



ELSEVIER

Journal of Public Economics 81 (2001) 51–71

JOURNAL OF
PUBLIC
ECONOMICS

www.elsevier.nl/locate/econbase

Redistribution when avoidance behavior is heterogeneous

Wojciech Kopczuk*

Department of Economics, University of Michigan, 611 Tappan, Ann Arbor, MI 48109, USA

Received 1 November 1999; received in revised form 1 March 2000; accepted 1 June 2000

Abstract

The optimal income taxation framework is extended to account for heterogeneity in avoidance behavior. The optimal tax formulae are shown to be simply and intuitively modified to reflect the value of taxable income as an indicator of the need for transfers at a given skill level. This effect is in addition to any effect implied by just considering additional margins of response to taxes. Alternative sources of heterogeneity and their likely consequences are discussed. When the choice of administration is added to the model, it is demonstrated that the optimal tax system may include some seemingly wasteful avoidance schemes, even if they could be eliminated without any cost. When an avoidance scheme is more heavily used by the needy, it may provide additional redistribution without affecting standard incentive constraints. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Optimal income taxation; Tax avoidance; Tax administration

JEL classification: H21; H24; H26

1. Introduction

The optimal income taxation literature trades off inefficiencies induced by distortionary taxation against redistributive gains. The usual source of inefficiency is the response of labor supply: a redistributive tax system decreases the relative price of leisure and thereby induces undesired substitution response. There is a consensus in the empirical literature that the labor supply response is small,

*Fax: +1-734-764-2769.

E-mail address: wojtek@umich.edu (W. Kopczuk).

seemingly implying that inefficiencies should be small as well. Clearly, however, the labor supply response is not the only behavioral response to taxation. There are numerous other margins which can be broadly described as tax avoidance.¹ Feldstein (1999) argues that the efficiency effects of all those behavioral responses can be summarized by the elasticity of taxable income.² There is a growing literature suggesting that this elasticity is substantial, at least for high-income individuals,³ implying that the efficiency costs of redistributive policies are likely to be high. Goolsbee (2000) and Gruber and Saez (1999) found that the elasticity of taxable income is much higher for the rich than for middle incomes, which suggests that tax avoidance is concentrated among the wealthier part of the population. However, there is also some conflicting evidence. Christian (1994) found that tax evasion, measured as a fraction of the 'true' income, decreases with the income level, and Sheffrin and Triest (1992) concluded that participation in tax evasion is not related to the income level. There are also types of tax avoidance that are concentrated among the poor, such as the black market economy. Lemieux et al. (1994) found that participation in this sector is significantly and negatively correlated with income. Finally, as I discuss in Section 2, there is quite convincing evidence that tax avoidance varies for reasons unrelated to the income level.

For the purpose of determining the optimal income tax structure, differences in avoidance behavior that vary systematically with the income level are simply accounted for by replacing the elasticity of labor supply by the (higher) elasticity of taxable income (as argued by Feldstein (1999) and Slemrod (1998)). However, the tax administration becomes of the foremost importance, because it may directly affect the excess burden of taxation, the point recently stressed by Slemrod and Kopczuk (2000). Finally, when the avoidance behavior is not homogeneous, the standard solution for the optimal taxes is no longer valid, and the possibility of using avoidance for redistribution arises.

Among other things, the effectiveness of redistributive policies that work through the income tax depends on how reliable taxable income is as a measure of the need for redistribution. The better it is an indicator, the better the income tax can serve its purpose of reducing inequalities. For example, when low taxable income may reflect either poverty (and the need for redistribution) or a high degree

¹I model tax avoidance as a choice under certainty. Cowell (1990) argues that tax evasion in the Allingham–Sandmo model may also be represented in such a way by introducing the explicit cost of risk taking. Concentrating on the choice under uncertainty, rather than tax avoidance, would obscure the feature of interest which is the behavioral response to taxation by fixing attention on the specific cost of risk-taking. It also makes one concentrate on particular, and not necessarily representative, policy instruments such as audits and penalties. Slemrod and Yitzhaki (1999) survey the literature on tax administration, evasion and avoidance.

²Slemrod (1998) discusses qualifications to this statement.

³Most of the studies find the elasticity between 0.5 and 1. See for example, Feldstein (1995), Moffitt and Wilhelm (1998), Auten and Carroll (1999), and Gruber and Saez (1999) (who also survey the literature).

of avoidance (because some individuals are better at tax sheltering than the others), this relation is weakened and, consequently, there should be less redistribution, as transfers will be directed not only to the poor, but also to the ‘avoiders’. Any other factors that affect the relation between taxable income and welfare have implications for the optimal progressivity. I consider alternative sources of heterogeneity in tax avoidance and analyze their consequences for the design of the optimal income tax. This is important for tax policy because it points out that concentrating on just vertical inequality may give misleading conclusions about the optimal level of progressivity.

When instruments other than the tax rates are added to the picture, it is demonstrated that it may be optimal for tax avoidance to persist, even if administrative costs of eliminating it are negligible. If a particular activity (even a seemingly wasteful one) is used by those who are ‘more deserving’ from the social point of view, it may be more efficient than redistribution via income tax. In the first-best optimum, the government would determine tax liabilities based on all characteristics of individuals, including skills and avoidance characteristics. The optimal taxation literature recognizes that skills are unobservable, so that one must instead rely on taxable income. Akerlof (1978) stresses that, if possible, one still should use other characteristics in the tax function. Avoidance characteristics are, however, unobservable, though they may affect individual behavior. Consequently, tax that an individual pays cannot depend on them, but an appropriate choice of avoidance possibilities may differentiate individuals’ opportunities, and, in some cases, provide additional transfers. Such a possibility does not arise if avoidance is only a function of income, so that the additional kind of heterogeneity is crucial for the result. Redistributive benefits of tax avoidance can help to explain why the real-life tax systems are complex and offer numerous opportunities for tax evasion and avoidance, even though clarifying changes may not be very expensive.⁴

In the models in this paper, society is differentiated along two dimensions: skill level and an exogenously given avoidance characteristic, which will be referred to as ‘vertical’ and ‘horizontal’ differences, respectively. In Section 2, possible sources of heterogeneity in avoidance and corresponding modeling strategies are discussed. In Section 3, the solution to the optimal linear income tax problem is characterized. It is interpreted as showing the trade-off between achieving horizontal and vertical equity. The key to determining how the optimal progressivity is affected by the ‘horizontal’ differences is how well the taxable income indicates who is deserving on the margin. Tax administration is added in Section 4. It is shown that, with positive administrative costs, avoidance should never be eliminated. On the other hand, when the skill level is the only feature varying across individuals (there is no heterogeneity in avoidance) and a general income tax is feasible, avoidance schemes have no value as a policy instrument,

⁴The administrative costs of removing deductions do not need to be high and that is what I am interested in here. Clearly, there are also public choice issues involved that I do not address.

and should be eliminated if administrative costs are negligible. More intriguingly, it is demonstrated by means of two simple examples that when avoidance behavior is heterogeneous, introducing avoidance schemes may be welfare improving even with nonlinear taxes and no administrative costs. This is because they can help sort out individuals according to their unobservable characteristics that cannot be present explicitly in the tax function. It is argued that similar logic stands behind the Stiglitz (1987) result on the possible desirability of random taxation in a redistributive context. Finally, a general characterization of optimal administration is presented. Section 5 concludes.

