The optimal grouping of commodities for indirect taxation

Pascal Belan\textsuperscript{1} Stéphane Gauthier\textsuperscript{2} Guy Laroque\textsuperscript{3}

April 4, 2007

\textsuperscript{1}LEN (Université de Nantes) and EUREQUA (Paris I).
\textsuperscript{2}CREST and CREM (Université de Caen).
\textsuperscript{3}CREST-INSEE and GRECSTA. We are most thankful to Ian Crawford, from the Institute for Fiscal Studies, who computed the elasticities and budget shares in the U.K., which we use in the empirical application. We have benefited from the comments of Stuart Adam, Richard Blundell, Martin Browning, Jean-Marc Robin and Stephen Smith. We also thank the participants in the seminars at CREST and IFS. The usual disclaimer applies.
Abstract

Indirect taxes contribute to a sizeable part of government revenues around the world. Typically there are few different tax rates, and the goods are partitioned into classes associated with each rate. The present paper studies how to group the goods in these few classes. We take as given the number of tax rates and study the optimal aggregation (or classification) of commodities of the fiscal authority in a second best setup. The results are illustrated on data from the United Kingdom.

Les impôts indirects forment une part notable des recettes fiscales. D’ordinaire, on observe un petit nombre de taux différents, et les biens sont répartis en classes associées à chacun de ces taux. On étudie ici comment grouper les biens au mieux. Le nombre de taux est supposé fixé de manière exogène, et on résout le problème d’agrégation (ou de classement) optimal des biens dans un cadre de second rang. Les résultats sont illustrés sur des données britanniques.

Keywords: indirect tax, Ramsey, aggregation.
JEL classification numbers: H21, H23
1 Introduction

The French government recently wanted to change the rate of the value added tax bearing on meals taken in restaurants. But the European Union did not accept France’s demand. The standard theory of indirect taxation would possibly recommend to tax restaurants at a higher rate than fast food places, e.g., because rich households spend a larger fraction of their income in restaurants than the less well-off, who tend to have instead a greater use of fast food. This theory, however, does not take into account a strong constraint imposed on EU members: according to the 92/77 directive, EU members are allowed to set only one or two reduced (low) rates in addition to the standard (high) tax rate, so that they are forced to impose the same rate on many different commodities. The purpose of this paper is to describe how different commodities should be grouped when there is such a constraint on the number of tax rates.

The Ramsey tax rule usually assumes that different commodities can be taxed at different rates. Then, in a partial equilibrium framework, when the demand for a good only depends on its own price, each commodity can be assigned two numbers: the elasticity of its own demand with respect to price and its social weight, which reflects its relative usage among the poor and wealthy in the population. For a given social weight, the optimal tax rate is inversely proportional to the price elasticity; and, given the elasticity, the tax rate decreases with the social weight. When the consumers have the same tastes and when their labor supplies are separable from their demands for commodities, nonlinear income taxation yields a uniform taxation of all the goods (Atkinson and Stiglitz (1976) and Mirrlees (1976)). We study a situation where consumers have heterogeneous tastes (Saez (2002))

When a constraint on the number of available tax rates binds, it seems natural to lump together the goods with similar price elasticities, with the less price elastic groups supporting the largest tax rates. Similarly, one could also group goods whose social weights are close to each other. But the problem has received little attention in the literature. Gordon (1989) studies it in a tax reform perspective and concludes that many different rearrangements are socially improving. In the absence of a redistribution motive, Belan and Gauthier (2004) and Belan and Gauthier (2006) study the case of low levels of collected tax in a framework with a finite number of goods.

In this paper, we consider an economy with a continuum of goods, each of

In addition indirect taxes are not trivial for certain types of production functions (Stiglitz (1982), Naito (1999) or Saez (2004)), if it is possible to evade tax (Boadway, Marchand, and Pestieau (1994)), in order to correct externalities (Green and Sheshinski (1976)), in presence of uncertainties (Cremer and Galvani (1995)), or when the authority implementing direct taxes is not perfectly coordinated with the one that designs indirect taxes, possibly because the decisions are taken at different points in time or in space (federal, state or city levels).
them being negligible with respect to the total, so that it is possible to remove an elementary commodity from one group and to insert it into another group while leaving unchanged the whole tax structure.

In this setup, the optimal grouping structure can be characterized with the help of a *purported tax rate*, defined for any elementary good as the rate which maximizes social welfare without constraints, all the other rates being held fixed. More precisely, the purported rate is the rate that the social planner would apply to the good under consideration if this good could be taxed freely, while keeping unchanged the tax rates supported by the other commodities, fixed at their constrained optimum values. Then, under some reasonable conditions, at the constrained optimum the good is taxed at one of the (typically two) rates that are closest to its purported rate.

It should be emphasized, however, that the purported tax rate is linked to the price elasticity and social weight of the good at the putative free optimum, and not at the actual observed (constrained) tax rate. In practice, it is unlikely that reliable information be available on how elasticities change with prices. In the particular case where the price elasticities and social weights do not vary with the tax rates, the Ramsey monotonicity properties are shown to be satisfied: given the price elasticity, the tax rate is non increasing with the social weight, and similarly, given the social weight, the tax rate is typically non increasing with the price elasticity.

In order to apply these results to data from the United Kingdom, the argument is adapted to the case in which cross price effects are not zero. Assuming that the observed rates on the existing groups are optimal yields constraints on the implicit redistributive aims of the government. The social weights that best fit the current tax scheme put most of the weight on the population segment associated with the fourth and fifth deciles of the consumption distribution. For these social weights, the actual commodity groupings do not look far from optimality. The main departures are Petrol and diesel and Beer, which should be much less taxed from an equity view point, but it is likely that other considerations (environment, public health) matter in such cases. In the U.K., Food out, which comprises restaurants and fast food places, is currently taxed at the standard rate, but appears to be too heavily taxed; our analysis actually would recommend to exempt both restaurants and fast food from any tax.

The paper is organized as follows. The bulk of the analysis assumes separability between goods. The general framework is laid out in the next section. Then the standard first-order conditions for optimality are presented when there are no constraints on the number of tax rates. Section 4 derives necessary conditions for optimality when a small given number of tax rates is allowed. The following section studies in some detail the case of constant elasticities. Section 6 extends the results to the case in which cross price effects matter. An application to data
from the United Kingdom is presented in the final section. In the UK, if not in France, it appears that restaurants should be exempted from value added tax. Technical proofs are gathered in the Appendix.

2 Consumers

There are a continuum of goods $g$ in the economy, $g$ in $G$, and a numeraire. The typical consumer in the economy is designated with an index $c$ in $C$, and her tastes are represented by an additively separable utility function. Consumer $c$ maximizes her utility function

$$\int_{G} u(x_g, g, c) \mu(g) dg + m$$

under her budget constraint

$$\int_{G} (1 + t_g) x_g \mu(g) dg + m = w_c.$$

The utility function $u$, defined over $\mathbb{R}_+ \times G \times C$, is assumed to be concave and twice continuously differentiable with respect to consumption $x_g$, $x_g$ in $\mathbb{R}_+$, and continuous with respect to the good $g$ and consumer $c$ characteristics. The sets $C$ and $G$ are subsets of some Euclidean space. The consumption of numéraire is denoted by $m$.

The relative importance of the various commodities $g$ is partially captured by their density $\mu(g)$ with respect to the Lebesgue measure. It should be emphasized that all commodities are small, in the sense that their measure is absolutely continuous with respect to Lebesgue, without mass points (see Belan and Gauthier (2006) for an analysis when there is a finite number of goods). The units of commodities are chosen so that all producer prices equal 1. Commodities are taxed linearly and the tax rate supported by commodity $g$ is denoted $t_g$ ($t_g$ is a number larger than $-1$); when $t_g$ is negative, the good in fact is subsidized. Finally, $w_c$ is the exogenous income of consumer $c$.

