The Tax Controversy or Should We Forsake Indirect Taxation? A Survey of the Literature

Luca Micheletto

Working Paper n. 85

July 2002
The Tax Controversy or, Should We Forsake Indirect Taxation? A Survey of the Literature.

Luca Micheletto*
“L. Bocconi” University, Milan

July 10, 2002

Abstract

We revise the more recent advances in the literature about the so-called “direct versus indirect controversy”.

In particular, we show how the result of “uniform commodity tax under non-linear income taxation” by Atkinson and Stiglitz (1976) has been amended in some contributions that recover a role for indirect taxation as an instrument of optimal tax policy even in the case where agents’ preferences are weakly separable between leisure and other goods.

Keywords: optimal non-linear income tax; optimal linear commodity taxes; Atkinson-Stiglitz theorem.

JEL Classification: H21

*Department of Economics, via U. Gobbi, 5, 20136 Milan, Italy. Tel.: ++39 02 58 36 53 24; fax: ++39 02 58 36 53 18; e-mail: luca.micheletto@uni-bocconi.it. I am grateful to Bruno Bosco and Alessandro Petretto for useful comments. Usual disclaimers applies.
1 Introduction

The choice of the most appropriate tax base is one of the oldest issues of taxation policy. In 1955 Walker claimed that “Without any hesitation, therefore, I would call the discussion (...) a sterile controversy. (...) It seems to me unfortunate that such a barren topic as the Direct-Indirect Tax Problem has occupied such an important place in Public Finance discussions in recent years (...)” (Walker, 1955, p.173). Since then, the analysis has made considerable progress and its policy relevance cannot be anymore ignored. The ambiguity about the notions of “direct taxation” and “indirect taxation”, which has marked the earlier debate, has been solved by referring to the chance to tailor a certain tax to the particular economic or social characteristics of the household being taxed. Nowadays, income taxation is viewed as direct because it can be made progressive, whereas commodity taxation is generally viewed as indirect, since it is based on anonymous transactions (rather than on personal levels of consumption). It has been clarified that the opinion, grounded on a Mundell-type assignment of instruments to targets, that assign direct taxation to the equity objective and indirect taxation to the goal of raising revenue efficiently, is fundamentally misleading; however, taken alone, this is not sufficient to establish the primacy of direct taxation or imply that, as claimed by some other authors, direct taxes are more desirable from an economic viewpoint, meeting the equity, efficiency and even administrative cost criteria better than indirect taxes. In spite of that, the role of commodity taxes as an instrument of optimal tax policy has been severely undermined by the publication of Atkinson and Stiglitz’s (hereafter AS) (1976) pathbreaking paper. In this contribution they showed that where the (common) utility function is weakly separable between leisure and all goods taken together (i.e. the marginal rate of substitution between any two goods is independent of labour supplied), then there is no need to employ indirect taxation in the optimum solution. As a consequence, it has been claimed (Stiglitz, 1988, p.494) that “It can be shown that if one has a well-designed income tax, adding differential commodity tax is likely to add little, if anything, to the ability to redistribute income”. Such a strong assertion obviously hinges on the empirical validity of the separability assumption (which in itself has been questioned and has found very little support in recent estimations, see e.g. Browning and Meghir, 1991) but above all, as emphasised by Saez (2002a), on the assumption that the

---

1 To be fair it should be mentioned that a result identical to the one that is here presented was also contained in Mirrlees (1976).

2 The result holds also in the sense that, if a subset of commodities is separable from labour, then they should be taxed at a uniform rate.

3 Because of this, some authors regard the AS result from the (opposite) perspective of stating that disturbing the Pareto conditions for the allocation between goods makes always part of an optimal designed tax policy.
aggregators over various commodities are ordinally equivalent, i.e. all individuals share the same subutility of consumption, which is \textit{per se} quite an implausible circumstance.

Leaving aside the problem of homogeneity/heterogeneity in tastes for goods across agents, as well as the assumption of interdependent utilities (as it happens with the “altruistic” and “jealous” agents of Oswald, 1983) and the existence of “merit/demerit” goods (which both provide cases for justifying differential commodity taxation), it has been recently developed a strand of literature that we can shortly label as “in defence of commodity taxation” and that tackles the problem of the usefulness of indirect taxation moving from the AS result and pursuing some alternative lines of attack to the AS result.

In this paper we expound the directions taken by this literature and present in the unifying framework of the Stiglitz-Stern (1982, 1982) two types model the main results that have been achieved, trying at the same time to shed some light on the mechanisms at work behind the recovery of a positive role for indirect taxation as an instrument of optimal tax policy also in the case where agents’ preferences are weakly separable between leisure and other goods.

The paper proceeds as follows. Section 2 presents the benchmark model. Section 3 offers a survey of the more interesting attempts to recover a role for (non-uniform) indirect taxation. Section 4 offers concluding remarks.

2 The Basic Model

The basic model, which will be used as a benchmark, is a straightforward extension of Stiglitz (1982) and Stern (1982). There are two types of individuals\footnote{In what follows, we will use indifferently the terms “individuals”, “agents” and “workers”; also, neglecting intra-generational redistribution within husband and wife, we will use the terms “individuals” and “household” synonymously. Optimal income taxation of couples is studied by Apps and Rees (1997) and Schroyen (2001).} differing only because of their innate (and inalterable) (market) ability: type 1 are low skilled agents whereas type 2 are high skilled ones. The difference in ability (output per unit of time spent working) is reflected in the difference of unitary pre-tax wage ($w$) the two types of agents are paid: $w^2 > w^1$. Without loss of generality we assume that there is one individual of each type. Both individuals have the same preferences described by the quasi-concave utility function $u(\bar{x}, L)$, where $L$ is the labour supply and $\bar{x}$ is a $n$-vector of commodities $x_i, i = 1, 2, ..., n$. Production is linear and uses labour as the only input; units are chosen to make all producer prices ($\bar{p}$) equal to one.

The government has two sets of instruments at its disposal. First, a general income tax schedule $T(Y)$, where $Y = wL$ is labour income; the
government knows the distribution of ability in the population, but can observe neither $L$ nor $w$, while it is able to observe their product $Y$. The problem of choosing the direct tax schedule can be equivalently stated as the problem of selecting two pairs of pre-tax and disposable incomes ($Y^h, B^H$), $B^H = Y^H - T(Y^H)$, $h = 1, 2$. Second, a set of linear (more exactly, proportional) commodity taxes/subsidies on $x$: linearity follows from the assumption that the government can only observe anonymous transactions.

Consumer prices are $q_i = p_i + t_i = 1 + t_i, i = 1, 2, ..., n$.

Following Christiansen (1984), it has become common practice, in models of optimal taxation with self-selection, to break the consumer’s optimisation problem into two stages. At the first, a fixed amount of expenditure $B$ is optimally allocated over the consumption goods, taking labour supply as given. This gives conditional indirect utility

$$ V(q, B, Y; w) = \max \left\{ u \left( \frac{x}{w} \right) \mid q x = B \right\}, \quad (1) $$

where $x$ denotes the vector $(x_1, x_2, ..., x_n)$. The first order conditions are:

$$ \frac{\partial u(\bullet)}{\partial x_i} = \alpha q_i, \quad i = 1, 2, ..., n, \quad (2) $$

where $\alpha$ is the marginal utility of income.

We assume that the level curves of $V(\bullet)$ in $(Y, B)$-space are shallower, other things being equal, the higher an individual’s wage rate. That is, for all $q, B, Y$,

$$ \frac{\partial}{\partial w} \left( \frac{V_Y}{V_B} \right) < 0, \quad (3) $$

where subscripts indicate derivatives. This requirement allows for what have been called in literature the “Mirrlees-Spence” single crossing property.

Solving (1) gives conditional demand functions

$$ x_i(q, B, Y; w) = -\frac{V_{q_i}}{V_B}, \quad i = 1, ..., n, \quad (4) $$

where $V_{q_i}$ stands for $\frac{\partial V(\bullet)}{\partial q_i}$. Differentiating the individual budget constraint, we get the so called “adding-up” conditions

---

5This informational restriction prevents the government from imposing, as it would be first best, lump sum taxes/transfers conditioned on ability. Since the government can only make tax payment contingent on income, which can be manipulated by the taxpayer, the problem faced by the policy maker becomes a classic revelation mechanism design one.

6It can also be equivalently stated as a problem of choosing two pairs of labour supplies and disposable incomes.

7Notice that the linearity assumption avoids any arbitrage opportunities, which seems especially relevant when private commodities are easily retradeable.
\[
\sum_{i=1}^{n} q_i \frac{\partial x_i}{\partial B} = 1, \tag{5}
\]
\[
\sum_{i=1}^{n} q_i \frac{\partial x_i}{\partial Y} = 0. \tag{6}
\]

At the second stage of the individual’s optimisation, labour supply \( L \) (or equivalently, given the identity \( L = \frac{Y}{w} \), gross income \( Y \)) is chosen to maximise \( V(q, B, Y; w) \), subject to the link between pre-tax earnings and post-tax earnings available for goods expenditure implied by the direct tax schedule: \( B = Y - T(Y) \). The first order conditions of this problem allow us to define implicitly the marginal income tax rate \( T'(Y) \) faced by an agent by means of the following equation:

\[
-\frac{V_Y}{V_B} = 1 - T'(Y). \tag{7}
\]

We shall refer to \( 1 + \frac{V_Y}{V_B} \) as the marginal income tax rate\(^8\).

