



Working Papers

www.cesifo.org/wp

Optimal Commodity Taxation with Varying Quality of Goods

Spencer Bastani
Sören Blomquist
Luca Micheletto

CESIFO WORKING PAPER NO. 5635

CATEGORY 1: PUBLIC FINANCE

DECEMBER 2015

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

ISSN 2364-1428

Optimal Commodity Taxation with Varying Quality of Goods

Abstract

A standard result in the optimal taxation literature is that, when agents differ in market ability and the government aims at redistributing from high- to low-skilled agents by means of an optimal nonlinear labor income tax and a set of commodity taxes, an optimally designed commodity tax structure should encourage (discourage) the consumption of goods/services that are complement with labor (leisure). In this paper we highlight that, when agents can choose both the quality and the quantity of a given good/service, this standard commodity tax result needs to be qualified. First, we show that it becomes relevant to distinguish between specific and ad valorem taxes/subsidies. Second, whether the standard result holds or not depends on how the concept of labor (leisure) complement is defined, namely, whether it is defined in terms of number of units or in terms of expenditure. We also show that levying specific and ad valorem taxes at opposite signs on a given good can be a feature of the second-best optimum.

JEL-Codes: H210, H420.

Keywords: nonlinear income taxation, commodity taxation, redistribution, varying quality.

*Spencer Bastani**
Department of Economics
Uppsala University / Sweden
spencer.bastani@nek.uu.se

*Sören Blomquist***
Department of Economics
Uppsala University / Sweden
soren.blomquist@nek.uu.se

*Luca Micheletto****
Department of Law
University of Milan / Italy
luca.micheletto@unimi.it

November 23, 2015

*Financial Support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

**Financial support from Riksbankens Jubileumsfond is gratefully acknowledged.

***Financial support from Riksbankens Jubileumsfond is gratefully acknowledged.

1 Introduction

A standard result in the optimal taxation literature is that, when agents differ in market ability and the government aims at redistributing from high- to low-skilled agents by means of a nonlinear income tax and a set of commodity taxes (either *specific* or *ad-valorem*), an optimally designed commodity tax structure should encourage (discourage) the consumption of goods/services that are complement with labor (leisure).¹

In a model with just two private consumption goods plus leisure, where one of the consumption goods is chosen as the untaxed *numéraire*, this requirement implies that the non-*numéraire* good should be subsidized (taxed at a positive rate) if it is a complement with labor (leisure), meaning that, for any given amount of disposable income, the demand for the non-*numéraire* good is an increasing function (a decreasing function) of labor supply.² The rationale for this policy prescription is that, by doing so, one can relax the binding self-selection constraint requiring high-skilled agents not to “mimic” low-skilled agents which descends from the fact that the government has only a statistical information about the distribution of agents’ types in the population and does not know “who is who”.

The intuition for the result is the following. Given that a high-skilled agent is paid a higher wage rate than a low-skilled agent, if a high-skilled agent behaves as a mimicker he/she will work fewer hours than a low-skilled agent; thus, he/she will consume more (less) of goods/services that are complements with leisure (labor). Starting from a pure income tax optimum where commodity taxes are not used, it is then possible to introduce a small tax (subsidy) on a good that is a leisure (labor) complement, while adjusting the income tax schedule in such a way that the overall tax reform is both budget-neutral for the government and welfare-neutral for all non-mimicking agents, in

¹The result is derived in a setting where all private goods consumed by the agents are purchased in the market. For a model allowing for some goods to be produced within the household, see Cremer and Gahvari (2015).

²See, for instance, Mirrlees (1976), Christiansen (1984), and Edwards et al. (1994).

such a way to make a high-skilled mimicker worse-off and thereby relaxing the binding self-selection constraint.^{3,4} This, in turn, will open the way for a further change in the income tax schedule that allows enhancing the redistribution in favor of low-skilled agents, and hence increasing social welfare.

The commodity tax result that we have described above has been obtained in the context of models where no attention has been devoted to the possibility that a given good/service is available in the market at different quality levels. In this paper, instead, we consider a model where, at least for some goods/services, agents have both a quantity and a quality choice. We show that in such a context the aforementioned commodity tax result needs to be qualified.

First of all, it becomes important to distinguish between *specific* and *ad valorem* taxes/subsidies, a distinction that is of no relevance in the standard model that has been analyzed in the previous literature.

Second, whether the standard result holds or not depends on how the concept of labor complement is defined, namely, whether it is defined in terms of the difference between the number of units purchased by two agents with identical disposable income but different labor supply, or in terms of the difference between their total expenditure on a given good.

Finally, a last result that we show is that jointly levying *specific* and *ad valorem* taxes at opposite signs on a given good can be a feature of a second-best optimum.

To appreciate the importance of distinguishing between *specific* and *ad valorem*

³As recently stated by Stiglitz (2015, p. 42), “...in the presence of an (optimal) income tax, ... commodity taxation can be viewed as a particular type of Pigouvian corrective tax. The focus is not on the impact on tax revenues, or even directly on dead weight losses (as usually conceived), but on impacts on the self-selection constraints that are central to the design of the optimal income tax. ‘Loosening’ the self-selection constraints has a first order effect on welfare, while the distortions associated with small commodity taxation have a second order effect on welfare.”

⁴This type of reasoning underpins the Atkinson-Stiglitz (1976) theorem on the redundancy of commodity taxes in the presence of an optimal nonlinear income tax. The theorem shows that, if individual preferences are weakly separable between leisure and other goods, an optimal nonlinear income tax is sufficient to implement any incentive-compatible Pareto-efficient allocation. On the Atkinson and Stiglitz (1976) theorem see, more recently, Boadway and Song (2015).

taxes/subsidies, and to properly define the concept of labor complement, consider the case of child-care services, one of the prominent examples in the literature of a good that ought to be subsidized (or taxed at a lower rate than other goods) due to its complementarity with labor supply.

The case for subsidizing child care expenditures has been derived in models where the hourly price of child-care services does not differ across child-care facilities and where, therefore, the only margin of choice for households pertains to the number of hours that their child is at a child-care center. Under these assumptions, it is quite reasonable to expect that, since a high-skilled mimicker works fewer hours than a true low-skilled parent, more hours of child-care services will be needed by the latter, and in this sense child-care services can unambiguously be regarded as a labor complement.

However, things differ if child-care services are available in the market at different quality levels, and households can choose both the quality of the facility and the number of hours that their child is at the child-care center. In this case, one cannot in general rule out the possibility that a mimicker would choose a higher quality of child-care services than a low-skilled. If that happens, it might be that a mimicker would spend more on child-care services than a low-skilled, while at the same time using child-care services for fewer hours. Child-care services would then still be a labor complement when the concept is defined in terms of number of hours that the service is used, but it would no longer characterize as a labor complement if the concept were to be defined in terms of total expenditure on the service. Also, one would still like to subsidize (or tax at a relatively low rate) the purchase of child-care services if the government uses *specific* taxes/subsidies; however, one would like to tax (or subsidize at a relatively low rate) the purchase of child-care services if the government resorts to *ad valorem* taxes/subsidies.

