



Working Papers

www.cesifo.org/wp

How Should Commodities be Taxed? A Counterargument to the Recommendation in the Mirrlees Review

Spencer Bastani
Sören Blomquist
Jukka Pirtilä

CESIFO WORKING PAPER NO. 4240
CATEGORY 1: PUBLIC FINANCE
MAY 2013

An electronic version of the paper may be downloaded

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

How Should Commodities be Taxed? A Counterargument to the Recommendation in the Mirrlees Review

Abstract

The Mirrlees Review recommends that commodity taxation should in general be uniform, but with some goods consumed in conjunction with labour supply (such as child care) left untaxed. This paper examines the validity of this claim in an optimal income tax framework. Contrary to the recommendation of the Review, our theoretical results imply that even if all goods other than the good needed for working are separable from leisure, the optimal tax on these goods should not be uniform. Instead, goods with larger expenditure elasticities should be discouraged relatively more by the tax system. If the government fully subsidises the cost of the good needed for working, then commodity taxation is uniform under the standard separability assumption. Our results imply that the optimal commodity tax system is dependent on the expenditure side of the government. A calibration exercise presented in the paper suggests that these results can be quantitatively important.

JEL-Code: H210, H420.

Keywords: income taxation, commodity taxation, public provision, separability.

Spencer Bastani
Uppsala Center for Fiscal Studies at the
Department of Economics
Uppsala University / Sweden
spencer.bastani@nek.uu.se

Sören Blomquist
Uppsala Center for Fiscal Studies at the
Department of Economics
Uppsala University / Sweden
soren.blomquist@nek.uu.se

Jukka Pirttilä
University of Tampere / Finland
jukka.pirttila@uta.fi

We are grateful to Bas Jacobs, Håkan Selin and participants of the CESifo Public Sector Area Conference of 2012 for useful comments on an earlier version (which circulated under a different title). Mike Brewer and Andrew Shephard gave valuable advice regarding UK data sources. Financial support from The Riksbankens Jubileumsfond and The Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

1. Introduction

In the authoritative treatment of the normative implications of tax theory, the Mirrlees Review, a uniform structure of commodity tax rates, with the possible exception of child care is recommended. These recommendations build on results presented in the empirical analysis of a background chapter to the Review by Crawford, Keen and Smith (2010).

'In sum, the efficiency arguments for differential tax rates are important but, in our view, can be very hard to operationalize in practical terms. The only exception to this is that there is probably a strong case for exempting childcare costs from VAT because, in many cases, spending on childcare is so closely related to the choice over how many hours to work.' (Mirrlees et al. 2011, p. 162)

'There are reasons other than equity for favouring differential tax rates, including a desire to tax more lightly the consumption of those goods associated with work. This is likely to provide a strong reason for a low (perhaps zero) VAT rate on childcare.' (Mirrlees et al. 2011, p. 166)

Economists working in the optimal tax tradition have examined the optimal commodity tax structure for years. The theoretical basis to study this question was laid out in Mirrlees (1976) and Atkinson and Stiglitz (1976).¹ According to the Mirrlees approach to optimal commodity taxation, tax rates on commodities should be set so as to help screen high skill persons from low skill. A high skill person choosing the same taxable income as a low skill person will have more leisure time as compared to the true low skill person. Thus taxing commodities whose demand is increasing in the amount of leisure time is a way to discourage high skill persons from reproducing the taxable income of low skill persons by a reduction in his/her labour effort. In this way commodity taxes can help improve the efficiency of the tax system.

The article by Atkinson and Stiglitz is well known for giving conditions under which uniform taxation is optimal and commodity taxes are not useful for screening purposes; if the utility function is such that leisure is weakly separable from commodities, then uniform taxation is optimal. This is because under this condition commodity demand will not depend on the amount of leisure available.

¹ Some of the other key contributions are Christiansen (1984), Edwards, Keen and Tuomala (1994) and Saez (2002).