We focus on second-best Pareto efficient taxation, taking the object of policy to be the maximisation of the indirect utility of low skilled agents subject to achieving a minimum level of indirect utility \( V \) for high skilled agents. In pursuing its aims, the government faces two further constraints. The first is that receipts from income and commodity taxes be sufficient to finance the exogenous revenue requirement\(^9\) (budget constraint). As regards the second additional constraint, remember that the wage rate is private information; then a household could be tempted to earn the same gross income as the other type in order to misrepresent its type and gain a more favourable tax treatment. A household that misrepresents its type is called a “mimicker”. Since the incentive to mimic would undermine the implementability of the allocation, the government must design the tax system so that each type (weakly) prefers the \((Y, B)\) bundle intended for it to that intended for the other (self-selection or incentive compatibility constraints). For brevity, we will consider only the “normal” case in which the government wishes to improve the welfare of the low skilled agents significantly relative to what they would have obtained at the \( laissez faire \) allocation and redistribution goes from the high ability type to the low, in the sense that the only binding self-selection constraint is the one ruling out the possibility that high-wage

\(^8\)Even if the direct tax schedule chosen by the government will not, in general, be differentiable, it can be proved that there always exists an optimal tax structure for which \( 1 + \frac{V_Y}{V_B} \) is the left-handed derivative of the tax function at the bundle intended for the agent.

\(^9\)For simplicity, we will assume that the fixed amount of revenue to be collected is equal to zero, i.e. that taxation is purely redistributive.
individuals mimic low-wage ones\textsuperscript{10}.

Combining these ingredients, the planner’s problem is

$$\max_{Y^1, Y^2, B^1, B^2} V^1 \left( q, B^1, Y^1; w^1 \right)$$

subject to:

$$V^2 \left( q, B^2, Y^2; w^2 \right) \geq V^1 \left( q, B^1, Y^1; w^1 \right), \quad (\delta)$$

$$V^2 \left( q, B^2, Y^2; w^2 \right) \geq V^2 \left( q, B^1, Y^1; w^2 \right), \quad (\lambda)$$

$$\sum_{i=1}^{2} (Y^i - B^i) + \sum_{i=1}^{n} t_i x^1_i + \sum_{i=1}^{n} t_i x^2_i \geq 0, \quad (\gamma)$$

where Lagrange multipliers are within parentheses and $x^h_i$ represents demand of commodity $i$ by an agent of type $h$.

Using the convention to distinguish all variables pertaining to a mimicker by a “hat”, the first order conditions\textsuperscript{11} with respect to $Y^1$, $Y^2$, $B^1$, $B^2$, $t$, are respectively:

$$V^1_Y = \lambda \frac{\partial V^1}{\partial Y} - \gamma \left( 1 + \sum_{i=1}^{n} t_i \frac{\partial x^1_i}{\partial Y} \right); \quad (8)$$

$$(\delta + \lambda) V^2_Y = -\gamma \left( 1 + \sum_{i=1}^{n} t_i \frac{\partial x^2_i}{\partial Y} \right); \quad (9)$$

$$V^1_B = \lambda \frac{\partial V^1}{\partial B} + \gamma \left( 1 - \sum_{i=1}^{n} t_i \frac{\partial x^1_i}{\partial B} \right); \quad (10)$$

\textsuperscript{10}In this way we are not tracing out the entire constrained utility possibility frontier: actually, this is made up also by regions where no self-selection constraint is binding (and second best utility possibility frontier coincides with first best one) and by regions where the self-selection constraint binds in the opposite direction (from low-wage households towards high-wage ones). Stiglitz (1982) shows that single crossing precludes the possibility of both self-selection constraints binding at an optimum. This means that, in such a framework, a pooling equilibrium is inefficient.

\textsuperscript{11}The literature on optimal taxation has been often criticised (see e.g. Hammond (1990)) because of the fact that it is reduced to a plethora of first order conditions which, since the problems are generally not well-behaved, cannot guarantee more than local optimality. This problem is no exception; it is simply supposed the existence of a solution and characterised conditionally on this assumption.
\[(\delta + \lambda) V_B^2 = \gamma \left(1 - \sum_{i=1}^{n} t_i \frac{\partial x_i^2}{\partial B^2}\right); \quad (11)\]

\[-V_B^1 x_i^1 - (\delta + \lambda) V_B^2 x_i^1 + \lambda V_B^2 x_i^2 + \gamma \left[\sum_{j=1}^{n} \left(\sum_{j=1}^{n} t_j \frac{\partial x_j^h}{\partial t_i} + x_i^h\right)\right] = 0, \quad i = 1, \ldots, n; \quad (12)\]

where, in deriving eq. (12), use has been made of Roy’s identity.

2.1 The Marginal Effective Tax Rates

The total amount of taxes paid by an agent at income \(Y\) is given by

\[\tau(Y) = T(Y) + \sum_{i=1}^{n} t_i x_i (q, Y - T(Y), Y; w). \quad (13)\]

Differentiating gives the marginal effective tax rate

\[\tau'(Y) = T'(Y) + \sum_{i=1}^{n} t_i \left[\frac{\partial x_i}{\partial B} (1 - T') + \frac{\partial x_i}{\partial Y}\right]. \quad (14)\]

The following proposition sums up the main results about the structure of the marginal effective tax rates:

**Proposition 1** (Edwards, Keen and Tuomala, 1994) Pareto efficient taxation, constrained by a binding self-selection constraint on the high ability individuals, requires that:

(a) the high ability agents face a marginal effective tax rate of zero \((\tau'(Y^2) = 0)\);

(b) the low ability agents face a marginal effective tax rate given by:

\[\tau'(Y^1) = \frac{\lambda V_B^2}{\gamma} \left(\frac{V_B^2}{V_B^1} - \frac{V_Y^2}{V_B^1}\right), \quad (15)\]

which is strictly positive due to the Mirrlees-Spence single-crossing condition.

**Proof.** See Edwards, Keen and Tuomala (1994). □

Proposition 1 (a) is often referred to as the “no distortion at the top” result and is one of the most striking in the optimal taxation literature. It means that, although the high ability agents will generally be distorted at an optimum by income and commodity taxation, these distortions will “average out” to zero in the sense that if a high skilled agent were to earn a little
more his (her) total tax liability would remain unchanged. In turn, it is reminiscent of other two “end-point” results, the one prescribing zero marginal taxation on the highest earner in models with only one consumption good and (nonlinear) income taxation (Seade, 1977), and the one requiring that there should be no distortionary taxation (at the margin) on the individual with the highest ability in the case where there are many consumption goods which can all be taxed non-linearly (Stiglitz, 1987). However, to explain such a result, we cannot simply recur to the intuition, typically provided for the other quoted cases, that distorting the decisions of the highest ability type is of no use because eliminating it would enable more revenue to be raised without affecting the utility of any individual. The difference relies on the circumstance that in our case the government is not unconstrained in the tax instruments at its disposal since, when it is prevented, due to informational reasons, from implementing nonlinear commodity taxes, it must satisfy the requirement that marginal rates of substitution between commodities be the same for all consumers. In this sense, wishing a government to leave the highest earner’s decisions entirely undistorted, it should set not only the marginal income tax rate faced by that agent to zero, but also, because of the linearity of the commodity taxes, should set the commodity tax rates faced by each agent to zero, in this way giving up recurring to indirect taxation. Nonetheless, it can be shown that, if the highest earner were not “globally” undistorted (in the sense prescribed before, referred to the sign of his (her) marginal effective tax rate) at an optimum, then there would be the possibility of a small Pareto improving reform which increases the total tax paid by the top ability individual whilst leaving their welfare unchanged.

2.2 The Commodity Tax Structure

Having denoted by \( h_i \) the hicksian demands, we can characterise the indirect tax structure by means of the following proposition:

**Proposition 2** (Edwards, Keen and Tuomala, 1994) Pareto efficient taxation, constrained by a binding self-selection constraint on the high ability types, requires that:

\[
\sum_{k=1}^{2} \sum_{j=1}^{n} t_j \frac{\partial h_k}{\partial t_j} = \frac{\lambda V_B^{-2}}{\gamma} \left( x_i^1 - x_i^2 \right), \quad i = 1, \ldots, n. \tag{16}
\]

