A New Approach to Optimal Commodity Taxation

by Stefan Homburg

DISCUSSION PAPER, NO. 299
JUNE 2004
ISSN: 0949-9962

Institute of Public Finance, School of Economics and Management, University of Hannover, Königsworther Platz 1, D-30167 Hannover, Germany. *email*: homburg@fiwi.uni-hannover.de.

ABSTRACT: This paper makes a fresh attempt at characterizing optimal commodity taxes. Under the usual assumptions, an extremely simple expression of second-best commodity taxes is derived, showing tax rates as functions of observable variables only, rather than as functions of unobservable variables such as compensated cross elasticities. The main formula is independent of special preferences, and independent of the number of commodities. It has a simple economic meaning and could be particularly useful for empirical research. Examples and remarks on the normalization problem are provided

Revised version published in:

FinanzArchiv **63**, September 2006

KEYWORDS: Optimal Commodity Taxation, Ramsey Rule.

JEL-CLASSIFICATION: H21

1. Introduction[1]

For more than one and a half centuries, economists have been trying to characterize commodity taxes which minimize the excess burden of taxation. Ramsey (1927) surmounted the partial equilibrium framework of Dupuit (1844) and others, laying the foundation for the modern treatment. Research on this topic is still continuing, and if one takes to hand a contemporary textbook[2] or a recent survey[3] in order to pick up the fruits of these efforts, it turns out that the theory of public finance essentially offers three answers to the question as to how a policy maker should set commodity taxes optimally. However, each of the three answers has so severe shortcomings that they are almost useless for practical purposes. In the order of decreasing popularity[4], the answers have become known as the Ramsey rule, the inverse elasticity rule, and the Corlett-Hague rule. Let us review these briefly.

- The *Ramsey rule* characterizes optimal quantities, stating that optimal taxes diminish all demands in the same proportion. In fact, this interpretation holds only approximately in the neighborhood of zero tax revenue.
- According to the *inverse elasticity rule*, optimal commodity tax rates are inversely proportional to elasticities of demand. This presupposes a very special form of the utility function and is far from a general result.
- Finally, the *Corlett-Hague rule* yields a substantive conclusion for any well-behaved preferences. It states that in a setting with two commodities and leisure, precisely that commodity which is more complementary with leisure should be taxed at a higher rate. But unfortunately, we live in a world with considerably more than two commodities.

This paper makes a fresh attempt at characterizing optimal commodity taxes. Adopting the standard general equilibrium framework without externalities, distributional concerns, or presence of a non-linear income tax[5], a very simple rule for optimal commodity taxes will be derived. This rule holds independently of special assumptions on preferences, and independently of the number of commodities. It entails an explicit formula for optimal tax rates in terms of observable

---

1 This paper was presented at the CESifo Area Conference on Public Sector Economics, which took place in April 2005 in Munich. I thank the conference participants in general, and Volker Grossmann in particular, for helpful hints. I am also grateful for constructive criticisms of two anonymous referees.

2 E. g. Myles (1995).

3 E. g. Auerbach and Hines (2002).

4 To get an approximate measure of popularity, I performed an internet search which, at the time it was conducted, entailed over 1.000 hits for “Ramsey rule”, roughly 500 hits for “inverse elasticity rule” and less than 100 hits for “Corlett-Hague rule”.

5 For a recent contribution to the direct-indirect-mix, see Kaplow (2006).

variables. Regarding empirical applications, an explicit solution in terms of observable variables seems important and useful; after all, the fact that most traditional results are formulated in terms of compensated elasticities is often considered as an eminent obstacle to putting the theory into practice.

The paper is organized as follows. Section 2 presents the standard approach in a nutshell, in order to facilitate comparison with the new approach. Section 3 contains our principal result. A complementarity rule which holds for an arbitrary number of commodities is derived in section 4. Some examples are provided in section 5. Section 6 considers optimal commodity taxes in the presence of an additional wage tax, section 7 contains a blueprint for classroom presentation, and section 8 concludes. The appendix clarifies some minor points of rigor.

The common underlying framework is this one: There are n commodities, produced using linear technologies with labor as the sole input and represented by non-negative vectors[6] $\boldsymbol{x}=(x_1, ..., x_n)$. Taking positive consumer prices $\boldsymbol{q}$ as given, consumers maximize utility $u(y, \boldsymbol{x})$ by selecting effort y and consumption subject to the budget constraint $y \geq \boldsymbol{qx}$. Effort and leisure sum up to a given time endowment, and are also referred to as commodity no. 0. Gross and net wage rates being normalized to unity, the variable y represents effort in hours as well as earnings in monetary units. Differentiating the Lagrangean $\mathcal{L}=u(y, \boldsymbol{x})+\mu(y-\boldsymbol{qx})$ with respect to consumers' choice variables yields $n+1$ first-order conditions of an individual optimum:

$$\frac{\partial u}{\partial y} = -\mu \quad \text{and} \quad \frac{\partial u}{\partial \boldsymbol{x}} = \mu \boldsymbol{q} . \tag{1}$$

