

# An alternative way to model merit good arguments\*

Fred Schroyen<sup>†</sup>

*First version:* September 2002

*This version:* October 2003

**Abstract** Besley (1988) uses a scaling approach to model merit good arguments in commodity tax policy. In this paper, I question this approach on the grounds that it produces 'wrong' recommendations—taxation (subsidisation) of merit (demerit) goods—whenever the demand for the (de)merit good is inelastic. I propose an alternative approach that does not suffer from this deficiency, and derive the ensuing first and second best tax rules, as well as the marginal cost expressions to perform tax reform analysis.

**JEL code:** H21

**Keywords:** merits goods, commodity taxation, tax reform analysis

---

\*This version of the paper was written while visiting the Dept. of Economics, Universitat Autònoma de Barcelona, which provided a very friendly and hospitable work environment. Financial support by the Programme for Health Economics in Bergen (HEB) through an SNF grant is gratefully acknowledged. I should like to thank Kjetil Bjorvatn, Bart Capéau, Agnar Sandmo, Bertil Tungodden and two anonymous referees for helpful comments and discussions. Remaining flaws are entirely mine.

<sup>†</sup>Dept. of Economics, Norwegian School of Economics & Business Administration, Helleveien 30, N-5045 Bergen (Norway). E-mail: fred.schroyen@nhh.no.

# 1 Introduction

Empirical tax reform analysis for European countries almost without exception reports on the very high welfare costs involved in marginally raising total tax revenue through an increase in the VAT or excise rate on alcoholic beverages and tobacco.<sup>1</sup> The coexistence of such a high marginal welfare cost with much lower values for other commodities and services seemingly points at the possibility of significant welfare increases by reducing of taxes on alcohol and tobacco, and raising those on, say, transport. Still, authors of such studies hasten to remark that their analysis does not take into account the demerit good arguments that probably motivated these high excise tax levels in the first place.

Although the (de)merit good argument dates back to Musgrave (1959), its introduction in the optimal taxation literature did not come before Besley's (1988) analysis.<sup>2</sup> The reason for this delay is that standard second-best analysis of economic policy takes place in a welfaristic framework, while (de)merit good arguments drive a wedge between the evaluations of citizens and those of policy makers.

The strength of Besley's (1988) model is twofold. First, it incorporates merit good arguments in an otherwise very standard welfaristic setting, and therefore allows to demonstrate clearly how putting one foot into non-welfaristic terrain affects the familiar first- and second-best policy rules. Second, Besley relates the (de)merit good considerations to where we would expect them to have their root: the consumption pattern of the individual agent, rather than the aggregate level of consumption of particular commodities.<sup>3</sup>

Unfortunately, Besley's first-best analysis suffers from a deficiency that puts the framework into question. As I will show below, his model in fact prescribes to subsidise demerit goods and to tax merit goods whenever the demand for those goods is inelastic. Keeping in mind the fact that habit formation is often responsible for an inelastic demand for cigarettes and alcoholic beverages, one is then led to the paradox that such demand should be encouraged rather than discouraged.

Although a normative model cannot be subjected to the same kind of falsification tests as positive models, if its policy prescriptions go exactly against one's gut-feeling, then one should ask whether the framework within which those prescriptions were derived is an appropriate one. I think one can question Besley's

---

<sup>1</sup>Decoster & Schokkaert (1989) for Belgium, Madden (1995) for Ireland, Kaplanoglou & Newbery (2002) for Greece, Schroyen & Aasness (2002) for Norway.

<sup>2</sup>Several years earlier, Sandmo (1983) outlined for different degrees of market completeness the policy implications of a divergence between the agent's beliefs about the future states of the world, and those of the social planner. Besley's (1988) model, and the present one, concern the case of *(de)merit wants*: the divergence between the agent's preferences and those of the social planner.

<sup>3</sup>See e.g. Pazner (1972).

framework, but at the same time that it can be easily remedied in a way that preserves its twofold strength.

To show my arguments in the sharpest way, I start by explaining them in a first best setting. This, I do in the next two sections. In section 4, I offer an alternative way of modeling (de)merit good arguments and derive the ensuing first best policy rules. Second-best rules are presented in section 5, and in section 6, I derive the marginal cost expressions for tax reform analysis. Section 7 concludes.

## 2 Consumer behaviour and government opinion

Consider a representative consumer whose preferences over the *numéraire* commodity ( $z \in R_+$ ), a (de)merit good ( $y \in R_+$ ), and a standard commodity ( $x \in R_+$ ) can be represented by the monotonic and strongly quasi-concave utility function  $u(z, x, y)$ . Let the consumer prices for these commodities be given by  $(1, q_x, q_y)$  and the consumer's exogenous disposable income equal to  $m - T$ , where  $T$  is the lump sum tax and  $m$  is pre-tax income.