2. Sources of heterogeneity

Before proceeding with the formal analysis, one needs to address the source of heterogeneity in avoidance behavior. This is important because both the normative implications and the modeling approach depend on it. Two alternative explanations that I consider are differences in tastes and differences in available avoidance technologies.

Opportunities for avoidance depend on many characteristics that are not perfectly related to the skill level.⁵ For example, it may be easier to avoid taxes for individuals who are self-employed.⁶ Similarly, it may be easier for better-educated individuals to find loopholes in the complicated tax laws. Access to the black-market economy may be related to one's family background and one's residence. A study by Clotfelter (1983) indicates that tax evasion is negatively related to age, and may depend positively on being married. As argued by Cowell (1990), differences in risk aversion across individuals may be interpreted as differences in avoidance technologies.⁷ None of these characteristics is systematically related to a person's skill level.

However, not all differences in tax compliance behavior can be explained by differences in opportunities. For example, it is not rare to observe what is apparently pathological honesty — individuals who pay every penny they ought to, regardless of any expected utility calculus. On the other hand, there are individuals who apparently will go to great lengths to avoid taxes.⁸ It is not too egregious to say that such behavior reflects, to some extent, differences in a taste for paying taxes. Individuals who are willing to contribute to charities may be less

⁵Many of such characteristics are potentially not exogenous, because it is possible that individuals take into account tax law when making decisions about, say, their job or education. I do not address it explicitly, but I recognize its importance and return to this issue in Section 3. The same objection applies to the models that treat skill level as exogenous, ignoring its endogenous determination.

⁶Carroll (1992, p. 46) gives a number of examples of occupation-specific tax avoidance.

⁷This feature may be quite reasonably thought of as being exogenous.

⁸Hessing et al. (1992) and Skinner and Slemrod (1985) concluded that there is a significant group of taxpayers who never evade, as well as a group of taxpayers who evade very often.

willing to avoid taxes than the others, if they believe that the purpose they serve is worth it.⁹ This kind of behavior would be hard to explain by tastes about public goods, because of the possibility of free-riding. Clotfelter's (1983) finding of an inverse relation between evasion and age could be explained by the existence of stigma associated with cheating which becomes more relevant as people get older, and may be modeled as differences in attitudes toward paying taxes rather than the physical cost of avoidance. One may also interpret paying taxes as consumption of 'giving', as in Andreoni (1990). Individuals enjoy (or dislike) giving and it enters their utility function.¹⁰ Finally, individuals with identical avoidance opportunities may behave differently due to some other, unobservable, differences. In this case, the amount of tax paid reflects these other characteristics. Consequently, allowing for differentiated tastes about paying taxes may be thought of as a reduced form of a more complicated model.

It is assumed throughout the paper that the social planner considers the level of utility of individuals in its objective function. This is a normative choice, particularly important when tastes differ. In this context, a tax cheater is considered to be a 'tax-averse' individual whose disutility from paying taxes should be taken into account, similarly as e.g., disutility from exposure to risk or disutility from supplying labor. One could of course argue that the utility of a 'leisure-lover' should not matter as much as the utility of a hard-working person. However, if the social-planner refrains from judging the 'right' and 'wrong' ways of enjoying life, what is relevant is the utility of an individual, and not the way in which it is attained. Factors influencing utility may be relevant to the extent that they affect others via the revenue constraint or the external effects, but only in these cases. Consequently, if this normative choice seems unacceptable, one should determine what are the negative consequences of tax avoidance that were ignored, and include them explicitly in the objective function. There is also another argument for considering the utility level as an appropriate objective. Recently, Kaplow and Shavell (1999) demonstrated that any social welfare function that is non-individualistic, i.e. one that has any arguments other than individual utilities, must violate the Pareto principle. Thus, adopting any other objective may lead to Pareto-inefficient choices.

Consider individuals that are characterized by two parameters: w , a skill level (equal to the wage rate) and θ , which characterizes the avoidance behavior (with a

⁹Alm et al. (1992) suggest that individuals may pay taxes, because they value the goods provided by the government, or as voluntary contributions to public goods. Carroll (1992) identifies three approaches to thinking about taxes: 'compliance with the law', 'payment for services', and 'giving to the common good', that may be interpreted as differences in tastes about avoidance behavior.

¹⁰What is different in the models discussed here is that 'giving' is not a choice variable, but instead it is forced by a government policy. When people enjoy 'giving' (as is the case in Andreoni's model), it is possible that they would wish to supplement taxes paid by individual donations. I ignore this possibility. On the other hand, if taxes are sufficiently high it is possible that they already exceed the amount an individual is willing to donate.

convention that high values of θ correspond to lower avoidance). They enjoy leisure and consumption goods (C) which are financed from compensation received for providing labor (L) on the market. Avoidance is introduced as in Mayshar (1991) and Slemrod (2000): the amount of income subject to taxation may be decreased by A at the cost of $D(wL, A)$.¹¹ The marginal tax rate is denoted by t and the lump-sum transfer is G . As a result, the budget constraint is

$$C = wL + G - t(wL - A) - D(wL, A). \quad (1)$$

The amount of taxable income is $I \equiv wL - A$. Finally, specification of the dependence of individual behavior on θ is necessary. I present two general specifications that fit into this framework and correspond to the two qualitatively different sources of avoidance differences. The derivation of the optimal tax formula in the next section is more general than these two models (in fact, it makes only use of the Roy's identity), but the qualitative discussion that follows is facilitated by this formalization.

- Differences in technology are represented by individualized avoidance functions. It is assumed that individual's type θ affects this cost; i.e. $D(\cdot) = D(wL, A; \theta)$. In this context, individual utility functions are assumed to be the same.
- Heterogeneity of preferences about paying taxes is modeled by allowing the amount of taxes paid, $T \equiv tI - G$, to affect the level of utility, in addition to its usual effect on disposable income, $U(\gamma(C, L), T; \theta)$. In this setup, the avoidance technologies are assumed to be identical. Taxes and utility from consumption and labor enter the utility function separately, so that individuals are identical except for their tastes about taxes — preferences between consumption and leisure are assumed to be identical and unaffected by the amount of taxes paid. In other words, when the tax is fixed (e.g. a uniform lump-sum taxation), θ does not affect decisions.