The separability assumptions imply that the overall consumer problem is equivalent to separate maximizations of $u(x_g, g, c) - (1 + t_g) x_g$ with respect to each quantity $x_g$ of good $g$, with $m$ determined by the budget constraint. Under the usual Inada conditions, the demand $\xi_g(t_g, c)$ of commodity $g$ by consumer $c$ is the unique solution of the first-order condition $u'_x(x, g, c) = 1 + t_g$. It is decreasing and continuously differentiable with respect to the tax rate. The indirect utility from consuming a good $g$ taxed at rate $t_g$, is

$$v_g(t_g, c) = u(\xi_g(t_g, c), g, c) - (1 + t_g) \xi_g(t_g, c),$$

and therefore the overall indirect utility of consumer $c$ is equal to

$$\int_{G} v_g(t_g, c) \mu(g) dg + w_c.$$
3 Optimal tax schedules

The aggregate quantities, summed over the set of consumers, are denoted with capital letters. The aggregate demand for good $g$ is

$$X_g(t_g) = \int_C \xi_g(t_g, c)d\nu(c),$$

where $\nu$ is the (probability) measure describing the distribution of consumer characteristics on the set $C$.

When choosing indirect taxes, the government takes as given market behavior. It seeks to maximize the sum of the utilities of the consumers in the economy, weighted by some \textit{a priori} weights $\alpha_c$, $\alpha_c \geq 0$ for all $c$, normalized so that

$$\int_C \alpha_c d\nu(c) = 1.$$

Using separability, the objective of the government can be written as the sum

$$\int_{G} V_g(t_g) \mu(g) dg,$$

where

$$V_g(t_g) = \int_C \alpha_c v_g(t_g, c)d\nu(c).$$

We shall often use the derivative of $V_g$ with respect to the tax rate,

$$\frac{dV_g}{dt_g}(t) = -a_g(t)X_g(t).$$

In this expression,

$$a_g(t) = \int_C \alpha_c \frac{\xi_g(t, c)}{X_g(t)} d\nu(c)$$

is a positive number which measures the \textit{social weight of good} $g$. Namely, it is large when the agents $c$ with the largest weights $\alpha_c$ consume relatively more of the good.

If there is no constraint on rates setting, when fiscal income to be collected is $R$, the government maximization problem can be written as

$$\max_t \int_{G} V_g(t_g) \mu(g) dg$$

under the budget constraint

$$\int_{G} t_g X_g(t_g) \mu(g) dg = R.$$
where the choice variable $t$ is the collection of tax rates $(t_g)$, $g$ in $G$.

Let $\lambda$ denote the multiplier associated with the budget constraint. The problem is equivalent to maximizing

$$\int_G \mathcal{L}_g(t_g)\mu(g)dg,$$

where the Lagrangian, the contribution of good $g$, after division by $\mu(g)$, to the welfare objective is equal to

$$\mathcal{L}_g(t_g) = V_g(t_g) + \lambda t_g X_g(t_g).$$

Under regularity conditions, at the optimum, one can interpret $\lambda$ as the marginal cost of public funds. If the authority freely chooses the tax rate bearing on good $g$, the necessary first-order condition for an interior optimum is

$$-a_g(t)X_g(t) + \lambda \left( X_g(t) + tX'_g(t) \right) = 0,$$

or, dropping the index $g$ to simplify notations,

$$\frac{t}{1 + t} = \frac{\lambda - a}{\lambda} \frac{X}{-(1 + t)X'}.$$

This is the celebrated Ramsey rule, in which the tax rate applying to a consumption good is inversely related to the price elasticity $-(1 + t)X'/X$ of the demand for this good.\(^2\)

Note that the program is not well behaved, since the maximand is convex in $t_g$. The following assumption helps to put in perspective the use of the first-order condition:

**Assumption 1** A good $g$ satisfies the single peaked assumption, given the marginal cost of public funds $\lambda$, when the function $\mathcal{L}_g$, defined on $(-1, +\infty)$, satisfies one of the following three properties:

1. It is increasing;
2. It is increasing from $-1$ to some $\tau_g(\lambda)$ and decreasing from then on;
3. It is decreasing.

\(^2\)Indirect taxation appears to be useless when $\lambda = a_g$ for all consumption goods $g$, a condition unlikely to be satisfied when the agents do not have the same tastes, as emphasized in the recent literature, e.g. in Saez (2002).
In the normal situation of Assumption 1.2, the Ramsey first-order condition has a unique solution which characterizes the optimum. The analysis is easily extended when the solution goes to the boundaries of the tax domain: under Assumption 1.1 (resp., Assumption 1.3), the optimal tax rate is equal to $+\infty$: the good is made infinitely expensive (resp., to -1: the good is made free).

One can identify a number of circumstances where the single peaked assumption holds. To this aim, note first that the derivative $L' = -aX + \lambda(X + tX')$ has the same sign as $(\lambda - a)/\lambda + tX'/X$. Thus, in the absence of redistributive motive, i.e., $a$ is equal to 1, whenever the elasticity of aggregate demand with respect to the tax rate, $tX'/X$, is non increasing in $t$, $L'$ at most has one change of sign, and so $L$ is single peaked. This is the case, in particular, when the price elasticity of demand is constant, $X = A(1 + t)^{-\varepsilon}$, since $tX'/X = -\varepsilon t/(1 + t)$ is then decreasing.

When there is a redistribution motive, $a$ typically varies with $t$, and single peakedness holds if, in addition of a decreasing tax elasticity of demand, $a$ increases with $t$. That is, the larger the tax rate, the larger the social weight of the good, i.e. the demand for the good of the socially unfavored (rich) agents relatively decreases in comparison to that of the socially favored (poor) agents.

### 4 Tax rule with a finite number of rates

When institutional constraints impose the use of a fixed and finite number of different tax rates, the Ramsey rule of the preceding section cannot be applied as is. We study the properties of the optimum when there is an a priori given finite number $K$ of different tax rates. We note these tax rates $t_k$, $k = 1, \ldots, K$, and we assume that they are ranked in increasing order, $t_k \leq t_{k+1}$ for all $k$. In some cases, we add the additional constraint that one of the tax rates is equal to zero, but this feature is inessential for most of the analysis. Let $G_k$ be the subset of goods which are taxed at rate $t_k$ and $G$ the collection of $G_k$. The government program becomes:

\[
\begin{align*}
\max_{(t_k, G_k)_{k=1}^K} & \sum_{k=1}^K \int_{G_k} V_g(t_k) \mu(g) dg \\
\sum_{k=1}^K \int_{G_k} t_k X_g(t_k) \mu(g) dg &= R \\
\bigcup_{k=1}^K G_k &= G.
\end{align*}
\]

\(3\) Of course, demand functions do not always satisfy Assumption 1, for instance, the property is not preserved under aggregation, since the sum of single peaked functions generally is not single peaked (while concavity is preserved by summation). In such cases, the Ramsey rule may not be relevant. See Section 5 for a further discussion on this point.
The government now has to choose the $K$ tax rates (or possibly $K - 1$, if one of them is constrained to be equal to zero) and the partition of the set of commodities associated with the various tax rates. Formally, this is a more complicated problem than the Ramsey problem, since it involves the variables $G$, to which the standard Lagrangian methods do not immediately apply.