It is an empirical question whether goods and labour supply are separable. Browning and Meghir (1991) found that they were not, but did not discuss whether the non-separabilities were large enough to motivate differential taxation. In Crawford et al. (2011), a chapter in the Mirrlees review, the authors claim that, although leisure is not separable from commodities, as a close approximation it is, and the policy recommendation is that there should be uniform commodity taxation with the exception that child care should be left untaxed. The reason for leaving child care untaxed is that there is a close association between hours of work and child care.

The purpose of the present article is to examine the validity of the recommendations of the Mirrlees Review in an optimal tax framework using the assumption that all goods are separable from leisure with the exception that child care is needed for work. These are the same assumptions as those used in the Mirrlees review. However, the conclusions about the optimal tax structure that we reach are quite different. We find that when all goods are weakly separable from leisure but there is a need to purchase a good in order to work (such as child care or elderly care) commodity taxation of those goods that are separable from labour supply should *not be uniform*.

We adopt the framework which lies behind the Mirrlees Review; as a first approximation leisure is weakly separable from *all commodities except one*. We think of this exception as, for example, child care. Elderly care is also a relevant example.² Thus we study commodity taxes under the assumption that the utility function takes the form $U[F(x_1, x_2, \dots, x_n), G(x_c, l)]$. In this sense we follow the Mirrlees approach to optimal taxation, but consider a preference structure not studied before. We do not study the full optimal commodity tax problem under these circumstances, but for simplicity we consider two polar situations where particularly clean results can be obtained. In the first case child care (or elderly care) is subject to a zero VAT tax rate, corresponding to the Mirrlees Review recommendation. In the other case child care (or elderly care) is publicly provided and provided free of charge which more or less is in accordance with the policy in the Nordic countries.

Why do we reach another conclusion than the Mirrlees review? The reason is that the Mirrlees review first considers the case where commodities are weakly separable from leisure. This leads to optimally uniform taxation. Then they note that, since child care is closely associated with hours of work, it should not be taxed. They reach this conclusion without

² With child care it is obvious that it is the parents who pay for child care. With elderly care it is less obvious who buys elderly care in order to work. What we have in mind are persons who feel responsible for the care of some elderly person and either takes care of the person him/herself (which would make working difficult) or buys elderly care (to be able to work). Even if the elderly person formally pays for the care himself/herself it is in some cases still reasonable to say that it is the son or daughter who ultimately pays for the elderly care, but in the form of a reduced inheritance.

formally analysing optimal commodity taxation in the presence of child-care use. In this paper we carry out such an analysis given the preference structure above.

Our main theoretical results can be summarized as follows. We find that when leisure is weakly separable from all commodities except one, a commodity that is needed in order to work, and individuals pay for this commodity themselves, taxes on all the other commodities should be differentiated. There should be higher taxes on goods with large income elasticities. This recommendation is basically in accordance with the wisdom prevailing long before the Mirrlees analysis. However, if the commodity needed in order to work is publicly provided and provided free of charge, then uniform commodity taxation on all other goods is optimal. Thus our paper demonstrates that the optimal structure of commodity taxation depends on the expenditure side of the government and, in particular, on the extent of public provision. We find that the presence of public provision affects the structure of optimal marginal (income) tax rates but entails the optimality of uniform commodity taxation.

We also conduct a simulation analysis that explores the practical relevance of our theoretical results. This is of key importance, since the Crawford et al. (2010) argument for favouring uniform commodity taxation was a practical one: they argued that the gains from differential taxation are likely to be small and, at the same time, maintaining a non-uniform commodity tax system is administratively cumbersome. Our simulation example, which also builds on UK data, shows that the commodity tax differentiation result can be of significant practical importance. Another important result from our simulations is that setting the VAT rate on child care to be zero, or lower than the current existing VAT rate on other goods in the U.K. (in accordance with the recommendations in the Mirrlees review) implies an implicit subsidy rate on child care which is far too low.