The remainder of the paper is organized as follows. In Section 2 we describe the structure of the model. Section 3 characterizes the optimal commodity tax structure for the general case when the government can jointly use *specific* and *ad valorem*

taxes/subsidies on the purchase of a good that is available in the market at different quality levels. Section 4 characterizes the optimal marginal income tax rates and the optimal marginal effective tax rates generated by the combined effect of income and commodity taxation. Section 5 provides a numerical example illustrating the possibility that, at a second-best optimum, the government may want to jointly impose on a given good *specific* and *ad valorem* taxes at opposite signs. Finally, section 5 offers concluding remarks.

2 The model

Consider an economy consisting of two types of agents differing only in terms of market ability (output produced per unit of time spent working). High-skilled workers are paid w^h and low-skilled workers w^ℓ , with $w^h > w^\ell$. Agents derive utility from the consumption of two private goods/services (denoted by x and y) and disutility from labor supply (denoted by L). To illustrate in the simplest way the fact that standard commodity tax rules may be suboptimal when a given good/service is available in the market at different quality levels, assume that good x is produced at a uniform quality level so that only for good y agents choose both the preferred quantity and the preferred quality. Choosing good x as the *numéraire* of our economy, assume also that the price of a unit of good y relates to its quality, denoted by θ , via the increasing function $p(\theta)$. Let agents' preferences be represented by the function $U = u(x, \theta, y, L)$ and let total population be normalized to one with π^j ($j = \ell, h$) representing the proportion of agents of type j in the population.

As usual in the optimal tax literature, we assume that the government knows the distribution of types in the population but can observe neither L nor w , while it can observe their product I , earned income. Thus, the government is prevented from using type-specific first-best lump-sum taxes/transfers. Instead, the government has at its disposal a general income tax $T(I)$. 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 (I^j, B^j) , where $B^j \stackrel{def}{=} I^j - T(I^j)$, $j = \ell, h$. We also assume, again in accordance with the bulk of the optimal tax literature, that the government can only observe anonymous transactions, i.e. it cannot identify the type of the consumer who makes a purchase; therefore nonlinear commodity taxes are not feasible, while linear commodity taxes/subsidies, either *specific* or *ad valorem*, are available. Finally, assume that, albeit commodity tax rates can be differentiated across goods, the tax rate on a given good/service does not depend on its quality type.⁵ With good x as the untaxed *numéraire* of our economy, this last assumption implies that the design of the indirect tax structure reduces to the selection of the appropriate commodity tax/subsidy on good y . While in a standard optimal tax model with nonlinear income taxation and linear commodity taxation it is of no importance for the qualitative results whether one assumes *specific* or *ad valorem* commodity taxes, this choice becomes relevant, as we will see, in our setting. For this reason, hereafter we adopt a general framework where good y can in principle be taxed/subsidized both in a *specific* and an *ad valorem* way, and we will denote the *specific* tax by t and the *ad valorem* tax by τ .

The government's problem is to design the tax policy that maximizes a weighted sum of individuals' utilities, given the assumed informational constraints and an exogenous revenue requirement. Due to the non-linearity of the income tax schedule, the government must design the tax system so that each ability type (weakly) prefers the (I, B) -bundle intended for it to that intended for some other type (self-selection constraints). An agent that misrepresents his type is called a mimicker.

We denote by $V(t, \tau, B, I; w^j)$ the conditional indirect utility for a type j agent

⁵This assumption is motivated by the circumstance that, whereas it is common to observe governments levying *specific* taxes on the quantity purchased of various goods as well as *ad valorem* taxes on the money spent on different goods, it is less common to observe commodity taxes set at rates that are, for a given good, differentiated according to the quality type of the good. Most importantly, even when this sometimes happen (think at instances where higher taxes are levied on "luxury" cars than on "regular" cars), the rate differentiation is usually quite rough and not fine-tuned with respect to the various quality types available in the market.

(with $j = \ell, h$) obtained, for a given value of I , by optimally allocating a fixed amount of expenditure B over the consumption goods:

$$V(t, \tau, B, I; w^j) = \max_{x, y, \theta} \left\{ u \left(x, \theta, y, \frac{I}{w^j} \right) \mid x + [(1 + \tau)p(\theta) + t]y = B \right\}.$$

The solution to this problem defines the conditional demand functions

$$x^j = x(t, \tau, B, I; w^j), \quad (1)$$

$$y^j = y(t, \tau, B, I; w^j), \quad (2)$$

$$\theta^j = \theta(t, \tau, B, I; w^j). \quad (3)$$

Labor supply is then determined by maximizing $V(t, B, I; w^j)$ subject to the link between pre-tax earnings and post-tax earnings available for goods expenditure implied by the direct tax schedule: $B = I - T(I)$. As has become common practice in the optimal tax literature, this allows to implicitly define the marginal income tax rate at the income level I for an agent of type j as:

$$T'(I) = 1 + \frac{\partial V(t, \tau, B, I; w^j) / \partial I}{\partial V(t, \tau, B, I; w^j) / \partial B} = 1 - MRS_{IB}^j, \quad (4)$$

where MRS_{IB}^j denotes the marginal rate of substitution between B and I for an agent of type j .

Denote by V^j the indirect utility of a j -type agent (with $j = \ell, h$) and by V^{jk} the indirect utility of a j -type agent mimicking a k -type agent,

$$\begin{aligned} V^j &= V(t, \tau, B^j, I^j; w^j), \\ V^{jk} &= V(t, \tau, B^k, I^k; w^j). \end{aligned}$$

Restricting attention to the so-called “normal” case where redistribution goes from the high- to the low-skilled agents, so that the only relevant self-selection constraint is

the one requiring high-skilled agents not to mimic low-skilled agents, a Pareto-efficient tax structure can be described as the solution to the following problem:

$$\max_{t, I^\ell, I^h, B^\ell, B^h} \sum_{j=\ell, h} \delta^j V^j$$

subject to:

$$V^h \geq V^{h\ell}, \quad (\lambda)$$

$$\sum_{j=\ell, h} \pi^j \{ I^j - B^j + [\tau p(\theta^j) + t] y^j \} \geq \bar{R}, \quad (\gamma)$$

where δ^j are the positive welfare weights used by the government ($\sum_{j=\ell, h} \delta^j = 1$), Lagrange multipliers are within parentheses, and \bar{R} denotes an exogenous revenue requirement.

The first order conditions for this problem are given in Appendix A.

3 Optimal commodity taxes

Denoting a compensated variable by a “tilde” symbol, the following Proposition characterizes the optimal structure of commodity taxation.