2. CHARACTERIZING OPTIMAL QUANTITIES

The prevailing dual approach does not aim to describe optimal taxes, but seeks to characterize optimal quantities. It derives ordinary demand functions $\boldsymbol{x}(\boldsymbol{q})$ from the first-order conditions (1). The budget constraint, which binds at the individual optimum, defines an earnings function $y(\boldsymbol{q})=\boldsymbol{qx}(\boldsymbol{q})$. Differentiating this earnings function entails the well-known Cournot aggregation $\partial y/\partial \boldsymbol{q}=\boldsymbol{x}+\boldsymbol{q}\,\partial \boldsymbol{x}/\partial \boldsymbol{q}$. The government, using consumer prices as controls, maximizes utilities subject to the revenue constraint $\boldsymbol{tx} \geq g$. Here, $g>0$ is a given revenue requirement, $\boldsymbol{t}=\boldsymbol{q}-\boldsymbol{p}$ are the taxes and $\boldsymbol{p}$ represents positive producer prices, which are exogenous due to the linear production technologies assumed above[7]. Differentiating the Lagrangean $\mathcal{L}=u(y(\boldsymbol{q}), \boldsymbol{x}(\boldsymbol{q}))+\lambda(\boldsymbol{tx}(\boldsymbol{q})-g)$ with respect to the controls yields the first-order condition of the government's second-best problem:

6 Following Mas-Colell et al. (1995, 926), the same notation is used for row and column vectors because both types are simply points in n-dimensional space.

7 It is well-known that all results remain unchanged in case of endogenous producer prices.

(2) $$\frac{\partial u}{\partial y}\frac{\partial y}{\partial \boldsymbol{q}}+\frac{\partial u}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{q}}+\lambda\left(\boldsymbol{x}+\boldsymbol{t}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{q}}\right)=\boldsymbol{0},$$

since $\partial \boldsymbol{t}/\partial \boldsymbol{q}=\mathbf{I}$. This leads to the uncompensated version of the Ramsey rule, if we substitute from (1) and the Cournot aggregation:

(3) $$\boldsymbol{t}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{q}}=-\frac{\lambda-\mu}{\lambda}\boldsymbol{x}.$$

Ramsey (1927) assumed utility to be quasi-linear in effort. Then, as there are no income effects on commodity demands, the expression on the left involves a symmetric matrix and may be re-written as $\partial \boldsymbol{x}/\partial \boldsymbol{q}\,\boldsymbol{t}$. According to the mean-value theorem, this equals the changes in demands, $\boldsymbol{dx}$, when taxes are increased from zero to $\boldsymbol{dt}$. Hence, all demands are diminished by the common factor $(\lambda-\mu)/\lambda$.

Samuelson (1951) modified Ramsey's rule in the following way: He used Slutsky's decomposition $\partial x_i/\partial q_j=\partial x_i^c/\partial q_j-x_j\,\partial x_i/\partial z$, where x_i^c represents the compensated demand for the i-th commodity and z denotes a virtual income, to derive a formula which holds also in the presence of income effects:

(4) $$\left(\frac{\partial x_i^c}{\partial q_j}\right)_{i,j=1..n}\boldsymbol{t}=\left(\boldsymbol{t}\frac{\partial \boldsymbol{x}}{\partial z}-\frac{\lambda-\mu}{\lambda}\right)\boldsymbol{x}.$$

Mirrlees (1976) called $\Sigma_j\,\partial x_i^c/\partial q_j\,t_j/x_i$ the "index of discouragement" because it gives the reduction in the i-th commodity following a tax increase. Thus, the compensated variant of Ramsey's rule states that the index of discouragement should be the same for all commodities, at the optimum.

3. CHARACTERIZING OPTIMAL TAXES

The primal approach seeks to characterize optimal taxes rather than optimal quantities. It uses *inverse demand functions* $\boldsymbol{q}(\boldsymbol{x})$ which, as shown in the appendix, exist for all $\boldsymbol{x}\in D$, where D denotes the image of the ordinary demand functions $\boldsymbol{x}(\boldsymbol{q})$. Now, the binding budget constraint defines an *offer curve* $y(\boldsymbol{x})=\boldsymbol{q}(\boldsymbol{x})\boldsymbol{x}$. Differentiating the offer curve with respect to the quantities[8] entails the *inverse Cournot aggregation* $\partial y/\partial \boldsymbol{x}=\boldsymbol{x}\,\partial \boldsymbol{q}/\partial \boldsymbol{x}+\boldsymbol{q}$. To understand what this means, observe that the government, by setting appropriate prices, can induce consumers to choose any demand $\boldsymbol{x}\in D$. In so doing, however, the government must ensure that consumers remain on their offer curves. In this sense, earnings are a function of the demands, and the derivative $\partial y/\partial x_i$ indicates the response of earnings to a small change in x_i.

8 After differentiation, the vector $\boldsymbol{x}$ must appear left to $\partial \boldsymbol{q}/\partial \boldsymbol{x}$, cf. Mas-Colell et al. (1995, 927).

Assuming a single commodity ($n=1$) for a moment facilitates a graphical exposition of the offer curve (Fig. 1). Each ray through the origin represents a certain consumer price q. In the dual approach, the government selects q and accepts the resulting choices $x(q)$ and $y(q)$. In the primal approach, the government selects a quantity x, supports it by $q(x)$ and accepts the resulting effort $y(x)=q(x)x$ which follows from the budget constraint. The offer curve, familiar from the theory of international trade, is simply the set of points of tangency of the indifference curves with the rays through the origin.