2. Individual behaviour

We consider an extension of the discrete optimal income tax model and analyse optimal income taxation and linear commodity taxation in a framework similar to Edwards et al. (1994). Each individual h consumes $i = 1 \dots n$, $i \neq c$ different consumption goods x_i^h . The index c is reserved for a commodity needed in order to work, which we denote by x_c . The driving force in our model is this commodity. To add realism, we assume that individuals differ not only with respect to their income-earning ability but also in terms of their needs/tastes for the commodity that is needed in order to be able to work. We thus decompose the population into users and non-users of this good. Heterogeneity in needs/tastes can be incorporated into the model in two different ways.

Heterogeneity could be introduced to the utility function directly: $U[F(x_1, x_2, \dots, x_n), G(\gamma x_c, l)]$ where l denotes labour supply and $\gamma = 1$ for users and $\gamma = 0$ for non-users of x_c . An alternative way is to retain similarity in preferences, but introduce the need only through the budget constraint (x_c does not bring utility, it is only needed to be able to work). We will concentrate on the latter formulation of heterogeneity in needs and make the following:

Assumption 1 *Individuals maximize a weakly separable utility function $U[F(x_1, x_2, \dots, x_n), l]$ subject to a needs constraint $x_c = \gamma f(l)$, for some non-decreasing function f and $\gamma \in \{0, 1\}$, and a budget constraint.*

The above formulation is equivalent to assuming that leisure is weakly separable from all commodities except x_c . For simplicity we will focus on the case $f(l) = l$.³

To be concrete we assume in the model that the commodity needed for work is child care, although the model should be interpreted broadly with ‘child care’ possibly replaced by other similar goods and services, out of which elderly care is an equally relevant example. Because only parents with small children need child care, we assume that the users and non-users of child care correspond to parents and non-parents.

Each individual h supplies l^h hours of work at a fixed wage rate w^h and earns income $Y^h = w^h l^h$ taxed according to a non-linear income tax schedule. The after-tax income for household h is given by $A^h = w^h l^h - T(w^h l^h) = Y^h - T(Y^h)$ where T denotes the income tax function. Households with children need to buy child care to be able to work and the hourly price of child care is denoted by π . The budget available for consumption goods (net of child care costs) is $B^h = A^h - (1 - \phi) \frac{\pi}{w^h} Y^h$ for parents and $B^h = A^h$ for non-parents where ϕ is the subsidy on child care. As mentioned in the introduction, we focus on the two polar cases $\phi = 0$ (child care untaxed) and $\phi = 1$ (child care publicly provided). Let the consumer price of each commodity be

³ The formulation $x_c = \gamma l$ then implies a perfect correlation between working hours and child care use. It also means that in principle the government would be able to observe parents’ working hours, which is not compatible with the informational assumptions of optimal tax models. To preserve asymmetric information, we assume that child care authorities and tax administration do not share information regarding working hours. The functional form is chosen because of analytical ease; it could also be a more complicated one, as in Blomquist et al. (2010), without changing any of the qualitative results. Then the assumption regarding no information sharing could also be dropped.

$q_i = p_i + t_i$, where p_i denotes the producer price and t_i is a linear commodity tax. The private budget constraint for an individual is thus

$$(1) \quad \sum_i q_i x_i^h = B^h,$$

We denote by x^h the vector of consumption goods and by q the vector of consumer prices. Individual optimisation is decomposed into a two-step process. In the first stage, the individual chooses x^h in order to maximize $U^h[F(x^h), l^h]$ subject to the budget constraint (1) where disposable income B^h , pre-tax income Y^h and consumer prices q all are treated as given. This maximisation yields the conditional indirect utility function $V^h[B^h, Y^h, q] = U^h[F(x^{*h}), l^h]$ where $x_i^{*h} = x_i^h[B^h, Y^h, q]$ is the demand function for good i . In the second stage, the individual chooses optimal labour supply (given the link between pre-tax income and disposable income implied by the tax schedule) by choosing Y^h in order to maximize $V^h[B^h, Y^h, q] = V^h[Y^h - T(Y^h) - (1 - \phi)(\pi/w^h)Y^h, Y^h, q]$.