### 4.1 Optimal tax rates for a given partition of the goods

Given the partition $G$, however, the problem is standard. Under usual regularity conditions, one can write the Lagrangian and the multiplier $\lambda$ associated with the government budget constraint is equal to the derivative of the objective function with respect to $R$. When differentiating with respect to the tax rates, to get the analogue of the Ramsey rule, it is natural to consider the aggregate commodity $G_k$, the demand of which is defined as

$$X_{G_k}(t) = \int_{G_k} X_g(t)\mu(g)dg.$$  

The necessary first-order condition associated with $t_k$, first derived in Diamond (1973), is:

$$\int_{G_k} \left[-a_g(t_k)X_g(t_k) + \lambda (X_g(t_k) + t_k X_g'(t_k))\right] \mu(g)dg = 0,$$

or

$$t_k \frac{X_{G_k}}{1 + t_k} = \frac{\lambda - a_{G_k}}{\lambda} \frac{X_{G_k}}{-(1 + t_k)X_{G_k}'} = \frac{\lambda - a_{G_k}}{\lambda} \frac{1}{\varepsilon_{G_k}}.$$  

The social weight $a_{G_k}$ of the aggregate good is the average of the social weights of the individual commodities,

$$a_{G_k} = \int_{G_k} \frac{X_g(t_k)}{X_{G_k}(t_k)}a_g(t_k)\mu(g)dg.$$  

The price elasticity $\varepsilon_{G_k}$ of the demand for the aggregate good $G_k$ is a weighted sum of the elementary price elasticities of the goods $g$ in the group,

$$\varepsilon_{G_k} = \int_{G_k} \frac{X_g(t_k)}{X_{G_k}(t_k)}\varepsilon_g(t_k)\mu(g)dg,$$

the weights being proportional to the quantities consumed $X_g$ of good $g$ multiplied by the density $\mu(g)$ of the good.

### 4.2 Optimal partition of the goods

The aim of the paper is to study the optimal partition $G$ of the goods. At the optimum, the goods are taxed at one of the $K$ optimal tax rates, and there is
an associated marginal cost of public funds $\lambda$. This cost obviously depends on institutional constraints: it is different from the marginal cost of public funds that would prevail if all goods were taxed freely. Consider an individual commodity $g$, small with respect to the whole economy. Under the continuum hypothesis, a change in its tax rate leaves $\lambda$ unchanged, and we have (the proof is in the Appendix):

**Theorem 1** A necessary condition for optimality is that, for almost every good $g$, good $g$ be attached to a group $k$ such that

$$\mathcal{L}_g(t_k) = \max_{h=1,...,K} \mathcal{L}_g(t_h).$$ (4)

Let $t^R_g$ be the tax rate that this good would support in the hypothetical situation where it would be taxed individually. For an interior solution, $t^R_g$ satisfies the Ramsey rule (1). A direct consequence of the single peakedness of the Lagrangian is

**Lemma 1** Under Assumption 1, at the optimum,

1. If $\mathcal{L}_g$ is increasing, good $g$ belongs to the more heavily taxed group $K$; if it is decreasing, it belongs to the less taxed group.

2. Otherwise, with $t^R_g$ the tax rate that maximizes $\mathcal{L}_g$,

   (a) if $t^R_g$ is larger than $t_K$, commodity $g$ supports the maximal rate;

   (b) if there exists $k, k < K$, such that $t_k \leq t^R_g \leq t_{k+1}$, then $g$ is taxed either at rate $t_k$ or at rate $t_{k+1}$;

   (c) if $t^R_g$ is less than $t_1$, $g$ is taxed at rate $t_1$.

This lemma helps to describe some features of the optimal groups of commodities. Define the demand elasticity $\varepsilon^R_g = \varepsilon_g(t^R_g)$ and the social weight $a^R_g = a_g(t^R_g)$ at the peak. When only efficiency matters ($a_g$ is identically equal to one for all $g$), the monotonicity of the Ramsey formula (1) in elasticities directly allows to translate Lemma 1 into:

**Theorem 2** At an optimum, in the absence of redistribution motive, if the Ramsey price elasticity of good $g$, $\varepsilon^R_g$, is smaller than $\varepsilon(G_K)$, good $g$ is taxed at the maximal rate $t_K$. If $\varepsilon^R_g$ is larger than $\varepsilon_{G_1}$, good $g$ is untaxed. Otherwise, $g$ is taxed at one of the $k$ or $k + 1$ rates such that

$$\varepsilon_{G_k} \geq \varepsilon^R_g \geq \varepsilon_{G_{k+1}}.$$
When the government has a redistributive objective, the social weights of the commodities typically differ from one. In the plan \((\varepsilon, a/\lambda)\), when the representative point \((\varepsilon_g, a_g/\lambda)\) of good \(g\) belongs to the cone delimited by the two half lines

\[
\frac{a}{\lambda} = 1 - \frac{t_k}{1 + t_k} \varepsilon \quad \text{and} \quad \frac{a}{\lambda} = 1 - \frac{t_{k+1}}{1 + t_{k+1}} \varepsilon,
\]

Lemma 1 and the first-order condition (1) imply that it should be taxed at one of the two rates \(t_k\) or \(t_{k+1}\) that correspond to the boundaries of this region (in the top left region, it is subsidized at the most favorable rate; in the bottom left, it is taxed at the maximal rate).

Figure 1, where there are four rates, \(t = -0.1, 0.2, 0.3\) and 0.4, highlights the relative roles of the social weight and price elasticity of the commodity. When there are exempted goods (the zero rate is allowed), a good is subsidized (or taxed) when its social weight \(a_g\) is larger (or smaller) than the marginal cost of public funds \(\lambda\). For \(a_g\) larger than \(\lambda\), smaller price elasticities tend to be associated with larger subsidies. For \(a_g\) smaller than \(\lambda\), smaller price elasticities tend to be associated with larger tax rates. Nevertheless, without further assumptions, we cannot say more on the shape of the regions in the \((\varepsilon, a/\lambda)\) space corresponding to a given tax rate.

Remark 1. Lemma 1 and Theorem 2 crucially rely on Assumption 1: if the Lagrangian function \(L_g\) has several peaks, the results may not hold. An example is depicted on Figure 2, which represents the demand function for the union of two goods with constant price elasticity: \(X = A_1(1 + t)^{-\varepsilon_1} + A_2(1 + t)^{-\varepsilon_2}\), with
Figure 2: A Lagrangian function with two peaks

\[ \lambda = 1.25, \quad \varepsilon_1 = 0.25, \quad A_1 = 1.8, \quad \varepsilon_2 = 4 \quad \text{and} \quad A_2 = 1. \] Then, the tax rate \( t^{R}_g \) which maximizes \( \mathcal{L}_g \) is between \( t_2 \) and \( t_3 \), but the optimal rate is \( t_1 \).

5 The case of constant elasticities

It is possible to derive a more precise characterization when all the consumers’ demands have the same constant price elasticities for each good. The utility functions which yield demand functions whose price elasticities are constant are of the form

\[
    u(x, g, c) = \begin{cases} 
        \left[ A_g(c) \right]^{1/\varepsilon_g(c)} \frac{x^{1-1/\varepsilon_g(c)}}{1-1/\varepsilon_g(c)} & \text{for } \varepsilon_g(c) > 0, \ \varepsilon_g(c) \neq 1 \\
        A_g(c) \ln \frac{x}{A_g(c)} & \text{for } \varepsilon_g(c) = 1.
    \end{cases}
\]

The associated indirect utility functions are

\[
    v_g(t, c) = \begin{cases} 
        \frac{A_g(c)}{\varepsilon_g(c) - 1} (1 + t)^{1-\varepsilon_g(c)} & \text{for } \varepsilon_g(c) > 0, \ \varepsilon_g(c) \neq 1 \\
        -A_g(c) - A_g(c) \ln(1 + t) & \text{for } \varepsilon_g(c) = 1.
    \end{cases}
\]