Proposition 1 *Under an optimal nonlinear income tax, when both specific and ad valorem commodity taxes are used, the optimal commodity tax structure satisfies the fol-*

lowing set of conditions:

$$t = \frac{\frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell}}{\Psi} \left\{ \begin{array}{l} (y^\ell - y^{h\ell}) \left[\sum_{j=\ell,h} \pi^j \left(\underbrace{p(\theta^j) \frac{\partial \tilde{y}^j}{\partial \tau} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial \tau} y^j}_{\Delta_\tau} \right) \right] \\ - [y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})] \left[\sum_{j=\ell,h} \pi^j \left(\underbrace{p(\theta^j) \frac{\partial \tilde{y}^j}{\partial t} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial t} y^j}_{\Delta_t} \right) \right] \end{array} \right\}, \quad (5)$$

$$\tau = \frac{\frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell}}{\Psi} \left[[y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})] \left(\sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial t} \right) - (y^\ell - y^{h\ell}) \left(\sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial \tau} \right) \right], \quad (6)$$

where Ψ is a positive term defined in the appendix.

When the non-numéraire good is only taxed via an ad valorem tax, the optimal value for τ is given by:

$$\tau = \frac{\frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell} [y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})]}{\sum_{j=\ell,h} \pi^j \left[p(\theta^j) \frac{\partial \tilde{y}^j}{\partial \tau} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial \tau} y^j \right]}. \quad (7)$$

When the non-numéraire good is only taxed via a specific tax, the optimal value for t is given by:

$$t = \frac{\frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell} (y^\ell - y^{h\ell})}{\sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial t}}. \quad (8)$$

Proof. See Appendix B. ■

Let's first consider the case when the non-numéraire good is subject to both *specific* and *ad valorem* taxation. Given that, depending on the signs of the various terms in (5)-(6), several possibilities may arise, for illustrative purposes we will here focus on two cases where good y is assumed to be a labor complement in the sense that $y^\ell > y^{h\ell}$.

With term labelled Δ_τ being negative,⁶ the assumption that $y^\ell > y^{h\ell}$ implies that the first term within curly brackets in (5) calls for setting $t < 0$. Suppose then that the term labelled Δ_t in (5) is also negative;⁷ this implies that the sign of the second term within curly brackets in (5) is negative when a high-skilled mimicker spends more on good y than a true low-skilled agent. In this case, with $y^\ell p(\theta^\ell) < y^{h\ell} p(\theta^{h\ell})$ and $y^\ell > y^{h\ell}$, the optimal t is negative, i.e. a specific subsidy. Let's now look at (6). With both $y^\ell p(\theta^\ell) < y^{h\ell} p(\theta^{h\ell})$ and $y^\ell > y^{h\ell}$, the sign of both terms within curly brackets in (6) is positive,⁸ thus implying a positive optimal value for τ . Summarizing, in such a case we would have that the government should optimally set a *specific* subsidy on the consumption of the labor-complement good y , while at the same time taxing its purchase via an *ad valorem* tax.

Alternatively, suppose again that $y^\ell > y^{h\ell}$ whereas this time $y^\ell p(\theta^\ell) > y^{h\ell} p(\theta^{h\ell})$. Then, the optimal signs of t and τ become ambiguous; however, if in both (5) and (6) the dominant term happens to be the one depending on the difference $y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})$, one would get the result that the government should optimally set a *specific* tax on the consumption of the labor-complement good y , while at the same time subsidizing its purchase via an *ad valorem* subsidy.

Consider now the cases when the non-*numéraire* good is subject to only one type of indirect taxation, either *specific* or *ad valorem*. If the consumption of good y is only taxed via an *ad valorem* tax (so that t is restricted to zero), the relevant optimality condition becomes eq. (7). With the denominator of (7) being negative, the sign of the optimal τ is positive (resp.: negative) when a high-skilled mimicker spends more (resp.: less) on good y than a true low-skilled agent. But with agents having the possibility to choose the preferred quality of good y , its unitary price will in general vary across agents. Therefore, when the purchase of the good is taxed only via an *ad valorem*

⁶Notice that $\frac{\partial \bar{y}^j}{\partial \tau} < 0$ and $\frac{\partial \bar{\theta}^j}{\partial \tau} < 0$. With $\frac{\partial p(\theta^j)}{\partial \theta^j}$ it follows that Δ_τ is necessarily negative.

⁷Notice that, while $\frac{\partial \bar{y}^j}{\partial t} < 0$, the sign of $\frac{\partial \bar{\theta}^j}{\partial t}$ is in principle ambiguous.

⁸Notice that both $\frac{\partial \bar{y}^j}{\partial t}$ and $\frac{\partial \bar{\theta}^j}{\partial \tau}$ are negative.

tax, assuming that good y is a labor complement, in the sense that $y^\ell > y^{h\ell}$, is not a sufficient condition for the desirability of subsidizing the purchase of good y . Instead, what is needed in order to obtain $\tau < 0$ is to redefine the concept of labor complement in terms of a comparison between the expenditure on a given good by two agents with identical disposable income but different labor supply, so that y is said to be a labor complement when $y^\ell p(\theta^\ell) > y^{h\ell} p(\theta^{h\ell})$.

On the other hand, if the consumption of good y is only taxed via a *specific* tax (so that τ is restricted to zero), the relevant optimality condition becomes eq. (8) and in this case assuming that good y is a labor complement, in the sense that $y^\ell > y^{h\ell}$, is a necessary and sufficient condition for the desirability of subsidizing the consumption of good y .

To get some insights on the conditions under which a high-skilled mimicker would spend more on good y than a true low-skilled agent, consider the individual optimization problem for a given value of disposable income B . Using subscripts to denote the variable(s) with respect to which derivatives are taken, so that for instance $u_x \equiv \partial u(x, \theta, y, L) / \partial x$ and $u_{xy} \equiv \partial^2 u(x, \theta, y, L) / \partial x \partial y$, the first order conditions of the individual maximization problem are:⁹

$$\begin{aligned} u_x &= \alpha, \\ u_\theta &= (1 + \tau) \alpha y p_\theta, \\ u_y &= [(1 + \tau) p(\theta) + t] \alpha, \\ x + [(1 + \tau) p(\theta) + t] y &= B, \end{aligned}$$

with α denoting the private marginal utility of income.