For both specifications of avoidance behavior I also assume that taxable income is increasing in the skill level, $\partial I / \partial w > 0$.¹²

3. The optimal tax

The problem of optimal redistribution should deal in its full generality with a general nonlinear income tax. Often, a simpler problem of choosing a linear tax is

¹¹The presence of the income term, wL , allows for what Slemrod (2000) calls the 'avoidance facilitating effect' of income.

¹²In the standard model this is guaranteed by the agent monotonicity condition $\partial / \partial L \{L(U_L / U_C)\} < 0$. I also make this assumption (in the case of taste differences this is an assumption about γ). The necessary and sufficient additional condition on the avoidance technology to guarantee that it is still the case here is $D_{AX} + D_{AA} > 0$. Derivation of this result may be obtained from the author.

solved, usually justified by numerical simulations showing that the optimal nonlinear tax is approximately linear (Myles, 1995). This is not an innocuous simplification, especially when other tax instruments are possible. However, in this section I consider the linear tax. My goal is not to characterize fully the optimal tax structure, but rather to understand the qualitative consequences of this new source of heterogeneity for the government's choice. Although the solution to the nonlinear problem would obviously be different, the intuition developed in this section should still apply. Added complexity could be justified if the goal was to make any quantitative statements about the shape of the optimal tax schedule. Then, however, one would need to specify the form of the cost of avoidance function, and make assumptions about the distribution of avoidance characteristics. Without further empirical work on this dimension of taxpayer behavior, all such assumptions would be arbitrary. Thus, I do not pursue this approach.¹³ Furthermore, in the next section, it is argued that one should think about avoidance schemes as endogenous policy instruments, so that any assumptions about their nature should follow from a more general optimization problem. I relax the assumption of linearity when talking about these issues, in order to avoid the impression that the restrictiveness of the tax instrument is responsible for the results.

In the optimal taxation literature it became conventional to concentrate on a single dimension of differences and to assume that all individuals have identical preferences. Usually, the dimension of interest is the skill level. Sandmo (1993) considered instead differences in tastes for leisure. The analysis in this paper demonstrates how the presence of an additional dimension of heterogeneity modifies the single-dimensional solution. Notably, an early treatment of the problem (a general model of linear taxation in Diamond and Mirrlees (1971)) does not assume identical utility functions and thus allows for types of heterogeneity other than just skill differences. However, generality of that approach makes it difficult to isolate consequences of distributional characteristics.

The social planner's problem is to choose a marginal tax rate t and a transfer G to maximize social welfare, subject to the revenue constraint and given exogenous revenue requirement R . Formally, the problem is

$$\max_{t, G} \int v(w, \theta; t, G) dF(w, \theta), \quad (2)$$

subject to the budget constraint

$$t \int I(w, \theta; t, G) dF(w, \theta) = R + G. \quad (3)$$

¹³Clearly, this choice is also motivated by an attempt to keep the problem simple. Mirrlees (1986) discusses complications that arise when one extends the problem of choosing the optimal nonlinear income tax to a multi-dimensional society.

Here $v(\cdot)$ denotes an indirect utility function, $I(\cdot) \equiv wL - A$ is taxable income, and $F(\cdot)$ is a joint cumulative distribution function of w and θ in the society (the population size is normalized to 1). The strength of egalitarian objectives is determined by the concavity of the cardinal representation of preferences.

The first-order conditions for G and t are standard:

$$0 = \int \frac{\partial v}{\partial G} + \lambda \left(t \frac{\partial I}{\partial G} - 1 \right) dF(w, \theta), \quad (4)$$

$$0 = \int \frac{\partial v}{\partial t} + \lambda \left(t \frac{\partial I}{\partial t} + I \right) dF(w, \theta). \quad (5)$$

One can use Roy's identity ($v_t = -I(\partial v / \partial G)$)¹⁴ and rearrange these conditions to obtain

$$1 = \int \alpha + t \frac{\partial I}{\partial G} dF(w, \theta), \quad (6)$$

$$t = \frac{\int I(\alpha - 1) dF(w, \theta)}{\int r dF(w, \theta)}, \quad (7)$$

where $\alpha = \lambda^{-1} \partial v / \partial G$ and r (responsiveness of taxable income) is $\partial I / \partial t$. The value of α is the marginal social utility of a dollar transferred to a given individual. This formula is a standard result. Although it does not provide an explicit solution for the marginal tax rate it may be used to gain qualitative understanding of the determinants of the optimal tax. A similar expression was derived by, e.g., Atkinson and Stiglitz (1980).

Formula (7) may be further decomposed to highlight the effect of an additional dimension of heterogeneity. Consider first its numerator. Denote as $F(w)$ the c.d.f. of w , and as $F(\theta|w)$ the marginal distribution of θ given w , and use x^w to denote the mean of variable x given w :

$$\begin{aligned} & \int I(\alpha - 1) dF(w, \theta) \\ &= \int \int I(\alpha - \alpha^w + \alpha^w - 1) dF(\theta|w) dF(w) \\ &= \int \int I(\alpha - \alpha^w) dF(\theta|w) dF(w) + \int (\alpha^w - 1) \int I dF(\theta|w) dF(w) \\ &= \int \int I(\alpha - \alpha^w) dF(\theta|w) dF(w) + \int I^w (\alpha^w - 1) dF(w). \end{aligned} \quad (8)$$

¹⁴It is straightforward to verify that it holds for each of the two specifications of heterogeneity in avoidance examined here.

Also, the denominator may be expressed as $\int r \, dF(w, \theta) = \int r^w \, dF(w)$. Formula (7) now yields

$$t = \frac{\int I^w(\alpha^w - 1) \, dF(w)}{\int r^w \, dF(w)} + \frac{\int \int I(\alpha - \alpha^w) \, dF(\theta|w) \, dF(w)}{\int r \, dF(w, \theta)}. \tag{9}$$

Note also that $\int I(\alpha - \alpha^w) \, dF(\theta|w) = E[(I - I^w)(\alpha - \alpha^w)|w] = \text{cov}(I, \alpha|w)$ is the covariance of I and α given w . It yields,

$$t = t^* + \frac{\int \text{cov}(I, \alpha|w) \, dF(w)}{\int r \, dF(w, \theta)}, \quad \text{and} \quad t^* \equiv \frac{\int I^w(\alpha^w - 1) \, dF(w)}{\int r^w \, dF(w)} \tag{10}$$

where I^w, r^w, α^w are the mean values of the appropriate variables given w , and $\text{cov}(I, \alpha|w)$ is the covariance of I and α conditional on the skill level.