This consumer then solves the problem

$$\begin{aligned} & \max_{z,x,y} u(z, x, y) & (1) \\ \text{s.t. } & z + q_x x + q_y y = m - T. \end{aligned}$$

The first order conditions are<sup>4</sup>

$$\frac{u_2}{u_1} = q_x, \frac{u_3}{u_1} = q_y, \quad (2)$$

and, together with the budget constraint, these are satisfied by the optimal commodity demands

$$z(q_x, q_y, m - T), x(q_x, q_y, m - T), \text{ and } y(q_x, q_y, m - T). \quad (3)$$

The government evaluates the allocation of resources according to the modified utility function

$$u^g(z, x, y). \quad (4)$$

It takes individual behaviour (3) as given and is concerned with solving:

$$\begin{aligned} & \max_{t_x, t_y, T} u^g(z(q_x, q_y, m - T), x(q_x, q_y, m - T), y(q_x, q_y, m - T)) & (5) \\ \text{s.t. } & t_x x(q_x, q_y, m - T) + t_y y(q_x, q_y, m - T) + T \geq \bar{R} \quad (\lambda) \end{aligned}$$

where  $q_x = p_x + t_x$  and  $q_y = p_y + t_y$ .

---

<sup>4</sup>Subscript  $i$  with a function denotes a partial derivatives w.r.t. the  $i$ th argument.

The first order conditions are shown in the appendix to provide us with the following tax rules:

$$t_x = \frac{u_2}{u_1}(z, x, y) - \frac{u_2^g}{u_1^g}(z, x, y) \quad (6a)$$

$$t_y = \frac{u_3}{u_1}(z, x, y) - \frac{u_3^g}{u_1^g}(z, x, y) \quad (6b)$$

who have the intuitive interpretation of driving a wedge between the consumer's marginal willingness to pay (MWP) for each good, and that of the government.

### 3 Besley's scaling approach

Besley (1988) proposes the following specification for the government's evaluation function:

$$u^g(z, x, y) = u(z, x, \theta y). \quad (7)$$

and defines the third commodity as a merit (demerit) good whenever  $\theta > (<)$ 1. This scaling approach dates back to Fisher & Shell (1967) who used it to construct an index for the true cost of living when people's tastes change or when products change in quality. In the present context, the government converts the quantity of the (de)merit good into efficiency units, but otherwise fully respects individual preferences.

With this specification, the earlier derived tax rules become<sup>5</sup>

$$t_x = \frac{u_2}{u_1}(z, x, y) - \frac{u_2}{u_1}(z, x, \theta y), \quad (8a)$$

$$t_y = \frac{u_3}{u_1}(z, x, y) - \theta \frac{u_3}{u_1}(z, x, \theta y). \quad (8b)$$

To see what these rules imply, consider the preference ordering represented by the CES utility function  $u(z, x, y) = (\alpha_1 z^\rho + \alpha_2 x^\rho + \alpha_3 y^\rho)^{1/\rho}$ , with  $\alpha_i > 0$  ( $i = 1, 2, 3$ ), and  $-\infty < \rho < 1$ . With such preferences,  $\frac{u_3}{u_1} = \frac{\alpha_3}{\alpha_1} \left(\frac{y}{z}\right)^{\rho-1}$ ,  $\frac{u_3^g}{u_1^g} = \frac{\alpha_3}{\alpha_1} \theta \left(\frac{\theta y}{z}\right)^{\rho-1}$  and  $\frac{u_2}{u_1} = \frac{u_2^g}{u_1^g} = \frac{\alpha_2}{\alpha_1} \left(\frac{x}{z}\right)^{\rho-1}$ . No tax should be levied on the standard commodity, while for the (de)merit good, we have

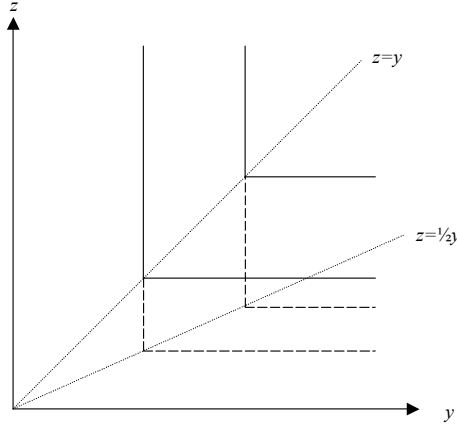
$$\frac{t_y}{q_y} = \frac{(1 - \theta^\rho)}{\theta}. \quad (9)$$

Whenever the elasticity of substitution is below one ( $-\infty < \rho < 0$ ), we get the paradoxical result that a merit good should be taxed while a demerit good should be subsidised!

---

<sup>5</sup>In deriving the first best rule for the tax on the (de)merit good (8b), Besley (1988) made a mistake. The correct rule was provided in a comment by Feehan (1990).

What is going on? Why does a seemingly natural way of modelling (de)merit good arguments lead to results whose intuitive appeal depends on the elasticity of substitution? Let us consider the limiting case of Leontief preferences, and ignore—for the sake of graphical representation—the standard non-*numéraire* good. Suppose that a person’s preferences can be represented by  $\min\{z, y\}$ . This person has L-shaped indifference curves with the corners lying on the  $45^\circ$  line, as shown in figure 1.



Leontief preferences and a demerit good

The government, on the other hand, thinks of good  $y$  as a demerit good and subscribes to the preference ordering represented by  $\min\{z, \frac{1}{2}y\}$ . The associated indifference curves are the dashed lines. Clearly, the government’s preferences are *more favourable* to commodity  $y$  than the agent’s preferences are!

