Fiscal Studies (1991), and was based on ideas formulated in Boadway and Bruce (1984) and Bond and Devereux (1995).

References

- Acemoglu, Daron and Robert Shimer (1999), "Efficient Unemployment Insurance", *Journal of Political Economy*, 107: 893-928.
- Akerlof, George A. (1978), "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Training", *American Economic Review*, 68: 8-19.
- Atkinson, Anthony B. and Joseph E. Stiglitz (1976), "The Design of Tax Structure: Direct vs. Indirect Taxation", *Journal of Public Economics*, 6: 55-75.
- Australian Treasury (2010), *Australia's Future Tax System* (The Henry Review), Canberra: Commonwealth of Australia.

- Azariadis, Costas (1975), "Implicit Contracts and Underemployment Equilibria", *Journal of Political Economy*, 83: 1183-1202.
- Baily, Martin (1974), "Wages and Employment under Uncertain Demand", *Review of Economic Studies*, 41: 37-50.
- Banks, James and Peter Diamond (2010), "The Base for Direct Taxation", in James Mirrlees, Stuart Adam, Timothy Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba (eds.), *Dimensions of Tax Design: The Mirrlees Review*, Oxford, Oxford University Press, 548-648.
- Besley, Timothy and Stephen Coate (1992), "Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs", *American Economic Review*, 82: 249-61.
- Blackorby, Charles and David Donaldson (1988), "Cash versus Kind, Self Selection and Efficient Transfers", *American Economic Review*, 78: 691-700.
- Blomquist, S ren and Vidar Christiansen (1995), "Public Provision of Private Goods as a Redistributive Device in an Optimum Income Tax Model", *Scandinavian Journal of Economics*, 97: 547-67.
- Blomquist, S ren and Vidar Christiansen (2008), "Taxation and Heterogeneous Preferences", *FinanzArchiv*, 64: 218-44.
- Blundell, Richard (2011), "Viewpoint: Empirical Evidence and Tax Policy Design: Lessons from the Mirrlees Review", *Canadian Journal of Economics*, 44: 1106-37.
- Boadway, Robin (2012), *From Optimal Tax Theory to Tax Policy: Retrospective and Prospective Views: the 2009 Munich Lectures*, Cambridge: MIT Press.
- Boadway, Robin and Neil Bruce (1984), "A General Proposition on the Design of a Neutral Business Tax", *Journal of Public Economics*, 24: 231-39.
- Boadway, Robin and Katherine Cuff (1999), "Monitoring Job Search as an Instrument for Targeting Transfers", *International Tax and Public Finance*, 6: 317-37.
- Boadway, Robin and Katherine Cuff (2001), "A Minimum Wage Can Be Welfare-Improving and Employment-Enhancing", *European Economic Review*, 45: 553-76.
- Boadway, Robin and Firouz Gahvari (2006), "Optimal Taxation with Consumption Time as a Leisure or Labor Substitute", *Journal of Public Economics*, 90: 1851-78.
- Boadway, Robin and Laurence Jacquet (2008), "Optimal Marginal and Average Income Taxation under Maximin", *Journal of Economic Theory*, 143: 425-41.
- Boadway, Robin and Michael Keen (1993), "Public Goods, Self-Selection and Optimal Income Taxation", *International Economic Review*, 34: 463-78.
- Boadway, Robin, Nicolas Marceau and Steeve Mongrain (2007), "Redistributive Taxation under Ethical Behaviour", *Scandinavian Journal of Economics*, 109: 505-29.
- Boadway, Robin, Nicolas Marceau and Motohiro Sato (1999), "Agency and the Design of Welfare Systems", *Journal of Public Economics*, 73: 1-30.
- 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, Maurice Marchand and Motohiro Sato (1998), "Subsidies Versus Public Provision of Private Goods as Instruments for Redistribution", *Scandinavian Journal of Economics*, 100: 545-64.
- Boadway, Robin and Pierre Pestieau (2003), "Indirect Taxation and Redistribution: The Scope of the Atkinson-Stiglitz Theorem", in Richard Arnott, Bruce Greenwald, Ravi Kanbur and Barry Nalebuff (eds.), *Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz*, Cambridge, MA: MIT Press: 387-403.
- Boadway, Robin and Pierre Pestieau (2006), "Tagging and Redistributive Taxation", *Annales d'Économie et de Statistique*, 83-84: 123-47.
- Boadway, Robin, Zhen Song and Jean-François Tremblay (2012), "Optimal Nonlinear Taxation with Skill-Specific Jobs", Central University of Finance and Economics, Beijing, mimeo.
- Bond, Stephen R. and Michael P. Devereux (1995), "On the Design of a Neutral Business Tax under Uncertainty", *Journal of Public Economics*, 58: 57-71.
- Boone, Jan, Peter Fredriksson, Bertil Holmlund and Jan C. van Ours (2007), "Optimal Unemployment Insurance with Monitoring and Sanctions", *Economic Journal*, 117: 399-421.
- Browning, Edgar K. (1976), "The Marginal Cost of Public Funds", *Journal of Political Economy*, 84: 283-98.
- Chetty, Raj (2008), "Moral Hazard versus Liquidity and Optimal Unemployment Insurance", *Journal of Political Economy*, 116: 173-234.
- Chetty, Raj (2009), "Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance", *American Economic Journal: Economic Policy*, 1: 31-52.
- Chetty, Raj and Adam Looney (2006), "Consumption Smoothing and the Welfare Consequences of Social Insurance in Developing Economies", *Journal of Public Economics*, 90: 2351-56.
- Chetty, Raj and Emmanuel Saez (2010), "Optimal Taxation and Social Insurance with Endogenous Private Insurance", *American Economic Journal: Economic Policy*, 2: 85-114.
- Chon, Philippe and Guy Laroque (2010), "Negative Marginal Tax Rates and Heterogeneity", *American Economic Review*, 100: 2532-47.
- Christiansen, Vidar (1981), "Evaluation of Public Projects under Optimal Taxation", *Review of Economic Studies*, 48: 447-57.
- Christiansen, Vidar (1984), "Which Commodity Taxes Should Supplement the Income Tax?", *Journal of Public Economics*, 24: 195-220.
- Coate, Stephen (2000), "An Efficiency Approach to the Evaluation of Policy Changes", *Economic Journal*, 110: 437-55.
- Coles, Melvyn G. (2008), "Optimal Unemployment Insurance in a Matching Equilibrium", *Labour Economics*, 15: 537-59.
- Conesa, Juan Carlos, Sagiri Kitao and Dirk Krueger (2009), "Taxing Capital? Not a Bad Idea After All", *American Economic Review*, 99: 25-48.
- Corlett, W.J. and D.C. Hague (1953), "Complementarity and the Excess Burden of Taxation", *Review of Economic Studies*, 21: 21-30.