For any individual \( c \) and good \( g \), for all non negative \( \lambda \), the function \( v_g(t, c) + \lambda t \xi_g(t, c) \) is single peaked; a sufficient assumption for the aggregate Lagrangian

\[
    \mathcal{L}_g(t) = \int_{C} (\alpha(c)v_g(t, c) + \lambda t \xi_g(t, c)) \, d\nu(c)
\]
to be also single peaked, an assumption which is maintained in the rest of this Section, is that the price elasticities are identical across consumers, i.e., $\varepsilon_g(c) = \varepsilon_g$. Then, the contribution of commodity $g$ to social welfare is

$$L_g(t) = \begin{cases} A_g \frac{(1 + t)}{\varepsilon_g - 1} [a_g(1 + t) + \lambda t (\varepsilon_g - 1)] & \text{for } \varepsilon_g > 0, \varepsilon_g \neq 1 \\ A_g \left[ a_g [-\ln (1 + t) - 1] + \frac{\lambda t}{1 + t} \right] & \text{for } \varepsilon_g = 1 \end{cases}$$

(5)

where $A_g = \int_c A_g(c) d\nu(c)$. For this specification, one can obtain a precise description of the optimal classification of goods in the different tax groups. Indeed, the inequality $L_g(t') > L_g(t)$ is equivalent to $a_g/\lambda < \phi(\varepsilon_g, t, t')$, where the function $\phi$, whose analytic expression is given in the Appendix, has the following properties

**Lemma 2** The function $\phi(., t, t')$ has the following two properties:

1. It is convex on $(0, +\infty)$;

2. When $\varepsilon$ goes to infinity, for $t' > t$, it is asymptote to $(1 - \varepsilon)t/(1 + t)$.

By Theorem 2 in the $(\varepsilon, a)$ plan, when the point $(\varepsilon_g, a_g)$ belongs to the cone delimited by the two straight (broken on Figures 3 and 4) half lines originating at the point of ordinate $\lambda$ on the $a$ axis and of slopes $-\lambda t_k/(1 + t_k)$ and $-\lambda t_{k+1}/(1 + t_{k+1})$, good $g$ is taxed either at rate $t_k$ or at rate $t_{k+1}$. By Lemma 2, inside this cone, there is a curve of equation $a/\lambda = \phi(\varepsilon, t_k, t_{k+1})$ (this is a thick solid line on the figures) which partitions the cone into two regions: the goods above the curve are taxed at the lower rate, those below the curve are taxed at the upper rate. We therefore have a full characterization of the optimal goods classification.

Figures 3 and 4 make clear that, given the price elasticity, tax rates are always decreasing with the social weight. Moreover, given the social weight, the tax (resp. subsidy) rates decrease (resp. increase) with the price elasticity, provided that some goods can be exempted, i.e., 0 is among the tax rates. When 0 is not among the tax rates, the optimal rate is not necessarily monotone in elasticity: in Figure 4, for instance, there is a finite interval of elasticities $(\varepsilon', \varepsilon'')$ for which the goods are subsidized, while the goods in $(0, \varepsilon')$ and in $(\varepsilon'', +\infty)$ are taxed.

### 6 Non separability

So far, it has been assumed that consumers’ preferences were separable. To bring theory closer to the data, one must dispense with this assumption and possibly, in addition, introduce labor supply together with direct taxes.

---

4In Figure 3, the three tax rates are $t_1 = 0, t_2 = 0.2$ and $t_3 = 0.4$, while in Figure 4, they are equal to -0.1, 0.3 and 0.4 (and $\lambda$ is equal to 1.5).
Figure 3: The efficient tax structure

Figure 4: Taxes may not be decreasing with elasticity
Let the tastes of agent $c$ be represented with the utility function $U(x, L_c, c)$, where $x$ describes the consumption of goods, a measurable mapping from the set of commodities $G$ into $\mathbb{R}_+$, and we work conditionally on the labor supply $L_c$. The budget constraint of the typical consumer is:

$$\int_G (1 + t_g)x_g\mu(g)dg = Y_c,$$

where $Y_c$ is after tax income, i.e. $Y_c = w_cL_c - T(w_cL_c)$.

Let $t$ be the collection of tax rates ($t_g$) for $g$ in $G$. For a given labor supply $L_c$ and after tax income $Y_c$, the conditional indirect utility function $V(t, L_c, Y_c, c)$ of consumer $c$ depends on the tax rates $t$ and on $(L_c, Y_c)$. Under regularity conditions, if $\rho_c$ is the marginal utility of income of consumer $c$ (the Lagrange multiplier associated with her budget constraint), an application of the envelope theorem yields:

$$\frac{\partial V}{\partial t_g} = -\rho_c \xi_g(t, L_c, Y_c, c),$$

where $\xi_g(t, L_c, Y_c, c)$ is her conditional (Marshallian) demand for good $g$.

The government chooses $t$ which maximize

$$\int_C \alpha_cV(t, L_c, Y_c, c)d\nu(c)$$

subject to the budget constraint

$$\int_C \int_G t_g\xi_g(t, L_c, Y_c, c)\mu(g)dg \ d\nu(c) = R.$$

With $\lambda$ the multiplier of the budget constraint, the government maximizes the social objective $L(t)$, equal to

$$\int_C \alpha_cV(t, L_c, Y_c, c)d\nu(c) + \lambda \left\{ \int_C \int_G t_g\xi_g(t, L_c, Y_c, c)\mu(g)dg \ d\nu(c) - R \right\}.$$

The Diamond first-order condition associated with the tax rate $t_k$ of group $k$ is (see Appendix C)

$$\int_{g \in G_k} \left\{ (-a_g + \lambda)X_g + \lambda \left[ t_g \frac{\partial X_g}{\partial t_g} + \int_{g' \neq g} t_{g'} \frac{\partial X_{g'}}{\partial t_g} \right] \right\} \mu(g)dg = 0, \quad (6)$$

where the social weight of good $g$

$$a_g = \int_C \frac{\xi_g}{X_g} \alpha_c\rho_c d\nu(c) \quad (7)$$

now depends on $\rho_c$. 

13
Similarly, if good $g$ could be taxed freely, the first-order condition associated with an interior maximum for the individual tax rate $t^R_g$ is

$$(-a_g + \lambda)X_g + \lambda \left[ t^R_g \frac{\partial X_g}{\partial t_g} + \sum_{k=1}^K t_k \frac{\partial X_{G_k \setminus \{g\}}}{\partial t_g} \right] = 0. \quad (8)$$

The analysis of Section 4.2 then can be adapted to this more general setup, provided Assumption 1 applies to the function $L$, seen as a function of each $t_g$ separately. Namely, rewriting the first-order condition (8) as

$$\frac{a_g}{\lambda} - b_g = 1 - \frac{t}{1 + t} \varepsilon_g, \quad (9)$$

where

$$b_g = \frac{1}{X_g} \sum_{k=1}^K t_k \frac{\partial X_{G_k \setminus \{g\}}}{\partial t_g}, \quad (10)$$

one can draw Figure 1 in the plan $(\varepsilon, a/\lambda - b)$ with a similar interpretation, provided that $L$ is single peaked (so that the first-order condition characterizes the optimal tax rate for good $g$ all other rates being kept fixed) and that the quantities in (9), $(a_g, \varepsilon_g, b_g)$, all are evaluated at the optimal tax rates solution to (6).

