Non-Welfarist Optimal Taxation and Behavioral Public Economics*

by

Ravi Kanbur
Cornell University

Jukka Pirttilä
Bank of Finland

and

Matti Tuomala**
University of Tampere

This version: 25 August 2004

* We are grateful to Eytan Sheshinski and other participants of the CESifo Behavioral Public Economics Workshop, Venice Summer Institute, July 2004, for very helpful comments.

** Corresponding author. Address: Department of Economics, 33104 University of Tampere, Finland. Email: matti.tuomala@kolumbus.fi

Abstract:

Research in behavioral economics has uncovered the widespread phenomenon of people making decisions against their own good intentions. In these situations, the government might want to intervene, indeed individuals might want the government to intervene, to induce behavior that is closer to what individual wish they were doing. The analysis of such corrective interventions, through taxes and subsidies, might be called ”behavioral public economics.” However, such analysis, where the government has an objective function that is different from that of individuals, is not new in public economics. In these cases the government is said to be ”non-welfarist” in its objectives, and there is a long tradition of non-welfarist welfare economics, especially the analysis of optimal taxation and subsidy policy where the outcomes of individual behavior are evaluated using a preference function different from the one that generated the outcomes. The object of this paper is to first of all present a unified view of the non-welfarist optimal taxation literature and, secondly, to present behavioral public economics as a natural special case of this general framework.

Key words: non-welfarism, optimal taxation, behavioral economics

1. Introduction

Behavioral economics has highlighted a widespread phenomenon. In different ways and in different contexts, individuals do not seem to behave in the manner of text book rational choice models. This has major implications for positive economic analysis, as the apparatus of behavioral economics has been brought to bear in explaining a number of empirical phenomena that are not consistent with standard rational choice models.[1] It also has implications for normative analysis. For example, limited self control may lead to overconsumption of alcohol and drugs and underinvestment in human capital. In situations like these individuals might benefit if an outsider induced them to behave according to preferences they wish they had. This outsider could be the government, and the inducements might be through tax and subsidy policies. A new kind of market imperfection, mistakes in individual behavior, brings us, then, to the realm of public economics—specifically, behavioral public economics.

Behavioral public economics is a rapidly expanding field whose central focus is on public policy when individual preferences differ from social ones.[2] O'Donoghue and Rabin (2003) consider optimal paternalistic taxes that the government imposes to correct individual behavior regarding consumption of harmful goods. Sheshinski (2003) proposes a general model with faulty individual decision making, where restricting individuals' choices leads to welfare improvements. Kanbur et al (2004) examine taxation under income uncertainty when individuals behave according to the tenets of prospect theory, but the government uses expected utility theory to evaluate the outcomes of this behavior. The situation in the normative part of this research agenda is, therefore, one where market behavior is generated by one set of preferences, but the society evaluates it with respect to another set of preferences.

---

[1] For surveys of the literature, see Camerer and Lowenstein (2004) and Rabin (2002).

In many respects, the situation described above is fairly common in welfare and normative public economics. Perhaps the most well-known example is the analysis of so-called merit goods (Sandmo 1983, Besley 1988). The consumption of these goods, in the viewpoint of the government, is meritorious and should be encouraged or imposed, ignoring individual choice. Optimal taxation when the government attempts to alleviate poverty (e.g. Kanbur et al 1994a) is another application of a much larger literature on "non-welfarist" public economics, where the social planner explicitly uses some other criterion for evaluating an individual's welfare than the preferences of that individual.[3]

The object of this paper is to provide a unified framework for non-welfarist optimal taxation, expanding the seminal work by Seade (1980), and to then view the recent interest in behavioral public economics in light of this framework. It will be seen that the general results of the non-welfarist public economic literature provide a useful guide and framework for developing the specific analysis called for by the new behavioral economics. We will not touch upon the question of how one can make reliable inference on individual utility when decision making contains mistakes and utility is time dependent. This serious and extremely difficult question is discussed in depth by Bernheim and Rangel (2004). Rather, we will take the two sets of preferences—the ones individuals have and the ones they wish they had, or in any event the ones the government evaluates outcomes with—as given and examine their consequences for optimal taxation.

---

[2] A general discussion is to be found in Camerer et al (2003).

[3] Perhaps at some level one could also argue that redistribution – where the government can evaluate individual welfare in a different way than the individuals themselves – and correction of externalities are additional examples in which the social welfare function differs from the individual utility.

The plan of this paper is as follows. Section 2 first presents a general model of non-welfarist optimal non-linear taxation. It highlights in particular the difference between the standard second best case for distortionary taxation from the paternalistic case when private and public preferences differ. It then illustrates specific analyses in the literature as special cases of the general formulation. Section 3 presents a general model of non-welfarist mixed taxation, where income is taxed on non-linear scale and commodities on a linear scale. It also discusses merit goods and commodity taxation. Section 4 turns to behavioral public economics and shows how recent discussions fit into the standard non-welfarist framework. Section 5 concludes.

2. Non-welfarist optimal non-linear income taxation with two goods

2.1 The general model

The purpose of this section is to provide a general non-welfarist formulation of income tax problem which unifies special cases which have been studied in non-welfarist tax literature. The aim is to bring out their common structure and results. We concentrate here on a general case of non-linear taxation, but to make the arguments clear, we examine a two good case (e.g. labor and leisure).[4] The analysis builds on the information-based approach to optimal tax policy, initiated by Mirrlees (1971), where the availability of instruments is restricted on the basis of what the government can observe. The income-earning ability of taxpayers is hidden information, but the government can observe income and design a general, non-linear, tax schedule based on that.

[4] Non-linear taxation with many goods would yield essentially similar results. The results differ more between completely non-linear and mixed taxation case. Mixed taxation is dealt with in Section 3.

There is a continuum of individuals, each having the same preference ordering, which is represented by a utility function $u = u(x, y)$ over consumption *x* and hours worked *y*, with $u_x > 0$ and $u_y < 0$. Individuals are otherwise identical, but they differ in their income-earning ability, or the wage rate, *n*. Workers differ only in the pre-tax wage *n* they can earn. There is a distribution of *n* on the interval $(\underline{n}, \bar{n})$ represented by the density function $f(n)$. Gross income is given by $z = ny$.