With Leontief preferences, no finite subsidy will be able to distort the agent’s budget allocation away from the *laissez-faire* solution, but once the degree of substitutability becomes positive, it will. The reason why discounting a commodity leads to subsidisation should now be clear. The government respects that red wine is complementary to a pasta meal. But it regards one bottle of wine only half as good as you do. In order to get the maximal utility out of a pasta dish, it wants you to drink *more* wine, not less.

This paradox holds true more generally. Notice that the government’s MWP for the (de)merit good ( $\theta \frac{u_z}{u_1}(z, x, \theta y)$ ) has the following elasticity w.r.t.  $\theta$ :

$$\frac{\partial \log\left(\frac{u_z}{u_1}\right)}{\partial \log \theta} = 1 - \left( -\frac{\partial \log \frac{u_y}{u_z}(z, x, \theta y)}{\partial \log(\theta y)} \right). \quad (10)$$

Thus, whether the government’s MWP exceeds or falls short of that of the consumer depends on whether the (own) demand price elasticity for the (de)merit good exceeds or falls short of 1 in absolute value. If the demand for tobacco, say, is inelastic, the demand price elasticity is likely to be large, and under the scaling

approach the government's MWP for tobacco will exceed that of the smoking agent.<sup>6</sup>

## 4 An alternative way of modeling merit good arguments

The previous analysis indicates the need for an approach that ties down in a more robust way the relationship between the government's MWP and that of the consumer. I propose to model this relationship by means of the total willingness-to-pay (TWP) function in terms of the *numéraire* commodity:

$$F(x, y, u). \quad (11)$$

This function gives the amount of the *numéraire* required to bring the consumer at the utility level  $u$  when consuming  $x$  and  $y$  units of the other two commodities; its graph is the indifference surface. The marginal willingness to pay (MWP) for the (de)merit good is then

$$-\frac{\partial F(x, y, u)}{y} = -F_2(x, y, u). \quad (12)$$

If the government is of the opinion that consumers appreciate the (de)merit good too (much) little, then a natural way of proceeding is to attribute to the government the MWP function

$$-F_2^g(x, y, u) = -F_2(x, y, u) + \mu(y), \quad (13)$$

where  $\mu(y)$  is (negative) positive for a (de)merit good.<sup>7</sup> Notice that (13) does not directly impose any single crossing in the commodity space because it is conditional on the utility level  $u$ : the amount of  $z$  the consumer believes is necessary to reach that utility level need not be the same as the amount the government believes is required. Indeed, integrating this MWP function to a TWP function gives

$$F^g(x, y, u) = F(x, y, u) - \int_{\underline{y}^g}^y \mu(\chi) d\chi, \quad (14)$$

where  $\underline{y}^g$  can be thought of as the level of consumption above which the government's marginal evaluation starts to deviate from the consumer's.

---

<sup>6</sup>With additive preferences quasi-linear in the *numéraire*, the (own) elasticity of the inverse demand schedule is exactly the reciprocal of the (own) Marshallian price elasticity.

<sup>7</sup>More complicated modifications of the MWP function are possible, but one should make sure that for a corresponding TWP function to exist, the cross partial derivatives should be symmetric:  $F_{21}^g = F_{12}^g$  (Frobenius theorem).

The interpretation of (14) is as follows. Suppose the agent has  $(x, y)$  available. Then in order to reach a utility level  $\bar{u}$ , she believes  $F(x, y, \bar{u})$  units of the numéraire are required, while the government, convinced of the merit properties of the third commodity, believes  $F^g(x, y, \bar{u})$  units are sufficient. What (13) then tells is that the slope of the government's indifference curve through  $(F^g(x, y, \bar{u}), x, y)$  differs from the slope of the consumer's indifference curve through  $(F(x, y, \bar{u}), x, y)$  with a factor  $\mu(y)$ .

To this TWP function corresponds a utility function  $u^g(\cdot)$ , defined as

$$u^g(F^g(x, y, \nu), x, y, \nu) \equiv \nu \quad (\text{all } \nu). \quad (15)$$

It is easy to show that

$$u^g(z, x, y) = u\left(z + \int_{\underline{y}^g}^y \mu(\chi) d\chi, x, y\right). \quad (16)$$

(see appendix). A sufficient condition for  $u^g(\cdot)$  to be strongly quasi-concave is that  $\mu'(\cdot) \leq 0$ .

The two MRS expressions necessary to compute the optimal tax rates are therefore

$$\frac{u_2^g}{u_1^g}(z, x, y) = \frac{u_2}{u_1}\left(z + \int_{\underline{y}^g}^y \mu(\chi) d\chi, x, y\right) \quad (17a)$$

$$\frac{u_3^g}{u_1^g}(z, x, y) = \frac{u_3}{u_1}\left(z + \int_{\underline{y}^g}^y \mu(\chi) d\chi, x, y\right) + \mu(y). \quad (17b)$$

Again, the reason (17b) is not exactly  $\mu(y)$  is that the evaluation here is at the same bundle, not at the same utility level as in (13). A sufficient condition for single crossing of indifference curves (in the sense that  $\text{sign}(MRS_{yz}^g - MRS_{yz}) = \text{sign}(\mu(y))$ ) is that the (de)merit good is a normal good (in the sense that  $\frac{\partial}{\partial z}(\frac{u_3}{u_1}) > 0$ ).