3. Government's problem

We now proceed with the government's optimisation problem. The government maximizes social welfare by designing an optimal non-linear income tax and optimal linear taxes on consumer goods subject to a revenue constraint and a set of self-selection constraints. Instead of choosing the income tax function $T(\cdot)$ directly, the government assigns pre-tax and after-tax income points (Y^h, A^h) for each agent (the tax schedule can implicitly be calculated as $T(Y^h) = Y^h - A^h$). The set of self-selection constraints ensure that each agent weakly prefers the income point assigned to him/her rather than the income point assigned to any other agent.

There are two possibilities for designing the income tax: either the social planner sets different income tax schedules for users and non-users of the good needed for work or the income tax is the same irrespective of this status. We will refer to these as the "tagging" and "no-tagging" cases respectively using the terminology of Akerlof (1978). If the good needed for work is child care, one could think that tagging would be relatively straightforward: benefits and taxes should be made contingent on having a child in the day care age in the family. However, even in this case, the tax schedule should be individualized: in a system of family taxation, one could end up subsidizing non-working spouses in families. Some of the real-world examples of in-work credits (such as the EITC in the US) operate, in fact, within a family tax system. If, on the other hand, the good needed for working is elderly care, then tagging is even more complicated. The tag

ought to be dependent on the health status (i.e. the need for care) of the parent of the working-age person, and such a tax system seems difficult to operationalize. In sum, it is important to make a difference between fully optimal tagging and the type of tagging which is feasible in the real world. Actual tax systems are mixtures of tagging and no-tagging schemes, and that is why we think it is important to cover both cases.

We assume there are two types of parents and non-parents: those with high skill level (type 2) and those with a low skill level (type 1). Altogether, there are 4 different kinds of households, type 1 parents (1P), type 1 non-parents (1NP), type 2 parents (2P) and type 2 non-parents (2NP). A parent of type i does not necessarily have to have the same wage rate as a non-parent of type i .

With tagging, there are only two self-selection constraints to consider: high-skilled non-parents should weakly prefer his/her own income point rather than the income point of the low-skilled non-parents and similarly high-skilled parents should not prefer to mimic the choice of the low-skilled parents. We label these self-selection constraints (2NP, 1NP) and (2P, 1P) respectively. Formally

$$\begin{aligned} (2NP, 1NP) \quad & V^{2,NP}(q, B^{2,NP}, Y^{2,NP}) \geq \hat{V}^{2,NP}(q, \hat{B}^{1,NP}, Y^{1,NP}) \quad \text{and} \\ (2P, 1P) \quad & V^{2,P}(q, B^{2,P}, Y^{2,P}) \geq \hat{V}^{2,P}(q, \hat{B}^{1,P}, Y^{1,P}), \end{aligned}$$

where \hat{V} is used to denote the indirect utility of a mimicker and \hat{B} is the amount of income the mimicker has available for private consumption. Note that \hat{B} is in general different from B because even though a mimicker and a true type person necessarily have the same after-income-tax income A , they do not purchase the same amount of child care because the mimicker and the true type work do not work the same number of hours. For example, in the case of the (2P, 1P) self-selection constraint, the mimicking high-ability parent has a disposable income of $B^{2P} = A^{1P} - (1 - \phi)(\pi / w^{2P})Y^{1P}$ which does not coincide with the disposable income of a true (low skill) ability type who has a disposable income equal to $B^{1P} = A^{1P} - (1 - \phi)(\pi / w^{1P})Y^{1P}$. As we will see later, this observation is crucial because it means that commodity demand will be different for the mimicker and the true type, to the extent that commodity demand depends on disposable income.