Individuals maximize utility subject to the budget constraint:

$$\max_{x,y} u(x, y) \text{ subject to } x = ny - T(ny), \tag{1}$$

where *T* depicts the non-linear tax schedule set by the government. The necessary condition of (1) is given by

$$u_x(1 - T') + u_y / n = 0, \tag{2}$$

where T' depicts the marginal tax schedule set by the government. This individual optimization condition gives the self-selection constraint for the government optimization problem. Totally differentiating utility with respect to *n*, and making use of workers utility maximization condition, we obtain the incentive compatibility constraints,

$$\frac{du}{dn} = -\frac{yu_y}{n} \equiv u_n(x, y, n),^{5} \tag{3}$$

[5] The first-order condition of individual's optimization problem is only a necessary condition for the individual's choice to be optimal, but we assume here that it is sufficient as well. Assumptions that assure sufficiency are provided by Mirrlees (1976). Note also that while we here presume an internal solution for *y*, (3) remains valid even if individuals were bunched at $y = 0$ since, for them, $du/dn = 0$.

In sum, the way the individual optimization is modelled is therefore completely similar to the approach in welfarist tax literature. Note that below we will use the same model to examine behavioural economics applications where individuals can make mistakes. Yet, it is assumed that individuals are perfectly rational when assessing the self-selection constraint. That individuals can make mistakes with respect to incentive compatibility constraint as well is clearly a somewhat different topic from the one we consider. This is examined further in Sheshinski (2002).

It is usual in optimal tax theory to assume an additively separable individualistic welfare function. One can of course allow for any increasing transformation of individual utilities here, so as to capture a greater or lesser concern with inequality on the part of the government. Suppose, therefore, that the aim of policy can be expressed as maximizing the following social evaluation criterion (allowing for non-individualistic preferences)

$$S = \int_{\underline{n}}^{\bar{n}} P(x, y, n) f(n) dn \,, \tag{4}$$

where $P = P(.,n)$, following Seade (1980), is "the social utility" derived from an *n*-individual's consumption and labor (leisure), which may in particular coincide with, or be related in some special form to, $u(.,n)$. S is restricted to be additively separable in individual utilities, but the formulation still allows e.g. the social welfare to depend on any linear form on utilities or on specific goods such as income.[6]

[6] The individualistic form of the welfare function has been criticized, most notably by Sen (1985), as unable to meet in many instances common-sense notion of equality, which would generally relate to distribution of consumption, i.e. directly to quantities not necessarily through utilities (non-welfarism).

The government cannot observe individuals' productivities and thus is restricted to setting taxes and transfers as a function only of earnings, $T(z(n))$. Inverting direct utility then gives $x = h(u, y)$, where

$$h_y = -\frac{u_y}{u_x} \quad , \qquad h_u = \frac{1}{u_x}, \tag{5}$$

Defining, too, $g(u, y, n) = u_n(h(u, y), y, n)$, it is straightforward to check that

$$g_y = -nu_x s_n, g_u = \frac{u_{nx}}{u_x}, \tag{6}$$

where we have defined the variable $s = -u_y(x, z/n)/nu_x(x, z/n) > 0$ to denote the marginal rate of substitution between x and y. Preferences are taken to satisfy the further restriction that $s_n < 0$. This is assumption B of Mirrlees (1971) and the Agent Monotonicity assumption of Seade (1982). It implies that indifference curves in consumption-gross income space become flatter the higher is an individual's wage rate, which in turn ensures that both consumption and gross earnings increase with the wage rate.

Since $T = ny - x$, we can think of government as choosing schedules *y(n)* and *x(n)*. In fact it is easier to think of it choosing a pair of functions, *u(n)* and *y(n)*, which maximizes index (4) subject to the incentive compatibility condition (3) and the revenue requirement $\int T[z(n)]f(n)dn = R$. Introducing multipliers λ and $\mu(n)$ for the budget constraint and incentive compatibility constraint, and integrating by parts, the Lagrangean becomes

$$L = \int_{\underline{n}}^{\bar{n}}((P(x, y) + \lambda(ny - x))f(n) - \mu' u - \mu g)dn + \mu(\bar{n})u(\bar{n}) - \mu(\underline{n})u(\underline{n}), \tag{7}$$

Differentiating with respect to *u* and *y* gives the first-order conditions

$$L_u = (P_x - \lambda)h_u f(n) - \mu'(n) - \mu(n)\frac{u_{nx}}{u_x} = 0\,, \tag{8}$$

$$L_y = (P_x h_y + P_y + \lambda(n - h_y))f(n) + \mu(n)nu_x s_n = 0\,, \tag{9}$$

Dividing (9) by λf, using (2) and (5) and rearranging, (9) becomes

$$T'(z(n)) = (P_x(s - \frac{P_y}{nP_x}))/\lambda - \mu(n)u_x s_n / \lambda f\,, \tag{10}$$

where

$$\mu(n) = \int_{\underline{n}}^{n}(P_x + \lambda)(1/u_x)\exp(-\int_{p}^{n}u_{nx}/nu_x)f(p)dp.\,, \tag{11}$$

is the multiplier on the incentive compatibility constraint. This latter satisfies the transversality conditions

$$\mu(\underline{n}) = \mu(\bar{n}) = 0\,, \tag{12}$$

and

$$\mu(n) > 0\,, \text{ for } \; n \in (\underline{n}, \bar{n})\,, \tag{13}$$

The optimal marginal tax rate formula (11) can be rewritten in a slightly different form in comparison to the original Mirrlees (1971) optimal tax model

$$T'(z(n)) = (P_x(s - s^p))/\lambda - \mu(n)u_x s_n / \lambda f\,, \tag{14}$$

where $s^p = -\frac{P_y}{nP_x}$ denotes the social (paternalist) marginal rate of substitution. The second term at the right is familiar from the welfarist literature, whereas the first term is novel. It cap-

tures the social value of divergence between private and social preferences, and is therefore called the paternalistic motive for taxation. It could also be called a first-best motive for taxation, as it corrects the individual activity to correspond to social preferences. The conventional term, the second at the right of (14), represents in turn the second-best motive for marginal distortion, arising from the asymmetric information.