7 Illustration with data from the United Kingdom

Professor Ian Crawford, from the Institute for Fiscal Studies, has provided us with uncompensated cross price elasticities for consumption in the UK, grouped into twenty categories, homogenous by tax rates, computed along the lines initiated by Blundell and Robin (1999), and with the budget shares by deciles of consumption expenditures in the population (the data is reproduced at the end of the Appendix). A large part of consumption, 49%, is subject to the ‘standard’ (17.5%) tax rate, and a substantial part, 27%, necessities including basic food, is either exempted or taxed at a zero rate. Our data do not separate exempted from zero rate items, and we treat the whole category as zero rated. Domestic fuel, 10% of consumption, is taxed at the ‘reduced’ (5%) rate. Tobacco, alcohol, and petrol and diesel bear large excise tax rates.

Given plausible redistributive aims of the UK government, represented by weights on the ten population deciles, we would like to see whether the actual grouping of commodities fits with the theory developed above.

$^5$In the analysis, we dropped children clothing, which represents less than 1% of aggregate consumption expenditure, because the estimated price elasticities are somewhat out of the ball park.
We proceed as follows. We assume that the tax authority takes as given after tax incomes, as well as the partition of commodities into groups subject to a common VAT rate. If VAT rates are optimally chosen, they satisfy the Diamond first order conditions, which provide some information on the social weights. We compute the individual purported rates and draw the commodity fan given these weights, which allow to assess the optimality of the composition of the commodity groups.

Note that if the Atkinson Stiglitz conditions hold (preferences are separable between commodities and labor, and the preferences for commodities are identical across individuals at the microeconomic level), when the government can freely tax incomes in a non linear way, we know that all the goods should be taxed at the same rate (see Atkinson and Stiglitz (1976), Kaplow (2006), or Laroque (2005)). To the best of our knowledge, there is no general agreement on the empirical relevance of the Atkinson Stiglitz conditions. Browning and Meghir (1991) find some evidence of non separability. Our exercise only makes sense if some of these assumptions fail to hold.

7.1 The government redistributive objectives

Since our data consists of consumption shares and elasticities by deciles, the redistributive stance of the government is represented by a vector of non negative weights associated with the ten population deciles, whose coordinates sum up to 1.

For consistency, these weights should be such that the Diamond first-order conditions (6) for the basic three commodity groups, exempted, reduced rate and standard rate, are satisfied. The tax rates on alcohol and tobacco on one side, petrol on the other, are likely to depend on other considerations than mere redistribution (here, public health or environmental issues), and therefore, in our opinion, do not give direct information on the redistributive stance. We keep ‘Public transport’ taxed at the standard rate, even though there are numerous employers subsidies towards its cost. All in all, we have four (in practice, linear) equations, three first order conditions and the normalization condition, that must be satisfied by the ten non negative weights and the marginal cost of public funds.
More precisely, given the observed tax rates, budget shares and price elasticities, the unknowns in (6) are the non negative social weights $\alpha_c \rho_c$ and marginal cost of public funds $\lambda$, where $a_g$ is linked to the social weights by (7). The equations are homogenous of degree one in $(\alpha_c \rho_c, \lambda)$, so that we only can recover the ratios $\alpha_c \rho_c / \lambda$.

The Diamond first order conditions do impose restrictions on the problem, but one cannot expect in general to recover the government objective from them. Indeed, if there is some interior (strictly positive) solution to the equations, the set of solutions is locally a manifold of dimension $11 - 4 = 7$. Still, the minimum of the squares of the three left-hand sides of the Diamond conditions is different from zero, and is obtained on the boundary of the simplex. This here gives a unique set of values for the ratios $(\alpha_c \rho_c / \lambda)$. We normalize the sum of $\alpha_c \rho_c$ over the deciles to unity, and compute $\lambda$ accordingly. The normalization yields an implicit choice of units: an increase of aggregate consumption of $dC$, uniformly distributed, gives $dC/10$ to each decile and therefore, for this choice of normalization, increases social welfare by $dC/10$: Social welfare is measured in tenths of aggregate consumption. This procedure gives $\lambda = 1.11$,

and puts most of the weight on the fourth and fifth deciles

$$\alpha_1 \rho_1 = 0.03 \quad \alpha_4 \rho_4 = 0.54 \quad \alpha_5 \rho_5 = 0.43,$$

with all the others equal to zero. The left-hand sides of the Diamond conditions are respectively equal to 0.003 for the exempted goods, -0.007 for domestic fuel.

6In practice, we have to rewrite (6), given the available statistics. There is a finite number of commodities, so that the equality becomes

$$\sum_{g \in G_k} \left\{ (-a_g + \lambda) X_g + \lambda \sum_{g' \in G} t_{g'} \frac{\partial X_{g'}}{\partial \theta} \right\} = 0.$$

The consumption, rather than production, price is the numeraire. Using tildas for the variables measured with the new numeraire:

$$\tilde{X}_g = (1 + t_g) X_g,$$

and

$$\frac{1}{1 + t_g} = 1 - \tilde{t}_g.$$

After some manipulations, with appropriate definition of the aggregates, (6) becomes

$$(-a_{G_k} + \lambda)(1 - \tilde{t}_k) \tilde{X}_{G_k} + \lambda \sum_{G_k} \tilde{t}_k \tilde{X}_{G_k} \tilde{e}_{G_k} a_k = 0.$$

Finally we work in shares of total consumption, dividing the equalities by total consumption.

7The social welfare function is normalized so that an equal lump-sum transfer of 1 monetary unit to each decile increases welfare by 1 unit.
(the only good taxed at the reduced rate), and to -0.0004 for goods taxed at the standard rate. These numbers are proportional, up to a positive factor, to the derivatives of the social objective \( \mathcal{L}(t) \) with respect to the corresponding tax rates. They are equal to the social values of marginal changes of the tax rates, measured as tenths of aggregate consumption. For instance, increasing by 1 point the standard rate, from 17.5% to 18.5%, would induce a social loss of \( 0.0004 \times 0.01 \times 10 = 0.004 \) of aggregate consumption. The Diamond first-order conditions therefore are close to be satisfied.

### 7.2 Is the grouping of commodities optimal?

Figure 5 allows to discuss the optimality of the commodity grouping, building on the previous results in the paper. It plots the discounted adjusted social weight \( \frac{a}{\lambda} - b \) as a function of the own price elasticity. From (9), the first-order condition for an individual commodity (supposed to be small enough, so as to have no influence on the marginal cost of the public fund) is

\[
\frac{a}{\lambda} - b = 1 - \tilde{t} \varepsilon,
\]
where $\tilde{t} = t/(1 + t)$ is the unknown optimal tax rate, computed as a fraction of the consumption price. We have set $a$, $b$ and $\varepsilon$ at their current observed values\(^8\) so that $\tilde{t}$ is the solution of a linear equation, read directly as a slope on the graph. The graph in the plan $(\varepsilon, a/\lambda - b)$ shows the half lines corresponding to the current tax rates. The representative points of eighteen\(^9\) commodities are also shown: in fact two points are drawn for each good, one in large bold type corresponds to the implicit social weights computed above, the other one in small italic type represents the good location for a Rawlsian government which would put all the social weight on the first population decile.

If Assumption 1 holds, optimality requires that the large bold representative points of all the exempted goods be above the reduced rate half line, the point associated with ‘Domestic Fuels’ (the only good supporting the reduced rate) be between the standard rate line and the horizontal, and all the goods bearing the standard rate be below the reduced rate half line. Excluding the goods subject to excise taxes, 87\% of total consumption expenditures are concerned. Of these 67\% appear to be taxed consistently with the optimality criterion\(^{10}\). The main departures from optimality are the following. A number of exempted goods should be taxed at the standard rate: (some of) ‘Dairy products’, ‘Fruits and Vegetables’, and ‘Other non VAT foods’\(^{11}\) (Some of) ‘Food out’ and ‘Public transport’, currently taxed at the standard rate, should be exempted\(^{12}\). At least in the U.K., if not in France, restaurants are too heavily taxed.