**Proof.** See Edwards, Keen and Tuomala (1994). □

The left hand side of eq. (16) is the discouragement index (multiplied by the aggregate demand of commodity \( i \)) of Mirrlees (1976), a linear approximation of the “reduction in compensated demand” induced by the tax
system; because of the symmetry of the substitution matrix, it can also be interpreted as the marginal dead-weight loss associated to the distortions of consumption prices. The rule for second best Pareto efficient commodity taxation requires that the change in the aggregate compensated demand for each good \( i \), induced by a small intensification of the commodity tax structure, be the same proportion of the amount by which the demand for good \( i \) of the low ability individual exceeds that of the high ability mimicker; that is to say, for a given value of the ratio \( \frac{\lambda V_2}{\beta} \), the proportionate reduction in the compensated demand for commodity \( i \) is greater, the greater the difference between the consumption of commodity \( i \) by the mimicker and the low ability type. The intuition is that, when the income tax schedule is optimally chosen, commodity taxation is only useful as long as it helps weakening the binding self-selection constraints. Suppose in fact to start from a situation in which only the income tax is in place and to introduce indirect taxation in a revenue- and welfare-neutral manner. If a mimicker has more taste for a good than the low skilled individual, then the discouragement inflicts a larger loss of real income on the former than on the latter: therefore, the compensation needed to restore the utility level of the low skilled agent is not enough for the mimicker, the incentive compatibility constraint is relaxed and we can tax the high ability individual more heavily without having him (her) trying to “disguise” himself (herself) as a low ability person. This can be better seen by multiplying eq. (16) by \( t_i \) and summing over \( i \) to find

\[
\sum_{k=1}^{2} \sum_{i=1}^{n} \sum_{j=1}^{n} t_i t_j \frac{\partial h_i^k}{\partial t_j} = \frac{\lambda V_2}{\beta} \left( \sum_{i=1}^{n} t_i x_i^1 - \sum_{i=1}^{n} t_i x_i^2 \right). \tag{17}
\]

If the Slutsky substitution matrix is of full rank and, therefore, negative definite, the left hand side of eq. (17) is negative and the mimicker will pay a higher amount of commodity taxes than the low ability type; since the mimicker will pay the same income tax as the mimicked, it follows that the total tax liability of a mimicker will exceed that of a mimicked. Following this reasoning, the right hand side of eq. (16) can be interpreted as a measure of the marginal benefit, in terms of “rent” extraction, of distorting consumption prices and eq. (16) as one requiring equality, at the optimum, between marginal costs and benefits.

Since the only difference between a mimicker and a mimicked regards the level of labour supply (leisure enjoyed), it is clear that the condition expressed by eq. (16) is strictly connected to the degree of complementarity (in the sense of conditional demand relations) of various goods with leisure. Roughly speaking, the condition requires that the discouragement must be

\[12\text{When the tax is negative, the “reduction” becomes an increase, but the argument does not change.}\]
greater the closer the complementarity between commodity $i$ and leisure (the closer the substitutability between commodity $i$ and labour). It is this link between unobservable ability type (labour supply) and the pattern of commodity consumption that is exploited in order to “screen” between agents of different types. This observation represents also the basis to provide an intuitive explanation for the well known AS theorem\textsuperscript{13}. According to their result, if preferences are weakly separable in labour supply and produced goods (that is, preferences can be represented by a utility function of the form $u[f(x),L]$), nonlinear income taxation does not need to be supplemented by commodity taxation: labour income is a “sufficient statistics” and any second best optimum can be achieved by income taxation alone. The reason is that weak separability makes conditional demands independent on labour supplied (and therefore independent on the wage rate), so that a mimicker will not only earn the same income as a low skilled individual but he (she) will also spend it over taxed commodities in exactly the same way. Indirect taxation can play no role in discriminating between agents.

3 The AS Theorem and the Role of Commodity Taxes

In this section we will trace out a brief survey of the main contributions that have been recently explored in the strand of literature “in defence of commodity taxation”. According to different authors, the relevance of the AS result, in a context where the set of instruments of the policy maker includes no other instruments than a nonlinear income tax and linear commodity taxes, is alternatively called into question by introducing tax evasion (Boadway, Marchand and Pestieau, 1994), uncertainty (Cremer and Gahvari, 1995), endogenous wages (Naito, 1999), household production (Anderberg and Balestrino, 2000) or bi-dimensional differentiation across agents (Cremer, Pestieau and Rochet, 2001).

3.1 Allowing for tax evasion (Boadway, Marchand and Pestieau, 1994)

The direct versus indirect tax controversy is addressed by Boadway, Marchand and Pestieau (1994) allowing for different taxes having different evasion characteristics. Their model exploits the idea that, if an otherwise ideal tax base can be evaded, obtaining some revenues from a parallel tax base that is less open to such a problem can mitigate the problem. More precisely, the authors assume, recovering an argument originally made by De Viti De Marco (1936), that indirect taxes may be more difficult to evade than direct\textsuperscript{13}We will take up this point again in the following section more in detail.
taxes. To avoid the complexities descending from the need to model how the benefits of commodity tax evasion are shared between producers and consumers, they consider the extreme case where evasion is possible only for direct taxes\textsuperscript{14}.

As in the basic model we exposed in the previous section, there are two types of agents, 1 (low skilled) and 2 (high skilled), the population of each type is normalised to unity, labour is the only input used in the economy, technology is linear and producer prices \( p = q - t \) are normalised to unity; instead of many different commodities, it is assumed that only two consumption goods, \( a \) and \( b \), are produced and that individuals have the same strictly quasi-concave utility functions, \( u \left( c^a_i, c^b_i, L_i \right) \), where \( c^a_i \) is consumption of good \( a \) by person \( i \) and \( L_i \) is his (her) labour supply. Tax evasion is introduced in a highly stylised way since the costs of it depend on the amount of income evaded, but it is also assumed that once these costs have been incurred, the tax evader is certain to escape detection by the tax authorities\textsuperscript{15}. Formally, individual \( i \) earns an amount of income \( w_i L_i \), of which a proportion \( \sigma^i \) is reported for income tax purposes. As the rates of marginal income taxation may be either positive or negative, both under- and overreporting of income (respectively \( \sigma^i < 1 \) and \( \sigma^i > 1 \)) is allowed for. Taking \( \sigma^i \) as given, it is assumed that the cost of concealing a unit of income is given by \( G(1 - \sigma^i) \), where \( G(\bullet) \) is a non-negative and strictly convex function with \( G(0) = 0 \) and \( G'(0) \) finite. Therefore, for an earned income of \( w_i L_i \) total concealment costs are given by \( (1 - \sigma^i) w_i L_i G(1 - \sigma^i) \); having defined \( g(1 - \sigma^i) \equiv (1 - \sigma^i) G(1 - \sigma^i) \), total concealment costs can be written as \( w_i L_i g(1 - \sigma^i) \), where \( g(\bullet) \) is non-negative and strictly convex with \( g(0) = g'(0) = 0 \). This means that the marginal cost of income concealment rises with the proportion of income under- or overreported and that the costs of concealment are proportional to gross income \( w_i L_i \).

The approach followed by Boadway, Marchand and Pestieau (1994) to solve the optimal taxation problem is to define a virtual budget constraint by linearising the after-tax budget constraint of an agent at the point on the tax schedule (optimally) chosen by him (her) through his (her) labour supply and income reporting decisions. In this way, to each household, it will be associated a specific marginal tax rate (denoted by \( 1 - \tau^i \)) and lump-sum component (denoted by \( T^i \)). The household’s problem can be stated as

\[ \text{subject to: } \]

\[ g(1 - \sigma^i) = (1 - \sigma^i) G(1 - \sigma^i) \]

\[ T^i = g'(0) = 0 \]

\[ \text{marginal cost of income concealment rises with the proportion of income under- or overreported and that the costs of concealment are proportional to gross income } w_i L_i. \]

\textsuperscript{14}The case where also commodity taxes can be evaded is considered by Hindriks (1999). However, contrary to what is common in optimal mixed taxation models, he assumes that personal consumption levels are to some extent observable and allows the government to use its (imperfect) observation of the purchases of certain goods to screen between low ability persons and mimickers. Nonlinear commodity taxation is however highly problematic at least for those goods which are easily retradeable (see footnote 7).

\textsuperscript{15}A drawback of such an approach is that what can be viewed as the fundamental feature of tax evasion, its riskiness to the taxpayer, is suppressed. It is worth noting that it was precisely the cost of risk-taking that conditioned evasive behaviour in the pioneering model by Allingham and Sandmo (1972).
follows:

$$\max_{L^1, \sigma, \gamma} \ u \left( c^i_a, c^i_b, L^i \right)$$

subject to

$$q_a c^i_a + q_b c^i_b = w^i L^i \left( 1 - \sigma^i + \sigma^i \tau^i \right) - T^i - w^i L^i g \left( 1 - \sigma^i \right).$$

Restricting the analysis to that part of the feasibility locus in which the planner redistributes income from the high- to the low skilled persons, the planner’s problem can be stated as follows:

$$\max_{\tau^1, T^1, \tau^2, T^2, q_a, q_b} V^1 \left( \tau^1, T^1, q_a, q_b \right)$$

subject to

$$V^2 \left( \tau^2, T^2, q_a, q_b \right) \geq V^i \left( \tau^1, T^1, q_a, q_b \right),$$

$$\sum_{i=1}^{2} \left[ \left( 1 - \tau^i \right) w^i L^i \sigma^i + T^i + (q_a - 1) c^i_a + (q_b - 1) c^i_b \right] \geq \mathcal{R},$$

where Lagrange multipliers are within parentheses, $V^i$ denotes indirect utilities and $\mathcal{R}$ the exogenous amount of revenue to be raised.

Notice that in this framework, a high skilled agent can try to masquerade his (her) true identity in two different ways: he (she) can either work many hours and conceal a large part of his (her) income or work few hours and conceal a small part of income.