In the end points of income distribution, the second term at the right is zero, and the marginal tax is completely determined by paternalistic motives. Suppose, for instance, that the social planner regards very high incomes unwanted *per se*. In this case $s^p > s$. Therefore, the marginal tax rate at the top is positive, despite the fact that this policy is not Pareto efficient. The marginal tax rate is used as a device to correct 'unwanted' social outcomes.

The sign of the marginal rate will depend on the interaction between these terms. We might think of a government with redistributive goals, but its views on working are more "Calvinistic" or "puritanical" than taxpayers so that it would like to see people work harder and earn more. In his case $s^p < s$. As is known from Mirrlees (1971) the second term implies a non-negative marginal tax rate. The first term in turn implies a marginal subsidy as a incentive to promote labor supply. At the top the marginal tax rate is negative. Hence the property of welfarist optimal income tax – the non-negativity of marginal rate – no longer holds.

2.2 Special cases

Poverty reduction

Much of the attention of non-welfarist approaches has focused on a particular form of non-welfarism, namely poverty reduction. Policy discussion on poverty alleviation and the target-

ing of social policy often concentrates almost exclusively on income. Little weight is typically given to issues like the disutility the poor experience when working. Indeed, sometimes work requirements are seen in a positive light, as is often the case with workfare. This is in marked contrast with conventional, utility-based, objectives in optimal income taxation literature. Therefore it is worthwhile to examine the implications of poverty reduction objectives on optimal income tax rules.[7] It must also be remembered that the dividing line between welfarism and non-welfarism is not very clear. Conventional tax analysis utilizes social welfare functions with inequality aversion, which already implies a deviation of assessing individual welfare with the same function which the individual uses himself. In some sense, the social objective functions form a continuum in the welfarism – non-welfarism scale.

Kanbur, Keen and Tuomala (1994a) examine the properties of the Mirrlees-type optimal income tax model, when the government objective is alleviation of income poverty. [8] Instead of social welfare maximization, the government aims to minimize an income-based poverty index of the general additively separable form

$$S = \int G[x(n), x^*] f(n) dn \,, \tag{15}$$

where x^* is the poverty line. G is non-negative for $x < x^*$ and zero otherwise. It satisfies the following properties

$$G_x < 0, G_{xx} > 0 \;\forall\; x \in (0, x^*) \,. \tag{16}$$

[7] The literature makes clear that it does not necessarily advocate these objectives; rather the aim is to explore their implications.

[8] Kanbur and Keen (1989) analyse what kind of linear income tax schedules could be used to alleviate poverty, while Besley and Kanbur (1988) analyse commodity tax/subsidy rules (when no income taxation is available) for poverty alleviation. Kanbur, Keen and Tuomala (1994b) and Bradbury (2002) offer surveys.

This specification captures a number of widely-used poverty measures, such as the headcount ratio and the Gini-based measure of Sen (1976). Note that while it has a similarity with a Rawlsian social welfare function (focusing on the poor), poverty index depends only on income. In the Rawlsian difference principle, an individual's well-being is judged according to an index of primary goods.[9] The social evaluation function (4) reduces to (15). That is $P(x, y, n) = G(x, x^*)$. Here $P_y = 0$ and $P_x = G_x$.

The government minimizes (15) subject to the self-selection constraint and the government budget constraint. The optimal marginal tax rate in (11) now becomes

$$T' = \frac{G_x s}{\lambda} - \frac{\mu u_x s_n}{\lambda f}, \tag{17}$$

where λ and μ denote the Lagrange multipliers of the budget constraint and incentive compatibility constraint, respectively, and s is the marginal rate of substitution between consumption and income. The second term at the right of (17) is similar to the marginal tax rule derived in a welfarist setting, with the exception that terms are evaluated at a different optimum. The first term at the right is novel and captures poverty minimization objectives. At the lower end of income distribution this term is negative ($G_x < 0$), pointing to lower marginal tax rates on the working poor. However, because the interaction with the other terms, one cannot at the analytical level compare the income tax rates to those derived in welfarist framework.

A clear-cut result emerges at the lowest end of the income distribution. If some amount of work is always desirable,[10] the second term at the right vanishes. This gives rise to the obser-

[9] Economists have, however, narrowed Rawls's theory into one which allocates according to 'maximin utility'.

[10] This is the so-called no bunching case.

vation in the welfarist model that the marginal tax rate at the bottom of the income distribution is zero.[11] However, in the poverty alleviation case, the first term at the right remains, and the marginal tax rate for the lowest earner is negative. Over some interval at the bottom of the wage distribution, the marginal tax rate derived in the poverty alleviation case is therefore lower than in the conventional welfarist case. This policy, via inducing the poor to work and earn more, contributes to poverty reduction. The finding is potentially important in policy terms, motivating the use of wage subsidies (such as the earned income tax credit in the US). Notice that the policy outlined above would not necessarily raise welfare, because of the forgone leisure. Its desirability arises from the fact that the social planner does not evaluate its policy based on individual utility, but uses a different, non-welfarist notion.

Bradbury (2002) points out that policy discussion often goes beyond this, giving a negative weight to leisure. One reason for this is paternalism. Compulsion to work may be seen as the individuals' best interests, for instance because of learning-by-doing reasons that the individuals fail to see. Another reason is related to notions of obligation and reciprocity. The recipients of the welfare benefits have 'no rights without responsibilities'. They may have a responsibility to work to be entitled to social welfare programs, irrespective of the desirability of the work for themselves.