Inserting (17) in (6) and using a Taylor expansion of  $\frac{u_i}{u_1}(z + \int_{\underline{y}}^y \mu(\chi) d\chi, x, y)$  around  $\frac{u_i}{u_1}(z, x, y)$  gives

$$t_x \simeq - \left( \frac{\partial \frac{u_2}{u_1}(z, x, y)}{\partial z} \right) \int_{\underline{y}^g}^y \mu(\chi) d\chi \quad (18a)$$

$$t_y \simeq - \left( \frac{\partial \frac{u_3}{u_1}(z, x, y)}{\partial z} \right) \int_{\underline{y}^g}^y \mu(\chi) d\chi - \mu(y) \quad (18b)$$

The round bracket terms on the *rhs* denotes the uncompensated effect of a marginal increase in  $z$  on the demand price for each of the non-numéraire commodities. The signs of these effects are related to the normality of these

commodities (although with 3 or more commodities, the income effects are more involved).

To fix ideas, suppose that the government considers  $y$  as a merit good,  $\mu(\chi) > 0$ ,  $\forall \chi \geq 0$ , and that both  $x$  and  $y$  are normal goods. Then both non-numéraire commodities should be subsidised. First, the merit good should be subsidised with a rate at which the government's MWP exceeds that of the consumer ( $\mu$ ). But in addition, *both* non-numéraire goods should be subsidised for the difference in willingness to pay for all inframarginal units. The intuition is as follows. The government is convinced that the consumer is better off than she herself believes she is (because she ignores the merit properties of the third good). To the extent that the non-numéraire commodities are normal goods, it considers both commodities as underconsumed (relative to the numéraire) and wants to stimulate their consumption by subsidisation.

## 5 Second-best rules

Suppose now that there are many consumers, indexed by  $h$ , with not necessarily the same preferences or income level, and that the government cannot differentiate commodity taxes across consumers. Let social preferences be given by

$$W = \sum_h \lambda^h u^{gh}(z^h, x^h, y^h) = \sum_h \lambda^h u^h \left( z^h + \int_{\underline{y}^g}^{y^h} \mu(\chi) d\chi, x^h, y^h \right). \quad (19)$$

To focus on the efficiency arguments, I assume that differentiated lump sum taxes are available. The government's problem is then<sup>8</sup>

$$\begin{aligned} \max_{t_x, t_y, \{T^h\}} \sum_h \lambda^h u^{gh}(z^h(q_x, q_y, m^h - T^h), x^h(q_x, q_y, m^h - T^h), y^h(q_x, q_y, m^h - T^h)) \\ \text{s.t.} \quad \sum_h t_x x^h(q_x, q_y, m^h - T^h) + \sum_h t_y y^h(q_x, q_y, m^h - T^h) + \sum_h T^h \geq \bar{R} \quad (\gamma) \end{aligned} \quad (20)$$

Defining  $\beta^h \stackrel{\text{def}}{=} \frac{\lambda^h u_1^{gh}}{\gamma}$  and proceeding in the same way as in the first best case, we get:

$$\sum_h \beta^h \begin{pmatrix} \frac{\partial \hat{x}^h}{\partial q_x} & \frac{\partial \hat{x}^h}{\partial q_y} \\ \frac{\partial \hat{y}^h}{\partial q_x} & \frac{\partial \hat{y}^h}{\partial q_y} \end{pmatrix} \begin{pmatrix} t_x^h \\ t_y^h \end{pmatrix} = \begin{pmatrix} \sum_h \frac{\partial \hat{x}^h}{\partial q_x} & \sum_h \frac{\partial \hat{x}^h}{\partial q_y} \\ \sum_h \frac{\partial \hat{y}^h}{\partial q_x} & \sum_h \frac{\partial \hat{y}^h}{\partial q_y} \end{pmatrix} \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (21)$$

where a  $\hat{\cdot}$  denotes a compensated price effect and  $t_x^h \stackrel{\text{def}}{=} \frac{u_2^h}{u_1^h} - \frac{u_2^{gh}}{u_1^{gh}}$  and  $t_y^h \stackrel{\text{def}}{=} \frac{u_3^h}{u_1^h} - \frac{u_3^{gh}}{u_1^{gh}}$

---

<sup>8</sup>I agree with Capéau and Ooghe (2003) that the reaction functions entering the social welfare function should be those of the real agents, and not, as Besley (1988) proceeded, those of phantom agents that are endowed with the 'right' preferences.