If the government wants to raise more money by creating a larger tax rate, (some of) ‘Adult Clothing’ and ‘Leisure Goods’ seem to be good candidates to

---

\(^8\)We have done some experimentation with more sophisticated computations for $b$ and $\varepsilon$, using (10) and a linear approximation of the equation around the observed point. In particular we have looked at cases where all the elasticities are constant, equal to their observed values, where demand functions are linear, and at a couple of other variants, including QAIDS which underlies the empirical estimation. Unsurprisingly, the results are quite sensitive to the specification of the shape of the demand functions: in particular single peakedness is easily lost, and the Lagrangian may be locally convex at the observed point. More work is needed in this area.

\(^9\)Theory would require partial optimization to compute the first-order conditions at the optimal tax rate for each considered commodity: for lack of better information, we assume that the observed elasticities are good enough approximations to be used to compute the graph coordinates. In the interest of readability, wine and spirits do not appear on the graph: its own price elasticity is (-)3, much larger than that of the other goods.

\(^{10}\)It is difficult to provide a statistical assessment of this result given the possible measurement errors on elasticities. If one chooses, as a simple benchmark, the case in which the representative point of each class of goods were drawn independently uniformly in the half plan, given the half lines associated with each tax rate, then only 43\% of the 87\% consumption expenditures taken into account would be taxed according to the theory. This can be seen as a weak validation of the theory.

\(^{11}\)In practice, we do not have access to data on a continuum of goods, contrary to the setup used in the theoretical part of the paper. We work on (small) aggregates of ‘elementary’, probably diverse, commodities. The qualifier ‘some of’ designates some of the elementary commodities composing the aggregate.

\(^{12}\)Both ‘Food out’ and ‘Public transport’ are complementary with labor supply.
enter its basis.

Finally, four specific categories appear to be taxed more heavily than the redistributive social objective would recommend. They are ‘Domestic Fuels’, the only one taxed at the reduced rate, and ‘Beer’, ‘Petrol and Diesel’ and ‘Tobacco’, which are subject to excise taxes. This may be justified on public health or environmental protection grounds. The differences are large: for instance beer would otherwise be either exempted or taxed at a lower rate, and ‘Domestic fuels’ would be strongly subsidized.\footnote{We have not reported the representative point for ‘Wine and spirits’ on the graph, because of its large own price elasticity. ‘Wine and spirits’ are subject to a 55\% rate, while according to our computation they should support a 36\% rate.}

It is also of interest to look at the impact of the redistributive stance of the government on the diagram. Going from the social objective just discussed to a Rawlsian government, whose only concern is the first decile, tends to spread out the figure. A quarter of consumption, ‘Petrol and Diesel’, ‘Food out’, ‘Adult Clothing’ and ‘Leisure Goods’, are taxed more heavily. A third, ‘Household Goods and Services’, ‘Leisure Services’ and ‘Tobacco’, are unaffected. The remainder of consumption, approximately 45\%, gets a reduced rate or, more often a subsidy. This rather surprising outcome indicates that differences in the consumption structure of the various deciles are large enough to make the optimal indirect tax rates vary substantially with the redistributive objective, at least in the absence of direct taxation. In particular, the fact that consumers of the first decile devote a low fraction of their income to ‘Food out’, relative to the fourth and fifth deciles, implies that a Rawlsian social planner would heavily tax both restaurants and fast foods.

All things considered, these results look plausible and may be worth independent confirmation and further refinement.

References


Appendix

A Proof of Theorem 1

There is no vector space structure on the variables $G$, and therefore no way to differentiate with respect to $G$. To put a differentiability structure on the set of variables we abstract from the economic context and do as if it were possible to tax parts of good $g$ at the various available rates. Let $\pi_k(g)$ be the fraction of good $g$ subject to rate $t_k$, where $\pi = [\pi_k(g), k = 1, \ldots, K]$, is a vector of positive measurable functions, defined on $G$, of square integrable with respect to the measure $\mu(g)dg$. The program (2) then becomes

$$\max_{t, \pi} \sum_{k=1}^{K} \int \pi_k(g) V_g(t_k) \mu(g) dg$$

$$\sum_{k=1}^{K} \int \pi_k(g) t_k X_g(t_k) \mu(g) dg = R$$

$$\pi_k(\cdot) \geq 0, \text{ for all } k = 1, \ldots, K, \text{ and } \sum_{k=1}^{K} \pi_k(\cdot) = 1,$$

where the variables maximized upon are $(t, \pi)$ in $\mathbb{R}^K \times L^2_2(G)$ instead of $(t, G)$. The only solutions of economic relevance are such that the functions $\pi$ take only two values, either 0 or 1. An adaptation of the Lagrangian approach can be used to derive necessary conditions satisfied by a solution to the program (Theorem 7.3 of Jahn (2004)). Both the function to be maximized and the government revenue are Fréchet differentiable with respect to the variables $(t, \pi)$. Let $\lambda, \rho = (\rho_k)$, and $\sigma$, respectively in $\mathbb{R}, L^2_2(G)$ and $L^2_2(G)$, be the multipliers associated with the government budget constraint, the positivity constraints and the normalization constraints. $\rho$ is nonnegative and the solution is a local extremum of

$$\sum_{k=1}^{K} \int [\pi_k(g) V_g(t_k) + \lambda \pi_k(g) t_k X_g(t_k) + \pi_k(g) \rho_k(g) + \pi_k(g) \sigma(g)] \mu(g) dg,$$

with

$$\int \pi_k(g) \rho_k(g) \mu(g) dg = 0 \text{ for all } k,$$

and

$$\int [\sum_{k=1}^{K} \pi_k(g) - 1] \sigma(g) \mu(g) dg = 0.$$

Taking the Fréchet derivative with respect to $\pi_k$ yields, for $\mu$ almost all $g$,

$$V_g(t_k) + \lambda t_k X_g(t_k) + \rho_k(g) + \sigma(g) = 0,$$
with \( \rho_k(g) \geq 0, \pi_k(g) \geq 0, \) and \( \rho_k(g)\pi_k(g) = 0. \) It follows that a necessary condition for optimality is

\[
\sigma(g) = \max_{k=1,\ldots,K} [V_g(t_k) + \lambda t_k X_g(t_k)],
\]

and that \( \pi_\ell(g) \) is equal to zero whenever

\[
\sigma(g) > [V_g(t_\ell) + \lambda t_\ell X_g(t_\ell)].
\]

There are typically several optima, and there is always an economically meaningful solution in the set of optima, i.e. one solution such that \( \pi_k(g) \) is everywhere either equal to 0 or to 1. This relies on the assumption that the space of commodities has no atoms, and directly follows from the following lemma:

**Lemma 3** Let \( \Gamma \) be a subset of goods such that, for \( k = 1, \ldots, n \), there are real \( \mu \) integrable functions \( \alpha_k \) and \( \beta_k \) defined on \( \Gamma \), verifying \( \sigma(g) = \alpha_k(g) + \beta_k(g) \). Consider measurable functions from \( \Gamma \) into \([0,1]\) such that \( \pi_k(g), \pi_k(g) \geq 0, \sum_k \pi_k(g) = 1 \).