Other non-welfarist optimal tax analysis

Schokkaert et al (2003) examine in more detail the consequences of non-utilitarian motives for optimal income taxation in a framework where individuals differ in two respects: their in-

[11] The marginal tax rate at the higher end of income distribution is also zero. This conclusion holds also in the poverty reduction framework (inasmuch the highest earner is not poor).

come-earning ability (as in the conventional tax model) and in their taste for leisure. Here the social planner may have a different idea than the individuals themselves about the 'correct' or 'reasonable' preferences for leisure. The social planner may, for instance, want to restrict the hours worked to protect the workers from exhaustion or to impose limits to work (and consumption) for ecological reasons. The latter motivation can also be related to quality-of-life vs. material welfare considerations.

They assume that individual preferences between income and labour supply take the following, quasi-linear, form:

$$u(x, ny) = ny - \frac{1}{e}\frac{\varepsilon}{1+\varepsilon} y_0^{-\frac{1}{\varepsilon}} y^{\frac{1+\varepsilon}{\varepsilon}}, \tag{18}$$

where ε is the constant elasticity of labor supply and e represents a idiosyncratic taste parameter for leisure. The social planner, on the other hand, evaluates welfare using an 'advantage' function

$$a(x, ny) = ny - \frac{1}{g}\frac{\varepsilon}{1+\varepsilon} y_0^{-\frac{1}{\varepsilon}} y^{\frac{1+\varepsilon}{\varepsilon}}, \tag{19}$$

where individual preferences for leisure, e, are replaced by social preferences, g. If g reaches infinity, the social welfare depends on income alone, while $g < e$ represents the case, discussed above, where social planner attaches larger weight to quality of life than the individual.

The purpose of Schokkaert et al (2003) is to compare how optimal linear income tax derived using the advantage function differs from a welfarist solution, calculated using Rawlsian social welfare function. A decrease in g leads to higher tax rates, because the social planner at-

taches a higher disadvantage to labor, which it therefore wants to discourage more. Using an illustration based on Belgian data, they demonstrate how these considerations can have a sizable effect on the desirable tax rate, if labor supply elasticity is small enough.[12]

In terms of our general non-linear non-welfarist formulation of income tax problem, the case studied by Schokkaert et al (2003) would mean that $P_x = u_x$ but $P_y \neq u_y$. Now the marginal tax formula (11) becomes

$$T'(z(n)) = (u_x(s - \frac{P_y}{nu_x})) / \lambda - \mu(n) u_x s_n / \lambda f \,. \qquad (20)$$

Now suppose $-\frac{u_y}{nu_x} < -\frac{P_y}{nu_x} \Leftrightarrow \frac{u_y}{nu_x} > \frac{P_y}{nu_x}$. Hence this leads to higher marginal rates, because the government discourages labor supply.

3. Non-welfarist optimal mixed taxation

3.1 The general model

This section considers a mixed taxation case where income is taxed in a non-linear fashion, but commodities are taxed on a linear scale. Thus, we analyze a similar situation than in Mirrlees (1976) but with a non-welfarist government objective. The tax policy tools include a non-linear income tax $T(ny)$ and commodity taxes (tax vector) $t = q - p$, where $p = (p_1, p_2, ...) =$ producer's prices and $q = (q_1, q_2, ...) =$ consumer's prices. An individual n's

[12] The paper also departs from the traditional welfarist literature by considering 'responsibility-sensitive' egalitarianism, due to Roemer (1998), where individuals should only be compensated for differences in their innate skill levels, while they should be responsible for their preferences for leisure. Introducing these concerns leads typically to smaller tax rates than in the welfarist case.

budget constraint is $qx = z - T(z)$, where x is a vector of commodities subject to linear taxation.

The consumer's optimization problem remains the same as above, with the modification that a given income can now be spent over multiple commodities. The government optimizes the non-welfarist objective function by choosing linear commodity taxes and non-linear income tax optimally, subject to a self-selection constraints of the individuals and a budget constraint $\int \{T[z(n)] + tx(q,n)\} f(n)dn = R$.

Household optimization will be used to generate the incentive compatibility constraint for the government optimization. In the case where one good only is subject to non-linear taxation, an n-individual maximizes u subject to $qx = y - T(z)$. Define utility as a function of the optimally chosen commodities (satisfying the first-order conditions of individual optimization) $v(n) = \max u(x,z,n)$. Differentiating this function with respect to n and combining this with the first order conditions of individual optimization, we have the familiar envelope condition as in (3)

$$\frac{du}{dn} = -\frac{yu_y}{n} \equiv u_n(x,z,n), \tag{21}$$

Because of the need to deal with both non-linear and linear price structures, it is helpful to apply dual techniques to solve the optimization problem. We utilize partial expenditure and indirect utility functions, first discussed by Mirrlees (1976). Let the expenditure function for household be $E(q,z,n,v) = \min\left[qx : u(x,y,n) = v\right]$ and the partially indirect utility function $v(q,b,z,n) = \max\left[u(x,z,n) : qx = b\right]$, where expenditure on linearly-taxed goods is $b = E$.

By substituting Hicksian demand ($E_q = x^c$) into (21) we can eliminate *x* from (21). The resource constraint for this economy is

$$\int (z - px^c) f dw = A, \tag{22}$$

where $x^c = x^c(q, z, v, n)$ $(= E_q)$. The Lagrangean of the government optimization problem can then be written as

$$\begin{aligned} L &= \int \{(P[x^c(q,z,v,n),z] + \lambda(z - px^c))f + \mu v' + \alpha u_n\} dn \\ &= \int \{(P[x^c(q,z,v,n),z] + \lambda(z - px^c))f - \mu' v + \alpha u_n\} dn - \mu(\underline{n})v(\underline{n}) + \mu(\bar{n})v(\bar{n}) \end{aligned}, \tag{23}$$

where the latter formulation follows from integrating $\mu v'$ by parts. Maximizing with respect to *q* yields the following first-order condition

$$\int P_x x_q^c f dw - \int \{\lambda p x_q^c f + \mu \partial(u_n / \partial q)\} dn = 0, \tag{24}$$

where $P_x = \dfrac{\partial P}{\partial x^c}$. Equation (24) can be rewritten as[13]

$$t \int x_q^c f dn = -\int \pi(n) x_n(q, b, y, n) dn - \int \frac{1}{\lambda} P_x x_q^c(q, y, v, n) f dn, \tag{25}$$

where $\pi = v_E \mu / \lambda > 0$.[14] The expression in (24) is an implicit formulation for the optimal commodity tax structure. The left-hand side of this formulation measures, as pointed out by Mirrlees (1976), the extent to which commodity taxation encourages/discourages consumption of different commodities. The first term on the right is similar than in Mirrlees (1976). It

[13] See Pirttilä and Tuomala (2004) for details.