are the tax rates that person  $h$  would face under first best.<sup>9</sup> By premultiplying (21) through by the inverse of the aggregate Slutsky matrix, we could isolate the vector of second best tax rates  $(t_y^x)$ , but it is only under Hicksian independence between the two non-*numéraire* commodities ( $\frac{\partial \hat{x}^h}{\partial q_y} = 0$ ) that this becomes illuminating:

$$t_y = \sum_h \beta^h \frac{y^h \hat{\varepsilon}_{yy}^h}{y^a \hat{\varepsilon}_{yy}^a} t_y^h \quad (22)$$

with  $\hat{\varepsilon}_{yy}^h \stackrel{\text{def}}{=} \frac{\partial \hat{y}^h}{\partial q_y} \frac{q_y}{y^h}$ ,  $y^a \stackrel{\text{def}}{=} \sum_h y^h$  and  $\hat{\varepsilon}_{yy}^a \stackrel{\text{def}}{=} \sum_h \frac{\partial \hat{y}^h}{\partial q_y} \frac{q_y}{y^a}$ . Expression (22) reflects the second best nature of the policy. As the government cannot individualise commodity tax rates, it chooses a uniform tax rate on the (de)merit commodity which is a weighted average of the individual first best rates, where the weight depend on the social marginal utility of income ( $\beta^h$ ), the share in aggregate consumption ( $\frac{y^h}{y^a}$ ) and the sensitivity of individual relative to aggregate demand ( $\frac{\hat{\varepsilon}_{yy}^h}{\hat{\varepsilon}_{yy}^a}$ ). A similar expression is true for  $t_x$ .

Under first best, the social marginal utilities of income,  $\beta^h$ , should all be equal to unity. Under uniform commodity taxation, this is no longer necessarily the case. The first order condition for the lump sum tax on consumer  $h$  can now be written as:

$$\beta^h \left( t_x^h \frac{\partial x^h}{\partial m^h} + t_y^h \frac{\partial y^h}{\partial m^h} - 1 \right) = \left( t_x \frac{\partial x^h}{\partial m^h} + t_y \frac{\partial y^h}{\partial m^h} - 1 \right) \quad (23)$$

showing that even at an optimal income distribution the social marginal utilities of income will deviate from one to the extent that (i) the individual first best tax rates differ from the uniform ones, and (ii) the income effects are different from zero. The latter case occurs with quasi linear preferences in the *numéraire*. I will now give an example where all individual first best tax rates are identical and used to implement the optimal allocation.

Suppose preferences are quasi-homothetic, meaning that for each agent there are some 'survival' quantities  $\underline{z}^h$ ,  $\underline{x}^h$ , and  $\underline{y}^h$  such that

$$u^h(z, x, y) = F^h[v^h(z - \underline{z}^h, x - \underline{x}^h, y - \underline{y}^h)]$$

with  $F^{h'}(\cdot) > 0$  and  $v^h(\cdot)$  homogenous of degree 1 in  $(z - \underline{z}^h, x - \underline{x}^h, y - \underline{y}^h)$ . We then have the following

**Lemma 1** *Suppose that consumers have (i) quasi homothetic preferences that are (ii) identical, with (iii) both  $\frac{\partial x^h}{\partial m^h}$  and  $\frac{\partial y^h}{\partial m^h}$  positive. Suppose that (iv)  $\mu(\cdot)$  is constant and (v)  $\underline{y}^g = \underline{y}^h$ . Then the vectors of individual first best tax rates are identical and the first best allocation can be implemented by this common vector and a set of lump sum taxes that set all  $\beta^h$  equal to unity.*

---

<sup>9</sup>More correctly would be to say that they *characterise* the first best tax rates, as the differences in MRS are all evaluated at the second best solution.

The proof, of which the details are spelled out in the appendix, goes roughly as follows. (i) means that the MWPs are homogenous of degree zero in its arguments. (ii), (iv) and (v) then imply that the individual (first best) tax rates are identical across consumers. With (i), all substitution effects are proportional in income, which makes the (common) vector of first best tax rates proportional to the vector the government uses, the factor of proportionality being "the average  $\beta$  plus the covariance of the  $\beta$ s with supernumerary incomes". But with (i), income effects are independent of income and by (iii) and (23)  $\beta^h = 1$  (all  $h$ ), which means the factor of proportionality is 1 as well.

Implementation of the first best allocation is thus not incompatible with an unequal income distribution. But the combination of assumptions for this to happen is quite stringent and in general second best tax rates will be a genuine weighted average of the first best ones.

## 6 Tax reform analysis

The tax rules derived in the previous sections characterise the optimal solution under first and second best. For a policy maker, these rules may not be of primary interest (i) because income distribution policy is not necessarily on the same agenda as commodity tax policy, and (ii) because the existing commodity tax structure puts a straightjacket on what can be achieved through a reform. More interesting are then the rules that indicate in which direction marginal tax changes should occur, and that can easily be expressed in terms of accounting and statistical data.

For this purpose, one is interested in the marginal cost in terms of social welfare,  $W$ , of raising government revenue,  $R$ , with one Euro by changing the tax on commodities  $x$  and  $y$ :

$$MC_x = -\frac{\partial W/\partial t_x}{\partial R/\partial t_x}, \quad MC_y = -\frac{\partial W/\partial t_y}{\partial R/\partial t_y}. \quad (24)$$

Expressions of this kind have been discussed in detail by Ahmad & Stern (1984), who show that a neat parameterisation is obtained by multiplying nominator and denominator by the respective after tax prices. Since the (de)merit good arguments only affect the nominators, I limit myself to this part of the MC-expressions. Let me for simplicity assume that  $\mu(y) \equiv \mu$  and  $\underline{y}^g = 0$ .