For our purposes, it will be sufficient to derive the first order conditions with respect to the lump-sum components $T^1$ and $T^2$. Denoting by $\alpha^i$ the marginal utilities of income and using a “hat” to denote a variable when referred to a mimicker, these are respectively:

$$-\alpha^1 - \lambda \alpha^2 \left\{ -1 + w^1 \sigma^1 \left[ g^i - (1 - \tau^1) \right] \frac{\partial L^1}{\partial T^1} \right\} + \gamma \left[ 1 + (1 - \tau^1) w^1 \sigma^1 \frac{\partial L^1}{\partial T^1} \right] +$$

$$+ \gamma \left[ (q_a - 1) \frac{\partial c^i_a}{\partial T^1} + (q_b - 1) \frac{\partial c^i_b}{\partial T^1} \right] = 0; \quad (18)$$
\[-(\delta + \lambda) \alpha^2 + \gamma \left[ 1 + (1 - \tau^2) w^2 \sigma^2 \frac{\partial L^2}{\partial T^2} + (q_a - 1) \frac{\partial c_a^2}{\partial T^2} + (q_b - 1) \frac{\partial c_b^2}{\partial T^2} \right] = 0. \tag{19}\]

Now, differentiate the Lagrangian of the planner’s problem with respect to \(q_a\) and \(q_b\); using conditions (18) and (19) and denoting an income compensated variable by a “tilde”, yield the following conditions:

\[-\lambda \alpha^2 \left\{ \tilde{c}_a^1 - \tilde{c}_a^2 + \left[ \tilde{g}' - (1 - \tau^1) \right] w^1 \sigma^1 \frac{\partial \tilde{L}^1}{\partial q_a} \right\} + \]
\[+ \gamma \sum_{i=1}^{2} \left\{ (1 - \tau^i) w^i \sigma^i \frac{\partial \tilde{L}^i}{\partial q_a} + (q_a - 1) \frac{\partial \tilde{c}_a^1}{\partial q_a} + (q_b - 1) \frac{\partial \tilde{c}_a^2}{\partial q_a} \right\} = 0, \tag{20}\]
\[-\lambda \alpha^2 \left\{ \tilde{c}_b^1 - \tilde{c}_b^2 + \left[ \tilde{g}' - (1 - \tau^1) \right] w^1 \sigma^1 \frac{\partial \tilde{L}^1}{\partial q_b} \right\} + \]
\[+ \gamma \sum_{i=1}^{2} \left\{ (1 - \tau^i) w^i \sigma^i \frac{\partial \tilde{L}^i}{\partial q_b} + (q_a - 1) \frac{\partial \tilde{c}_b^1}{\partial q_b} + (q_b - 1) \frac{\partial \tilde{c}_b^2}{\partial q_b} \right\} = 0. \tag{21}\]

To show that AS result does not hold in this context, we search for a set of conditions under which the optimal commodity tax structure is uniform so the planner would not want to differentiate \(q_a\) from \(q_b\). At this point a clarification is needed. In the absence of evasion, an optimal tax structure calling for no differentiation would mean that efficiency in redistribution can be equivalently attained relying solely on the general income tax (commodity taxation is redundant), since the effect of proportional commodity taxation can be replicated by adjusting the level of the income tax. Conditions for uniform commodity taxation are at the same time conditions for redundancy of commodity taxation. This is no longer true if, as we supposed, direct and indirect taxation have different evasion characteristics and the equivalence between reported income and consumption is broken. In this case, it is intuitive and easy to show that there is a role for commodity taxation even under sufficient conditions for uniform taxation. Therefore, in a certain sense, we have already an argument against the AS result. However, what we are trying to prove here is slightly different: we want to show that sufficient conditions provided by AS for proportional taxation are no longer sufficient for this aim once, as in the present context, tax evasion is allowed for\textsuperscript{16}.

\textsuperscript{16} Intuitively, in order to mimic the low skilled type and report the same income, a high skilled agent will combine lower labour supply with higher concealment. A mimicker will therefore have more of both income and leisure than the mimicked and differentiated commodity taxation turns out to be a useful screening instrument even if, as it happens in the context of the AS theorem, the marginal rates of substitution between each pair of goods are not affected by differences in the amounts of leisure enjoyed.
Coming back to eq. (20) and (21), notice that the homogeneity property of consumer demand implies that at $q_a = q_b = q$:

$$q \left( \frac{\partial \tilde{c}_a}{\partial q_a} + \frac{\partial \tilde{c}_b}{\partial q_a} \right) = w^i \left[ 1 - \sigma^i + \sigma^i \tau^i - g \left( 1 - \sigma^i \right) \right] \frac{\partial \tilde{L}^i}{\partial q_a}. \tag{22}$$

Substituting eq. (22) into (20) and (21) and defining

$$x \equiv w^i \sigma^1 \left[ \hat{g} - (1 - \tau^1) \right],$$

$$z^i = \frac{w^i}{q} \left\{ \sigma^i (1 - \tau^i) + (q - 1) \left[ 1 - g \left( 1 - \sigma^i \right) \right] \right\}, \quad i = 1, 2,$$

one obtains:

$$-\lambda \alpha^2 \left( c^a - \tilde{c}^a + x \frac{\partial L^i}{\partial q_a} \right) + \gamma \sum_{i=1}^{2} z^i \frac{\partial \tilde{L}^i}{\partial q_a} = 0, \tag{23}$$

$$-\lambda \alpha^2 \left( c^b - \tilde{c}^b + x \frac{\partial L^i}{\partial q_b} \right) + \gamma \sum_{i=1}^{2} z^i \frac{\partial \tilde{L}^i}{\partial q_b} = 0. \tag{24}$$

Dividing (23) by (24) gives

$$\frac{c^a - \tilde{c}^a + x \frac{\partial L^i}{\partial q_a}}{c^b - \tilde{c}^b + x \frac{\partial L^i}{\partial q_b}} = \frac{\sum_{i=1}^{2} z^i \frac{\partial \tilde{L}^i}{\partial q_a}}{\sum_{i=1}^{2} z^i \frac{\partial \tilde{L}^i}{\partial q_b}}. \tag{25}$$

Assuming weak-separability between goods and leisure in the utility function allows us to write

$$\frac{\partial \tilde{L}^i}{\partial q_a} = m^i_a K^i, \tag{26}$$

$$\frac{\partial \tilde{L}^i}{\partial q_b} = m^i_b K^i, \tag{27}$$

where $m^i_a$ and $m^i_b$ are respectively the marginal propensity of agent $i$ to consume commodity $a$ and $b$ out of income, and $K^i$ is a common term not depending on good $a$ or $b$. Dividing (26) by (27) gives

$$\frac{\partial \tilde{L}^i}{\partial q_a} = \frac{m^i_a}{m^i_b} K^i, \quad i = 1, 2. \tag{28}$$

Substituting (28) into (25), we obtain:
A condition for eq. (29) to be satisfied is that

\[
\frac{c_a^1 - c_a^2}{c_b^1 - c_b^2} + \frac{m_a^1}{m_b^1} \frac{\partial L}{\partial p_a} + \frac{m_a^2}{m_b^2} \frac{\partial L}{\partial p_b}
\]

\[
= \sum_{i=1}^{2} \alpha_i \frac{\partial L^{i}}{\partial p_b}.
\]  

(29)

A condition for eq. (29) to be satisfied is that

\[
\frac{c_a^1 - c_a^2}{c_b^1 - c_b^2} = \frac{m_a^1}{m_b^1} = \frac{m_a^2}{m_b^2}.
\]

(30)

Condition (30) requires that the ratio of the marginal propensities to consume is independent of income. It descends that weak separability between leisure and other goods in the utility function is clearly not sufficient to justify a uniform tax structure; actually, condition (30) is satisfied if and only if, besides being weakly separable between leisure and other goods, individual preferences are also quasi-homothetic in goods.

### 3.2 Allowing for Uncertainty (Cremer and Gahvari, 1995)

In the contribution by Cremer and Gahvari, the two ability type Stiglitz-Stern model is reinterpreted in the sense that all agents are *ex ante* equal but face uncertainty about the returns to their labour which gives rise to an unequal distribution of the earning capabilities *ex post*. Population is composed of a continuum of agents, normalised to unity, earning a random wage which has two possible realisations: low, \( w^1 \), and high, \( w^2 \), occurring respectively with probabilities \( \pi^1 \) and \( \pi^2 = 1 - \pi^1 \). A particular individual’s post-uncertainty wage and labour supply\(^\text{17}\) are not publicly observable, but his (her) income \( (Y = wL) \) is. There are two groups of commodities produced in the economy: those the consumption of which must be committed to before the resolution of the uncertainty (pre-committed goods\(^\text{18}\) denoted by vector \( \mathbf{z} = (z_a, z_b) \)), and those the consumption of which can be decided after the resolution of the uncertainty (post-uncertainty goods denoted by \( \mathbf{x} = (x_a, x_b) \)). All producer prices are set equal to one. Time endowment is normalised to one. Preferences are separable between leisure and other goods and can be represented by the utility function

\[
U = u(z_a, z_b, x_a, x_b) + \phi(1 - L),
\]

where \( U \) is assumed to be twice continuously differentiable, strictly increasing in \( \mathbf{z} \) and \( \mathbf{x} \), and strictly decreasing in \( L \); moreover, to ensure risk aversion, \( u \) is assumed to be strictly, and \( \phi \) non strictly, concave.

\(^{17}\)Labour supply is decided after the resolution of the uncertainty. Assuming otherwise would allow to levy lump-sum and first-best efficient taxes (see Lundholm, 1992).