[14] The income tax is also assumed to be optimally chosen.

links the 'index of discouragement' at the left to the differences in consumption of a particular good among people with different abilities, *n*.

3.2 The poverty minimization case

The second term at the right is novel. To interpret it we take the case of minimization of poverty as in Pirttilä and Tuomala (2004). The social welfare maximization is now equal to minimizing a poverty/deprivation index, which must now be extended the capture the many-good situation and it is given by

$$P = -\int D\left[c, \pi x(q,n)\right] f(n)dn\,, \tag{26}$$

where $c = \pi^* x^*$ is a reference consumption bundle to which actual consumption level πx is compared. Consumer prices are depicted by *q,* and π denotes the shadow prices used in poverty measurement.[15] As earlier, $D_x < 0$.

Consider a case where good *i* is included in the deprivation measure and the tax (consumer price) of good *j* is increased. Then the index of discouragement at the left measures the discouragement of the consumption of *j*. If these goods are complements, then $x_q^c < 0$, and the consumption of good *j* is encouraged. Likewise, if *i* and *j* are substitutes, i.e. $x_q^c > 0$, the consumption of good *j* is discouraged through the tax system. Finally, since the compensated own price effect is always negative, the consumption of goods that itself enter the deprivation measure should be encouraged.

---

[15] Technological reasons would suggest using producer prices *p*, so that $p_x = p = q - t$, where *t* denotes commodity taxes. Emphasis on the purchasing power of the poor would support the use of consumer prices. But there can be other weights attached to different commodities. One may include only some necessities with their producer prices, but goods that are not included in the target vector have zero weights.

The intuition for the second term at the right of (25) is straightforward. If a good is included in the deprivation index, a decrease in its price leads to an increase in its consumption, and thus to a reduction in poverty. Likewise, setting a relatively low (high) tax for goods that are complements (substitutes) with goods in the deprivation measure reduces poverty indirectly. The interpretation of the first term of the right hand side of (25) is completely similar to earlier tax analysis. The government is still constrained by asymmetric information, and it must design its tax schedules so that individuals' incentive compatibility constraints are not violated. [16]

In terms of tax rates, commodity taxes should be the highest for goods for which the high-ability household have a relatively strong taste and that are substitutes with goods in the poverty measure. Commodity taxes should be the lowest for goods for which the low-ability households have a relatively strong taste and that are included in the deprivation measure or are complements with goods in the poverty measure.

3.3 The Atkinson-Stiglitz separability result

As shown originally by Atkinson and Stiglitz (1976), the incentive effect vanishes if consumer preferences are separable between goods and leisure. In this case, the demand of different commodities does not vary with the wage rate (or labour supply), and the first term at the right of (25) is always zero. However, even with separable preferences, the second term at the right in (25) is still positive or negative.

The Atkinson-Stiglitz result is often used as an argument against the use of differentiated commodity taxation as a redistributive device. Direct income transfers (as a part of an optimal income tax scheme) would be sufficient instead. In the present context, there is no reason to suppose that influencing income is better than affecting the consumption of the commodities. The poverty index depends directly on the consumption of some the commodities, and it is in the interest in the government to promote their consumption. This also implies that income-based targeting is not necessarily superior to targeting based on consumption goods.[17]

The fact that the Atkinson and Stiglitz (1976) separability result does not hold remains valid also in other non-welfarist formulations. While the second-best arguments would not require differentiated commodity taxation, the first-best term implied by non-welfarist objectives is still needed to correct differences between private and social value of consumption.

3.4 Effective marginal tax rates

To obtain the necessary conditions for the effective marginal tax rates, (23) is differentiated with respect to v and z:

---

[16] Christiansen (1984) shows that goods that are negatively related to labour supply should be taxed relatively more. Holding income constant, a reduction in hours worked can be achieved by an increase in skills. Therefore, a good for which people with higher abilities have stronger taste is negatively related to labor supply.

[17] Note finally that these results can also be linked to the taxation of savings. When different commodities are interpreted as consumption in different points in time, the Atkinson and Stiglitz (1976) result implies that savings should not be taxed. But when the government objective is poverty minimization, the tax schedule of savings also depends on which commodities are included in the poverty measure. A plausible case in practice is one where the poverty index is measured based on current consumption. This measurement, which can be defended at least if poverty is transitory, would imply a relative encouragement of present over future consumption, in other words, a positive tax rate on savings.

$$P_x x_v^c f - \lambda p x_v^c f + \mu\{\partial(u_n)/\partial v\} - \mu' = 0\,, \tag{28}$$

$$\mu(\underline{n}) = \mu(\bar{n}) = 0\,, \tag{29}$$

$$(P_x x_z^c + P_z + \lambda(1 - p x_z^c))f + \mu\{\partial(u_n)/\partial z\} = 0\,. \tag{30}$$

The main condition for optimality, (30), may be rewritten as

$$(1 - t x_b)s + 1 + t x_z = -\frac{1}{f}\pi s_n + \frac{1}{\lambda}P_x(s - s^p)\,, \tag{31}$$

where s is defined to be the marginal rate of substitution between z (=ny) and expenditure on goods; b, that are taxed on linear scale, i.e. $s(x,z,n) = v_z / v_b = -E_z(q,v,z,n)$ and $s^P = P_z / P_x$ is the paternalist marginal rate of substitution. As in Mirrlees (1976), the left-hand side of (31) measures the total increase in the tax liability (including commodity taxes and the income tax), or the effective marginal tax rate, of a household when income increases.