For commodity  $x$ , we then have that

$$-\frac{\partial W}{\partial t_x} = -\sum_h \lambda^h \tilde{u}_1^h \left\{ \frac{\partial z^h}{\partial q_x} + \frac{\tilde{u}_2^h}{\tilde{u}_1^h} \frac{\partial x^h}{\partial q_x} + \left[ \frac{\tilde{u}_3^h}{\tilde{u}_1^h} + \mu \right] \frac{\partial y^h}{\partial q_x} \right\}, \quad (25)$$

where a  $\tilde{\cdot}$  above a marginal utility denotes that it is evaluated at the bundle  $(z^h + \mu y^h, x^h, y^h)$ . Using the first order Taylor approximation  $\frac{\tilde{u}_i^h}{\tilde{u}_1^h} \simeq q_i +$

$\frac{\partial u_i^h}{\partial z} \frac{\partial u_i^h}{\partial x} (z, x, y) y^h \mu$  ( $i = 2, 3$ ) and the adding up condition  $\frac{\partial z^h}{\partial q_x} + q_x \frac{\partial x^h}{\partial q_x} + q_y \frac{\partial y^h}{\partial q_x} = -x^h$ , this expression can be rewritten as

$$-\frac{\partial W}{\partial t_x} \simeq \sum_h \lambda^h \tilde{u}_1^{gh} \left\{ x^h - \frac{\partial x^h}{\partial q_x} \frac{\partial u_1^h(z, x, y)}{\partial z} y^h \mu - \frac{\partial y^h}{\partial q_x} \left[ \mu + \frac{\partial u_1^h(z, x, y)}{\partial z} y^h \mu \right] \right\}. \quad (26)$$

Multiplying through by  $q_x$ , and turning derivatives into elasticities, we finally get

$$-q_x \frac{\partial W}{\partial t_x} \simeq \sum_h \lambda^h \tilde{u}_1^{gh} \left\{ q_x x^h - \mu y^h \left( \varepsilon_{xx}^h \delta_{xz}^h \frac{q_x x^h}{z^h} + \varepsilon_{yx}^h \left[ 1 + \delta_{yz}^h \frac{q_y y^h}{z^h} \right] \right) \right\}, \quad (27)$$

where the Marshallian price elasticities and the demand price elasticities are defined as

$$\varepsilon_{xx}^h = \frac{\partial x^h}{\partial q_x} \frac{q_x}{x^h}, \varepsilon_{yx}^h = \frac{\partial y^h}{\partial q_x} \frac{q_x}{y^h}, \delta_{xz}^h = \frac{\partial u_1^h}{\partial z} \frac{z}{u_1^h}, \delta_{yz}^h = \frac{\partial u_3^h}{\partial z} \frac{z}{u_1^h}. \quad (28)$$

Similar operations lead to an analogue expression for  $-q_y \frac{\partial W}{\partial t_y}$ :

$$-q_y \frac{\partial W}{\partial t_y} \simeq \sum_h \lambda^h \tilde{u}_1^{gh} \left\{ q_y y^h - \mu y^h \left( \varepsilon_{xy}^h \delta_{xz}^h \frac{q_x x^h}{z^h} + \varepsilon_{yy}^h \left[ 1 + \delta_{yz}^h \frac{q_y y^h}{z^h} \right] \right) \right\}, \quad (29)$$

with similar definitions for  $\varepsilon_{xy}^h$  and  $\varepsilon_{yy}^h$ . Notice how the existence of a (de)merit good argument corrects both MC expressions in a similar way.

Writing  $\mu y^h = \eta q_y y^h$ , the curly bracket terms can in principle be calculated using expenditure data (expenditures on  $x$  and on  $y$ , and expenditures on  $x$  and  $y$  relative to  $z$ ) and uncompensated price and quantity elasticities. If these elasticities are not available at the household (or income decile) level, they can as an approximation be replaced by the aggregate elasticities. One can then produce different rankings of the marginal costs, for different values of the (de)merit parameter  $\eta$ .

## 7 Discussion and concluding remarks

In this paper, I have questioned the scaling approach proposed by Besley (1988) to model merit good arguments on the ground that it often leads to counterintuitive policy prescriptions. I have proposed a different approach which directly interferes with the marginal willingness to pay for a (de)merit good.

Which approach to choose? If one is convinced that a (de)merit good argument should at least in a first best world prescribe the subsidisation (taxation) of that good, then the approach presented here meets this criterion better than

the scaling approach: only when the merit good is a very inferior one will the policy rule (18b) prescribe taxation; the scaling approach would do so whenever the good is price inelastic.

But even if the present approach prescribes taxation, then a *rationale* can be provided: insofar the merit good considerations also apply to the inframarginal units consumed, the government believes the agent is richer than she thinks she is. Respecting the strong inferiority of the good, it wants the agent to consume less of it.

The analysis has been performed for three commodities. However, its generalisation to  $n$  commodities should be straightforward.