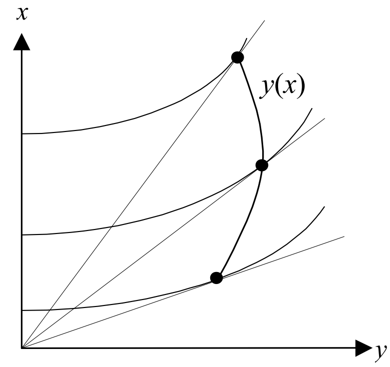

Fig 1: The Offer Curve.

Returning to the many-commodity case, taxes now follow from the identity $\boldsymbol{t}(\boldsymbol{x})=\boldsymbol{q}(\boldsymbol{x})-\boldsymbol{p}$ which implies $\partial \boldsymbol{t}/\partial \boldsymbol{x}=\partial \boldsymbol{q}/\partial \boldsymbol{x}$. Taking quantities as controls, the government maximizes utilities subject to its revenue constraint. Differentiating the Lagrangean $\mathcal{L}=u(y(\boldsymbol{x}),\boldsymbol{x})+\lambda(\boldsymbol{t}(\boldsymbol{x})\boldsymbol{x}-g)$ with respect to the controls, one obtains:

$$(5) \qquad \frac{\partial u}{\partial y}\frac{\partial y}{\partial \boldsymbol{x}}+\frac{\partial u}{\partial \boldsymbol{x}}+\lambda\left(\boldsymbol{x}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{x}}+\boldsymbol{t}\right)=\boldsymbol{0}\,.$$

Using (1) and the inverse Cournot aggregation, this becomes

$$(6) \qquad -\mu \boldsymbol{x}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{x}}-\mu \boldsymbol{q}+\mu \boldsymbol{q}+\lambda \boldsymbol{x}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{x}}+\lambda \boldsymbol{t}=\boldsymbol{0}\,.$$

Solving for $\boldsymbol{t}$, we obtain what could be called the inverse Ramsey rule:

$$(7) \qquad \boldsymbol{t}=-\frac{\lambda-\mu}{\lambda}\boldsymbol{x}\frac{\partial \boldsymbol{q}}{\partial \boldsymbol{x}}\,.$$

Comparison of (7) and (3) highlights the duality of the two approaches in a particularly nice fashion. Apart from this aesthetic virtue, the inverse Ramsey rule is as useless as the Ramsey rule itself. However, performing a simple operation that has no analogue in the dual approach, one can replace $-\boldsymbol{x}\,\partial \boldsymbol{q}/\partial \boldsymbol{x}$ by $\boldsymbol{q}-\partial y/\partial \boldsymbol{x}$, using again the inverse Cournot aggregation. This yields our main result:

$$(8) \qquad \boldsymbol{t}=\frac{\lambda-\mu}{\lambda}\left(\boldsymbol{q}-\frac{\partial y}{\partial \boldsymbol{x}}\right).$$

After a trivial re-calculation, an alternative version of the result reads:

$$t = \frac{\lambda - \mu}{\mu}\left(\boldsymbol{p} - \frac{\partial y}{\partial \boldsymbol{x}}\right). \tag{9}$$

The first version is more useful when it comes to applications, as will be shown below. From a theoretical point of view, the second version is more satisfactory because producer prices are exogenous. The crucial elements in determining optimal taxes are the derivatives $\partial y/\partial x_i$, which indicate the responses of earnings to changes in demands. Since $\lambda > \mu$, as demonstrated in the appendix, the rule asserts that a commodity which stimulates earnings should be taxed at a low rate, and *vice versa*.

To get to grips with formula (8), consider an original demand $\boldsymbol{x}^0$ and suppose that the governments induces an increase in x_i through an appropriate change in taxes, keeping all other demands constant. If the i-th commodity is time consuming, the time constraint requires a reduction in effort, $\partial y/\partial x_i$ is negative, and the formula suggests taxing the respective commodity at a high rate. Conversely, if the i-th commodity stimulates working, such as a work-related expenditure, $\partial y/\partial x_i$ is positive, and the corresponding tax rate should be low. In any case it is the impact on earnings as the genuine tax base which dictates the optimal tax rate.

Before concluding this section, some historical remarks are in order. The first economist adopting the primal approach to optimal commodity taxation was, of course, Ramsey himself, who worked in the Marshallian tradition of what we today call inverse demand. Indeed, Ramsey's (1927) central equation (3) is an archaic form of (7). Assuming $\partial \boldsymbol{q}/\partial \boldsymbol{x}$ to be symmetric and taxes to be infinitesimal, he used it to derive his principal result "that optimal taxes diminish the production of all commodities in the same proportion" (Ramsey 1927, p. 54).

Atkinson and Stiglitz (1972) also used a primal approach, treating earnings as an additional control and putting the condition (1) that consumers remain on their offer curves as an additional restriction into the Lagrangean. This facilitates describing optimal taxes in terms of the bordered Hessian of the utility function but does not yield much insight. Deaton (1979) essentially decomposed (7) into compensated prices changes and inverse income effects, using the distance function and the Antonelli matrix, and expressed optimal tax rates in terms of what he called compensated flexibilities. Comparing these approaches in an elegant survey, Stern (1986) pointed out that the analyst has the choice between three different matrices: Slutsky, Hessian or Antonelli.

The strategy followed above was not to add yet a fourth matrix type, but to let matrices completely disappear from the scene. This entails the simple vector equation (8). Of course, the new formulae describe one and the same optimum as the traditional formulae: They characterize allocations maximizing utility or, which amounts to the same thing (Kay 1980), minimize the aggregate excess burden of taxation. But (8) and (9) yield an intuitive interpretation of second-best commodity taxes, whereas the message of the Ramsey rule is obscure, and the other traditional results hold only in special cases.