Consider the end point at the top of income distribution. Then the transversality condition in (29) implies that the first term at the right of (31) is zero. Assuming that $s^p > s$, then the second term is positive in (31). In other words the effective marginal tax rate is positive. When the government minimizes poverty, the second term at the right of (31) takes the form $-D_x p c_z^c$. Pirttilä and Tuomala (2004) show that the standard result in optimal tax analysis – there should be no distortion at the top – carries over to the present case with poverty minimization, whereas at the bottom of the income distribution, the marginal effective tax rate should be negative.

3.5 Merit goods and commodity taxation

Familiar arguments for public intervention include distributional concerns and the existence of market failures. The notion of merit goods, initiated by Musgrave (1959), is used as another motivation for public intervention that is distinct from those cases above. Examples for merit good arguments are easy to find in reality. Compulsory education is perhaps the most-well known example of merit goods, whereas banning drug use is used to protect consumers from a harmful demerit good. In all such arguments, the principle of consumer sovereignty is ignored. The government's intervention is thought to be justified, since consumers make faulty choices. Public policy is then designed to correct consumers' choice, often against their will.

First-best commodity tax rules for merit goods, derived in the situation where there is no need to resort to distortionary taxation, are directly targeted to correct the difference between private and social valuations of these goods. In second-best situation with distortionary linear taxation, Ramsey-type rules emerge. Consumption of commodities that are complements with the merit goods should be encouraged, while substitutes should be discouraged (see e.g. Besley 1988).

Racionero (2001) considers linear commodity taxation in the presence of merit goods when the government has access to non-linear income taxation as well.[18] She utilises a merit good modeling due to Besley (1988), where individuals disregard the beneficial impact of con-

[18] Racionero (2000) examines the case where individuals also differ in their preferences over the merit good, but government only utilises income taxation.

sumption of one good on health, whereas the health effect is taken into account in the government's assessment of individual welfare. Assuming that preferences are weakly separable between consumption and leisure – when no commodity taxes would be needed without merit good considerations – there should still be a subsidy on the consumption of the merit good. The size of the subsidy is shown to be a sum of two elements. It depends, first, on the average of the marginal effects on health over individuals of different income level. Second, a covariance term emerges, which measures the dispersion of the marginal effects on health across population. If, for instance, workers with low income-earning ability are more sensitive to the subsidy (increase the consumption of the merit good relatively more when subsidized), the subsidy tends to be higher.[19]

Using the technique of section 2 we can formulate the merit good optimization procedure. Suppose that the individuals do not care about additional positive effects of certain goods on health, while the government does. This divergence can be expressed in the following way[20]

$$u^g = u(x, m, y) + h(m), \tag{32}$$

where u^g reflects government's preferences and u refers to individuals' preferences. $h(m)$ denotes the health function ($h'>0$ and $h''<0$).

Using partially indirect utility functions we write the government's welfare function as follows

[19] Racionero (2001) also demonstrates how merit good concerns affect the optimal (effective) marginal tax rates on income. The effective marginal tax rate at the top of the income distribution should be negative, while its sign is ambiguous at the bottom of the distribution.

[20] Similar modelling has been used by Racionero (2001) and, more generally, by Besley (1988).

$$W = \int (v(q, y, b, n) + h(m)) f(n) dn \,, \tag{33}$$

where $b = \sum q_i x_i + q_m m$. Now with weakly separable preferences we can derive the implicit commodity tax formula for a merit good

$$t^m \int m_q^c f dn = -\int \frac{1}{\lambda} h'(m) m_q^c f dn \,, \tag{34}$$

where m^c is compensated demand. The left hand side of (34) measures the extent to which commodity taxation encourages/discourages consumption of merit good. The term on the right hand side measures the impact of health effect of merit good. Since $m_q^c < 0$, the term is positive, suggesting that the consumption of merit good should be encouraged. In terms of tax rates, commodity tax on merit good should be low or negative (a subsidy).

4. Behavioral Public Economics

This section discusses some recent ideas in normative behavioral economics and attempts to show their direct connection to the general non-welfarist structure developed in the paper. We shall see that the key feature of exercises in behavioral public economics has been to highlight a term in optimal taxation formulae that captures the impact of actual individual preferences being different from what the individual would wish them to be (and therefore what he or she would want the government to use in formulating policies to change behavior).

4.1. Pensions

One area where public economists have traditionally built on behavioral assumptions is analysis of pensions. Insufficient savings by workers for their retirement can be seen as one key argument for public pension systems or compulsory pension contributions.

Diamond (2003) offers an excellent synthesis on public economics viewpoints on pension policy. Therefore, very brief notes on some aspects on the literature suffice here. Diamond (2003, chapter 4) and Diamond and Mirrlees (2000) consider a benchmark situation where individuals do not save at all. Workers are otherwise identical, but their skills differ (as in Mirrlees 1971), and the government's objective is to design optimal redistributive policy for the working age and for the retired. Another assumption is myopic labor supply by young workers, who simply ignore the implications of their earnings when young on the retirement income.

A specification that gives rise to striking conclusions is one where individual utility is additive in the following way:

$$u = v_1(x) + v_2(c) + w(1 - y), \tag{35}$$

where x and c denote consumption when young and when retired, respectively, n is the wage rate and y is labor supply when young. Myopic labor supply implies that retirement consumption does not enter the incentive compatibility constraint. Therefore, if the social welfare function exhibits inequality aversion, the optimal retirement consumption is shown to be higher for those whose lifetime income has been smaller.

A more plausible policy rule arises from a framework where another assumption with behavioral motivations is made. Suppose that preferences are not additive over time; moreover there

is a standard-of-living effect from the first period consumption on the utility from the retirement consumption. Preferences could then be given by

$$u = x + v(c, x) + w(1 - y)\,. \tag{36}$$

Equation (36) is used by the government when transforming individual welfare to social welfare. Individuals, in turn, ignore the impacts on second-period consumption when making labor supply decisions. They therefore maximize apparent utility given by

$$u = x + w(1 - y)\,, \tag{37}$$

The government's optimization is therefore constrained by a self-selection constraint that depends on the apparent utility alone. In an extreme case, the second period utility could only depend on the replacement rate c/x. It is shown that in this case, the optimal replacement rate is decreasing in n.[21] There are also more refined formulations on the impacts of realistic, behavioral, assumptions on pension policy. An example is Diamond and Köszegi (2002) who explicitly model the underlying reason for myopia by building on quasi-hyperbolic discounting.