Assume that the measure \( \mu \) has no atoms on \( \Gamma \). Then there exists a partition \((\Gamma_k)_{k=1,\ldots,n}\) of \( \Gamma \) such that:

\[
A = \int_{\Gamma} \sum_k \pi_k(g)\alpha_k(g)\mu(g)dg = \sum_k \int_{\Gamma_k} \alpha_k(g)\mu(g)dg,
\]

and

\[
B = \int_{\Gamma} \sum_k \pi_k(g)\beta_k(g)\mu(g)dg = \sum_k \int_{\Gamma_k} \beta_k(g)\mu(g)dg.
\]

**Proof:** For every \( g \), let \( \bar{\alpha}(g) = \max_k \alpha_k(g) \) and \( \underline{\alpha}(g) = \min_k \alpha_k(g) \). Note also \( \bar{k}(g) \) the smallest \( k \) such that \( \bar{\alpha}(g) = \alpha_k(g) \), and similarly \( \underline{k}(g) \) for the minimum. Of course:

\[
\bar{A} = \int_{\Gamma} \bar{\alpha}(g)\mu(g)dg \geq A \geq \underline{A} = \int_{\Gamma} \underline{\alpha}(g)\mu(g)dg.
\]

The non negative integral \( \int_\gamma (\bar{\alpha}(g) - \underline{\alpha}(g))\mu(g)dg \), where \( \gamma \) is a measurable subset of \( \Gamma \), defines a nonnegative atomless measure on \( \Gamma \). By Lyapunov (see e.g. Hildenbrand (1974), p.45), its range is the convex interval \([0, \bar{A} - \underline{A}]\). There is therefore a set \( \gamma \) such that

\[
A - \underline{A} = \int_\gamma (\bar{\alpha}(g) - \underline{\alpha}(g))\mu(g)dg.
\]

For all \( k \), define

\[
G_k = \{ g \in G | (g \in \gamma \text{ and } k = \bar{k}(g)) \text{ or } (g \notin \gamma \text{ and } k = \underline{k}(g)) \}.
\]
By construction the $G_k$’s form a partition of $G$, and
\[ \sum_k \int_{G_k} \alpha_k(g)\mu(g)dg = \int_G \bar{\alpha}(g)\mu(g)dg + \int_{G\setminus\gamma} \alpha(g)\mu(g)dg = A. \]

The second equality of the lemma is an immediate consequence of the equality $\alpha(g) = \sigma(g) - \beta(g)$.

The result follows from applying the lemma successively to all the subsets of tax rates $\chi = (k_1, ..., k_n)$ for which there exists a non-negligible set of goods such that $L_k$ is constant on $\chi$, and strictly smaller than $L_{k_i}$ for $k$ not in $\chi$. The construction yields the same value of welfare and the same government receipts.

This completes the proof of the Theorem. Program (2) has more restrictive constraints than Program (2') and we have exhibited an admissible allocation for (2) that maximizes (2'). It satisfies the necessary conditions for optimality:

\[ L_g(t_k) = \max_{h=1,...,K} L_g(t_h). \]

### B Constant elasticity

Under the assumptions of Section 6, one can derive the following results:

**Lemma 4** For any commodity $g$, with price elasticity $\varepsilon > 0$ and distributional characteristic $a$, the inequality $L_g(t') > L_g(t)$ is equivalent to

\[ \frac{a}{\lambda} < \phi(\varepsilon, t, t') \]

where

\[ \phi(\varepsilon, t, t') = (1 - \varepsilon) \left[ 1 + \frac{1}{\sqrt{(1+t')(1+t)}} \sinh \left( \frac{r\varepsilon}{2} \right) \right], \text{ and } r = \ln \left( \frac{1+t'}{1+t} \right). \]

**Proof:** Using $(1 + t')^{1-\varepsilon} = (1 + t')^{1-\varepsilon} - (1 + t')^{-\varepsilon}$, the inequality $L_g(t') > L_g(t)$ rewrites

\[ \left[ (1 + t')^{1-\varepsilon} - (1 + t)^{1-\varepsilon} \right] \left( \frac{a}{\varepsilon - 1} + \lambda \right) \frac{1}{\lambda} > (1 + t')^{-\varepsilon} - (1 + t)^{-\varepsilon}. \]

Note that, for any real number $\sigma$,

\[ (1 + t')^\sigma - (1 + t)^\sigma = 2 (1 + t)^{\sigma/2} (1 + t')^{\sigma/2} \sinh \left( \frac{\sigma r}{2} \right) \]

where $r$ is as defined in the Lemma. Thus, we get that $L_g(t') > L_g(t)$ is equivalent to

\[ \left[ (1 + t') (1 + t) \right]^{\frac{1}{2}} \sinh \left( r \frac{1 - \varepsilon}{2} \right) \left( \frac{a}{\varepsilon - 1} + \lambda \right) \frac{1}{\lambda} + \sinh \left( r \frac{\varepsilon}{2} \right) > 0. \]
Since sinh \((r \frac{1-\varepsilon}{2})\) has the same sign as \(1 - \varepsilon\), the last inequality rewrites

\[
\frac{a}{\lambda} < \phi(\varepsilon, t, t')
\]

where

\[
\phi(\varepsilon, t, t') = (1 - \varepsilon) \left[ 1 + \frac{1}{\sqrt{(1 + t')(1 + t)}} \frac{\sinh \left( \frac{r}{2} \right)}{\sinh \left( r \frac{1-\varepsilon}{2} \right)} \right].
\]

In the particular case \(\varepsilon = 1\), it is sufficient to show that \(\frac{a}{\lambda} < \phi(1, t, t')\) is equivalent to \(\mathcal{L}_g(t') > \mathcal{L}_g(t)\) for \(\varepsilon = 1\). Since \(\sinh x\) is equivalent to \(x\) in the neighborhood of \(x = 0\), we have

\[
\phi(1, t, t') = \frac{1}{\sqrt{(1 + t')(1 + t)}} \frac{\sinh \left( \frac{r}{2} \right)}{\frac{r}{2}} = \frac{(1 + t)^{-1} - (1 + t')^{-1}}{\ln(1 + t') - \ln(1 + t)}
\]

and \(\mathcal{L}_g(t') > \mathcal{L}_g(t)\) is equivalent to

\[
\frac{\lambda t'}{(1 + t')} - a \ln(1 + t') > \frac{\lambda t}{(1 + t)} - a \ln(1 + t)
\]

\[
\Leftrightarrow \lambda - \frac{\lambda}{(1 + t')} - a \ln(1 + t') > \lambda - \frac{\lambda}{(1 + t)} - a \ln(1 + t)
\]

\[
\Leftrightarrow \lambda \left[ \frac{1}{(1 + t)} - \frac{1}{(1 + t')} \right] > a [\ln(1 + t') - \ln(1 + t)]
\]

\[\Box\]

**Lemma 5** For \(t' > t\), the function \(\phi(\varepsilon, t, t')\) is convex in its first argument.

1. Its slope at the origin is

\[
\frac{\partial \phi}{\partial \varepsilon} = \frac{1}{t' - t} \left[ \ln \left( \frac{1 + t'}{1 + t} \right) - (t' - t) \right].
\]

2. When \(\varepsilon\) goes to \(\infty\), \(\phi\) is equivalent to

\[
(1 - \varepsilon) \frac{t}{1 + t}.
\]

**Proof:** 1) Using the identity

\[
\sinh a \cosh b + \sinh b \cosh a = \sinh(a + b),
\]

a direct computation yields

\[
\frac{\partial \phi}{\partial \varepsilon} = -1 - \frac{1}{\sqrt{(1 + t')(1 + t)}[\sinh (1 - \varepsilon)r/2]^2} \left[ \sinh \left( \frac{1 - \varepsilon}{2} \right) \sinh \frac{\varepsilon r}{2} - \frac{(1 - \varepsilon)r}{2} \sinh \frac{\varepsilon r}{2} \right].
\]

24
The desired formula follows when \( \varepsilon = 0 \), using the equality \( \sqrt{(1 + t')(1 + t)} \sinh \frac{r}{2} = \frac{(t' - t)}{2} \).

2) One can rewrite

\[
\phi(\varepsilon, t, t') = (1 - \varepsilon) \left[ 1 - \frac{1}{1 + t} \frac{1 - \exp(-\varepsilon r)}{1 - \exp[(1 - \varepsilon) r]} \right].
\]

When \( \varepsilon \) goes to infinity, the result follows.