4. A COMPLEMENTARITY RULE FOR ARBITRARILY MANY COMMODITIES

Equation (8) describes optimal taxes in terms of consumer prices, earnings, demands, and an unobservable common factor. To get rid of the common factor, let us define the *index of encouragement* for each commodity:

(10) $$e_i = \frac{\partial y}{\partial x_i}\frac{1}{q_i}.$$

The index of encouragement indicates the response of earnings to an increase in the respective expenditure. It may be positive, negative or zero. Dividing some row in (8) by the consumer price, inserting (10), and dividing two such rows yields our second principal result which describes the optimal tax structure in terms of observable variables only, as the Lagrangean multipliers drop out:

(11) $$\frac{t_i/q_i}{t_j/q_j} = \frac{1-e_i}{1-e_j}.$$

This new complementarity rule states *that commodities with a high index of encouragement should be taxed at low rates, and vice versa.* To illustrate this proposition, consider brandy and coffee. Brandy is likely to reduce earnings, at least if consumed in larger quantities. If this is true, brandy has a negative index of encouragement, and (11) suggests taxing it at a relatively high rate. By the same token, as long as drinking coffee stimulates earnings, the coffee tax should be low.

The economic logic behind these findings is as follows. Consumers have positive time endowments which cannot be taxed directly. Commodities with a high index of encouragement induce consumers to supply large portions of their time endowments in the market. A government restricted to taxing only market demands will stimulate consuming such commodities, as this broadens the tax base. Note, the impossibility of taxing time endowments is the driving force behind second-best taxes[9]. Without this restriction, the government could establish a first-best tax system consisting of a tax on endowments only. The literature sometimes confounds untaxed endowments with untaxed market demands. Clearly, any market demand can be taken as untaxed, because only relative prices matter. For a clarification of this point, see Stern (1986, p. 298) and section 6 below. Formula (11)

9 The implicit assumption that leisure is the only good with an untaxable endowment is not restrictive. If there were several such goods, these could be combined into a single Hicksian composite, since the government cannot change their relative prices by hypothesis.

shows that the optimal tax structure is determined by the slopes $\partial y/\partial x_i$ of the offer curve in multi-dimensional space.

We now wish to relate our complementarity rule to the famous Corlett-Hague rule, which can be derived from (4) within the dual approach. Defining compensated cross elasticities $\varepsilon_{ij} = \partial x_i^c / \partial q_j \, q_j / x_i$, the Corlett-Hague rule reads[10]

$$\frac{t_1/q_1}{t_2/q_2} = \frac{\varepsilon_{20} + \varepsilon_{21} + \varepsilon_{12}}{\varepsilon_{10} + \varepsilon_{21} + \varepsilon_{12}} \tag{12}$$

and states that the commodity which is stronger complementary with leisure (has the smaller value of ε_{i0}) should be taxed harder. Equating the (12) and (11) reveals an important equivalence:

$$\varepsilon_{10} < \varepsilon_{20} \quad \Leftrightarrow \quad e_1 < e_2 \, . \tag{13}$$

Thus, commodity no. 1 is stronger complementary with leisure (in the Corlett-Hague sense) if and only if its index of encouragement is lower. The traditional approach and the approach introduced here are equivalent in the two-commodity case. However, the Corlett-Hague rule fails in the presence of three or more commodities[11], whereas (11) holds for arbitrarily many commodities. As a further difference, our rule involves observable variables only instead of compensated cross elasticities that are very hard to estimate (Deaton, 1987).

The new complementarity rule provides the following intuition: Commodities with a high index of encouragement should be taxed at low rates, because consuming these commodities stimulates the supply of the untaxable time endowment as the genuine tax base. In this view, it seems sensible to have special taxes on alcoholic beverages, and to tax work-related expenses only moderately, or make them deductible under an income tax. The crucial force behind optimal taxation is the interaction of taxes with effort, not with commodity demands as such. The importance of selecting taxes so as to stimulate labour supply has also been stressed in an interpretative paper by Munk (2002) within the dual approach. The dual approach, however, does not yield intuitive formulae, except in special cases which will be considered next.

5. Examples

In this section, we would like to provide some examples by considering concrete functional forms. All examples presume a total time endowment equal to unity, so that $1-y$ denotes leisure.

1. Cobb-Douglas Utility: First consider the economist's workhorse

---

10 For a simple derivation, cf. Auerbach (1985, 92).

11 Intuitive formulae for the many-commodity case cannot be obtained, see Hatta (2004, p. 6).

(14) $$u(y,\boldsymbol{x}) = (1-y)^{\alpha_0} x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n} ,$$

where the numbers α_i sum up to unity. The textbook solutions read $y=1-\alpha_0$ and $x_i=\alpha_i/q_i$ for all i. Changes in commodity demands do not affect earnings since all expenditure shares, including the share of leisure itself, remain constant. The derivatives $\partial y/\partial x_i$ vanish for all commodities, hence $e_i=0$, and (8) implies taxation at the uniform rate $(\lambda-\mu)/\lambda$.