## References

- Besley T (1988) A simple model for merit good arguments, *Journal of Public Economics* **35**, 371-384.
- Capéau B & E Ooghe (2003) Merit goods and phantom agents, *Economics Bulletin* **8**(8), 1-5.
- Decoster A & E Schokkaert (1989) Equity and efficiency of a reform of Belgian indirect taxes, *Recherches Économiques de Louvain* **55**, 155-173.
- Feehan J (1990) A simple model for merit good arguments—a comment, *Journal of Public Economics* **43**, 127-129.
- Fisher F A & K Shell (1967) Taste and quality change in the pure theory of the true cost-of-living index, in J N Wolfe (ed) *Value, Capital and Growth: Papers in Honour of Sir John Hicks*, (Edinburgh: Edinburgh UP), 97-138.
- Kaplanoglou G & D Newbery (2002) Indirect taxation in Greece: evaluation and possible reform, CES working paper 661, Munich.
- Madden D (1995) Labour supply, commodity demand and marginal tax reform, *Economic Journal* **105**, 485-497.
- Musgrave R (1959) *The Theory of Public Finance* (New York: McGraw-Hill).
- Pazner E (1972) Merit wants and the theory of taxation, *Public Finance* **27**, 460-472.
- Sandmo A (1983) Ex post welfare economics and the theory of merit goods, *Economica* **50**, 19-33
- Schroyen F & J Aasness (2002) Indirect marginal tax reform analysis for Norway, mimeo, Norwegian School of Economics.

## Appendix

### • Derivation of the first best tax rules.

The first order conditions for an interior maximum are given by:

$$\frac{\partial \mathcal{L}}{\partial t_x} = u_1^g \left\{ \frac{\partial z}{\partial q_x} + \frac{u_2^g}{u_1^g} \frac{\partial x}{\partial q_x} + \frac{u_3^g}{u_1^g} \frac{\partial y}{\partial q_x} \right\} + \gamma \left\{ x + t_x \frac{\partial x}{\partial q_x} + t_y \frac{\partial y}{\partial q_x} \right\} = 0 \quad (30a)$$

$$\frac{\partial \mathcal{L}}{\partial t_y} = u_1^g \left\{ \frac{\partial z}{\partial q_y} + \frac{u_2^g}{u_1^g} \frac{\partial x}{\partial q_y} + \frac{u_3^g}{u_1^g} \frac{\partial y}{\partial q_y} \right\} + \gamma \left\{ y + t_x \frac{\partial x}{\partial q_y} + t_y \frac{\partial y}{\partial q_y} \right\} = 0 \quad (30b)$$

$$\frac{\partial \mathcal{L}}{\partial T} = -u_1^g \left\{ \frac{\partial z}{\partial m} + \frac{u_2^g}{u_1^g} \frac{\partial x}{\partial m} + \frac{u_3^g}{u_1^g} \frac{\partial y}{\partial m} \right\} - \gamma \left\{ 1 + t_x \frac{\partial x}{\partial m} + t_y \frac{\partial y}{\partial m} \right\} = 0 \quad (30c)$$

Performing  $\frac{\partial \mathcal{L}}{\partial t_x} - \frac{\partial \mathcal{L}}{\partial T} x$  and  $\frac{\partial \mathcal{L}}{\partial t_y} - \frac{\partial \mathcal{L}}{\partial T} y$  gives:

$$\left\{ \frac{\partial \hat{z}}{\partial q_x} + \frac{u_2^g}{u_1^g} \frac{\partial \hat{x}}{\partial q_x} + \frac{u_3^g}{u_1^g} \frac{\partial \hat{y}}{\partial q_x} \right\} + \tilde{\gamma} \left\{ t_x \frac{\partial \hat{x}}{\partial q_x} + t_y \frac{\partial \hat{y}}{\partial q_x} \right\} = 0 \quad (31a)$$

$$\left\{ \frac{\partial \hat{z}}{\partial q_y} + \frac{u_2^g}{u_1^g} \frac{\partial \hat{x}}{\partial q_y} + \frac{u_3^g}{u_1^g} \frac{\partial \hat{y}}{\partial q_y} \right\} + \tilde{\gamma} \left\{ t_x \frac{\partial \hat{x}}{\partial q_y} + t_y \frac{\partial \hat{y}}{\partial q_y} \right\} = 0 \quad (31b)$$

where  $\hat{\cdot}$  denotes a compensated price effect and  $\tilde{\gamma} \stackrel{\text{def}}{=} \frac{\gamma}{u_1^g}$ . Using the homogeneity condition on the compensated price effects, (31) may be rewritten as

$$\begin{bmatrix} 1 & \frac{u_2^g}{u_1^g} & \frac{u_3^g}{u_1^g} \end{bmatrix} \begin{bmatrix} -q_x & -q_y \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{x}}{\partial q_x} & \frac{\partial \hat{x}}{\partial q_y} \\ \frac{\partial \hat{y}}{\partial q_x} & \frac{\partial \hat{y}}{\partial q_y} \end{bmatrix} = \tilde{\gamma} \begin{bmatrix} t_x & t_y \end{bmatrix} \begin{bmatrix} \frac{\partial \hat{x}}{\partial q_x} & \frac{\partial \hat{x}}{\partial q_y} \\ \frac{\partial \hat{y}}{\partial q_x} & \frac{\partial \hat{y}}{\partial q_y} \end{bmatrix}. \quad (32)$$