Child care untaxed ($\phi = 0$) and tagging

The government objective is defined as a weighted sum of individual utilities

$$(4) \quad W = \rho^{1,P} V^{1,P} \left(q, A^{1,P} - (\pi/w^{1P}) Y^{1,P}, Y^{1,P} \right) + \rho^{2,P} V^{2,P} \left(q, A^{2,P} - (\pi/w^{2P}) Y^{2,P}, Y^{2,P} \right) \\ + \rho^{1,NP} V^{1,NP} \left(q, A^{1,NP}, Y^{1,NP} \right) + \rho^{2,NP} V^{2,NP} \left(q, A^{2,NP}, Y^{2,NP} \right)$$

where $\rho^{i,j}, i=1,2; j=P,NP$ are exogenous welfare weights indicating the importance of each agent's utility in the social objective. The government problem is the maximization of (4) subject to the self-selection constraints (2NP, 1NP) and (2P, 1P) and the revenue constraint

$$(5) \quad \sum_h (Y^h - A^h) + \sum_h \sum_i t_i x_i^h \left(q, A^h - (\pi/w^h) Y^h, Y^h \right) = R,$$

where R is an exogenous revenue requirement of the public sector. The Lagrange multipliers of the constraints (2NP, 1NP), (2P, 1P), and (5) are denoted by λ^{NP} , λ^P and γ , respectively. The first-order conditions with respect to the commodity tax and the after-tax income are presented in the appendix. These are standard and similar to those in Edwards et al (1994). Using the first-order conditions the optimal commodity tax rule can be derived:

$$(6) \quad \sum_h \sum_i t_i \frac{\partial s_k^h}{\partial q_i} = \lambda^{*P} (x_k^{1P} - \hat{x}_k^{2P}) + \lambda^{*NP} (x_k^{1NP} - \hat{x}_k^{2NP}),$$

where $\lambda^{*P} = (\lambda^P / \gamma) \hat{V}_B^{2P}$ and $\lambda^{*NP} = (\lambda^{NP} / \gamma) \hat{V}_B^{2NP}$ are both positive and s_k^h denotes individual h 's compensated demand for good k . The left-hand side of (6) is the aggregate compensated change in the desired demand of good k as a result of the tax policy. If it is negative, taxes discourage the consumption of the particular good.

The right hand side consists of two terms and is non-zero when a tax on good k can be used to screen high skill from low skill persons. A non-parent mimicker and a true type-1 non-parent have the same disposable income for consumption (because they do not pay for child care). Hence when commodity demand is independent of working hours (this is the case with separable utility) $x_k^{1NP} = \hat{x}_k^{2NP}$ which implies that the second term in (6) is zero. Now contrast the demand for good k between a mimicking parent of type 2, $\hat{x}_k^{2P} [q, Y^{1P}, A^{1P} - (\pi/w^{2P}) Y^{1P}]$, and a true type 1 parent, $x_k^{1P} [q, Y^{1P}, A^{1P} - (\pi/w^{1P}) Y^{1P}]$. Since the wage rate of type 2 is higher, he or she needs to buy less child care, and therefore the disposable income (net of child care purchases) for the type 2 mimicker is larger than for the true type 1 parent ($B^{1P} < \hat{B}^{2P}$). If the good k is

normal then the type 2 mimicker will buy more of good k as compared to the true type 1 parent i.e. $x_k^{1P} < \hat{x}_k^{2P}$. This implies that the consumption of good k should be discouraged by the tax system. The extent of discouragement should be the greater, the higher is the expenditure elasticity. Thus, if there is one commodity which needs to be consumed in order to work, then even if the utility from leisure is separable from other commodities, non-uniform commodity taxation should be used. Commodities with higher income elasticities should be subject to a higher tax burden and the Atkinson-Stiglitz result does not hold.