- Cremer, Helmuth and Firouz Gahvari (1995), "Uncertainty, Optimal Taxation and the Direct Versus Indirect Tax Controversy", *Economic Journal*, 105: 1165-79.
- Cremer, Helmuth, Maurice Marchand and Pierre Pestieau (1995), "The Optimal Level of Unemployment Insurance Benefits in a Model of Employment Mismatch", *Labour Economics*, 2: 407-20.
- Cremer, Helmuth, Pierre Pestieau and Jean-Charles Rochet (2001), "Direct Versus Indirect Taxation: The Design of the Tax Structure Revisited", *International Economic Review*, 42: 781-99.
- Crossley, Thomas F. and Hamish Low (2011), "Borrowing Constraints, the Cost of Precautionary Saving and Unemployment Insurance", *International Tax and Public Finance*, forthcoming.
- Cuff, Katherine (2000), "Optimality of Workfare with Heterogeneous Preferences", *Canadian Journal of Economics*, 33: 149-74.
- Deaton, Angus (1979), "Optimally Uniform Commodity Taxes", *Economics Letters*, 2: 357-61.
- Diamond, Peter A. and James A. Mirrlees (1971), "Optimal Taxation and Public Production I: Production Efficiency and II: Tax Rules", *American Economic Review*, 61: 8-27 and 261-78.
- Diamond, Peter (1980), "Income Taxation with Fixed Hours of Work", *Journal of Public Economics*, 13: 101-10.
- Diamond, Peter A. (1998), "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, 88: 83-95.
- Diamond, Peter A. (2007), "Comment on Golosov et al", *NBER Macroeconomics Annual 2006*, 365-79.
- Diamond, Peter A. and Johannes Spinnewijn (2011), "Capital Income Taxes with Heterogeneous Discount Rates" *American Economic Journal: Economic Policy*, 3: 52-76.
- Ebert, Udo (1992), "A Reexamination of the Optimal Nonlinear Income Tax", *Journal of Public Economics*, 49: 47-73.
- Edwards, Jeremy, Michael Keen and Matti Tuomala (1994), "Income Tax, Commodity Taxes and Public Good Provision: A Brief Guide", *Finanzarchiv*, 51: 472-97.
- Feldstein, Martin S. (1978), "The Effect of Unemployment Insurance on Temporary Layoff Unemployment", *American Economic Review*, 68: 834-46.
- Feldstein, Martin S. (1999), "Tax Avoidance and the Deadweight Loss of the Income Tax", *Review of Economics and Statistics*, 81: 674-80.
- Fleurbaey, Marc and François Maniquet (2011), *A Theory of Fairness and Social Welfare*, New York: Cambridge University Press.
- Garnaut, Ross and Anthony Clunies Ross (1975), "Uncertainty, Risk Aversion and the Taxing of Natural Resource Projects", *Economic Journal*, 85: 272-87.
- Guesnerie, Roger and Kevin Roberts (1984), "Effective Policy Tools and Quantity Controls", *Econometrica*, 52: 59-82.
- Golosov, Mikhael, Aleh Tsyvinski and Iván Werning (2007), "New Dynamic Public Finance: A User's Guide", *NBER Macroeconomics Annual 2006*, 317-63.
- Gordon, Roger H. and Julie Berry Cullen (2012), "Income Redistribution in a Federal System of Governments", *Journal of Public Economics*, forthcoming.