\(^{18}\)Think at purchase of durable goods in general and housing in particular.
The optimal tax structure is the one implementing the Pareto-efficient allocation corresponding to the maximum of a utilitarian social welfare function\(^{19}\). The optimal allocation descends from the solution to the following problem:

$$\max_{z_a, z_b, x_a^i, x_b^i} \sum_{i=1}^{2} \pi^i \left[ u(z_a, z_b, x_a^i, x_b^i) + \phi \left( 1 - \frac{Y^1}{w^1} \right) \right]$$

subject to

$$u(z_a, z_b, x_a^i, x_b^i) + \phi \left( 1 - \frac{Y^2}{w^2} \right) \geq u(z_a, z_b, x_a^1, x_b^1) + \phi \left( 1 - \frac{Y^1}{w^1} \right), \quad (\lambda)$$

$$\sum_{i=1}^{2} \pi^i (Y^i - x_a^i - x_b^i) - z_a - z_b \geq \overline{R}, \quad (\gamma)$$

where Lagrange multipliers are within parentheses, \(x_k^i\) denotes consumption of commodity \(x_k\) by an agent being paid a wage rate \(w^i\), and \(\overline{R}\) the exogenous amount of revenue to be raised.

Denoting \(u(z_a, z_b, x_a^1, x_b^1)\) and \(u(z_a, z_b, x_a^2, x_b^2)\) by \(u^1\) and \(u^2\) respectively, the first order conditions are:

$$\pi^1 \frac{\partial u^1}{\partial z_a} + \pi^2 \frac{\partial u^2}{\partial z_a} + \lambda \left( \frac{\partial u^2}{\partial z_a} - \frac{\partial u^1}{\partial z_a} \right) - \gamma = 0, \quad (32)$$

$$\pi^1 \frac{\partial u^1}{\partial z_b} + \pi^2 \frac{\partial u^2}{\partial z_b} + \lambda \left( \frac{\partial u^2}{\partial z_b} - \frac{\partial u^1}{\partial z_b} \right) - \gamma = 0, \quad (33)$$

$$\left( \pi^1 - \lambda \right) \frac{\partial u^1}{\partial x_a^i} - \gamma \pi^1 = 0, \quad (34)$$

$$\left( \pi^1 - \lambda \right) \frac{\partial u^1}{\partial x_b^i} - \gamma \pi^1 = 0, \quad (35)$$

$$\left( \pi^2 + \lambda \right) \frac{\partial u^2}{\partial x_a^i} - \gamma \pi^2 = 0, \quad (36)$$

$$\left( \pi^2 + \lambda \right) \frac{\partial u^2}{\partial x_b^i} - \gamma \pi^2 = 0, \quad (37)$$

$$\pi^1 \frac{\partial \phi}{\partial Y^1} \left( 1 - \frac{Y^1}{w^1} \right) - \lambda \frac{\partial \phi}{\partial Y^1} \left( 1 - \frac{Y^1}{w^1} \right) + \gamma \pi^1 = 0, \quad (38)$$

\(^{19}\)Cremer and Gahvari claim that, since all agents are identical \textit{ex ante}, this seems the more natural concept of optimality to employ.
\[(\pi^2 + \lambda) \frac{\partial \phi}{\partial Y^2} \left(1 - \frac{Y^2}{w^2}\right) + \gamma \pi^2 = 0.\]  

(39)

To implement the optimal allocation, the government must properly design the nonlinear income tax schedule \(T(Y)\) and choose the set of linear commodity taxes. We will use the convention to denote by \(q_a, q_b, \bar{q}_a\) and \(\bar{q}_b\) the consumer prices of \(z_a, z_b, x_a\) and \(x_b\) respectively (remember that producer prices are set equal to one). The individual is an expected utility maximiser and solves the following optimisation problem:

\[
\max_{z_a,z_b,x_a^i,x_b^i,\bar{x}_a^i,\bar{x}_b^i,Y^1,Y^2} \sum_{i=1}^{2} \pi^i \left[ u(z_a,z_b,x_a^i,x_b^i) + \phi \left(1 - \frac{Y^i}{w^i}\right)\right]
\]

subject to \(Y^j - T(Y^j) = q_a z_a + q_b z_b + \bar{q}_a x_a^i + \bar{q}_b x_b^i, \quad j = 1, 2.\)

The individual’s optimal behaviour is expressed by the following conditions obtained by manipulating the first order conditions of the problem above:

\[
\frac{\partial u^1}{\partial x^1} = \frac{\partial u^2}{\partial x^2} = \frac{\bar{q}_a}{\bar{q}_b} \tag{40}
\]

\[
\frac{\pi^1 \partial u^1}{\partial z_a} + \frac{\pi^2 \partial u^2}{\partial z_a} = \frac{q_a}{q_b} \tag{41}
\]

\[
\frac{\pi^1 \partial u^1}{\partial z_b} + \frac{\pi^2 \partial u^2}{\partial z_b} = \frac{q_b}{q_b} \tag{42}
\]

\[
\frac{\partial \phi \left(1 - \frac{Y^1}{w^1}\right)}{\partial Y^1} = 1 - \frac{T'(Y^1)}{\bar{q}_i}, \quad i = a, b, \tag{43}
\]

\[
\frac{\partial \phi \left(1 - \frac{Y^2}{w^2}\right)}{\partial Y^2} = 1 - \frac{T'(Y^2)}{\bar{q}_i}, \quad i = a, b. \tag{44}
\]

To establish a case for commodity taxation it is sufficient to show that in general, despite the existence of a nonlinear income tax and the separability between leisure and other goods, a commodity belonging to the pre-committed group should be taxed differently from a commodity belonging to the other group. Algebraic manipulation of eq. (32), (33), (36) and (37) yields
\[
\frac{\pi^1 \frac{\partial u^1}{\partial z_i} + \pi^2 \frac{\partial u^2}{\partial z_j}}{\pi^1 \frac{\partial u^1}{\partial x_j} + \pi^2 \frac{\partial u^2}{\partial x_j}} = 1 + \lambda \left( \pi^1 - \lambda \right) \left( \pi^2 + \lambda \right) \frac{\left( \frac{\partial u^1}{\partial z_i} - \frac{\partial u^2}{\partial z_j} \right) - \gamma \lambda}{\gamma \left[ \pi^1 \lambda + \pi^2 \left( \pi^1 - \lambda \right) \right]}.
\]

(45)

Putting together eq. (45) and eq. (42) we have

\[
\frac{q_i}{q_j} = 1 + \lambda \frac{\left( \pi^1 - \lambda \right) \left( \pi^2 + \lambda \right) \left( \frac{\partial u^1}{\partial z_i} - \frac{\partial u^2}{\partial z_j} \right) - \gamma \lambda}{\gamma \left[ \pi^1 \lambda + \pi^2 \left( \pi^1 - \lambda \right) \right]}.
\]

(46)

According to eq. (46), it is not in general true that the ratio between the consumer prices of \(z_i\) and \(x_j\) should be set equal to one (uniform taxation); this can only happen by chance, for a particular utility function and for a particular value of \(\pi^1\) and \(\pi^2\), when the following condition is satisfied:

\[
\frac{\partial u^1}{\partial z_i} - \frac{\partial u^2}{\partial z_j} = \frac{\gamma \lambda}{\left( \pi^1 - \lambda \right) \left( \pi^2 + \lambda \right)}.
\]

This would be sufficient to conclude that the AS result no longer holds; however, Cremer and Gahhari are also able to prove other interesting results that we mention without going through the details of the proofs. In particular, they show that the post-uncertainty goods should go either tax free or be subject to a uniform tax rate and that the commodities belonging to the pre-committed group must be taxed differently from one another unless the following condition holds:

\[
\frac{\partial u^1}{\partial z_a} - \frac{\partial u^2}{\partial z_a} = \frac{\partial u^1}{\partial z_b} - \frac{\partial u^2}{\partial z_b}.
\]

Intuitively, the usefulness of commodity taxation can be explained by referring to two different reasons. On the one hand, the distinction between committed and uncommitted goods may be exploited by the government in order to use commodity taxation as a basis of separation between low- and high-wage households, and this despite the separability between leisure and other goods. On the other hand, uncertainty and commitment make commodity taxation useful as an insurance mechanism even if it cannot be used as a separation device.

3.3 Allowing for Endogenous Wages (Naito, 1999)

The contribution by Naito (1999) considers a two agent types, two factors and two goods model of a closed perfectly competitive economy where the assumption of constant marginal cost of production is abandoned and different individuals are not considered as being perfect substitutes for one another in production process. The economy consists of two types of decision making agents: (i) consumers choosing their consumption of private
goods and labour supply, and (ii) a government choosing tax rates. The tax instruments facing the government are a nonlinear income tax and linear commodity taxes. Since our economy is made up of only two goods, we can choose one as *numéraire* and set it untaxed so as to reduce the design of the indirect tax structure to the selection of the appropriate commodity tax (subsidy) \( t \) on good two.

There are two agent types represented by low skilled workers (denoted by superscript 1) and high skilled workers (denoted by superscript 2). For simplicity we assume that the population of both type of workers is the same and we normalize it to one. The low skilled workers supply low skilled labour while the high skilled workers supply high skilled labour. Both types of individuals have the same quasiconcave utility function \( u = u(z, x, L) \) depending on the consumption of the two private goods \((z, x)\) and on leisure \((L\) denotes labour supply, i.e. time subtracted to leisure); besides, we assume that both private goods and leisure are normal goods.