Using the tools of non-linear income taxation Diamond (Ch 4, 2003) derives the marginal tax formula for the first-period income

$$T' = \frac{W' v_x s}{\lambda} - \frac{\mu u_x s_n}{\lambda f}\,, \tag{38}$$

where W' is the derivative of the social welfare function with respect to individual utility. The rule above is similar to one presented in equation (14) of section 2. Again the difference be-

tween the optimal tax rate in this type of world and that under conventional theory is the first term on the right hand side. It is a first-best motive for taxation. In (41) this term corrects internality because individuals ignore the impacts on second period consumption in their labor supply decisions.

4.2. Reference incomes

Boskin and Sheshinski (1978), Oswald (1983) and Tuomala (1990), for example, consider the implication of utility interdependence (or 'envy') – the situation in which individual's utility is negatively affected by others' income – on optimal income taxation. There is nowadays ample evidence that people indeed care about their relative positions (see e.g. Blanchflower and Oswald 2004).

However, it is not clear whether utility interdependence should be allowed to enter the social welfare function: is envy a trait one wants to honor? For example, Harsanyi (1982) does not accept antisocial preferences such as envy, malice etc in a utilitarian social welfare function. Unlike earlier studies (Boskin and Sheshinski(1978), Oswald (1983) and Tuomala (1990)), we avoid here this criticism. Utility interdependence affects the way people behave, which the government must take into account as a constraint when designing tax schedules, but envy is not included in the government objective function.

All individuals are supposed to have the same tastes, represented by the utility function $u(x, y, \mu)$, where x is consumption, y is the amount of work done and μ a reference income

[21] Diamond (2003, chapter 6) also incorporates myopic behaviour to the analysis of retirement incentives when

level which depends on the aggregate income in the society. The government objective function takes the form $\int_{\underline{n}}^{\bar{n}} \hat{u}(x,y) f(n) dn$. In other words, it does not include reference income. Now we can reinterpret our model in section 2.

The optimal marginal tax rate formula can be written as follows

$$T'(z(n)) = (\hat{u}_x (s - \hat{s})) / \lambda - \mu(n) u_x s_n / \lambda f \,, \tag{39}$$

where s is again the (individual) marginal rate of substitution between consumption and income (including envy effect) and $\hat{s} = -\frac{\hat{u}_y}{n\hat{u}_x}$ denotes the social marginal rate of substitution. The second term at the right is again familiar from the welfarist literature, whereas the first term is novel. It captures the social value of divergence between private (including envy effect) and social preferences (no envy). It corrects the envy effect to correspond to social preferences.

4.3. Sin taxes

One reason why people can end up making choices against their own good is excessive discounting of future. This may result in e.g. overconsumption of goods which offer initial satisfaction but belated suffering. O'Donoghue and Rabin (2003) consider how a paternalistic government could respond to such a situation by designing appropriate, corrective, 'sin' taxes.[22]

---

all workers have the same skill level, but their disutility of labour differs.

[22] They use a variant of Ramsey taxes, i.e. linear commodity taxation.

We can capture some of the arguments developed by O'Donoghue and Rabin (2003) in the present, general, framework. Consider a case where all consumers have self-control problem. Utility is $u = u^*(x, a, z, n)$, where a is a "sin" good. (x is untaxed). All consumers have some degree of self-control problem so that there is an over-consumption of a. By contrast, optimal behavior maximizes $u = u^{**}(x, a, z, n)$, so that $a^* > a^{**}$. Otherwise the model is the same as the one used in section 3. Now we have

$$t_a \int a_q^c f dn = -\int \pi(n) a_n dn - \int \frac{1}{\lambda} P_a a_q^c f dn, \tag{40}$$

With weakly separable preferences (the first term on the right hand side is zero) we have $t_a > 0$, i.e. the consumption of the sin good should be taxed. If the first term of the right is non zero, the optimal commodity taxes are a combination of traditional welfarist concerns and the need to influence the consumption of harmful good.

An alternative formulation of sin goods might be one where the degree of irrationality is assumed to vary across individuals. As optimal taxation exercises where agents differ in two respects (as ability and tastes) are difficult, we concentrate on a simpler case where individuals do not differ in terms of their income-earning ability. Utility may now be defined by $u(x, a, \beta)$, where β is an index of irrationality, with density f. The government objective function takes the non-welfarist form $NW = \int_{\underline{\beta}}^{\bar{\beta}} \hat{u}(x, a) f(\beta) d\beta$. In other words $\hat{u}$ is the social utility derived from a β individual's consumption. Now we can reinterpret our model in section 2.

The optimal marginal tax rate formula can be written as follows

$$T' = (\hat{u}_x(s - \hat{s}))/\lambda - \mu(\beta)u_x s_\beta / \lambda f \ , \tag{41}$$

where s is again the (individual) marginal rate of substitution between a and x and $\hat{s} = -\frac{\hat{u}_a}{\beta \hat{u}_x}$ denotes the social marginal rate of substitution. The second term at the right is again familiar from the welfarist literature, whereas the first term is novel. It captures the social value of divergence between private and social time preferences. Suppose that for the most irrational individual we have $\hat{s}$>s so that society would like to see him to consume less of the sin good than he would choose to do at any given prices. At the optimum the relative price of x faced by this individual is lowered to discourage his consumption of a.