- Gruber, Jonathan and Emmanuel Saez (2002), "The Elasticity of Taxable Income: Evidence and Implications", *Journal of Public Economics*, 84: 1-32.
- Hellwig, Martin F. (2009), "A Note on Deaton's Theorem on the Undesirability of Nonuniform Excise Taxation", *Economics Letters*, 105: 186-88.
- Hellwig, Martin F. (2010), "A Generalization of the Atkinson-Stiglitz (1976) Theorem on the Undesirability of Nonuniform Excise Taxation", *Economics Letters*, 108: 156-58.
- Hoon, Hian Teck and Edmund S. Phelps (2003), "Low-Wage Employment Subsidies in a Labor-Turnover Model of the 'Natural Rate'", in Phelps (2003), 16-43.
- Hosios, Arthur (1990), "On the Efficiency of Matching and Related Models of Search and Unemployment", *Review of Economic Studies*, 57: 279-98.
- Hubbard, R. Glenn and Kenneth L. Judd (1987), "Social Security and Individuals Welfare: Precautionary Saving, Liquidity Constraints, and the Payroll Tax", *American Economic Review*, 77: 630-46.
- Hungerb hler, Mathias and Etienne Lehmann (2009), "On the Optimality of a Minimum Wage: New Insights from Optimal Tax Theory", *Journal of Public Economics*, 93: 464-81.
- Hungerb hler, 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.
- Immonen, Ritva, Ravi Kanbur, Michael Keen and Matti Tuomala (1998), "Tagging and Taxing: The Optimal Use of Categorical and Income Information in Designing Tax/Transfer Schemes", *Economica*, 65: 179-92.
- Institute for Fiscal Studies (1991), *Equity for Companies: A Corporation Tax for the 1990s*, Commentary 26, London: Institute for Fiscal Studies.
- Jacquet, Laurence, Etienne Lehmann and Bruno Van der Linden (2010), "Optimal Redistributive Taxation with both Extensive and Intensive Responses", CESifo Working Paper no. 3308, December 2010.
- Jacquet, Laurence, Etienne Lehmann and Bruno Van der Linden (2011), "Optimal Redistributive Taxation with both Labor Supply and Labor Demand Responses", IZA Working Paper no. 5642, April 2011.
- Jacquet, Laurence and Bruno Van der Linden (2006), "The Normative Analysis of Tagging Revisited: Dealing with Stigmatization", *FinanzArchiv*, 62: 168-98.
- Kaplow, Louis (2006), "On the Desirability of Commodity Taxation Even When Income Taxation Is Not Optimal", *Journal of Public Economics*, 90: 1235-50.
- Kocherlakota, Narayana R. (2004), "Wedges and Taxes", *American Economic Review*, 94: 109-13.
- Kocherlakota, Narayana R. (2010), *The New Dynamic Public Finance*, Princeton: Princeton University Press.
- Konishi, Hideo (1995), "A Pareto-Improving Commodity Tax Reform under a Smooth Nonlinear Income Tax", *Journal of Public Economics*, 56: 413-46.
- LaLonde, R. J. (2007), *The Case for Wage Insurance*, Council on Foreign Relations Special Report No. 30, New York: Council on Foreign Relations.