2. Inverse Elasticity Utility: Next we derive the inverse elasticity rule:

(15) $$u(y,\boldsymbol{x}) = 1 - y + u_1(x_1) + \ldots + u_n(x_n) .$$

Quasi-linearity in y excludes income effects, so that $\mu=1$. Differentiating the right-hand equation in (1) yields $\partial q_i/\partial x_j=0$ for all $j\neq i$, implying $\partial y/\partial x_i=\partial q_i/\partial x_i x_i+q_i$. Therefore, the index of encouragement equals

(16) $$e_i = \frac{\partial q_i}{\partial x_i}\frac{x_i}{q_i} + 1 ,$$

from which the inverse elasticity rule $t_i/q_i=-(\lambda-\mu)/\lambda\cdot\partial q_i/\partial x_i\cdot x_i/q_i$ follows immediately. Standard textbook wisdom declares that commodities whose demands react stronger to price changes should be taxed at lower rates because quantity reactions induce excess burdens. Though this is literally true under the very stringent assumption on preferences made here, a more convincing explanation, which already comes close to the intuition behind (11), was offered by Sandmo (1987, p. 92): If the compensated cross elasticities ε_{ij} vanish for all $i\neq j$, the adding-up property implies $\varepsilon_{0i}=-\varepsilon_{ii}$. Therefore, taxing a commodity with a high own elasticity of demand entails a strong reduction in earnings, and hence in tax revenue. From this perspective, the inverse elasticity rule only paraphrases the above contention that commodities which stimulate earnings should be taxed at low rates. In fact, the impact on earnings is again the driving force behind the result, and the utility function (15) presents but a theoretical curiosity.

3. Weakly Separable Utility: The following assumption covers a broader class of utility functions.

(17) $$u(y,\boldsymbol{x}) = u(y,v(\boldsymbol{x})),$$

where $v(\boldsymbol{x})$ is a linear-homogenous sub-utility function. The first order conditions (1) become:

(18) $$\frac{\partial u(y,v(\boldsymbol{x}))}{\partial y} = -\mu \quad \text{and} \quad \frac{\partial u(y,v(\boldsymbol{x}))}{\partial v}\frac{\partial v}{\partial \boldsymbol{x}} - \mu \boldsymbol{q} = \boldsymbol{0} .$$

Substituting the first equation into the second and multiplying by $\boldsymbol{x}$ yields:

(19) $$\frac{\partial u(y,v(\boldsymbol{x}))}{\partial v} v(\boldsymbol{x}) + \frac{\partial u(y,v(\boldsymbol{x}))}{\partial y} y = 0 ,$$

where Euler's equation $v(\boldsymbol{x})=\partial v/\partial \boldsymbol{x}\boldsymbol{x}$ and the identity $y=\boldsymbol{q}\boldsymbol{x}$ have been exploited. The desired result now follows from differentiating this equation implicitly (demonstrating how neat some proofs become under the primal approach):

$$\text{(20)} \qquad \frac{\partial y}{\partial x_i} = -\frac{\dfrac{\partial^2 u}{\partial v^2}\dfrac{\partial v}{\partial x_i}v + \dfrac{\partial u}{\partial v}\dfrac{\partial v}{\partial x_i} + \dfrac{\partial^2 u}{\partial y \partial v}\dfrac{\partial v}{\partial x_i}y}{\dfrac{\partial^2 u}{\partial v \partial y}v + \dfrac{\partial^2 u}{\partial y^2}y + \dfrac{\partial u}{\partial y}} = \gamma \frac{\partial v}{\partial x_i}.$$

The common factor γ is defined by the right-hand equation. As (18) asserts that $\partial v/\partial x_i = -\partial u/\partial y/(\partial u/\partial v)\, q_i$, we have $e_i = -\partial u/\partial y/(\partial u/\partial v)\gamma$. The index of encouragement does not vanish in general, but is identical across commodities, implying uniform taxation. In the present case, the marginal rates of substitution between two arbitrary commodities are independent of earnings and *vice versa*. Therefore, distorting demands along an indifference hyper-surface increases the economic costs of the commodity bundle but leaves earnings unchanged, which cannot be optimal. Differentiated taxes and the associated distortions are optimal only as long as differentiation stimulates earnings.

4. Stone-Geary Utility: The last example regards utility functions of the form

$$\text{(21)} \qquad u(y,\boldsymbol{x}) = (1-y)^{\alpha_0}(x_1-\beta_1)^{\alpha_1}\cdot\ldots\cdot(x_n-\beta_n)^{\alpha_n},$$

the numbers α_i again summing up to unity. Using $\hat{x}_i = x_i - \beta_i$, we are back in the Cobb-Douglas case and obtain $\hat{x}_i = \alpha_i/q_i$ or $q_i = \alpha_i/(x_i - \beta_i)$, the familiar linear expenditure system. As cross elasticities vanish, $\partial y/\partial x_i = -\alpha_i x_i/(x_i-\beta_i)^2 + \alpha_i/(x_i-\beta_i)$. Dividing the last expression by the consumer price yields the index of encouragement:

$$\text{(22)} \qquad e_i = -\frac{\beta_i q_i}{\alpha_i}.$$

As tax rates are proportional to $1-e_i$, they are increasing in β_i. This is often taken to mean that taxes on basic needs such as food should be high. Considering the possibility of home production, Sandmo (1990) suggests a somewhat different interpretation: Tax theory focuses on *market demands* rather than on demands in the literal sense. If consumers can avoid taxes on specific market demands by home production, these demands should be taxed moderately. On the other hand, complicated commodities like computers or cars cannot be produced at home, and taxing them at high rates forces consumers to increase earnings. Following this logic, a high value of β_i indicates that consumers cannot easily dispense with the respective market demand.