Totally differentiating the above system with respect to the wage rate w , we get, in matrix form:

⁹Notice that we also write p_θ and $p_{\theta\theta}$ for, respectively, $\partial p(\theta) / \partial \theta$ and $\partial^2 p(\theta) / \partial \theta^2$.

$$\begin{aligned}
& \begin{bmatrix} u_{xx} & u_{x\theta} & u_{xy} & -1 \\ u_{\theta x} & u_{\theta\theta} - (1 + \tau) \alpha y p_{\theta\theta} & u_{\theta y} - (1 + \tau) \alpha p_{\theta} & -(1 + \tau) y p_{\theta} \\ u_{yx} & u_{y\theta} - (1 + \tau) \alpha p_{\theta} & u_{yy} & -[(1 + \tau) p(\theta) + t] \\ 1 & (1 + \tau) y p_{\theta} & [(1 + \tau) p(\theta) + t] & 0 \end{bmatrix} \begin{bmatrix} dx/dw \\ d\theta/dw \\ dy/dw \\ d\alpha/dw \end{bmatrix} \\
= & \begin{bmatrix} u_{xL} I/w^2 \\ u_{\theta L} I/w^2 \\ u_{yL} I/w^2 \\ 0 \end{bmatrix}
\end{aligned}$$

To get a simpler expression for $d(y p(\theta))/dw$, assume that $t = \tau = 0$ and that the individuals' utility function is quasi-linear in the *numéraire* good x , so that $\alpha = 1$. We would then get the following system in matrix form:

$$\begin{bmatrix} 0 & u_{\theta\theta} - y p_{\theta\theta} & u_{\theta y} - p_{\theta} \\ 0 & u_{y\theta} - p_{\theta} & u_{yy} \\ 1 & y p_{\theta} & p(\theta) \end{bmatrix} \begin{bmatrix} dx/dw \\ d\theta/dw \\ dy/dw \end{bmatrix} = \begin{bmatrix} u_{\theta L} I/w^2 \\ u_{yL} I/w^2 \\ 0 \end{bmatrix}$$

Solving the system for dy/dw and $d\theta/dw$ gives:

$$dy/dw = \frac{(u_{\theta\theta} - y p_{\theta\theta}) u_{yL} - (u_{y\theta} - p_{\theta}) u_{\theta L}}{u_{\theta\theta} u_{yy} - y p_{\theta\theta} u_{yy} - [u_{y\theta} - p_{\theta}]^2} I/w^2, \quad (9)$$

$$d\theta/dw = \frac{[u_{yy} u_{\theta L} - (u_{y\theta} - p_{\theta}) u_{yL}]}{u_{\theta\theta} u_{yy} - y p_{\theta\theta} u_{yy} - [u_{y\theta} - p_{\theta}]^2} I/w^2, \quad (10)$$

where the denominator of the expressions on the right hand side of (9)-(10) takes a positive sign from the second-order conditions of the individual maximization problem.

To further simplify, consider the case when $u_{\theta L} = 0$, and assume that $u_{yL} > 0$ so that from (9) we can infer that $dy/dw < 0$.

Given that $d(y p(\theta))/dw$ is defined as

$$d(y p(\theta))/dw = (dy/dw) p(\theta) + (d\theta/dw) y p_{\theta}, \quad (11)$$

one can see that a necessary condition for having $d(y p(\theta)) / dw > 0$ is that $d\theta / dw > 0$, which in turn requires (see eq. (10)) $u_{\theta y} < p_{\theta}$.

Notice also that, substituting in eq. (11) the values for dy/dw and $d\theta/dw$ provided by (9)-(10), one gets (when $u_{\theta L} = 0$):

$$d(y p(\theta)) / dw = \frac{(u_{\theta\theta} - y p_{\theta\theta}) p(\theta) - (u_{\theta y} - p_{\theta}) y p_{\theta}}{u_{\theta\theta} u_{yy} - y p_{\theta\theta} u_{yy} - [u_{y\theta} - p_{\theta}]^2} u_{yL} I / w^2. \quad (12)$$

From (12) one can see that, when good y is a labor complement in terms of units purchased (in the sense that $dy/dw < 0$, which requires $u_{yL} > 0$), it will at the same time be a leisure complement in terms of budget shares (in the sense that $d(y p(\theta)) / dw > 0$) when the following condition is satisfied:

$$(p_{\theta} - u_{\theta y}) y p_{\theta} > (y p_{\theta\theta} - u_{\theta\theta}) p(\theta).$$

In turn, the condition above is more likely to be satisfied when $u_{\theta y} < 0$, so that quality and quantity of good y are substitutes, and when the curvature of price-of-quality function is low, so that $p_{\theta\theta}$ is small.

4 Optimal marginal income and effective tax rates

The total amount of taxes paid at income I by an agent of type j (with $j = \ell, h$) is given by $\Upsilon(I) \equiv T(I) + [\tau p(\theta^j) + t] y^j$, with θ^j and y^j being defined by eqs. (2)-(3). We can then define the marginal *effective* tax rate faced by an agent of type j (with $j = \ell, h$) as:

$$\begin{aligned} \Upsilon'(I) = & T'(I) + [\tau p(\theta^j) + t] \left\{ \frac{\partial y^j}{\partial I} + \frac{\partial y^j}{\partial B} [1 - T'(I)] \right\} \\ & + \tau y^j \frac{\partial p(\theta^j)}{\partial \theta^j} \left\{ \frac{\partial \theta^j}{\partial I} + \frac{\partial \theta^j}{\partial B} [1 - T'(I)] \right\}. \end{aligned} \quad (13)$$

The quantity defined in eq. (13) captures the overall distortion on an individual's labor supply generated by the combined effect of income and commodity taxation.

The formula given in eq. (13) applies more generally if the implicit marginal income tax rate $1 - MRS_{IB}^j$ (see eq. (4)) is substituted for $T'(I)$. Thus, we can equivalently rewrite (13) as:

$$\begin{aligned} \Upsilon'(I) = & 1 - MRS_{IB}^j + [\tau p(\theta^j) + t] \left(\frac{\partial y^j}{\partial I} + \frac{\partial y^j}{\partial B} MRS_{IB}^j \right) \\ & + \tau y^j \frac{\partial p(\theta^j)}{\partial \theta^j} \left(\frac{\partial \theta^j}{\partial I} + \frac{\partial \theta^j}{\partial B} MRS_{IB}^j \right), \end{aligned} \quad (14)$$

The following Proposition characterizes the optimal marginal income and effective tax rates for our two-type model.

Proposition 2 Define $(dy^j/dI)_{dV^j=0}$ and $(d\theta^j/dI)_{dV^j=0}$ as:

$$\begin{aligned} (dy^j/dI)_{dV^j=0} & \equiv \partial y^j / \partial I + (\partial y^j / \partial B) MRS_{IB}^j, \\ (d\theta^j/dI)_{dV^j=0} & \equiv \partial \theta^j / \partial I + (\partial \theta^j / \partial B) MRS_{IB}^j. \end{aligned}$$

At a second-best optimum the marginal income tax rates faced by the high-skilled and the low-skilled agents are given, respectively, by:

$$T'(I^h) = - [\tau p(\theta^h) + t] \left(\frac{dy^h}{dI^h} \right)_{dV^h=0} - \tau y^h \frac{\partial p(\theta^h)}{\partial \theta^h} \left(\frac{d\theta^h}{dI^h} \right)_{dV^h=0}, \quad (15)$$

$$\begin{aligned} T'(I^\ell) = & \frac{\lambda \frac{\partial V^{h\ell}}{\partial B^\ell}}{\mu \pi^\ell} (MRS_{IB}^\ell - MRS_{IB}^{h\ell}) \\ & - [\tau p(\theta^\ell) + t] \left(\frac{dy^\ell}{dI^\ell} \right)_{dV^\ell=0} - \tau y^\ell \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \left(\frac{d\theta^\ell}{dI^\ell} \right)_{dV^\ell=0}, \end{aligned} \quad (16)$$

where the difference $MRS_{IB}^\ell - MRS_{IB}^{h\ell}$ is evaluated at the (I^ℓ, B^ℓ) -bundle intended by the government for the low-skilled agents.