In Eq. (10), t^* is the marginal tax rate that would be calculated if one ignored horizontal differences¹⁵ and replaced all agents at any given w by their averages, and thus it already reflects the effect of skill level on avoidance behavior. In addition, any behavioral response due to the availability of avoidance is accounted for by an increased taxable income response (compared to the standard model with only consumption/leisure decisions) in the denominator of t^* . This is where the standard analysis stops.

When horizontal differences are present, t is no longer equal to t^* . It is natural to assume that $\int r \, dF(w, \theta) < 0$, i.e. that an increase in the tax rate decreases the aggregate taxable income. In this case, t depends negatively on the covariances of α and I at any given w . When these covariances are positive ($\text{cov}(I, \alpha|w) > 0$ for any w), higher incomes at a given w correspond to high marginal utility, i.e. a higher need for redistribution. As a result, progressive income taxation aggravates horizontal differences, which should lead to a reduction of the marginal tax rate below what could be suggested by considering only the vertical inequality (i.e. $t < t^*$). Similarly, $\text{cov}(I, \alpha|w) < 0$ for any w implies that $t > t^*$. This simply reflects that vertical and horizontal goals are complementary: high taxable income characterizes both individuals with high skill levels, and the ones that have lower social marginal utility of income than their peers with the same w . Thus, the motivation for redistributive taxation is strengthened.¹⁶

The issue is whether those who evade more, given w , are more or less

¹⁵Note that the formula for t^* has the same structure as the one for t in Eq. (7).

¹⁶Note that this discussion does not rely in any way on the specific models analyzed in this paper. In particular, it applies regardless of the interpretation of θ : horizontal differences are important for evaluating optimal progressivity.

‘deserving’ at the margin. The answer to this question depends on the source of differences in avoidance behavior and the normative assumptions. To trace the implications of different sources of heterogeneity for the optimal progressivity, one needs to determine in each case the signs of the covariance terms in Eq. (10). In what follows I discuss this question for the two hypotheses regarding the source of avoidance differences.

Heterogeneity in avoidance technologies. This is the situation when the cost of avoidance depends on θ , $D(wL, A, \theta)$. This may apply, for example, when sheltering opportunities depend on the type of job one has (and one may not change it easily). Assume that $D_\theta \geq 0$. In general, high θ people need not avoid less or have higher taxable income, but these are natural assumptions to make, $\partial I / \partial \theta > 0$.¹⁷ Then, the key to determining the sign of covariance terms in Eq. (10) is understanding how the marginal utility of income α depends on θ at a constant skill level. This turns out to be ambiguous in general, due to conflicting forces at play. Firstly, the concavity of the welfare functions makes marginal utility increase with higher avoidance cost (high θ), because individuals are worse off. Secondly, differences in the cost of avoidance usually induce shifts between consumption and leisure which may act to decrease the marginal utility of income. The more concave is the welfare function the stronger is the first type of effect, and it is natural to concentrate on the situation when it dominates. Then, a person with low taxable income may be either truly poor or else able to avoid a lot. Only the former is a reason to transfer income. One would wish to redistribute income from the rich to the poor and, at the same time, from the good avoiders to the bad ones. However, in one case high taxable income indicates individuals who are better off, but in the latter case it indicates individuals who are worse off. The quality of taxable income as an indicator of well-being is lower, as reflected by positive covariance terms, and Eq. (10) implies that progressivity ought to be reduced.

Differences in tastes. The second source of heterogeneity considered is due to differences in tastes, represented by the utility function $U(\gamma(C, L), T, \theta)$, where $T = wL - G$. To be more specific, consider individuals who usually consider tax cheating as unethical, but they may be willing do it as a last resort. The amount of taxes paid becomes an indicator of a good luck — those who pay more are successful ones.¹⁸ Similarly as in the case of technology differences, the key issue

¹⁷Together with the assumptions necessary for taxable income to increase in w , the following assumptions guarantee it: $D_{A\theta} > 0$, $D_{X\theta}D_{AA} - D_{XA}D_{A\theta} < 0$, and the normality of consumption. In fact, these are the weakest possible assumptions that guarantee the result and do not involve simultaneous restrictions on the utility function and the avoidance technology. Details are available from the author.

¹⁸A model with similar implications was developed by Andreoni (1992). When individuals face binding borrowing constraints, they use tax evasion as a high-risk loan, which they would not utilize otherwise. Tax evasion is then, at least to some extent, an indicator of a low realization of income, and thus economic hardship (though, at the same time, it indicates higher permanent income).

is how the marginal utility depends on θ . As before, the direction of this dependence may not be determined without additional assumptions. An interesting special case is that of additive separability: $U(\gamma(C, L), T, \theta) = V(C, L) + \phi(T, \theta)$. In this case, $\partial\alpha/\partial\theta = \phi_{T\theta}(tw(\partial L/\partial G) - 1)$.¹⁹ Assuming, as it is usual, that labor is an inferior good ($\partial L/\partial G < 0$), and that the marginal tax rate t is positive, the sign of this expression is the same as the sign of $-\phi_{T\theta}$. Incidentally, in this case, $\phi_{T\theta} > 0$ is a necessary and sufficient condition for $\partial I/\partial\theta > 0$. As a result, the marginal utility of income and taxable income move in the opposing direction, and the covariance terms in Eq. (10) are negative. High taxable income is an indicator of being better off at a given skill level, and consequently progressivity ought to be increased. People who have lower taxable income are, *ceteris paribus*, worse off, and that is why they decided to make use of avoidance even though they bear a psychic cost of doing so. According to the expression (10), the optimal policy should recognize this by increasing the marginal tax rates and the demogrant, in effect reducing the tax burden of individuals with lower taxable income.

In practice, the technology explanation for the differences in tax avoidance is likely to be more compelling and it is easier to accept from the normative standpoint. The analysis indicates that the standard approach to the optimal taxation is likely to overestimate the optimal marginal tax rates in this context. The optimal policy needs to take into account that high marginal tax rates target people who are just bad in avoiding taxes, and consequently that they exacerbate horizontal inequities. The discussion of taste differences, although more controversial from the normative point of view, introduces a qualification to this statement. If some people derive higher satisfaction from contributing to the common good than the others, their relatively high taxable income indicates higher level of welfare and thus the optimal policy should tax them more (by increasing the marginal tax rates for everyone). Practical evaluation of the importance of this bias requires more evidence about the distribution of differences in tax sheltering among individuals with similar skill levels. However, the evidence cited in Section 2 suggests that it is likely to be non-trivial.