We finally show the convexity of \( \phi \) with respect to \( \varepsilon \), which is derived from the fact that its second derivative is positive. Indeed differentiating the expression obtained in 1) for the first derivative gives:

\[
\frac{\partial^2 \phi}{\partial \varepsilon^2} = \frac{r \sinh \frac{r}{2}}{\sqrt{(1 + t')(1 + t)}} \left[ \sinh \frac{r}{2} \left[ -1 + \frac{(1 - \varepsilon) r \cosh \frac{(1 - \varepsilon) r}{2}}{2 \sinh \frac{(1 - \varepsilon) r}{2}} \right] \right].
\]

It is positive since \( (x / \tanh x) \) is larger than 1 for all \( x \) (the \( \tanh \) curve is below the \( 45^\circ \) line for positive \( x \), and above for negative \( x \)).

C First-order condition of the planner problem without separability of individual preferences

For a (small) group of goods \( \{g\} + dG \) around \( g \), define \( \tau_{\{g\} + dG}(t, s) \) to be the set of tax rates \( t' \) such that \( t' = t \) for \( \gamma \) not in \( \{g\} + dG \) and \( t' = s \) for \( \gamma \) in \( \{g\} + dG \). Adapting Theorem 1, a necessary condition for the optimality of a partition \( G \) associated with tax rates \( t \) is that, for all \( k \), for all \( g \) and all small enough \( dG \) such that \( \{g\} + dG \) is in \( G_k \), and for all \( h \)

\[
\mathcal{L}[\tau_{\{g\} + dG}(t, t_k)] \geq \mathcal{L}[\tau_{\{g\} + dG}(t, t_h)].
\]

We prove formula (8) of the text. The first-order condition for an interior maximum is

\[
0 = - \int_C \alpha_c \rho_c \int_{\{g\} + dG} \xi_s[\tau_{\{g\} + dG}(t, s), L_c, Y_c, c] \mu(\ell) d\ell \ d\nu(c)
+ \lambda \int_C \int_{\{g\} + dG} \xi_s[\tau_{\{g\} + dG}(t, s), L_c, Y_c, c] \mu(\ell) d\ell \ d\nu(c)
+ \lambda \int_C \int_G t \frac{\partial \xi_s[\tau_{\{g\} + dG}(t, s), L_c, Y_c, c]}{\partial s} \mu(\ell) d\ell \ d\nu(c).
\]
We want to get the limit of the above expression when $dG$ goes to zero, after division by the weight $\mu(dG) = \int_{dG} \mu(\ell) d\ell$. The two first terms, as well as the last one, are easily dealt with. Indeed, define, with some abuse of notation:

$$\xi_g[\tau_g(t, s), L_c, Y_c, c] = \lim_{dG \to 0} \int_{\{g\} + dG} \frac{\xi(\tau_{(g)} + dG(t, s), L_c, Y_c, c)}{\mu(dG)} \mu(\ell) d\ell.$$ 

Then, the two first terms tend to

$$\{-a_g[\tau_g(t, s)] + \lambda\} X_g[\tau_g(t, s)].$$

The third term needs some more care. When taking the limit, one must separate the own price effect from the substitution effect on other goods:

$$\frac{\partial \xi_g[\tau_g(t, s), L_c, Y_c, c]}{\partial s} = \lim_{dG \to 0} \int_{\{g\} + dG} \frac{1}{\mu(dG)} \frac{\partial \xi(\tau_{(g)} + dG(t, s), L_c, Y_c, c)}{\partial s} \mu(\ell) d\ell,$n

$$\frac{\partial \xi_{G_k \setminus \{g\}}[\tau_g(t, s), L_c, Y_c, c]}{\partial s} = \lim_{dG \to 0} \int_{G_k \setminus \{g\} + dG} \frac{1}{\mu(dG)} \frac{\partial \xi(\tau_{(g)} + dG(t, s), L_c, Y_c, c)}{\partial s} \mu(\ell) d\ell.$$

The former limit is the own price elasticity, while the latter is the average substitution effect on the commodities in the set $G_k \setminus \{g\}$, which only exists when substitution between commodities is not too ‘large’. Finally, summing up on agents, define

$$\frac{\partial X_g[\tau_g(t, s)]}{\partial s} = \int_C \frac{\partial \xi_g[\tau_g(t, s), L_c, Y_c, c]}{\partial s} d\nu(c),$$

and

$$\frac{\partial X_{G_k \setminus \{g\}}[\tau_g(t, s)]}{\partial s} = \int_C \frac{\partial \xi_{G_k \setminus \{g\}}[\tau_g(t, s), L_c, Y_c, c]}{\partial s} d\nu(c).$$

The equations (6) and (8) of the text follow with $s$ standing for the optimal rate applied to group $G_k$ or commodity $g$.

---

14We use the notation $G_k \setminus \{g\} + dG$ as a short hand for $G_k \setminus [G_k \cap \{g\} + dG]$. Note that since, by construction, $\{g\} + dG$ is contained in a single member of the partition, say $G_h$, all the $G_k \setminus \{g\} + dG$’s coincide with $G_h$, for all $k$ different from $h$. 

26
| Decile 1 | Decile 2 | Decile 3 | Decile 4 | Decile 5 | Decile 6 | Decile 7 | Decile 8 | Decile 9 | Decile 10 | Overall Avg. Tax Rate |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----------|--------------------|
| Bread & Cereals | -0.391 | -0.085 | -0.154 | -0.02 | -0.175 | 0.385 | -0.441 | 0.812 | -0.227 | -0.551 | 0.151 | 0.46 | 0.146 | 0.005 | -0.471 | -0.489 | 0.471 |
| Meat & Fish | -0.026 | 0.038 | -0.019 | -0.107 | 0.072 | -0.082 | 0.003 | -0.029 | -0.187 | -0.05 | -0.088 | -0.173 | 0.084 | -0.024 | 0.171 | 0.064 | -0.066 | 0.325 |
| Other Fruits & Vegetables | -0.32 | -0.266 | -0.037 | -0.089 | -0.322 | 0.312 | -0.338 | 0.086 | -0.034 | -0.159 | 0.159 | 0.159 | 0.159 | -0.088 | 0.159 | 0.159 | -0.088 | 0.159 |
| Standard VAT Food | -0.045 | -0.129 | 0.359 | -0.173 | -0.04 | 0.275 | 0.081 | -0.005 | -0.045 | -0.135 | 0.274 | 0.165 | 0.165 | -0.016 | 0.165 | 0.165 | -0.016 | 0.165 |
| Tobacco | 0.073 | 0.282 | 0.191 | 0.166 | 0.146 | 0.025 | 0.065 | 0.036 | 0.138 | 0.09 | 0.148 | 0.089 | 0.089 | 0.089 | 0.089 | 0.089 | 0.089 | 0.089 |
| Petrol and Diesel | -0.006 | 0.259 | -0.125 | 0.215 | 0.076 | 0.006 | 0.023 | 0.023 | -0.044 | -0.281 | 0.175 | -0.047 | 0.367 | -0.101 | 0.367 | -0.101 | 0.367 | -0.101 |
| Books & Newspapers | -0.062 | 0.002 | 0.013 | 0.045 | 0.027 | 0.055 | 0.083 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 |
| Leisure Services | 0.061 | 0.041 | 0.012 | 0.045 | 0.027 | 0.055 | 0.083 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 | 0.061 |

**I. Elasticities of the demand for goods (rows) with respect to prices (columns)**

**II. Budget shares by consumption deciles and aggregate: indirect VAT and excise tax rate (fraction of consumption price)**