The optimal marginal effective tax rates are instead given by:

$$\Upsilon' (I^h) = 0, \quad (17)$$

$$\Upsilon' (I^\ell) = \frac{\lambda \frac{\partial V^{h\ell}}{\partial B^\ell}}{\mu \pi^\ell} (MRS_{IB}^\ell - MRS_{IB}^{h\ell}). \quad (18)$$

Proof. See Appendix C. ■

Eqs. (17)-(18) reproduce results that are fairly standard in the optimal tax literature.¹⁰ In particular, they show that a distortion must be imposed on the labor-leisure choice of agents of type j whenever the government needs to discourage other agents from choosing the (I, B) -bundle intended for agents of type j only.

With redistribution going from high- to low-skilled agents, the latter are not tempted to choose the bundle intended for the former and, according to eq. (17), the combined effect of income and commodity taxes should be such that no distortion is imposed at the margin on the labor-leisure choice of high-skilled agents. Notice also that, as shown by eq. (15), the optimal marginal income tax rate faced by high-skilled agents will in general need to be non-zero in order to achieve the no-distortion-at-the-top result when commodity taxes are used.

For low-skilled agents, instead, the combined effect of income and commodity taxes will result in a distortion of the labor-leisure choice, at least when the redistributive goals of the government are strong enough to generate a positive value for the Lagrange multiplier λ . Moreover, assuming that the level curves of $V(t, \tau, B, I; w^j)$ in the (I, B) -space are flatter, other things being equal, the higher an individual's wage rate,¹¹ we can see that, according to (18), the optimal marginal effective tax rate faced by low-skilled agents should be positive.¹² This means that, at a second-best optimum, the

¹⁰See, for instance, Edwards et al. (1994).

¹¹Formally, this monotonicity condition means that, for all t, τ, B and I , $\frac{\partial}{\partial w} \left(-\frac{\partial V / \partial I}{\partial V / \partial B} \right) < 0$.

¹²The difference $MRS_{IB}^\ell - MRS_{IB}^{h\ell}$ appearing in eq. (18) is the difference between the marginal rate of substitution for a low-skilled agent and a high-skilled mimicker at a same (I, B) -bundle, the one intended for low-skilled agents only. Thus, under the assumption that the level curves of $V(t, \tau, B, I; w^j)$ in the

combined effect of income and commodity taxes will distort downwards the labor supply of low-skilled agents.

5 A numerical example

To illustrate the possibility that it might be optimal for the government to levy *specific* and *ad valorem* taxes on good y at opposite signs, consider the following example where individual preferences are of the form $U = u(F(y, \theta, L), x, L)$, implying that the marginal rate of substitution between y and θ does not depend on the amount consumed of the *numéraire* good x . In particular, assume that skilled and unskilled workers have identical preferences represented by:

$$U = \left[(4y^{-1} + \theta^{-1})^{-1} - 2\frac{L}{y} - x^{-2} - \frac{1}{25}\frac{L}{x} - \frac{L^2}{1000} \right] k, \quad (19)$$

where k is a constant set equal to 5.

Further, regarding the price function $p(\theta)$, assume that

$$p(\theta) = \theta/5.$$

The government has a max-min social welfare function, and there is an equal proportion of low- and high-skilled workers so that $\pi^h = \pi^\ell = 1/2$. Their wage rates, reflecting their productivities, are set equal to $w^\ell = 5$ and $w^h = 9$. As far as the government's external revenue is concerned, we set $\bar{R} = 0$ so that optimal taxes are purely redistributive.

(I, B) -space are flatter, other things being equal, the higher an individual's wage rate, we can conclude that $MRS_{IB}^\ell - MRS_{IB}^{h\ell} > 0$.

Under these assumptions the solution to the government's problem is given by:

$$\begin{aligned}\tau &= 13.25\%, & t &= -0.112, \\ T'(I^\ell) &= 38.16\%, & T'(I^h) &= -11.68\%, \\ \Upsilon'(I^\ell) &= 43.62\%, & \Upsilon'(I^h) &= 0.\end{aligned}$$

The optimal tax policy results in the following values for the arguments of the utility function:¹³

$$\begin{aligned}x^\ell &= 6.317, & y^\ell &= 60.836, & \theta^\ell &= 12.856, & L^\ell &= 13.350, \\ x^h &= 13.359, & y^h &= 100.844, & \theta^h &= 20.007, & L^h &= 56.985.\end{aligned}$$

Calculating the optimal choices for a high-skilled mimicker we get:

$$x^{h\ell} = 5.609, \quad y^{h\ell} = 59.218, \quad \theta^{h\ell} = 13.247, \quad L^{h\ell} = 7.416,$$

and we can then see that good y is a labor complement in the sense that $y^\ell > y^{h\ell}$. However, with $p(\theta^\ell) = \theta^\ell/5 = 2.571$ and $p(\theta^{h\ell}) = \theta^{h\ell}/5 = 2.649$, we also have that $p(\theta^\ell)y^\ell < p(\theta^{h\ell})y^{h\ell}$, meaning that, in terms of budget shares, good y is a leisure complement.¹⁴

At the second-best optimum, the utility values are given by $u^h = u^{h\ell} = 33.005$ and $u^\ell = 31.202$ resulting in a value for the social welfare equal to 31.202. By way of comparison, imposing $\tau = 0$, while choosing $t = -0.009$ optimally, lowers the value of social welfare to 31.1983,¹⁵ whereas imposing $t = 0$, while choosing $\tau = 11.78\%$

¹³We also have:

$$x^{h\ell} = 5.609, \quad y^{h\ell} = 59.218, \quad \theta^{h\ell} = 13.247, \quad L^{h\ell} = 7.416.$$

¹⁴The same qualitative result, i.e. $y^\ell > y^{h\ell}$ and $p(\theta^\ell)y^\ell < p(\theta^{h\ell})y^{h\ell}$, is obtained at the solution to the government's problem when the commodity taxes are restricted to zero.