4. Results about the optimal amount of avoidance

This section extends the previous models by abandoning the assumption that the avoidance technology is given, and instead introducing a policy instrument s that affects it. The choice of optimal administration in a redistributive context has been considered in the literature. Slemrod (1994) adopted a simple formulation of avoidance technology, and argued that ignoring administration may lead to

¹⁹This is because, by the envelope theorem, $\partial\alpha/\partial\theta = \partial^2 v(\cdot)/\partial\theta\partial G = d\phi_\theta/dG$.

misleading conclusions about the optimal level of progressivity. Cremer and Gavhari (1996) considered the problem of optimal redistribution with tax evasion, and demonstrated several propositions about the optimal level of enforcement and optimal fines. These are, however, very specific instruments, so that their conclusions are not general results about tax administration. Slemrod and Kopczuk (2000) allowed for general administrative instruments, and analyzed a particular channel through which they operate — an effect on the strength of behavioral response. None of these papers considered heterogeneity in evasive behavior.

Assume that the cost of avoiding taxes depends on a parameter s , $s \in \mathbb{R}$ so that $D(wL, A; \theta; s)$, with $D_s > 0$. Thus, a small s represents high opportunities for avoidance. Assume for now that there is some resource cost of choosing s , $M(s)$ (with $M'(s) \geq 0$).

Firstly, I demonstrate that with non-negligible avoidance costs avoidance should not be eliminated completely, even if it is possible. Denote by $\mu(s)$ a measure of the set of individuals who engage in avoidance ($A > 0$) when the government chooses s . Suppose that there is a value of s^* such that $\mu(s^*) = 0$. The following proposition may be established.

Proposition 4.1. *Suppose that the left derivative of $\mu(s)$ exists at s^* ,²⁰ and assume that taxable income and indirect utilities are differentiable. Then, if $M'(s) > 0$, it is not optimal to eliminate all avoidance.*

Proof. Suppose that s^* is the optimum. Consider reducing it by ds . By assumption, only a small number of individuals, $\mu'(s^*) ds$, begins to engage in avoidance behavior, and the effect on their utilities and taxable incomes is, by differentiability, at most of the first order. Consequently, the effects on aggregate welfare and total tax revenue are of the second order (they are products of the two first-order effects), while the effect on administrative costs is of the first order. Thus, it is welfare improving, proving the proposition. \square

The benefit of prohibiting avoidance falls to zero for sufficiently large enforcement, and thus it is not optimal to remove it completely.²¹ Is there some intrinsic value of having an option of avoidance, in addition to the usual tax

²⁰This assumption means that as s falls slightly, only a ‘small’ fraction of individuals begins to engage in avoidance. Whether or not it holds depends on both the assumptions about individual behavior and on the assumptions about the distribution of individual characteristics. An important case when this assumption does not hold arises when θ is not differentiated and the avoidance technology depends just on A , i.e. it is $D(A, s)$. In that case, all individuals begin to avoid at the same time.

²¹Note that this result is conceptually different than Baldry’s (1984) result of non-complete enforcement of income tax in the presence of tax evasion. He showed, in a representative-individual model, that a small degree of tax evasion has a second-order effect on tax revenue, and thus it is not optimal to eliminate it in the presence of first-order administrative costs. In our setup, even small avoidance could have a first-order effect on tax revenue in a representative agent model (and one can show that usually it will), but the result follows from the heterogeneity of the society.

function? To consider this question, suppose that the administrative costs are negligible ($M(s) = 0$). One can then demonstrate the following proposition.

Proposition 4.2. *Suppose that skill differences are the only source of heterogeneity, non-linear taxes are feasible, administration is costless, and it is possible to eliminate all avoidance. Then it is not optimal to have a positive amount of avoidance.*

Proof. An individual chooses A to maximize $wL - T(wL - A) - D(wL, A)$, thus defining optimal A , $A = A(wL)$. Consider some mix of T and D . Define $J(wL) = T(wL - A(wL)) + D(wL, A(wL))$ and consider using J in place of the tax function T with all avoidance prohibited. By definition, it does not change the individual's problem, and thus individual utilities are not affected. However, as long as some individuals avoided taxes previously, $J(wL) > T(wL - A(wL))$ for some wL , and revenue increases. Thus, the mix of T and D is dominated by J with avoidance eliminated completely, and it is not optimal to have avoidance present. \square

With just skill differences, avoidance is a purely wasteful activity. It is a redundant instrument because everything that may be accomplished by using it may be also accomplished by non-linear taxes with a greater effect on revenue. It is important to note that this result crucially depends on the assumed avoidance technology. If, for example, the skill level directly affected the cost of avoiding (separately from its impact on taxable income), the proposition would no longer obtain. But, then individuals would no longer have identical avoidance opportunities so that one could speak of heterogeneity in avoidance behavior.

Proposition 4.2 cannot be extended to the heterogeneous avoidance case. In what follows, I present two examples that show, to the contrary, that avoidance might be a useful instrument of redistribution. The common feature of these examples is that a nonlinear tax system is feasible, and that it costs nothing to eliminate all avoidance, but some heterogeneity in avoidance behavior is present. The incentive constraints limit the extent of redistribution. When it is pushed too far, the rich individuals start to mimic lower types, and the policy is no longer effective. Observe, however, that in some circumstances there are avoidance schemes that are available to the poor and that are less accessible to the rich. For example, the stigma attached to having an 'underground' job may be much greater for a better-qualified person than for a less qualified one, because it's more acceptable to have this kind of job among peers of the poor. If so, the social planner may want to allow for this kind of activity: the rich do not engage in it, while the poor can. The transfer to the poor takes place, but standard incentive constraints are unaffected.

4.1. Examples of welfare-improving avoidance

Technology differences. Consider a society consisting of two individuals: a 'poor'

and a ‘rich’ individual. Call the ‘poor’ individual a ‘waiter’ and the rich one a ‘professor’. Suppose that the social planner has already solved the optimal redistribution problem, and the optimal non-linear tax function does not redistribute income completely (as will usually be the case), resulting in the social marginal valuations of transfers b_p and b_R , $b_p > b_R$. At this optimum, further redistribution is limited by the incentive constraint, i.e. any additional transfer would make the professor mimic the waiter.