6. The Normalization Problem

The above analysis uses the classical framework of a pure excise tax system. This specific normalization of consumer prices, which assumes the absence of a wage tax, is not as innocuous as it appears, nor does it present a serious obstacle to developing a meaningful theory of optimal commodity taxes. In what follows, producer prices $\boldsymbol{p}$ are all set equal to unity through appropriate choices of units. The above analysis involves a vector $(q_0, \boldsymbol{q})$ of consumer prices, where effort is treated as commodity no. 0. As effort and commodity demands are homogenous of degree zero in $(q_0, \boldsymbol{q})$, the first component of this vector, the consumer price of leisure, can also be chosen arbitrarily. In the above analysis it was set to unity, implying that earnings remained untaxed. Thus, the variable y represented both earnings and effort, and we obtained concrete optimal tax rates – which may be the relevant ones in case of a less developed country that raises all tax revenue through excises.

To generalize, let us now write earnings more explicitly as $y=-q_0x_0$, where $-x_0$ represents effort in hours; as usual in general equilibrium theory, the minus sign indicates a supply. Introducing a wage tax means reducing q_0 somewhat. With a given tax revenue, this change does in no way influence the allocation. In the presence of a single commodity, for instance, it is immaterial whether we set $q_1=2$ and $q_0=1$ (a pure excise tax) or $q_1=1$ and $q_0=½$ (a pure wage tax). But the relative tax rates described by (11) are not invariant with respect to the normalization. This finding led Mirrlees (1976) to dismiss altogether the idea of characterizing optimal tax rates. However, this appears to push the point too far because, once we know the wage tax rate – and there is nothing to prevent us from doing so – we can infer the optimal commodity tax rates. These are uniquely defined, relative to the actual wage tax. The upshot is that calculating relative tax rates becomes possible once the wage tax is known. As the latter is public knowledge, an applied economist can easily use it. However, it is important to keep in mind – especially regarding Europe with its largely harmonized VAT regulations – that optimal commodity taxes cannot be determined without regard to the actual mix of direct and indirect taxation.

For a theorist not interested in concrete numerical solutions, there is yet an easier way to cope with the normalization problem. Normalizing all producer prices to unity, which can be done without loss of generality, the relationship $q_i-q_j>0$ asserts unambiguously that commodity i bears a higher tax than commodity j. The normalization of prices affects only the size of the difference, not its sign. Substituting $t_i=q_i-1$ into (9), subtracting two such equations and rearranging yields

$$q_i - q_j = -\frac{\lambda-\mu}{\mu}\left(e_i - e_j\right). \tag{23}$$

It remains true that commodities are taxed the more heavily, the smaller their index of encouragement. Indeed, the index of encouragement itself is *invariant* with respect to normalizations of consumer prices, and so is the difference $e_i - e_j$, which provides an abstract measure of optimal relative tax burdens. Normalizations influence the difference of consumer prices solely through their effects on the Lagrangean multipliers. Since scaling up the price vector does not affect the sign of $q_i - q_j$, a theorist interested in qualitative properties can forget about normalization.

7. A Blueprint for Classroom Presentation

Up to this point, every effort was made to cast the theory in the accustomed form, in order to be able to relate it to the traditional results. This strategy may obscure the fact that the basic derivation is trivial indeed. We now wish to present a much simpler proof, for which one does not need duality theory, Jacobians, inverse Cournot aggregations, and the like. With an elementary knowledge of Lagrangean multipliers, optimal commodity taxes can be derived in two steps.

First step: Form the Lagrangean $\mathcal{L} = u(y, \boldsymbol{x}) + \mu(y - \boldsymbol{qx})$ from individual utility maximization and differentiate with respect to some demand:

$$\frac{\partial u}{\partial y} = -\mu \quad \text{and} \quad \frac{\partial u}{\partial x_i} = \mu q_i . \tag{24}$$

Second step: The identity $y = \boldsymbol{qx} = \boldsymbol{px} + \boldsymbol{tx}$ allows writing tax revenue $\boldsymbol{t}(\boldsymbol{x})\boldsymbol{x}$ as $y(\boldsymbol{x}) - \boldsymbol{px}$. Thus, tax revenue, or government expenditure, equals the value of total production minus the value of private expenditure. The government maximizes utilities subject to its budget constraint. Differentiating the Lagrangean $\mathcal{L} = u(y(\boldsymbol{x}), \boldsymbol{x}) + \lambda[(y(\boldsymbol{x}) - \boldsymbol{px} - g]$ with respect to one of the controls yields

$$\frac{\partial u}{\partial y}\frac{\partial y}{\partial x_i} + \frac{\partial u}{\partial x_i} + \lambda\left(\frac{\partial y}{\partial x_i} - p_i\right) = 0 . \tag{25}$$

Substituting from (24) and replacing p_i by $q_i - t_i$, this becomes

$$-\mu\frac{\partial y}{\partial x_i} + \mu q_i + \lambda\frac{\partial y}{\partial x_i} - \lambda q_i + \lambda t_i = 0 , \tag{26}$$

which is easily solved to yield the principal result:

$$t_i = \frac{\lambda - \mu}{\lambda}\left(q_i - \frac{\partial y}{\partial x_i}\right) . \tag{27}$$

8. Conclusion

In this paper, an alternative route to characterizing second-best commodity taxes has been developed. It has been shown that optimal tax rates depend directly on the impact of commodity demands on earnings. If a specific commodity like coffee induces the consumer to work harder, then this commodity should be taxed only moderately. Conversely, time-consuming activities should be taxed heavily. The intuition behind this result is straight-forward: As the government cannot tax time endowments directly, it should encourage demands which induce the consumers to work harder because this move increases earnings as the genuine tax base. Noting that actual tax systems often treat work-related expenditures generously, whereas alcoholic beverages or time consuming activities are often taxed at high rates, it appears that legislators around the world are aware of this fact.