¹⁵The corresponding optimal income tax rates are $T'(I^\ell) = 45.08\%$ and $T'(I^h) = 1.09\%$; the marginal effective tax rates are $\Upsilon'(I^\ell) = 43.72\%$ and $\Upsilon'(I^h) = 0$.

optimally, lowers the value of social welfare to 31.1958.^{16,17}

6 Concluding remarks

A standard result in the optimal taxation literature is that, when agents differ in market ability and the government aims at redistributing from high- to low-skilled agents by means of a nonlinear income tax and a set of commodity taxes (either *specific* or *ad-valorem*), an optimally designed commodity tax structure should encourage (discourage) the consumption of goods/services that are complements with labor (leisure).

In this paper we have highlighted that, when agents can choose both the quality and the quantity of a given good/service, the aforementioned commodity tax result needs to be qualified.

First, it becomes relevant to distinguish between *specific* and *ad-valorem* taxes/subsidies. Second, whether the standard result holds or not depends on how the concept of labor complement is defined, namely, whether it is defined in terms of the difference between the number of units purchased by two agents with identical disposable income but different labor supply, or in terms of the difference between their total expenditure on a given good.

Another result that we have highlighted is that jointly levying on a given good *specific* and *ad valorem* taxes at opposite signs can be a feature of the second-best optimum.

Two final remarks are in order.

First, our results are not in conflict with the celebrated Atkinson-Stiglitz (1976) theorem on the redundancy of commodity taxes in the presence of an optimal nonlinear income tax. Specifically, even in our setting with (some) goods being available in

¹⁶The corresponding optimal income tax rates are $T'(I^\ell) = 37.49\%$ and $T'(I^h) = -11.54\%$; the marginal effective tax rates are $\Upsilon'(I^\ell) = 43.78\%$ and $\Upsilon'(I^h) = 0$.

¹⁷Relying only on the income tax and setting $t = \tau = 0$ results in a value for social welfare equal to 31.1932. The corresponding optimal income tax rates are $T'(I^\ell) = 43.86\%$ and $T'(I^h) = 0$.

the market at different quality levels, commodity taxes are a redundant policy instrument when preferences are separable between leisure and other goods. The intuition is straightforward: under separability a high-skilled mimicker and a true low-skilled agent would spend their disposable income in exactly the same way, both in terms of units purchased of the various goods/services and in terms of quality chosen for each good/service. Thus, neither *specific* nor *ad-valorem* taxes can be used as a screening device to relax the binding self-selection constraint.

Second, in our setting it is still true that commodity taxes are used for efficiency reasons to stimulate labor supply and thereby offset the distortions of the income tax on labor supply. Also, the intuition that goods should be taxed/subsidized if they are more/less complementary with leisure than the untaxed *numéraire* good is still valid; however, it requires to define the concept of labor-complement in two different ways depending on whether *specific* or *ad valorem* taxes/subsidies are used.

Appendix A

The first order conditions to the government's problem

Introduce V^j as a shorthand for $V(t, \tau, B^j, I^j; w^j)$ and V^{jk} as a shorthand for $V(t, \tau, B^k, I^k; w^j)$. The first order conditions of the government's problem with respect to $I^h, I^\ell, B^h, B^\ell, t$ and τ are, respectively:

$$(\delta^h + \lambda) \frac{\partial V^h}{\partial I^h} + \mu \pi^h \left\{ [(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial I^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial I^h} y^h + 1 \right\} = 0, \quad (\text{A1})$$

$$\delta^\ell \frac{\partial V^\ell}{\partial I^\ell} - \lambda \frac{\partial V^{h\ell}}{\partial I^\ell} + \mu \pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] \frac{\partial y^\ell}{\partial I^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial I^\ell} y^\ell + 1 \right\} = 0, \quad (\text{A2})$$

$$(\delta^h + \lambda) \frac{\partial V^h}{\partial B^h} + \mu \pi^h \left\{ [(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial B^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial B^h} y^h - 1 \right\} = 0, \quad (\text{A3})$$

$$\delta^\ell \frac{\partial V^\ell}{\partial B^\ell} - \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} + \mu \pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] \frac{\partial y^\ell}{\partial B^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial B^\ell} y^\ell - 1 \right\} = 0, \quad (\text{A4})$$

$$\begin{aligned} & (\delta^h + \lambda) \frac{\partial V^h}{\partial t} + \delta^\ell \frac{\partial V^\ell}{\partial t} - \lambda \frac{\partial V^{h\ell}}{\partial t} \\ & + \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial t} + \left[\tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial t} + 1 \right] y^j \right\} \\ = & 0, \end{aligned} \quad (\text{A5})$$

$$\begin{aligned} & (\delta^h + \lambda) \frac{\partial V^h}{\partial \tau} + \delta^\ell \frac{\partial V^\ell}{\partial \tau} - \lambda \frac{\partial V^{h\ell}}{\partial \tau} \\ & + \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial \tau} + \left[\tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial \tau} + p(\theta^j) \right] y^j \right\} \\ = & 0. \end{aligned} \quad (\text{A6})$$

Appendix B

Proof of Proposition 1

With $\partial V^j / \partial t = -(\partial V^j / \partial B^j) y^j$, $\partial V^{jk} / \partial t = -(\partial V^{jk} / \partial B^k) y^{jk}$, $\partial V^j / \partial \tau = -(\partial V^j / \partial B^j) y^j p(\theta^j)$ and $\partial V^{jk} / \partial \tau = -(\partial V^{jk} / \partial B^k) y^{jk} p(\theta^{jk})$, we can rewrite the first order conditions for t and τ as follows:

$$\begin{aligned}
 & -(\delta^h + \lambda) y^h \frac{\partial V^h}{\partial B^h} - \delta^\ell y^\ell \frac{\partial V^\ell}{\partial B^\ell} + \lambda y^{h\ell} \frac{\partial V^{h\ell}}{\partial B^\ell} \\
 & + \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial t} + \left[\tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial t} + 1 \right] y^j \right\} \\
 = & 0
 \end{aligned} \tag{B1}$$

$$\begin{aligned}
 & -(\delta^h + \lambda) y^h p(\theta^h) \frac{\partial V^h}{\partial B^h} - \delta^\ell y^\ell p(\theta^\ell) \frac{\partial V^\ell}{\partial B^\ell} + \lambda y^{h\ell} p(\theta^{h\ell}) \frac{\partial V^{h\ell}}{\partial B^\ell} \\
 & + \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial \tau} + \left[\tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial \tau} + p(\theta^j) \right] y^j \right\} \\
 = & 0
 \end{aligned} \tag{B2}$$