5. Conclusion

We have shown that non-welfarist optimal tax rules have an essentially simple common structure, with two key components. The first component captures the "first best" or "paternalistic" motive for taxation, because it arises from differences between social and private preference. The second component is the second best motive for taxation, to correct market distortions or to raise revenue in the least distortionary manner. Viewed in this light, exercises in behavioral public economics are seen to be applications of general non-welfarist public economics, with the focus on the first of the two components mentioned above. For whatever reason, individuals do not pursue their own best interests, which opens up the case for the government to intervene in order to induce them to do so. Thus the government uses a different set of preference from those generating individual behavior, which is precisely what is meant by non-welfarist welfare economics. Since behavioral public economics is one manifestation of non-welfarist public economics, it is not surprising that optimal behavioral tax rules have the same general structure as optimal non-welfarist tax rules. As behavioral economics expands, and as more results are derived for specific cases, we hope that our exposition will serve to provide a broad framework in which new results can be better appreciated, and better related to earlier results and to each other.

References

Atkinson, A.B. and Stiglitz, J.E. (1976) 'The design of tax structure: direct versus indirect taxation'. *Journal of Public Economics 6*, 55-75.

Bernheim, D.G. and Rangel, A. (2004) 'Behavioral public economics: welfare economics when individuals can make mistakes' Stanford University, mimeo.

Besley, T. (1988) 'A simple model for merit good arguments' *Journal of Public Economics* 35, 371-84.

Besley, T. and Kanbur, R. (1988) 'Food subsidies and poverty alleviation', *The Economic Journal 98*, 701-719.

Blanchflower, D and Oswald, A. (2004), 'Well-Being over Time in Britain and in the USA', *Journal of Public Economics 88*, 1359-86.

Boskin, M and E. Sheshinski (1978) 'Optimal income redistribution when individual welfare depends on relative income', *Quarterly Journal of Economics 92*, 589-602.

Bradbury, B. (2002) 'Targeting social assistance', Mimeo, University of New South Wales.

Camerer, C.F. and G. Loewenstein (2004) 'Behavioral economics: Past, present, future', in C.F. Camerer, G. Loewenstein and M. Rabin (eds.) *Advances in Behavioral Economics*, Princeton University Press.

Camerer, C.F. , S. Issacharoff, G. Loewenstein, T. O'Donoghue and M. Rabin (2003) 'Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism',' University of Pennsylvania Law Review, 151(3), January 2003, 1211-1254 .

Diamond, P.A. (2003) *Taxation, incomplete markets, and social security*. MIT Press.

Diamond, P.A. and J. Mirrlees (2000) 'Adjusting one's standard of living: Two period models', in P.J. Hammond and G.D. Myles (eds.) *Incentives, organisation and public economics, Essays in Honour of Sir James Mirrlees*, Oxford: University Press.

Diamond, P.A. and B. Köszegi (2002) 'Quasi-hyperbolic discounting and retirement', mimeo, MIT.

Harsanyi, J. (1982) 'Morality and the theory of rational behaviour', in A. Sen and B. Williams (eds.) Utilitarianism and Beyond, Cambridge University Press.

Kanbur, R. and Keen, M. (1989) 'Poverty, incentives and linear income taxation', in Dilnot, A. and Walker, I (eds.) *The economics of social security*, Oxford: Clarendon Press.

Kanbur, R., Keen, M. and Tuomala, M. (1994a) 'Optimal non-linear income taxation for the alleviation of income-poverty', *European Economic Review 38*, 1613-32.

Kanbur, R., Keen, M. and Tuomala, M. (1994b) 'Labor supply and targeting in poverty alleviation programs', *The World Bank Economic Review 8*, 191-211.

Kanbur, R., Pirttilä, J. and Tuomala, M. (2004) 'Moral hazard, income taxation, and prospect theory', University of Tampere, mimeo.

Mirrlees, J.A. (1971) 'An exploration in the theory of optimal income taxation', *Review of Economic Studies 38*, 175-208.

Mirrlees, J.A. (1976) 'Optimal tax theory. A synthesis'. *Journal of Public Economics 6*, 327-58.

Musgrave, R. (1959) *The theory of public finance*. New York: McGraw Hill.

O'Donoghue, T. and M. Rabin (2003) 'Studying optimal paternalism, illustrated by a model of sin taxes', *American Economic Review 93*, 186-191.

Oswald, A. (1983) 'Altruism, jealousy and the theory of optimal nonlinear taxation', *Journal of Public Economics 20*, 77-87.

Pazner, E. (1972) 'Merit wants and the theory of taxation', *Public Finance 27*, 460-472.

Pirttilä. J. and M. Tuomala (2004) 'Poverty alleviation and tax policy', *European Economic Review* 48, 1075-1090.

Rabin, M. (2002) 'A perspective on psychology and economics', *European Economic Review 46*, 657-685.

Racionero, M. del Mar (2000) 'Optimal redistribution with unobservable preferences for an observable merit good', *International Tax and Public Finance 7*, 479-501.

Racionero, M. del Mar (2001) 'Optimal tax mix with merit goods', *Oxford Economic Papers* 53, 628-641.

Roemer, J.E. (1998) *Equality of opportunity*, Harvard University Press.

Sandmo, A. (1983) 'Ex post welfare economics and the theory of merit goods', *Economica* 50, 19-33.

Schokkaert, E., D. Van de Gaer, F. Vandenbroucke and R. Luttens (2003) 'Responsibility-sensitive egalitarianism and optimal linear income taxation', Katholieke University Leuven, mimeo.

Seade, J. (1980) 'Optimal non-linear policies for non-utilitarian motives', in D. Collard, R. Lecomber and M. Slater (eds.) *Income distribution: the limits to redistribution*, Bristol: Scientechnica.

Seade, J. (1982) 'On the sign of optimum marginal income taxation', *Review of Economic Studies 49*, 637-43.

Sen, A. (1976) 'Poverty: An ordinal approach to measurement'*, Econometrica 44*, 219-231.

Sen, A. (1985) *Commodities and capabilities*, Amsterdam: North-Holland.

Sheshinski, E. (2002) 'Bounded rationality and socially optimal limits on choice in a self-selection model', The Hebrew University of Jerusalem, mimeo.

Sheshinski, E. (2003) 'Optimal policy to influence individual choice probabilities', The Hebrew University of Jerusalem, mimeo.

Tuomala, M. (1990) *Optimal income tax and redistribution*, Oxford: Clarendon Press.