The new tax rule (11) can be considered as a generalization of the celebrated Corlett-Hague rule. It differs from the latter in three respects: Firstly, the new rule holds for arbitrarily many commodities, not just in the two-commodity case. Secondly, optimal tax rates are characterized in terms of observable variables only, rather than in terms of unobservable compensated elasticities. And finally, the primal approach in the form invented here allows deriving optimal tax rates in a few lines, whereas the standard method follows a pretty roundabout course: It starts with characterizing optimal quantities using the dual approach, then obtains the Ramsey rule, and finally re-translates the result into primal terms in order to describe the optimal taxes.

The last section has shown that optimal commodity tax rates can be determined without ambiguity once the wage tax is known. Since data on earnings, expenditures and prices are readily available, and since a researcher can determine the optimal tax structure using (11), it remains to be hoped that the approach proposed here will stimulate empirical research in the field.

Appendix

In order to render the mathematics air-tight, it is most convenient to make the following assumptions both within the traditional approach and the new approach: The utility function is strictly decreasing in effort, strictly increasing in its other arguments, smooth, and strictly quasi-concave. Moreover, the analysis relies on an appropriate constraint qualification in order to have necessary optimality conditions, and rules out corner solutions. The latter assumption is not restrictive, because in case of corner solutions one can remove the commodities which are in zero demand, as they yield no tax revenue, and then optimize again, thus arriving at an interior solution.

Under the preceding assumptions, there exist differentiable ordinary demand functions $\boldsymbol{q}\to\boldsymbol{x}(\boldsymbol{q})$, mapping the positive orthant of n-dimensional Euclidian space into itself. Note that the domain is a convex set.

We also assume the so-called *Law of Demand*, which requires $\partial\boldsymbol{x}/\partial\boldsymbol{q}$ to be negative quasi-definite on its domain[12]. This premise is often made in general equilibrium theory to ensure uniqueness of market equilibrium[13]. We maintain that the Law of Demand is also useful for public economic theory and has not yet received sufficient attention in this realm. Let us first interpret it. The Law of Demand essentially rules out Giffen goods, as it states that prices and quantities move in opposite directions. This is because, for all $\boldsymbol{q}\neq\boldsymbol{q}^0$, negative quasi-definiteness implies $(\boldsymbol{q}-\boldsymbol{q}^0)\,\partial\boldsymbol{x}/\partial\boldsymbol{q}\,(\boldsymbol{q}-\boldsymbol{q}^0)<0$, hence $(\boldsymbol{x}(\boldsymbol{q})-\boldsymbol{x}(\boldsymbol{q}^0))(\boldsymbol{q}-\boldsymbol{q}^0)<0$, according to the mean value theorem.

Lemma: Under the Law of Demand, inverse demand functions $\boldsymbol{q}(\boldsymbol{x})$ exist, the offer curve $y(\boldsymbol{x})$ exists, and λ exceeds μ.

Proof: Existence of the inverse demand functions $\boldsymbol{q}(\boldsymbol{x})$ follows directly from the Gale-Nikaidô-Theorem, which states that a function defined on a convex set is invertible if its Jacobian is negative quasi-definite everywhere[14]. Existence of the offer curve is immediate from the identity $y(\boldsymbol{x})=\boldsymbol{q}(\boldsymbol{x})\boldsymbol{x}$. Finally, the relationship $\lambda<\mu$ is obtained from post-multiplying (3) by the tax vector, or from post-multiplying (7) by the consumption vector, observing that the resulting quadratic forms are strictly negative, whereas λ, μ and $\boldsymbol{tx}$ are strictly positive.

Without the Law of Demand, λ need not exceed μ and the usual economic interpretations get lost: If $\lambda<\mu$, for instance, Ramsey's rule (3) states that optimal commodity taxes *increase*, rather than diminish, all demands by a common factor – a perverse result stemming from the presence of Giffen goods. Likewise, the Law of Demand implies that the marginal cost of public funds exceed one[15], and it is equivalent to the "agent monotonicity" assumption usually made in models of non-linear taxation.

The above analysis, resting on the powerful Gale-Nikaidô-Theorem, is satisfactory but not necessarily comprehensible. Therefore, we may also outline a more traditional argument which proceeds from the first-order conditions associated with the individual optimization problem. Sticking to the one-commodity case which makes reading easier, and abbreviating derivatives by subscripts, the first-order conditions associated with the variables μ, y and x are:

---

12 A n×n matrix $\mathbf{A}$ is negative quasi-definite if $\mathbf{x}\,\mathbf{A}\mathbf{x}<\mathbf{0}$ for all $\mathbf{x}\neq\mathbf{0}$; it need not be symmetric.