There are two industries in the private sector of this economy: the first produces good one (and the quantity produced is denoted by \( y_1 \)) while the second produces good two (and the quantity produced is denoted by \( y_2 \)). Each one of the two production functions is concave and exhibits constant returns to scale; besides, each industry uses low skilled as well as high skilled labour. So we have

\[
\begin{align*}
  y_1 &= F^1 (L^1_1, L^2_1), \\
  y_2 &= F^2 (L^1_2, L^2_2),
\end{align*}
\]

where superscript denotes type while subscript denotes sector so that \( L^i_k \) is type \( i \) labour used in sector \( k \). Let \( w^1 \) and \( w^2 \) denote wages for low skilled and high skilled workers, respectively. Each industry maximises its profit taken as given the price of goods and wages. We assume that one of the two industries is always high skilled-labour intensive for each pair of wages. Since the economy is closed, we assume that both goods are produced in equilibrium (the equilibrium is not “specialised”) without loss of generality. If \( C_k (w^1, w^2) \) is the cost function to produce one unit of good \( k \), then perfect competition and constant returns to scale imply:

\[
\begin{align*}
  C_1 (w^1, w^2) &= 1, \\
  C_2 (w^1, w^2) &= p.
\end{align*}
\]

Given a price \( p \) the above equations determine \( w^1 \) and \( w^2 \) uniquely and so it is possible to write the wage ratio \( \Omega = \frac{w^1}{w^2} \) as a function of \( p \): \( \Omega = \Omega (p) \). According to the Stolper-Samuelson theorem, if \( p \) increases, then the wage of the labour force intensively used in the sector producing good two will increase while the other will decrease.
By using Shephard’s lemma, factor demands are

\[ L_1^k = y_k \frac{\partial C_k (w^1, w^2)}{\partial w^1}, \]
\[ L_2^k = y_k \frac{\partial C_k (w^1, w^2)}{\partial w^2}. \]

We assume that labour is perfectly mobile among the two private sectors. Thus, the labour market equilibrium conditions are

\[ L_1 = L_1^1 + L_1^2, \]
\[ L_2 = L_2^1 + L_2^2. \]

By Walras’ law we can restrict our attention to one of the two goods market equilibrium conditions. Choosing for this aim the commodity two market and denoting with \( x^1, x^2 \) the demand for good two expressed by low- and high skilled agents respectively, we have

\[ y_2 = x^1 + x^2. \]

We can write the output function of good two as:

\[ y_2 = Y \left( p; L^1, L^2 \right), \]

which can be seen (cfr. Dixit and Norman, 1980, p.326), by a special case of the envelope theorem, as the derivative with respect to \( p \) of the economy’s revenue function.

Since the technology is convex and factor intensity is different between the two sectors, the production possibility set is a strictly convex set and so

\[ \frac{\partial Y \left( p; L^1, L^2 \right)}{\partial p} > 0 \quad \text{if} \quad Y > 0. \]

In this 2x2 framework we can also invoke the Rybczynski theorem. Accordingly, given a fixed price \( p \), if the high skilled (low skilled) labour force increases, then the production of the high skilled (low skilled) labour intensive good will increase, while the production of the low skilled (high skilled) labour intensive good will decrease.

Finally, the market equilibrium price \( p \) is determined by the following equation:

\[ x^1 (p + t, B^1, L^1) + x^2 (p + t, B^2, L^2) = Y \left( p; L^1, L^2 \right). \quad (47) \]

From the assumption of non-separability it follows that, given a certain income, the choice between \( z \) and \( x \) is dependent on the quantity of labour.
an agent is supplying. Denoting by \( q \) the consumer price of good two \( (q = p + t) \), and remembering that we set good one as the untaxed good and as numéraire, the planner’s problem can be stated as

\[
\max_{L^1, L^2, B^1, B^2, t} V^1 (p (\cdot) + t, B^1, L^1)
\]

subject to:

\[
V^2 (p (\cdot) + t, B^2, L^2) \geq \overline{V}, \tag{6}
\]

\[
V^2 (p (\cdot) + t, B^2, L^2) \geq V^2 (p (\cdot) + t, B^1, \Omega (p (\cdot)) L^1), \tag{\lambda}
\]

\[
\sum_{i=1}^{2} (w^i (p (\cdot)) L^i - B^i + tx^i) \geq \overline{R}, \tag{\gamma}
\]

where \( \overline{V} \) is a pre-set utility level, \( \overline{R} \) is the public revenue requirement and Lagrange multipliers are within parentheses.

For simplicity, in what follows we write \( V^i_L, V^i_B, V^i_q, x^i_L, x^i_B, x^i_q \) instead of respectively \( \frac{\partial V^i}{\partial L}, \frac{\partial V^i}{\partial B}, \frac{\partial V^i}{\partial q}, \frac{\partial x^i}{\partial L}, \frac{\partial x^i}{\partial B}, \frac{\partial x^i}{\partial q} \), while a hat will characterise a variable when referred to a mimicker. Primes denote derivatives.

Manipulating the relevant first order condition, it can be shown that the optimal commodity tax rate obeys the following rule:

\[
t = \frac{\lambda V^2_L \Omega' (p) L^1}{\gamma Y_p} - \frac{\lambda V^2_B \left( \hat{x}^2 - x^1 \right)}{\gamma (h_q^1 + h_q^2)}. \tag{48}
\]

Since the difference between the non-separable and the weakly separable case relies on the fact that in the latter the way in which a given amount of income is allocated over the consumption goods is not related to the amount of labour an agent is supplying and therefore a mimicker and a true low skilled agent would have the same pattern of consumption, in eq. (48) the second term on the right hand side vanishes, being \( \hat{x}^2 = x^1 \). It turns out that, even in the case of weakly separable utility function, the optimal commodity tax structure is not uniform and that in this special case we would get:

\[
t = \frac{\lambda V^2_L \Omega' (p) L^1}{\gamma Y_p}. \tag{49}
\]

Since \( \frac{\lambda V^2_L L^1}{\gamma Y_p} < 0 \) always, we also have that the commodity tax is properly a tax (subsidy) if \( \Omega' (p) < (>) 0 \), which from Stolper-Samuelson theorem
means that the sector producing good two is high (low) skilled labour intensive. We can conclude that the sign of the commodity tax rate depends solely on technological factors.

As pointed out in Naito (1999), equation (49) shows that the AS result no longer holds if the production side of the economy is explicitly taken into account and general equilibrium effects arise. In particular, the numerator of the right side term of equation (49) represents the social valuation of the weakening of the self-selection constraint brought about by a change of the wage ratio through a change of the producer price: it measures the additional disutility in terms of additional labour a mimicker is experiencing in trying to pretend to be recognised as a low skilled agent.

This result has been recently called into question in a paper by Saez (2002b), where it is argued that the validity of the AS theorem would be only affected in the short-run, when people react to fiscal policy by adjusting their optimal choice of labour supply within occupations. On the other hand, if people react in the long-run through the “occupation choice margin”, then the AS theorem would be robust to the relaxation of the assumption of exogenous wages. The analysis of the possibly different effects which can arise in alternative temporal settings is of course of great importance in order to fully evaluate the desirability of a certain policy instrument. However, it will not be worthless to stress that the main reason driving the Saez’s result seems to be that in his long-run context self-selection problems have been definitely overcome and are not constraining the government’s policy anymore; but, then, we know that the non-linear tax scheme is essentially a lump-sum tax.

Moreover, Naito (1999) shows that production efficiency may not be Pareto optimal. Developing the Naito’s analysis, Blackorby and Brett (2000) show that, given an optimal nonlinear income tax, production inefficiency is Pareto optimal if the aggregate technology set is strictly concave and wages are fixed. They also prove that if there is production inefficiency then the AS conditions are neither necessary nor sufficient for zero commodity taxation; therefore, commodity taxes may be needed even if they don’t affect relative wages.

Weakening the assumption of exogenous wages (ratio) is in itself, as shown by Naito (1999), a sufficient way to reestablish indirect taxation. Depending on the specific mechanism one builds up in order to endogenise wages, there can also stand out other reasons to justify the use of differentiated commodity taxation. This is for example what happens in the model by Aronsson and Sjögren (2001), one of the very few papers dealing with optimal taxation in a non-Walrasian labour market, where the two types Stiglitz-Stern model of optimal taxation is adapted to the case of a small open economy with unemployment generated by union wage setting. In their case Pareto efficient taxation requires to take also into account what they call a “tax base effect” that basically works exploiting the way tax instruments affect the unemployment rate.
3.4 Allowing for Household Production (Anderberg and Balestrino, 2000)

The paper by Anderberg and Balestrino (2000) readdresses the direct versus indirect tax controversy by introducing household production into the framework. The fundamental consequence descending from the explicit consideration of this new element and the basic reason why some of the standard tax policy prescriptions are affected is that a wedge arises between “pure leisure” and “non-market use of time”. Ruling out household production as it has been so far common practice (with few exceptions) in the optimal taxation literature implies that talking about something called “leisure” is without any risk of confusion, even if for some authors the word carries the connotation of “pure leisure” (see e.g. Gronau, 1977), while by others it is used as a synonymous with “labour” (see e.g. Myles, 1995). The feature distinguishing the model by Anderberg and Balestrino from the basic model presented in the first section is that each household allocates its endowment of time (normalised to unity) to market labour (\(L\)) and two non-market uses, called “pure leisure” (\(l\)) and “home-time” (\(h\)). Pure leisure enters directly in the utility function, whereas home-time is combined with \(m\) marketed commodities (“inputs”) to produce (home production) a final consumption good \(x\). Other \(n\) final consumption goods are bought directly on the market. As usually assumed, technologies are linear, producer prices are set equal to one, the population is composed by two types of agents, differing for market ability (\(w^1 < w^2\)), and the size of each ability group is normalised to unity. For our purposes it will be sufficient to focus on the case where \(m = n = 1\) and pure leisure enters the utility function additively\(^{22}\). The household production technology is represented by the constant returns to scale production function \(x = x(z, h)\), where \(z\) is the marketed commodity used as input, and households’ preferences are represented by the common utility function
\[ u = f(c, x(z, h)) + \phi(1 - L - h), \]
where \(c\) is the only consumption good directly bought on the market and \(f\) and \(\phi\) are increasing concave functions. The general income tax schedule is denoted by \(T(Y)\), disposable income by \(B = Y - T(Y)\) and per-unit tax rates on commodities \(c\) and \(z\) by \(t_c\) and \(t_z\) respectively.