Because the substitution matrix  $\begin{pmatrix} \frac{\partial \hat{x}}{\partial q_x} & \frac{\partial \hat{x}}{\partial q_y} \\ \frac{\partial \hat{y}}{\partial q_x} & \frac{\partial \hat{y}}{\partial q_y} \end{pmatrix}$  is negative definite, (32) reduces to

$$\tilde{\gamma} t_x = q_x - \frac{u_2^g}{u_1^g}, \quad \text{and} \quad \tilde{\gamma} t_y = q_y - \frac{u_3^g}{u_1^g}. \quad (33a)$$

Inserting these conditions back in the FOC for  $T$ , shows that  $\tilde{\gamma} = 1$ . Since  $q_x = \frac{u_2}{u_1}$  and  $q_y = \frac{u_3}{u_1}$ , the first best tax rates are as in the text.

### • Derivation of $u^g(z, x, y)$

By definition of  $F(\cdot)$

$$z + \int_{\underline{y}^g}^y \mu(\chi) d\chi = F \left[ x, y, u \left( z + \int_{\underline{y}^g}^y \mu(\chi) d\chi, x, y \right) \right] \quad (34)$$

$\Updownarrow$

$$z = F \left[ x, y, u \left( z + \int_{\underline{y}^g}^y \mu(\chi) d\chi, x, y \right) \right] - \int_{\underline{y}^g}^y \mu(\chi) d\chi \quad (35)$$

By definition of  $F^g(\cdot)$ ,  $F^g[x, y, u^g(z, x, y)] = z$ , which means the above expression can be written as

$$F^g[x, y, u^g(z, x, y)] = F \left[ x, y, u \left( z + \int_{\underline{y}^g}^y \mu(\chi) d\chi, x, y \right) \right] - \int_{\underline{y}^g}^y \mu(\chi) d\chi \quad (36)$$

From (14)

$$F^g[x, y, u^g(z, x, y)] = F^g \left[ x, y, u \left( z + \int_{\underline{y}^g}^y \mu(\chi) d\chi, x, y \right) \right] \quad (37)$$

Therefore

$$u^g(z, x, y) = u \left( z + \int_{\underline{y}^g}^y \mu(\chi) d\chi, x, y \right). \quad (38)$$

• **Proof of lemma 1.**

Expression  $\left(\frac{u_2^h}{u_1^h}\right)^g$  can be written as

$$\frac{v_2^h}{v_1^h} (z^h - \underline{z}^h + \int_{\underline{y}^g}^{y^h} \mu(\chi) d\chi, x^h - \underline{x}^h, y^h - \underline{y}^h). \quad (39)$$

$\left(\frac{u_2^h}{u_1^h}\right)^g$  looks similar but has the extra term  $\mu(y^h)$ .

For these preferences, it can be shown that the Marshallian demands for commodity  $x$  is of the form

$$x^h(q, m^h) = \underline{x}^h + \frac{\frac{\partial b^h(q)}{\partial q_x} - \underline{x}^h}{b(q) - \underline{m}^h(p)} [m^h - \underline{m}^h(p)]. \quad (40)$$

where  $b(q)$  is minimal expenditure necessary to generate one unit of utility and  $\underline{m}^h(p) \stackrel{\text{def}}{=} \underline{z}^h + q_x \underline{x}^h + q_y \underline{y}^h$  denotes survival income. Similar expressions hold for the other two commodities.

Under the lemma's assumptions, expression (39) reduces for every agent  $h$  to

$$\frac{v_i}{v_1} \left( \frac{\partial b(q)}{\partial q_z} - \underline{x} + \mu \left[ \frac{\partial b(q)}{\partial q_y} - \underline{y} \right], \frac{\partial b(q)}{\partial q_x} - \underline{x}, \frac{\partial b(q)}{\partial q_y} - \underline{y} \right), \quad (41)$$

since the expression is homogenous of degree 0 in its arguments and the term  $\frac{m^h - \underline{m}(p)}{b(q) - \underline{m}(p)}$  thus drops out. Every agent will then have the same vector of (first best)

tax rates,  $t^c$ , say. Since all substitution effects are proportional to supernumerary income, (21) reduces to

$$\frac{\sum_h \beta^h (m^h - \underline{m}(p))}{\sum_h (m^h - \underline{m}(p))} t^c = t, \text{ or} \quad (42)$$

$$[\beta^{av} + cov(\beta^h, \frac{m^h - \underline{m}(p)}{m^{av} - \underline{m}(p)})] t^c = t \quad (43)$$

On the other hand, because the income effects are independent of income, (23) can be written as  $\beta^h r(t^c) = r(t)$ , where  $r(\cdot)$  denotes the net increase in government revenue from a marginal increase in the lump sum grant to a consumer. Combining both results gives:

$$\frac{r(t)}{r(t^c)} t^c = t \quad (44)$$

If then  $r(t) > (<) r(t^c)$  then  $t^c \gg (<<) t$ . Because both income effects are positive, this implies that  $r(t^c) > (<) r(t)$ , a contradiction. Therefore  $t^c = t$  and  $\beta^h = 1$  (all  $h$ ).