Now suppose that the waiter can avoid dA dollars (by, say, not reporting tips) at the marginal cost of c , and that this avoidance scheme is not available to the professor. In other words, assume that the availability of avoidance ‘tags’ the poorer individual and it is not related to the actual income — when the professor tries to mimic the waiter, he is still not able to avoid (because he does not receive any tips). The resulting loss of revenue is $t dA$. To make it up, the tax liability of the waiter is increased lump-sum by $\alpha t dA$ and that of the professor by $(1 - \alpha)t dA$, where $\alpha < 1$ is selected so that the incentive constraints are not violated.²² Then the total marginal effect on social welfare is

$$b_p(t - c) dA - t(\alpha b_p + (1 - \alpha)b_R) dA = b_p \left(t(1 - \alpha) \left(1 - \frac{b_R}{b_p} \right) - c \right) dA,$$

and when $c < t(1 - \alpha)(1 - b_R/b_p)$, allowing for avoidance is welfare improving.

Taste differences. Consider a society consisting of two individuals with inelastic labor supply,²³ whose respective utility functions are $u_1(C, T) = C$ and $u_2(C, T) = C - \beta T$, where C is consumption and T is the amount of taxes paid and β is some positive constant. Individual 2 is ‘tax-averse’: any taxes that he pays decrease his utility in addition to their effect on consumption. Assume also that the income of each individual is identical and equal to y .

Suppose that the goal of the social planner is to maximize the sum of utilities $u_1(C_1, T_1) + u_2(C_2, T_2)$, subject to the exogenous revenue constraint $T_1 + T_2 = R$. The types of individuals are not observable so that optimization is also subject to incentive compatibility constraints. Consequently, the first-best solution — an individual-specific lump sum tax — is not feasible. Any non-linear tax system in this setup can be reduced to a single lump-sum tax (because individuals always

²²The incentive constraint of the low type was not binding, and thus it won’t be affected by any small changes in the policy. Consider a small lump-sum increase in the tax liability assigned to the low type. It makes the incentive constraint of the high-type slack (because it reduces the attractiveness of deviation to the high type without affecting his own allocation). Consequently, the tax of the high-type may also be somewhat increased lump-sum without violating the constraint. Thus, a small amount of revenue may be collected by strictly increasing lump-sum taxes imposed on both types. Parameter $0 < \alpha < 1$ is selected to implement this mix.

²³This assumption is not essential, but it makes the example much simpler.

choose the smallest tax) and the optimal regime features a tax of $T = R/2$ with the resulting social welfare of $2y - R - (\beta/2)R$.

Now suppose that there is an available avoidance scheme where one can pay c to avoid one dollar of taxes, where $1 < c < 1 + \beta$. Clearly, individual 1 will never use it, while individual 2 wants to avoid as much taxes as possible (each dollar of avoidance increases utility by $1 + \beta - c$). It also changes the solution to the optimal tax problem. Now, one should impose the tax of $T = R$ which will be paid by the first individual and fully avoided by the second one. The utility of the first individual is $y - R$, the utility of the second one is $y - cR$, and the social welfare is $2y - R - cR$.

The optimum with avoidance results in a greater welfare level if $c < \beta/2$.²⁴ Under these circumstances it is optimal not to eliminate avoidance, even though it might be costless to do so. Note also that the social welfare function considered was linear (weakly concave), seemingly implying very weak incentives for redistribution. With a concave welfare function, a condition for welfare improvement would be less restrictive.

What do these examples show? As long as avoidance is more easily accessible to individuals to whom one wants to redistribute income, its existence may be welfare improving. Redistribution via income tax is limited by the standard incentive constraints. Avoidance may also have a redistributive aspect, and the relevant incentive constraints, though still present, might be weaker than the standard ones, allowing for additional redistribution. The loss of revenue may well be exceeded by the redistributive gains. As a result, even in the presence of the nonlinear taxes and negligible administrative costs, one may want to use avoidance as an instrument of redistribution.

This result is related to the argument of Stiglitz (1987) that ex post random taxation may be welfare improving.²⁵ His argument relies on differences in risk aversion between low and high skill individuals. If high-income individuals are very risk averse, introducing randomization at low taxable income weakens incentive of the high types to imitate low types, and makes it feasible to provide additional transfers to the poor. When the possibility of uncertainty is added to the standard model with just skill differences, it is as if there were two different characteristics varying across individuals, namely skill level and risk aversion. The traditional tax function is based only on taxable income, which is an indicator of the skill level, and Stiglitz argued that instruments which can use differences in the other characteristic may be useful, just as it is the case in the examples presented here. He did not introduce horizontal differences (varying degree of risk aversion

²⁴Given that $c > 1$ it also requires $\beta > 2$.

²⁵Stiglitz (1982) discusses at length the possibility of random taxes being welfare improving when all individuals are identical. More elastic risk aversion makes this result more likely to hold. It suggests that ex post heterogeneity in risk aversion introduced by randomization plays an important role also in this case.

at a given skill level), and thus, on the surface, there was just one dimension of differences among individuals. Consequently, this intuition was not transparent. In terms of the models in this paper, one can represent the Stiglitz's (1987) problem by introducing the cost of avoidance function $D(\cdot)$ that depends on a measure of risk aversion, which is, possibly, related to the skill level.²⁶ When the problem is represented in this way, individuals with different skill levels have different opportunities for avoidance and the first example (technology differences) can be easily modified to apply to it.

The main argument is related to the argument of Akerlof (1978): if possible, the tax function should also depend on variables other than taxable income. The point made here is that appropriately designed non-tax instruments can be designed to make use of unobservable dimensions of taxpayer behavior.

4.2. The optimal administration

I turn now to the characterization of the optimal policy. Mayshar (1991) characterized the criteria for the choice of optimal administration in the presence of tax sheltering, but he did not consider heterogeneity.

Recall that administration is introduced as a parameter s affecting the cost of avoidance, and its cost is given by $M(s)$. The first-order condition for s is (analogously to Eqs. (4) and (5))

$$0 \geq \int \frac{\partial v}{\partial s} + \lambda \left(t \frac{\partial I}{\partial s} - M'(s) \right) dF(w, \theta). \quad (11)$$

This condition holds with equality at an interior solution and it may be slack when $s = 0$. Using the envelope theorem $\partial v / \partial s = -D_s(\partial v / \partial G)$, and rearranging yields

$$\lambda M'(s) + \int D_s \frac{\partial v}{\partial G} dF(w, \theta) \geq \lambda t \int \frac{\partial I}{\partial s} dF(w, \theta). \quad (12)$$

Denote as $\bar{\alpha} \equiv \int \alpha dF(w, \theta)$, where $\alpha = \lambda^{-1} \partial v / \partial G$, and notice that

$$\begin{aligned} \int D_s \frac{\partial v}{\partial G} dF(w, \theta) &= \lambda \int D_s \alpha dF(w, \theta) \\ &= \lambda \text{cov}(D_s, \alpha) + \lambda \bar{\alpha} \int D_s F(w, \theta). \end{aligned} \quad (13)$$