The consumer’s optimisation problem is
\[
\max_{c,z,h} \quad f(c, x(z, h)) + \phi(1 - L - h) \\
\text{subject to} \quad (1 + t_c)c + (1 + t_z)z = wL - T(wL).
\]

Denoting by \(V\) the indirect utility, the government’s problem becomes:
\[
\max_{t_c,t_z,B^1,B^2,L_1,L_2} V^1 \left(1 + t_c, 1 + t_z, B^1, L^1\right) \\
\]
\(^{22}\)This is the case dealt with in section 4 of the Anderberg-Balestrino’s paper.
subject to

\[ V^2(1 + t_c, 1 + t_z, B^2, L^2) \geq \nabla, \]  \hspace{1cm} (\delta)

\[ V^2(1 + t_c, 1 + t_z, B^2, L^2) \geq V^2\left(1 + t_c, 1 + t_z, B^1, \frac{Y^1}{w^2}\right), \]  \hspace{1cm} (\lambda)

\[ \sum_{i=1}^{2} (w^i L^i - B^i + t_c c^i + t_z z^i) \geq \overline{R}, \]  \hspace{1cm} (\gamma)

where Lagrange multipliers are within parentheses, superscripts refer to individuals, \( \nabla \) is a pre-set utility level and \( \overline{R} \) is the exogenous revenue requirement.

Manipulating the first order conditions of the government’s problem, using a “hat” to distinguish a variable when referred to a mimicker and denoting by a “tilde” a compensated variable, it can be shown that the optimal commodity tax rates obey the following rules:

\[ \sum_{i=1}^{2} \left( t_c \frac{\partial \tilde{c}^i}{\partial t_c} + t_z \frac{\partial \tilde{c}^i}{\partial t_z} \right) = \frac{\lambda \frac{\partial V^2}{\partial B^2}}{\gamma} \left( c^1 - \tilde{c}^2 \right), \]  \hspace{1cm} (50)

\[ \sum_{i=1}^{2} \left( t_c \frac{\partial \tilde{z}^i}{\partial t_c} + t_z \frac{\partial \tilde{z}^i}{\partial t_z} \right) = \frac{\lambda \frac{\partial V^2}{\partial B^2}}{\gamma} \left( z^1 - \tilde{z}^2 \right). \]  \hspace{1cm} (51)

From conditions (50) and (51) we have that uniform commodity taxation hinges on the impossibility to screen among agents belonging to different ability types. This is the standard message derived analysing the basic model presented in the first section: same consumption pattern implies indiscernibility of agents which in turn implies uselessness of commodity taxation. However, as we will show, once household production is allowed for, separability between pure leisure and other goods is no longer sufficient to establish redundancy of indirect taxation, since what would actually be required is a more restrictive and rather implausible assumption, namely separability of all goods from all uses of non-market time.

Consider a mimicker: he (she) has a longer non-market time to allocate between pure leisure and home-time than does a true low skilled agent. This in turn means, by the assumptions made about the utility function, that he (she) will allocate more time to both activities. Now look at the way the home-time affects the marginal rate of substitution between \( c \) and \( z \); with obvious notation it is:
Denoting by $\varepsilon$ the elasticity of substitution between $c$ and $x$ in consumption and by $\eta$ the elasticity of substitution between $z$ and $h$ in the household production process, we can write:

$$\frac{\partial \ln MRS^u_{xc}}{\partial h} = \frac{\partial \ln MRS^f_{xc}}{\partial h} - \frac{\partial \ln MRT^x_{hz}}{\partial h} - \frac{\partial \ln x_h}{\partial h}. \quad (52)$$

Since we have assumed that the household production technology exhibits constant returns to scale ($x = x_h h + x_z z$), the following properties hold:

$$0 = x_{hh} h + x_{zh} z, \quad \eta = (x_{zh} x)^{-1} x_z x_h. \quad (55)$$

Making use of (55), the two last terms on the right hand side of (54) vanish and it remains:

$$\frac{\partial \ln MRS^u_{zc}}{\partial h} = (\varepsilon^{-1} - \eta^{-1}) x_h. \quad (56)$$

The possibility to use indirect taxation as an instrument to weaken the binding self-selection constraint depends on the circumstance that the marginal rate of substitution between the two marketed commodities $c$ and $x$ changes with the amount of time devoted to household production; otherwise a mimicker and a true low skilled agent appear indiscernible. Therefore, what condition (56) shows is that differentiated commodity taxation remains a useful instrument, despite separability between pure leisure and other goods in the utility function, unless the elasticity of substitution in consumption between $c$ and $x$ happens to be equal to the elasticity of substitution in the household production process between $z$ and $h$. In particular, we have:
Intuition would suggest that the results referred to the validity of the AS theorem obtained in the case where agents are equally skilled in the household production could be strengthened assuming bi-dimensional individuals differentiation. This has been shown to be the case by Balestrino, Cigno and Pettini (1999) in a model where the domestic production function is given by $x = x(z, h; b)$, where $b$ denotes a productivity parameter varying across agents. Even if the individual’s utility function is taken to be weakly separable between leisure and other goods, so long as there are differences in the home-productivity parameters, the demand of a mimicker will be different from that of the mimicked, thus restoring a role for indirect taxation to help weakening the binding incentive-compatibility constraint.

A special case of the models allowing for differences in market as well as non-market abilities has been developed in a number of recent papers (Cigno, 2001; Balestrino, Cigno and Pettini, 2002) dealing with the optimal design of family taxation when adult household members are treated as a single agent but households are differentiated both by ability to raise children and by ability to raise income. This strand of literature belongs to the field of household production models focusing on the home-production of non-tradeable goods as opposed to that concerned with the home-production of (tradeable) goods for which it makes sense to talk about market substitutes. In particular, children play the role of a non-tradeable good. Moreover, what makes especially attractive the analysis of this type of non-tradeable good is that it displays another valuable (from the policy maker’s perspective) feature, namely observability of the quantity “consumed” (number of children) at no (or rather negligible) cost for the fiscal authority.

### 3.5 Allowing for Differences in Endowments (Cremer, Pestieau and Rochet, 2001)

Cremer, Pestieau and Rochet (2001) weaken what they consider the most fundamental shortcoming of the AS analysis, that individuals differ in one single characteristic, and consider a setting where households differ not only in earning abilities but also in endowments. There are two types of individuals\(^{23}\) who differ in their labour productivities $w^i (i = 1, 2)$ and their initial endowments $\omega^i$ (“underbars” denote vectors) of $n$ consumption goods. The proportion of types $i$ in the population is denoted by $\pi^i$. All individuals have the same strictly quasi-concave utility function $u(c^i) - v(L^i)$, where $c^i$ is the $n$-dimensional vector of consumption and $L^i$ denotes labour supply.

\(^{23}\)In the original paper the authors allow for $N$ different types of individuals. We consider the case where $N = 2$ in order not to depart from the two types model used so far.
Technologies are as usual taken linear and producer prices normalised to one. Tax policy consists of a general income tax $T(\bullet)$ on labour income $Y$ and of $\underline{t}$, the vector of linear commodity taxes\footnote{Having an extra degree of freedom in selecting commodity tax rates, we can set without prejudice the tax rate on the first commodity equal to zero.} which determine $q$, the vector of consumer prices.

Denoting by $B^i$ the disposable income, given by after tax labour income plus market value of endowments, of individual $i$, and by $V(q,B)$ the indirect utility function associated with $u(\bullet)$, the problem of individual $i$ can be formalised as follows:

$$\max V(q,B^i) - v \left( \frac{Y^i}{w^i} \right)$$

subject to $B^i = Y^i - T(Y^i) + q^t i$.

The income tax schedule can be summarised by the two points $(Y^i,T^i)$ representing the individuals' respective choices. Incentive compatibility constraints require that the utility obtained by each individual $i$ when consuming $(Y^i,T^i)$ be at least equal to the utility level achievable mimicking individual $j$. Formally:

$$U^i = V(q,B^i) - v \left( \frac{Y^i}{w^i} \right) \geq U^{i(j)} = V(q,B^{i(j)}) - v \left( \frac{Y^{j}}{w^{j}} \right),$$

where $B^{i(j)} = B^j + q (\omega^j - \omega^j)$.