Multiply (A3) by y^h and (A4) by y^ℓ and then add the resulting equations to (B1). we get:

$$\begin{aligned}
 & (\delta^h + \lambda) y^h \frac{\partial V^h}{\partial B^h} + \mu \pi^h \left\{ [(\tau p(\theta^h) + t)] y^h \frac{\partial y^h}{\partial B^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial B^h} (y^h)^2 - y^h \right\} + \\
 & \delta^\ell y^\ell \frac{\partial V^\ell}{\partial B^\ell} - \lambda y^\ell \frac{\partial V^{h\ell}}{\partial B^\ell} + \mu \pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] y^\ell \frac{\partial y^\ell}{\partial B^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial B^\ell} (y^\ell)^2 - y^\ell \right\} \\
 & - (\delta^h + \lambda) y^h \frac{\partial V^h}{\partial B^h} - \delta^\ell y^\ell \frac{\partial V^\ell}{\partial B^\ell} + \lambda y^{h\ell} \frac{\partial V^{h\ell}}{\partial B^\ell} \\
 & + \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial t} + \left[\tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial t} + 1 \right] y^j \right\} \\
 = & 0.
 \end{aligned}$$

Simplifying terms gives:

$$\begin{aligned}
& \mu \sum_{j=\ell,h} \pi^j \left\{ [(\tau p(\theta^j) + t)] y^j \frac{\partial y^j}{\partial B^j} + \tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial B^j} (y^j)^2 \right\} \\
& + \mu \sum_{j=\ell,h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial t} + \left[\tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial t} \right] y^j \right\} \\
& = \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} (y^\ell - y^{h\ell})
\end{aligned}$$

Use a Slutsky-type decomposition to write $\partial y^j / \partial t$ and $\partial \theta^j / \partial t$ as $\partial y^j / \partial t = \partial \tilde{y}^j / \partial t - (\partial y^j / \partial B^j) y^j$ and $\partial \theta^j / \partial t = \partial \tilde{\theta}^j / \partial t - (\partial \theta^j / \partial B^j) y^j$, where a “tilde” symbol is used to denote a compensated variable. We can then simplify the equation above and end up with:

$$\begin{aligned}
& \mu \sum_{j=\ell,h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial \tilde{y}^j}{\partial t} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial t} \tau y^j \right\} \\
& = \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} (y^\ell - y^{h\ell}). \tag{B3}
\end{aligned}$$

Multiply (A3) by $p(\theta^h) y^h$ and (A4) by $p(\theta^\ell) y^\ell$ and then add the resulting equations to (B2). we get:

$$\begin{aligned}
& (\delta^h + \lambda) p(\theta^h) y^h \frac{\partial V^h}{\partial B^h} + \delta^\ell \frac{\partial V^\ell}{\partial B^\ell} p(\theta^\ell) y^\ell - \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} p(\theta^\ell) y^\ell \\
& + \mu \pi^h \left\{ [(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial B^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial B^h} y^h - 1 \right\} p(\theta^h) y^h \\
& + \mu \pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] \frac{\partial y^\ell}{\partial B^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial B^\ell} y^\ell - 1 \right\} p(\theta^\ell) y^\ell \\
& + (\delta^h + \lambda) y^h p(\theta^h) \frac{\partial V^h}{\partial B^h} + \delta^\ell y^\ell p(\theta^\ell) \frac{\partial V^\ell}{\partial B^\ell} - \lambda y^{h\ell} p(\theta^{h\ell}) \frac{\partial V^{h\ell}}{\partial B^\ell} \\
& + \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial \tau} + \left[\tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial \tau} + p(\theta^j) \right] y^j \right\} \\
& = 0.
\end{aligned}$$

Simplifying terms gives:

$$\begin{aligned}
& \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial B^j} + \tau \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial B^j} y^j \right\} p(\theta^j) y^j \\
& + \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial y^j}{\partial \tau} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \theta^j}{\partial \tau} \tau y^j \right\} \\
& = \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} [y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})].
\end{aligned}$$

Use a Slutsky-type decomposition to write $\partial y^j / \partial \tau$ and $\partial \theta^j / \partial \tau$ as $\partial y^j / \partial \tau = \partial \tilde{y}^j / \partial \tau - (\partial y^j / \partial B^j) y^j p(\theta^j)$ and $\partial \theta^j / \partial \tau = \partial \tilde{\theta}^j / \partial \tau - (\partial \theta^j / \partial B^j) y^j p(\theta^j)$, where a “tilde” symbol is used to denote a compensated variable. We can then simplify the equation above and end up with:

$$\begin{aligned}
& \mu \sum_{j=\ell, h} \pi^j \left\{ [(\tau p(\theta^j) + t)] \frac{\partial \tilde{y}^j}{\partial \tau} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial \tau} \tau y^j \right\} \\
& = \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} [y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})]. \tag{B4}
\end{aligned}$$

We then have the following system of equations in matrix form:

$$\begin{bmatrix} \sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial t} & \sum_{j=\ell,h} \pi^j \left(p(\theta^j) \frac{\partial \tilde{y}^j}{\partial t} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial t} y^j \right) \\ \sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial \tau} & \sum_{j=\ell,h} \pi^j \left(p(\theta^j) \frac{\partial \tilde{y}^j}{\partial \tau} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial \tau} y^j \right) \end{bmatrix} \begin{bmatrix} t \\ \tau \end{bmatrix} = \begin{bmatrix} \frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell} (y^\ell - y^{h\ell}) \\ \frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell} [y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})] \end{bmatrix}$$

Solving the system for t and τ gives:

$$t = \frac{\frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell}}{\Psi} \left\{ \begin{array}{l} (y^\ell - y^{h\ell}) \left[\sum_{j=\ell,h} \pi^j \left(p(\theta^j) \frac{\partial \tilde{y}^j}{\partial \tau} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial \tau} y^j \right) \right] \\ - [y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})] \left[\sum_{j=\ell,h} \pi^j \left(p(\theta^j) \frac{\partial \tilde{y}^j}{\partial t} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial t} y^j \right) \right] \end{array} \right\},$$

$$\tau = \frac{\frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell}}{\Psi} \left[\left(\sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial t} \right) [y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})] - \left(\sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial \tau} \right) (y^\ell - y^{h\ell}) \right],$$

where we have defined Ψ as:

$$\begin{aligned} \Psi &\equiv \left(\sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial t} \right) \left[\sum_{j=\ell,h} \pi^j \left(p(\theta^j) \frac{\partial \tilde{y}^j}{\partial \tau} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial \tau} y^j \right) \right] \\ &\quad - \left(\sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial \tau} \right) \left[\sum_{j=\ell,h} \pi^j \left(p(\theta^j) \frac{\partial \tilde{y}^j}{\partial t} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial t} y^j \right) \right] \\ &> 0. \end{aligned}$$

If the government were to use only the *specific* tax/subsidy t as a commodity tax instrument, substituting $\tau = 0$ in (B3) would deliver the following expression for the optimal t :

$$t = \frac{\frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell} (y^\ell - y^{h\ell})}{\sum_{j=\ell,h} \pi^j \frac{\partial \tilde{y}^j}{\partial t}}.$$