- Landais, Camille, Pascal Michaillat and Emmanuel Saez (2010), "Optimal Unemployment Insurance over the Business Cycle", NBER Working Paper 16526.
- Laroque, Guy (2005a), "Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof", *Economics Letters*, 87: 141-44.
- Laroque, Guy (2005b), "Income Maintenance and Labor Force Participation", *Econometrica*, 73: 341-76.
- Lehmann, Etienne, Alexis Parmentier and Bruno Van der Linden (2011), "Optimal Income Taxation with Endogenous Participation and Search Unemployment", *Journal of Public Economics*, 95: 1523-37.
- Mankiw, N. Gregory and Matthew Weinzierl (2010), "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution", *American Economic Journal: Economic Policy*, 2: 155-76.
- Marceau, Nicolas and Robin Boadway (1994), "Minimum Wage Legislation and Unemployment Insurance as Instruments for Redistribution", *Scandinavian Journal of Economics*, 96: 67-81.
- Marchand, Maurice, Pierre Pestieau and Mar a Racionero (2003), "Optimal Redistribution When Different Workers Are Indistinguishable", *Canadian Journal of Economics*, 36: 911-22.
- Meade Report (1978), *The Structure and Reform of Direct Taxation*, Report of a Committee chaired by Professor J. E. Meade, London: George Allen and Unwin.
- Mirrlees, James A. (1971), "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies*, 38: 175-208.
- Mirrlees, Sir James, Stuart Adam, Timothy Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles and James Poterba (2011), *Tax by Design: The Mirrlees Review*, London: Institute for Fiscal Studies.
- Mortensen, Dale T., and Christopher A. Pissarides (1999), "New Developments in Models of Search in the Labor Market", in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Amsterdam: North Holland, 2567-627.
- Myles, Gareth D. (1995), *Public Economics*, Cambridge, UK: Cambridge University Press.
- Nichols, Albert L. and Richard J. Zeckhauser (1982), "Targetting Transfers through Restrictions on Recipients", *American Economic Review*, 72: 372-77.
- Parsons, Donald O. (1996), "Imperfect 'Tagging' in Social Insurance Programs", *Journal of Public Economics*, 62: 183-207.
- Pereira, Joana (2009), *Essays on Time-Consistent Fiscal Policy with Unbalanced Budgets: The Role of Public Debt*, PhD Thesis, European University Institute, Florence, Italy.
- Phelps, Edmund S. (1968), "Money-Wage Dynamics and Labor-Market Equilibrium", *Journal of Political Economy*, 76: 678-711.
- Phelps, Edmund S. (1997), *Rewarding Work: How to Restore Participation and Self-Support to Free Enterprise*, Cambridge, MA.: Harvard University Press.
- Phelps, Edmund S. (ed.) (2003), *Designing Inclusion: Tools to Raise Low-end Pay and Employment in Private Enterprise*, Cambridge, UK: Cambridge University Press.
- Pissarides, Christopher A. (2000), *Equilibrium Unemployment Theory*, Second edition, Cambridge, MA: MIT Press.

- President's Advisory Panel on Federal Tax Reform (2005), *Simple, Fair, and Pro-Growth: Proposals to Fix America's Tax System*, Washington: U.S. Treasury.
- Roemer, John E. (1998), *Equality of Opportunity*, Cambridge, MA: Harvard University Press.
- Rogerson, Richard, Robert Shimer and Randall Wright (2005), "Search-Theoretic Models of the Labor Market: A Survey", *Journal of Economic Literature*, 43: 959-88.
- Saez, Emmanuel (2002a), "The Desirability of Commodity Taxation under Non-linear Income Taxation and Heterogeneous Tastes", *Journal of Public Economics*, 83: 217-30.
- Saez, Emmanuel (2002b), "Optimal Income Transfer Programs: Intensive vs Extensive Labor Supply Responses", *Quarterly Journal of Economics*, 117: 1039-73.
- Sheshinski, Eytan (1972), "The Optimal Linear Income Tax", *Review of Economic Studies*, 39: 297-302.
- Snower, Dennis J. (1994), "Converting Unemployment Benefits into Employment Subsidies", *American Economic Review Papers and Proceedings*, 84: 65-70.
- Spinnewijn, Johannes (2010), "Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs", London School of Economics, mimeo.
- Tuomala, Matti (1990), *Optimal Income Tax and Redistribution*, Oxford: Clarendon Press.
- Weinzierl, Matthew C. (2011), "The Surprising Power of Age-Dependent Taxes", *Review of Economic Studies*, 78: 1490-518.

Resumen

En este trabajo se revisan las contribuciones recientes en la teoría de la imposición óptima, particularmente aquellas que resultan relevantes para el diseño de políticas. Estas contribuciones incluyen refinamientos del modelo de imposición óptima de Mirrlees, redistribución óptima cuando los trabajadores toman decisiones en el mercado de trabajo a lo largo del margen disponible, generalizaciones de los Teoremas de Atkinson-Stiglitz y Deaton en la imposición uniforme de bienes, implicaciones del desempleo involuntario en la redistribución y el seguro de desempleo, imposición óptima dinámica e imposición de los ingresos del capital, instrumentos no impositivos para la redistribución y los problemas que surgen cuando las preferencias son heteróneas.

Palabras clave: Imposición óptima, margen intensivo, margen extensivo.

Clasificación JEL: H21, H24, H25.