Child care untaxed ($\phi = 0$) and no tagging

Without tagging, the pattern of self-selection constraints is more complicated, since in addition to the (2NP, 1NP) and (2P,1P) self-selection constraints we need to consider the possibility that a non-parent might be tempted to pick an income point available on the tax schedule for parents or vice versa. If income levels are ordered so that $Y^{1,P} < Y^{2,P} < Y^{1,NP} < Y^{2,NP}$ then in addition to self-selection constraints (2NP, 1NP) and (2P,1P) there is also the constraint

$$(1NP, 2P) V^{1,NP}(q, B^{1,NP}, Y^{1,NP}) \geq \hat{V}^{1,NP}(q, \hat{B}^{2,P}, Y^{2,P}).$$

If income levels alternatively are ordered so that $Y^{1,P} < Y^{1,NP} < Y^{2,P} < Y^{2,NP}$, one needs to consider the additional constraint:

$$(2P, 1NP) V^{2,P}(q, B^{2,P}, Y^{2,P}) \geq \hat{V}^{2,P}(q, \hat{B}^{1,NP}, Y^{1,NP}).$$

Suppose first that the additional constraint that needs to be taken care of is (1NP, 2P) and let δ be the Lagrange multiplier associated with this constraint. The Lagrangean incorporating all three self-selection constraints is presented in the appendix. The rule for commodity taxation is in this case

$$(8) \quad \sum_h \sum_i t_i \frac{\partial s_k^h}{\partial q_i} = \lambda^{*P} (x_k^{1P} - \hat{x}_k^{2P}) + \lambda^{*NP} (x_k^{1NP} - \hat{x}_k^{2NP}) + \delta^* (x_k^{2P} - \hat{x}_k^{1NP}),$$

where $\delta^* = (\delta/\gamma)\hat{V}_B^{1NP} > 0$. Similar arguments as above imply that $x_k^{1NP} = \hat{x}_k^{2NP}$ whereas $x_k^{1P} < \hat{x}_k^{2P}$ (if k is a normal good). These two terms together mean that the consumption of the good in question should be discouraged. Now compare $x_k^{2P} [q, Y^{2P}, A^{2P} - (\pi/w^{2P})Y^{2P}]$ and

$\hat{x}_k^{1NP} [q, Y^{2P}, A^{2P}]$. Since the mimicker does not have to buy child care, the money left to buy good 1 is greater for him ($B^{2P} < \hat{B}^{1NP}$). If the good k is normal, this means that $x_k^{2P} < \hat{x}_k^{1NP}$. This term also works towards levying a positive tax burden on good k . In this case, a positive effective tax helps to relax not only the self-selection constraint (2P, 1P) but also the additional self-selection constraint (1NP, 2P) which arises in the no-tagging case when income levels are ordered as $Y^{1,P} < Y^{2,P} < Y^{1, NP} < Y^{2, NP}$. In this case, the consumption of goods with high income elasticities should unambiguously be discouraged more heavily relative to other goods.

If instead the additional self-selection constraint which needs to be taken care of is (2P, 1NP) (and again letting δ be the Lagrange multiplier associated with this constraint) the optimal commodity tax rule is (see the appendix)

$$(9) \quad \sum_h \sum_i t_i \frac{\partial s_k^h}{\partial q_i} = \lambda^{*P} (x_k^{1P} - \hat{x}_k^{2P}) + \lambda^{*NP} (x_k^{1NP} - \hat{x}_k^{2NP}) + \delta^\# (x_k^{1NP} - \hat{x}_k^{2P}),$$

where $\delta^\# = (\delta/\gamma)\hat{V}_B^{2P} > 0$. Consider the third term on the right hand side of (9). In contrast to the case above, the mimicker's income for consumption is now less than that of a true type 1 non-parent, since the mimicker is a parent and needs to buy child care and the mimicked agent is a non-parent who does not need to buy child care. Therefore $\hat{x}_k^{2P} < x_k^{1NP}$, and the last term implies that consumption of the good k should be encouraged.