The government aims at maximising a weighted sum of utilities, subject to a revenue constraint and the incentive compatibility constraints. The weights of types $i$ are denoted by $\alpha^i$, with $\alpha^i \geq 0$ and $\alpha^1 = 1 - \alpha^2$. The problem can be stated as follows:

$$\max \sum_{i=1}^{2} \alpha^i \pi^i U^i$$

subject to

$$U^i \geq U^{i(j)}, \quad i, j = 1, 2,$$

$$\sum_{i=1}^{2} \pi^i (Y^i - B^i + t e^i + 1 \mu^i) \geq \bar{R},$$

where Lagrange multipliers are within parentheses and $\bar{R}$ is the exogenous revenue requirement.
Denoting by $\Lambda$ the Lagrangian of the government’s problem, the relevant (for our purposes) first order conditions are:

\[
\frac{\partial \Lambda}{\partial B^1} = \left( \alpha^1 \pi^1 + \lambda^{1(2)} \right) \frac{\partial V(q, B^1)}{\partial B} - \gamma^1 \pi^1 \left( 1 - t_c^1 \right) - \lambda^{2(1)} \frac{\partial V(q, B^{2(1)})}{\partial B} = 0,
\]

(57)

\[
\frac{\partial \Lambda}{\partial B^2} = \left( \alpha^2 \pi^2 + \lambda^{2(1)} \right) \frac{\partial V(q, B^2)}{\partial B} - \gamma^2 \pi^2 \left( 1 - t_c^2 \right) - \lambda^{1(2)} \frac{\partial V(q, B^{1(2)})}{\partial B} = 0,
\]

(58)

\[
\frac{\partial \Lambda}{\partial q_k} = \left( \alpha^1 \pi^1 + \lambda^{1(2)} \right) \frac{\partial V(q, B^1)}{\partial q_k} + \gamma^1 \pi^1 \left( c_k^1 + \frac{\partial \varphi^1}{\partial q_k} \right) +
+ \left( \alpha^2 \pi^2 + \lambda^{2(1)} \right) \frac{\partial V(q, B^2)}{\partial q_k} + \gamma^2 \pi^2 \left( c_k^2 + \frac{\partial \varphi^2}{\partial q_k} \right) +
- \lambda^{2(1)} \left[ \frac{\partial V(q, B^{2(1)})}{\partial q_k} + \frac{\partial V(q, B^{1(2)})}{\partial B} \left( \omega^1_k - \omega^2_k \right) \right] +
- \lambda^{1(2)} \left[ \frac{\partial V(q, B^{2(2)})}{\partial q_k} + \frac{\partial V(q, B^{1(2)})}{\partial B} \left( \omega^1_k - \omega^2_k \right) \right] = 0, \quad k = 2, \ldots, n,
\]

(59)

where $c_k^i$ and $\omega_k^i$ are respectively the $k$-th component of the vector $\varphi^i$ and of the vector $\omega^i$.

Define the compensated derivative of the Lagrangian with respect to $q_k$ as

\[
\frac{\partial \tilde{\Lambda}}{\partial q_k} = \frac{\partial \Lambda}{\partial q_k} + \sum_{i=1}^{2} c_k^i \frac{\partial \Lambda}{\partial B^i}.
\]

(60)

Substituting (57), (58) and (59) into (60), using Roy’s identity and Slutsky decomposition, and rearranging terms yields

\[
\frac{\partial \tilde{\Lambda}}{\partial q_k} = \frac{\partial \tilde{\Lambda}}{\partial B^1} \left( \sum_{i=1}^{2} \pi^i \frac{\partial \tilde{\varphi}^i}{\partial q_k} \right) + \lambda^{2(1)} \frac{\partial V(q, B^{2(1)})}{\partial B} \left[ z_k^{2(1)} - z_k^{1(1)} \right] + \lambda^{1(2)} \frac{\partial V(q, B^{1(2)})}{\partial B} \left[ z_k^{1(2)} - z_k^2 \right],
\]

where $z_k^{(j)}$ denotes the $k$-th component of the net consumption vector $(z_k^{(j)} = c_k^{(j)} - \omega_k^j)$ of individual $i$ mimicking individual $j$ and $z_k^i$ the $k$-th component of the net consumption vector of individual $i$.

At the optimum, $\forall k = 2, \ldots, n$ it must be
Denote by $S$ the aggregate reduced Slutsky matrix with general element
\[
\sigma_{jk} = \sum_{i=1}^{2} \pi_i \frac{\partial c_i}{\partial q_{jk}}, \quad j, k = 2, ..., n.
\]
Rearranging condition (61) allows to express the vector of optimal commodity tax rates as
\[
t = -S^{-1} \left[ \frac{1}{\gamma} \left( \lambda^{2(1)} \frac{\partial V(q, B^{2(1)})}{\partial B} [z^{2(1)}_k - z^1_k] + \lambda^{1(2)} \frac{\partial V(q, B^{1(2)})}{\partial B} [z^{1(2)}_k - z^2_k] \right) \right].
\]

Condition (62) implies that separability between leisure and other goods is no more a sufficient condition to ensure that a second best Pareto optimum could be attained relying only on the general income tax schedule. For the redundancy of commodity taxation a more restrictive condition is required, namely that for all binding self-selection constraints the net consumption bundle of the mimicker equals the net consumption bundle of the mimicked; basically, this only happens when mimicker and mimicked have the same initial endowments for all goods.

4 Concluding Remarks

The role of differential indirect taxation in redistributive models of optimal mixed taxation has been revised starting from the influential result by Atkinson and Stiglitz (1976) of labour income as a “sufficient statistics” when it can be taxed in a nonlinear way and commodities are separable from labour in the common utility function of all taxpayers. It has been emphasised that indirect taxation is useful as long as it helps weakening the binding self-selection constraints that prevent the government from taking redistribution as far as is deemed desirable. In a “standard world” where market ability is the only source of difference among agents, once it has been assumed separability of leisure from all other goods in the individual’s utility function, it has also been prevented any possibility to screen a mimicker from the agent who is being mimicked since, when consumption of various goods is not related to the amount of leisure enjoyed, both agents will spend their (identical amount of) disposable income over commodities in the same way. Displaying no screening power, indirect taxation or, more precisely, non-uniform indirect taxation cannot improve upon an optimally set nonlinear labour income tax schedule. Therefore, the first and more obvious way to restore a role for indirect taxation keeping in place the separability assumption is to allow people to have heterogeneous tastes. By doing
so, a mimicker and a mimicked will again spend differently their common disposable income over commodities and screening becomes once more feasible exploiting differences in tastes. Leaving aside this trivial “solution”, and also the one represented by referring to the concept of “merit/demerit” goods, in this paper we have provided a survey of the main arguments that have been recently considered in the optimal taxation literature to attack the Atkinson-Stiglitz result and strengthen the role of commodity taxation.

Our aim has not been to give a quantitative estimation of the potential role of indirect taxation in pursuing efficiently the redistributive goals of the government. Less ambitiously, we tried to illustrate the possible channels through which the use of indirect taxation, despite the skepticism that often surrounds its use, can improve upon an optimal designed nonlinear labour income tax schedule. As long as we move away from the simplifying assumptions of the Atkinson-Stiglitz’s theorem the chances that indirect taxation can play an important role are more and more increased and it can be recognised that non-uniform commodity taxes are a characteristic of almost all Pareto optima. This has been confirmed by all the contributions we have presented, the main results of which can be synthesized as follows:

- when the possibility of income tax evasion is taken into account there is no longer equivalence between after tax (reported) earnings and consumption; this means that a mimicker is less constrained in his (her) choice of labour supply and can be successful in misrepresenting his (her) type by selecting different combinations of labour supply/amount of income hidden to the fiscal authorities. Since a mimicker and a mimicked will in general have different amounts of disposable income, separability of goods from leisure is not sufficient to guarantee the optimality of fully uniform indirect taxes;

- when agents face uncertainty about the wage they will earn and some consumption decisions take place before uncertainty is solved and agents know their own type, commodity taxes, besides being useful as an insurance mechanism, can still be used as a basis of separation despite the separability of preferences between leisure and other goods;

- when different agents are not perfect substitutes for one another in the production process, the result stating the redundancy of commodity taxation (under assumption of separability of goods from leisure) is overcome since taxing different commodities at different rates can enlarge the feasible extent of redistribution through its effect on the wage ratio;

- when household production is considered and the possibility arises to have an alternative use (other than pure leisure) of non-market time, individuals’ (conditional) preference orderings over marketed
commodities will in general depend on the available non-market time even under the weak separability assumption of the Atkinson-Stiglitz’s theorem. Recognizing that non-market time does not necessarily coincides with leisure restores the chance to discern among agents with different market skills and in this way allows to get the optimality of a non-uniform commodity tax structure.

- multidimensional heterogeneity of agents provides a strong case for differential commodity taxation; when agents differ in the initial endowment of various goods, weak separability between leisure and other goods is not sufficient to guarantee that a mimicker and a mimicked have identical net demands for marketed commodities. Except when all pairs of types who are linked by a binding self-selection constraint have identical endowments, commodity taxation shows screening power and can be seen as an indirect way to tax the unobservable endowments.
References


33