Finally, observe that the marginal cost of funds, or MCF, equals $\lambda / \bar{\alpha}$. This is the ratio of the cost of collecting a dollar using available instruments to the cost of a lump-sum tax. Using these definitions, one can express Eq. (12) as

²⁶This approach is in the spirit of Cowell (1990). He shows that the cost of tax evasion may be approximated using a deterministic cost of tax sheltering.

$$\text{MCF} \times M'(s) + \int D_s \, dF(w, \theta) + \text{cov}\left(D_s, \frac{\alpha}{\bar{\alpha}}\right) \geq \text{MCF} \times \text{MR}_s, \quad (14)$$

where MR_s denotes the marginal revenue defined as $\text{MR} \equiv t \int (\partial I / \partial s) \, dF(w, \theta)$. The left-hand side of this formula represents the costs of increasing s , and the right-hand side represents the benefits.²⁷ The benefit of an increase in s is the revenue raised, which needs to be multiplied by the cost of alternative sources of revenue, i.e. MCF. On the cost side, there is an increase in administrative costs (again multiplied by the cost of collecting revenue), and a direct increase in the cost of avoidance imposed on consumers. The cost is also modified by the covariance between the change in the cost of avoiding and the marginal social utility of income that measures the need for redistribution. When this covariance is negative, reflecting this additional cost being concentrated among individuals who are better off, the cost of using s decreases. If it is positive (i.e. the reform affects individuals who are worse off), its cost increases. Recall that increased s is, loosely speaking, assumed to reduce tax avoidance. The presence of the covariance term in (14) reflects the point made by the examples considered earlier: that one may want to allow for avoidance if it is correlated with the need for redistribution. The policy variable s has an aspect of non-uniform lump-sum taxation: the direct cost that it imposes may vary across individuals (i.e. D_s may vary with w and θ). It is not exactly a non-uniform lump-sum tax because it does not collect revenue and, less importantly, it may have a distortionary effect by affecting the avoidance-facilitating role of income. Nevertheless, the benefits of using this imperfect version of the first-best instrument may exceed the costs.

Finally, recall the results of Section 4. Proposition 4.1 states that, with $M'(s) > 0$, avoidance should not be completely eliminated. In terms of Eq. (14), this says that when we are close to complete removal of avoidance, all terms but $\text{MCF} \times M'(s)$ are 0, because only a small subset of individuals avoids. Proposition 4.2 states that, without administrative cost and with non-linear taxes, avoidance should be eliminated completely. When it applies, one should not study Eq. (14), because there is a corner solution. There are two qualifications, however. Firstly, the instrument considered in this section was a linear tax, and an aspect of non-linearity that avoidance introduces may still be beneficial.²⁸ Secondly, the point made by previous examples is that differences in avoidance behavior may make some avoidance optimal even when non-linear taxes are feasible. There is some flexibility gained by having an additional instrument available because in certain

²⁷This equation is similar to the condition for optimal administration derived by Mayshar (1991). The original result was derived in the context of the representative agent model, and it is modified here to account for distributional differences.

²⁸This applies to any 'constrained' instrument, as the proof of Proposition 4.2 relied on the feasibility of an arbitrary nonlinear income tax. As a consequence, even avoidance just among the rich may be optimal, e.g. as a way of implementing the Seade's (1977) 'zero rate at the top' result.

circumstances it allows for more redistribution than just by the mere income tax. This is because in the standard problem the choice of optimal tax is subject to incentive constraints. The incentive constraints relevant for the avoidance instrument need not be identical, and thus they may allow for more redistribution.

5. Conclusions

Heterogenous avoidance behavior is important for tax policy design for at least two reasons. Firstly, recent empirical papers indicate that margins of response other than those usually considered (labor/leisure) are significant. Thus, tax avoidance deserves serious treatment and its redistributive implications are likely to be empirically important. The paper stresses that not only the vertical, but also the horizontal distribution of avoidance behavior is of importance. Secondly, the presence of tax avoidance introduces additional tax policy instruments that were previously often ignored such as enforcement regimes, designing avoidance schedules, and personalized deductions.

The standard optimal income tax results need to be modified in the presence of tax sheltering because it affects the relationship between taxable income and the social marginal utility of income. With avoidance, low income may indicate either a low skill level, a high degree of avoidance or some combination of the two. The effect on optimal progressivity depends on the relative welfare of avoiders. If people have different opportunities for avoidance, good avoiders are better off than the others and under reasonable assumptions it acts to decrease the marginal tax rates at the optimum. When tax avoidance reflects economic hardship, avoiders are the ones who are worse off and the marginal tax rate should increase (with a corresponding increase of the demogrant) in order to reduce their share of tax burden. The key decomposition of the tax rate, expression (10), shows that these considerations are simply captured by the covariance, conditional on the skill level, of taxable income and the marginal utility of income. This intuition also applies to other kinds of heterogeneity.

Introducing avoidance makes instruments other than the tax rate of primary interest. The paper demonstrates that allowing for some avoidance has no intrinsic value as a policy instrument when there are only skill differences and nonlinear taxes are feasible, though if administration is costly avoidance should not be eliminated completely. More importantly, however, avoidance may be welfare improving when compliance behavior is heterogeneous even if nonlinear taxes are feasible and administration is costless, because it may serve as a cheaper, though still wasteful, instrument of redistribution. In some circumstances, avoidance may serve to redistribute income even though further redistribution by standard instruments is inefficient due to incentive constraints.

On a more general level, this discussion suggests that many apparently wasteful

activities may be in fact welfare improving.²⁹ For example, observe that black market activities are highly concentrated among low-income people. Their existence can be a cheap instrument of redistribution that does not require making the tax system more progressive: the marginal tax rates may be held low via relatively high lump-sum tax component, with redistribution taking place through illegal activity.³⁰ The gains from reduced tax distortions may in some cases outweigh losses from illegal behavior. Similarly, imperfect enforcement of tax laws may be welfare improving if people are usually honest, but they turn to avoidance when they experience negative shocks to their income. Arguably, these kinds of instruments have been consciously used by policy makers, even though economists generally consider them as inefficient. Seemingly wasteful policies, such as implementing complicated tax laws, or imperfect enforcement of existing rules, may be welfare improving, and need not always reflect economic inefficiency of the underlying political system.

Acknowledgements

I am grateful to Joel Slemrod for his advice and detailed comments. I also benefited from the comments I received from the anonymous referee, Roger Gordon, Lutz Kilian, Rohini Somanathan, Michelle White, and participants of the Public Finance seminar at the University of Michigan.