If instead only the *ad valorem* tax/subsidy τ were used, substituting $t = 0$ in (B4) would deliver the following expression for the optimal τ :

$$\tau = \frac{\frac{\lambda}{\mu} \frac{\partial V^{h\ell}}{\partial B^\ell} [y^\ell p(\theta^\ell) - y^{h\ell} p(\theta^{h\ell})]}{\sum_{j=\ell, h} \pi^j \left[p(\theta^j) \frac{\partial \tilde{y}^j}{\partial \tau} + \frac{\partial p(\theta^j)}{\partial \theta^j} \frac{\partial \tilde{\theta}^j}{\partial \tau} y^j \right]}.$$

Appendix C

Proof of Proposition 2:

Rewrite eq. (A1) and eq. (A3) as:

$$(\delta^h + \lambda) \frac{\partial V^h}{\partial I^h} = -\mu\pi^h \left\{ [(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial I^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial I^h} y^h + 1 \right\}, \quad (C1)$$

$$(\delta^h + \lambda) \frac{\partial V^h}{\partial B^h} = -\mu\pi^h \left\{ [(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial B^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial B^h} y^h - 1 \right\}, \quad (C2)$$

and then divide eq. (C1) by eq. (C2) to obtain:

$$\frac{\partial V^h}{\partial I^h} / \frac{\partial V^h}{\partial B^h} = \frac{[(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial I^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial I^h} y^h + 1}{[(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial B^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial B^h} y^h - 1}. \quad (C3)$$

Multiplying both sides of (C3) by $[(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial B^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial B^h} y^h - 1$ gives:

$$\begin{aligned} & \frac{\frac{\partial V^h}{\partial I^h}}{\frac{\partial V^h}{\partial B^h}} \left\{ [(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial B^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial B^h} y^h - 1 \right\} \\ &= [(\tau p(\theta^h) + t)] \frac{\partial y^h}{\partial I^h} + \tau \frac{\partial p(\theta^h)}{\partial \theta^h} \frac{\partial \theta^h}{\partial I^h} y^h + 1. \end{aligned} \quad (C4)$$

Taking into account that $MRS_{IB}^h = -\frac{\partial V^h}{\partial I^h} / \frac{\partial V^h}{\partial B^h}$, the result stated in eq. (17) is then easily obtained by using the definition of marginal effective tax rate provided by (14) to collect terms in eq. (C4).

Rewrite eq. (A2) and eq. (A4) as:

$$\delta^\ell \frac{\partial V^\ell}{\partial I^\ell} = \lambda \frac{\partial V^{h\ell}}{\partial I^\ell} - \mu\pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] \frac{\partial y^\ell}{\partial I^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial I^\ell} y^\ell + 1 \right\}, \quad (\text{C5})$$

$$\delta^\ell \frac{\partial V^\ell}{\partial B^\ell} = \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} - \mu\pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] \frac{\partial y^\ell}{\partial B^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial B^\ell} y^\ell - 1 \right\}, \quad (\text{C6})$$

and then divide eq. (C5) by eq. (C6) to obtain:

$$\frac{\partial V^\ell / \partial I^\ell}{\partial V^\ell / \partial B^\ell} = \frac{\lambda \frac{\partial V^{h\ell}}{\partial I^\ell} - \mu\pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] \frac{\partial y^\ell}{\partial I^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial I^\ell} y^\ell + 1 \right\}}{\lambda \frac{\partial V^{h\ell}}{\partial B^\ell} - \mu\pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] \frac{\partial y^\ell}{\partial B^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial B^\ell} y^\ell - 1 \right\}}. \quad (\text{C7})$$

Multiplying both sides of (C7) by $\lambda \frac{\partial V^{h\ell}}{\partial B^\ell} - \mu\pi^\ell \left\{ [(\tau p(\theta^\ell) + t)] \frac{\partial y^\ell}{\partial B^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial B^\ell} y^\ell - 1 \right\}$ gives:

$$\begin{aligned} & \frac{\frac{\partial V^\ell}{\partial I^\ell}}{\frac{\partial V^\ell}{\partial B^\ell}} \left\{ \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} - \mu\pi^\ell \left[((\tau p(\theta^\ell) + t)) \frac{\partial y^\ell}{\partial B^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial B^\ell} y^\ell - 1 \right] \right\} \\ &= \lambda \frac{\partial V^{h\ell}}{\partial I^\ell} - \mu\pi^\ell \left[((\tau p(\theta^\ell) + t)) \frac{\partial y^\ell}{\partial I^\ell} + \tau \frac{\partial p(\theta^\ell)}{\partial \theta^\ell} \frac{\partial \theta^\ell}{\partial I^\ell} y^\ell + 1 \right]. \end{aligned} \quad (\text{C8})$$

Taking into account that $MRS_{IB}^\ell = -\frac{\partial V^\ell}{\partial I^\ell} / \frac{\partial V^\ell}{\partial B^\ell}$, the result stated in eq. (18) is then easily obtained by using the definition of marginal effective tax rate provided by (14) to collect terms in eq. (C8).

To derive eqs. (15)-(16), notice that, by using the definitions of $(dy^j/dI)_{dV^j=0}$ and $(d\theta^j/dI)_{dV^j=0}$, one can rewrite (14) as:

$$\Upsilon'(I) = 1 - MRS_{IB}^j + [\tau p(\theta^j) + t] \left(\frac{dy^j}{dI} \right)_{dV^j=0} + \tau y^j \frac{\partial p(\theta^j)}{\partial \theta^j} \left(\frac{d\theta^j}{dI} \right)_{dV^j=0}. \quad (\text{C9})$$

Eqs. (15)-(16) are then easily obtained using the implicit definition of marginal income tax rate provided by (4), and combining eq. (C9) first with eq. (17) and then with eq. (18).

References

- [1] Atkinson, A. B., and J. E. Stiglitz, 1976. The design of tax structure: direct versus indirect taxation. *Journal of Public Economics* 6, 55–75.
- [2] Boadway, R., and Z. Song, 2015. Indirect taxes for redistribution. Should necessity goods be favored? *Research in Economics*, this issue.
- [3] Christiansen, V., 1984. Which commodity taxes should supplement the income tax? *Journal of Public Economics* 24, 195–220.
- [4] Cremer, H., and F. Gahvari, 2015. Atkinson and Stiglitz theorem in the presence of a household production sector. *Economics Letters* 126, 91-95.
- [5] Edwards, J., Keen, M., and M. Tuomala, 1994. Income tax, commodity taxes and public good provision: a brief guide. *Finanzarchiv* 51, 472-487.
- [6] Mirrlees, J. A., 1976. Optimal tax theory: a synthesis. *Journal of Public Economics* 6, 327-358.
- [7] Stiglitz, J. E., 2015. In praise of Frank Ramsey’s contribution to the theory of taxation. *Economic Journal* 125, 235–268.