13 See Hildenbrand (1983), for instance.

14 Cf. Gale and Nikaido (1965), Theorem 6. A good textbook reference is Sydsæter et al. (2005, p. 103).

15 Cf. also Auerbach and Hines (2002, p. 1387), who derive $\lambda>\mu$ from the premise that the tax base is not Giffen.

$$(28) \qquad \begin{aligned} & y - qx = 0, \\ & u_y(y,x) + \mu = 0, \\ & u_x(y,x) - \mu q = 0. \end{aligned}$$

Traditional theory sees this as a system of three equations determining the three endogenous variables μ, y and x at given prices q. With the conditions of the implicit function theorem fulfilled, there exists an ordinary demand function with derivative

$$(29) \qquad \frac{dx}{dq} = -\frac{\begin{vmatrix} 0 & 1 & -x \\ 1 & u_{yy} & 0 \\ -q & u_{xy} & -\mu \end{vmatrix}}{\begin{vmatrix} 0 & 1 & -q \\ 1 & u_{yy} & u_{yx} \\ -q & u_{xy} & u_{xx} \end{vmatrix}}.$$

The primal approach considers (28) as a system of three equations determining the three endogenous variables μ, y and q at given quantities x. Now, the implicit function theorem states that a differentiable inverse demand function $q(x)$ exist if the Jacobian formed with respect to μ, y and q happens to be non-singular. But this Jacobian, of course, is the matrix shown in the numerator of (29). Ruling out Giffen goods implies that its determinant does not vanish. Hence the desired implicit function $q(x)$ exists in fact, its derivative being just the reciprocal of (29).

Ramsey himself assumed vanishing cross-derivatives. Then, the offer curve becomes $u_y(y)y + u_x(x)x = 0$, as can be seen upon equating the last two equations in (28), multiplying by x and substituting from the first. As $u_y(y)y$ is strictly decreasing in y, each choice of x determines y uniquely, hence μ from the second equation in (28) and q from the third. This holds because separability rules out Giffen goods, as is well known.

References

Atkinson, A. B. and J. E. Stiglitz (1972) The Structure of Indirect Taxation and Economic Efficiency. *Journal of Public Economics* **1**, pp. 97–119.

Auerbach, A. J. (1985) The Theory of Excess Burden and Optimal Taxation. In: Auerbach, A. J. und M. Feldstein (Eds.) *Handbook of Public Economics, Vol. I.* Amsterdam: North-Holland.

Auerbach, A. J. and J. R. Hines (2002) Taxation and Economic Efficiency. In: Auerbach, A. J. and M. Feldstein (eds.) *Handbook of Public Economics, Vol. III.* Amsterdam: Elsevier.

Deaton, A. S. (1979) The Distance Function and Consumer Behaviour with Applications to Index Numbers and Optimal Taxation. *Review of Economic Studies* **46**, pp. 391–405.

Deaton, A. S. (1987) Econometric Issues for Tax Design in Developing Countries. In: Newbery, D. M. and N. H. Stern (Eds.) *The Theory of Taxation for Developing Countries*. New York: Oxford University Press.

Dupuit, A. (1844) De la mesure de l' utilité des travaux publics. *Annales des ponts et chausseés*. English in: Arrow, K. J. and T. Scitovsky (1969) *Readings in Welfare Economics*. Homewood: Irwin, pp. 255–283.

Gale, D. and H. Nikaido (1965) The Jacobian and Global Univalence of Mappings. *Mathematische Annalen* **159**, pp. 81–93.

Hatta, T. (2004) A Theory of Commodity Tax Reform under Revenue Constraint. *Japanese Economic Review* **55**, pp. 1–16.

Hildenbrand, W. (1983) On the "Law of Demand". *Econometrica* **51**, pp. 997–1019.

Kaplow, L. (2006) On the Undesirability of Commodity Taxation even when Income Taxation is not Optimal. *Journal of Public Economics (forthcoming)*.

Kay, J. A. (1980) The Deadweight Loss from a Tax System. *Journal of Public Economics* **13**, pp. 111–120.

Mas-Colell, A. et al. (1995) *Microeconomic Theory*. New York: Oxford University Press.

Mirrlees, J. A. (1976) Optimal Tax Theory – A Synthesis. *Journal of Public Economics* **6**, pp. 327–358.

Munk, K. J. (2002) What Determines the Optimal Commodity Tax Structure from an Intuitive Point of View? Working paper, University of Copenhagen.

Myles, G. D. (1995) *Public Economics*. Cambridge: Cambridge University Press.

Ramsey, F. P. (1927) A Contribution to the Theory of Taxation. *Economic Journal* **37**, pp. 47–61.

Samuelson, P. A. (1951) Theory of Optimal Taxation. Unpublished Memorandum for the U. S. Treasury. Published 1986 in: *Journal of Public Economics* **30**, pp. 137–143.

Sandmo, A. (1987) A Reinterpretation of Elasticity Formulae in Optimum Tax Theory. *Economica* **54**, pp. 89–96.

Sandmo, A. (1990) Tax Distortions and Household Production. *Oxford Economic Papers* **42**, pp. 78–90.

Stern, N. (1986) A Note on Commodity Taxation: The Choice of Variable and the Slutsky, Hessian and Antonelli Matrices (SHAM). *Review of Economic Studies* **53**, pp. 293–299.

Sydsæter, K., P. Hammond, A. Seierstad and A. Strøm (2005) *Further Mathematics for Economic Analysis*. Harlow, England: Pearson.