References

- Akerlof, G.A., 1978. The economics of 'tagging' as applied to optimal income tax, welfare programs, and manpower planning. *American Economic Review* 68 (1), 8–19.
- Alm, J., Jackson, B.R., McKee, M., 1992. Deterrence and beyond: toward a kinder, gentler IRS. In: Slemrod, J. (Ed.), *Why People Pay Taxes: Tax Compliance and Enforcement*. University of Michigan Press, Ann Arbor, pp. 311–329.
- Andreoni, J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. *Economic Journal* 100 (401), 464–477.
- Andreoni, J., 1992. IRS as loan shark: tax compliance with borrowing constraints. *Journal of Public Economics* 49 (1), 35–46.
- Atkinson, A.B., Stiglitz, J.E., 1980. *Lectures on Public Economics*. McGraw-Hill, London.
- Auten, G., Carroll, R., 1999. The effect of income taxes on household behavior. *Review of Economics and Statistics* 81 (4), 681–693.

²⁹Boadway and Keen (2000) survey literature on welfare improving, though seemingly inefficient, instruments that may be used to mitigate problems caused by the existence of interest groups.

³⁰Notably, a recent calibration of the optimal tax system by Gruber and Saez (1999) features a significant lump-sum transfer component that is quickly taxed away, and relatively low marginal tax rates for higher levels of income. In such a system non-tax instruments would be especially valuable.

- Baldry, J.C., 1984. The enforcement of income tax laws: efficiency implications. *Economic Record* 60 (169), 156–159.
- Boadway, R., Keen, M., 2000. Redistribution. In: Atkinson, A.B., Bourguignon, F. (Eds.), *Handbook of Income Distribution*. Elsevier/North Holland, Amsterdam, In press.
- Carroll, J.S., 1992. How taxpayers think about their taxes: frames and values. In: Slemrod, J. (Ed.), *Why People Pay Taxes: Tax Compliance and Enforcement*. University of Michigan Press, Ann Arbor, pp. 43–63.
- Christian, C.W., 1994. Voluntary compliance with the individual income tax: results from 1988 TCMP study. In: *IRS Research Bulletin 1993/1994 Publication 1500*. Internal Revenue Service, Washington DC.
- Clotfelter, C.T., 1983. Tax evasion and tax rates: an analysis of individual returns. *Review of Economics and Statistics* 65 (3), 363–373.
- Cowell, F.A., 1990. Tax sheltering and the cost of evasion. *Oxford Economic Papers* 42 (1), 231–243.
- Cremer, H., Gahvari, F., 1996. Tax evasion and the optimum general income tax. *Journal of Public Economics* 60 (2), 235–249.
- Diamond, P.A., Mirrlees, J.A., 1971. Optimal taxation and public production II: tax rules. *American Economic Review* 61 (3), 261–278.
- Feldstein, M.S., 1995. The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act. *Journal of Political Economy* 103 (3), 551–572.
- Feldstein, M.S., 1999. Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics* 4 (81), 674–680.
- Goolsbee, A., 2000. What happens when you tax the rich? Evidence from executive compensation. *Journal of Political Economy* 108 (2), 352–378.
- Gruber, J., Saez, E., 1999. *The Elasticity of Taxable Income: Evidence and Implications*, Mimeo.
- Hessing, D.J., Elffers, H., Robben, H.S.J., Webley, P., 1992. Does deterrence deter? Measuring the effect of deterrence on tax compliance in field studies and experimental studies. In: Slemrod, J. (Ed.), *Why People Pay Taxes: Tax Compliance and Enforcement*. University of Michigan Press, Ann Arbor, pp. 291–305.
- Kaplow, L., Shavell, S., 1999. Any non-individualistic social welfare function violates the Pareto principle. National Bureau of Economic Research, Working paper 7051.
- Lemieux, T., Fortin, B., Frechette, P., 1994. The effect of taxes on labor supply in the underground economy. *American Economic Review* 84 (1), 231–254.
- Mayshar, J., 1991. Taxation with costly administration. *Scandinavian Journal of Economics* 93 (1), 75–88.
- Mirrlees, J.A., 1986. The theory of optimal taxation. In: Arrow, K.J., Intriligator, M.D. (Eds.), *Handbook of Mathematical Economics*, Vol. 3. Elsevier/North Holland, Amsterdam, Chapter 24.
- Moffitt, R., Wilhelm, M., 1998. Taxation and the labor supply decisions of the affluent. Office of Tax Policy Research, University of Michigan, Working paper 98-8.
- Myles, G.D., 1995. *Public Economics*. Cambridge University Press, Cambridge.
- Sandmo, A., 1993. Optimal redistribution when tastes differ. *Finanz Archiv* 50, 149–163.
- Seade, J., 1977. On the shape of optimal tax schedules. *Journal of Public Economics* 7 (2), 203–235.
- Sheffrin, S.M., Triest, R.K., 1992. Can brute deterrence backfire? Perceptions and attitudes in taxpayer compliance. In: Slemrod, J. (Ed.), *Why People Pay Taxes: Tax Compliance and Enforcement*. University of Michigan Press, Ann Arbor, pp. 193–218.
- Skinner, J., Slemrod, J., 1985. An economic perspective on tax evasion. *National Tax Journal* 38 (3), 345–353.
- Slemrod, J., 1994. Fixing the leak in Okun's bucket. Optimal progressivity when avoidance can be controlled. *Journal of Public Economics* 55 (1), 41–51.
- Slemrod, J., 1998. Methodological issues in measuring and interpreting taxable income elasticities. *National Tax Journal* 51 (4), 773–788.
- Slemrod, J., 2000. A general model of the behavioral response to taxation. *International Tax and Public Finance*, In press.

- Slemrod, J., Kopczuk, W., 2000. The optimal elasticity of taxable income. National Bureau of Economic Research, Working Paper 7922.
- Slemrod, J., Yitzhaki, S., 1999. Tax avoidance, evasion and administration. In: Auerbach, A., Feldstein, M.S. (Eds.). *Handbook of Public Economics*, Vol. 3. Elsevier/North Holland, Amsterdam, In press.
- Stiglitz, J.E., 1982. Utilitarianism and horizontal equity: the case for random taxation. *Journal of Public Economics* 18 (1), 1–33.
- Stiglitz, J.E., 1987. Pareto efficient and optimal taxation and the new new welfare economics. In: Auerbach, A.J., Feldstein, M.S. (Eds.). *Handbook of Public Economics*, Vol. 2. North-Holland, Amsterdam, pp. 991–1042.