Recognizing the presence of self-selection constraints of both types, i.e. (1NP, 2P) where a non-parent mimics a parent and of type (2P, 1NP) where a parent mimics a nonparent, it is not clear whether the commodity tax should be positive or negative, since relaxing different self-selection constraints requires commodity tax changes in opposite directions. Gathering the results on commodity taxation without tagging we can state the following:

Proposition 1 *If the individual optimization problem satisfies Assumption 1, the commodity tax should be non-uniform and the Atkinson-Stiglitz result does not hold. This holds irrespective of the uses of tagging in conjunction with income taxation.*

Proposition 1 establishes the general result that when there is a commodity needed in order for work and agents pay for this good themselves, uniform taxation is not optimal, contrasting the Mirrlees review recommendation.

Proposition 2 *If the individual optimization problem satisfies Assumption 1, and tagging is feasible and used, commodities with higher income elasticities should be subject to a higher tax burden. If tagging is not used, or is infeasible, depending on which self-selection constraints bind in the government's optimum, commodity tax rates will be either higher or lower for commodities with higher income elasticities.*

Proposition 2 highlights the fact, that our argument for differentiated commodity taxation differs from the traditional case, i.e. when leisure is not weakly separable from goods, then those goods whose demand is increasing in the amount of leisure should be taxed more heavily. This is because mimickers have more leisure time. With the preference structure we study here, commodity demand is independent of leisure. Instead, mimickers have more disposable income than those mimicked and to deter mimicking behaviour, those goods whose demand increase in income should be taxed more heavily.

Child care publicly provided ($\phi = 1$)

Underlying Proposition 1 and 2 was that individuals need to pay for child care themselves. Now suppose instead, following Blomquist et al. (2010), that child care is publicly provided free of charge ($\phi = 1$). When child care is provided free of charge, the disposable income is the same as after-tax income for all mimickers $A^h = B^h$ even for parents, and with separable preferences $x_k^{1P} = \hat{x}_k^{2P}$ (and as before under separability $x_k^{1NP} = \hat{x}_k^{2NP}$). Thus, nothing can be gained from non-uniform commodity taxation under public provision. In the social planner's optimisation problem, what changes is that the government's budget constraint now includes the cost of public provision as follows:

$$(7) \quad \sum_h (Y^h - A^h) + \sum_h \sum_i t_i x_i^h(q, A^h, Y^h) - \frac{\pi}{w^{1P}} Y^{1P} - \frac{\pi}{w^{2P}} Y^{2P} - R.$$

This constraint replaces the original budget constraint in the government optimisation problem. Notice that this change does not affect the first-order conditions for after-tax income or the commodity tax. Therefore, the commodity tax rule (7) continues to hold.⁴ This leads to our second main result:

⁴ Instead, as emphasized by Blomquist et al. (2010), the presence of public provision affects the optimal structure of labor income taxation.

Proposition 3 *If the individual optimization problem satisfies Assumption 1, and the commodity that must be consumed when working is provided free of charge by the government, the Atkinson-Stiglitz result continues to hold, irrespective of the use or feasibility of tagging.*

In our view, this is a novel point: the optimal structure of commodity taxation depends on the expenditure side of the government and, in particular, on the extent of public provision. This could mean that the case for uniform commodity taxation is stronger in countries with extensive public provision (as in Scandinavia) than countries with a more limited public provision (such as the UK).

Effective marginal tax rates

As is customary in the literature, we derive an expression for the marginal tax rate in terms of the slope of an individuals' indifference curve in the (Y, A) -space at the individual optimum. The second stage of an individuals' maximization problem presented at the end of section 2 implies that $V^h[B^h, Y^h, q]$ is maximized with respect to Y^h where $B^h = Y^h - T(Y^h) - (\pi/w^h)Y^h$. The first order condition for this maximization yields the following (implicit) expression for the marginal tax rate:

$$(10) \quad T'(Y^h) = 1 + \frac{V_Y^h}{V_B^h} - \frac{\pi}{w^h}.$$

In the original model by Edwards et al (1994), the effective marginal tax rate (the joint increase in the tax burden via both the income tax and the commodity taxes as income increases) was shown to be zero for the high-ability type and positive for the low-ability type. The former result is one interpretation of the well-known 'no-distortion at the top result'. In the appendix we show that in our model without public provision, the no-distortion at the top result still holds. However, when the good needed to work is publicly provided free of charge by the government, the result in Blomquist et al (2010) is reproduced. A non-distortive, positive element π/w^h appears in the marginal tax rate formula (the formula for the conventional marginal income tax rates in their case, the formula for the *effective* marginal income tax rate in our case). In particular, it appears also for the highest ability type. This term acts like a corrective tax for public provision. The corrective tax internalizes the additional resource cost (in terms of publicly provided child care) incurred in the government budget constraint when an additional unit of earned (pre-tax) income

is supplied by a private agent. Thus, the presence of public provision affects the structure of optimal marginal (income) tax rates but entails the optimality of uniform commodity taxation.

4. Simulation results

The purpose of this section is to examine the quantitative/empirical significance of our main finding that non-uniform taxation is desirable when preferences are separable between leisure and consumption goods with the exception of one good (such as child care or elderly care). Since the recommendation for uniform commodity taxation in the Mirrlees Review was built on the background paper by Crawford et al. (2010) that used UK data, our simulation analysis is also based on UK data. Note that the purpose is to provide first steps towards illustrating the possible size of the issues at stake, and not to provide a full-fledged numerical analysis of optimal commodity taxation.⁵

Our theoretical analysis has highlighted three factors of relevance for our tax differentiation result. First, as indicated by equation (6), the degree of tax differentiation depends on the size of the difference in the disposable income between the mimicker and the true type.⁶ This difference is large when child care expenditures represent a sizable fraction of total consumption (for instance when child care is expensive). According to the OECD report "Doing Better for Families" (2011) the UK child care costs as a fraction of net family income is estimated to be as high as 26.6% which is higher than all other OECD countries except Switzerland. Hence child care costs represent a sizable fraction of total household expenditure.

Second, the optimal degree of tax differentiation depends on the composition of users and non-users of child care in the population. In order for a differentiated commodity tax system to be quantitatively important, the benefits pertaining to relaxing the incentive constraints must outweigh the distortions it imposes on the price system. Efficiency gains are possible for users of child care in the economy but for non-users of child care differentiated commodity taxation is purely distortionary.⁷ Third, unless income tax schedules can be tagged, there might be non-standard self-selection constraints which occur when the mimicker has a *lower* disposable income

⁵ Simulations of the income tax schedule are widespread in the optimal tax literature, but there are few simulations of optimal mixed tax systems (with linear commodity taxes).

⁶ The difference in the disposable income between the mimicker and the true type depends on the size of the reduction in labor supply required for a mimicker to reproduce the earned income of the agent being mimicked. The required labor supply reduction of the mimicker will be a function of the distance between the wage rate of the true type and the mimicker.

⁷ Unless of course tagging applies to commodity taxes so that different commodity tax systems can be used for the two groups.

than the agent being mimicked.⁸ To tell if goods with higher income elasticities should be taxed more heavily in the no-tagging case, numerical simulations are needed.

Our numerical exercise considers a standard neoclassical "unitary" model of the household where each household supplies labour along one dimension and maximizes a single utility function. To illustrate our point we set up a demand system with two commodities differing in their respective income elasticities of demand. For this purpose we let the sub-utility of consumption goods be represented by a Stone-Geary utility function yielding a linear expenditure system. This demand system, which results from a generalization of the Cobb-Douglas utility function, allows us to introduce different income elasticities for different goods in the simplest possible way. The disutility of labour supply is assumed to be of the standard iso-elastic form and separable from consumption goods.

Preferences are represented by the utility function

$$(10) \quad U(c_1, c_2, h) = (c_1 - c_{10})^{\beta_1} (c_2 - c_{20})^{\beta_2} - \frac{h^{1+k}}{1+k},$$

where $\beta_1 + \beta_2 = 1$, $\beta_i > 0, i = 1, 2$, and $c_1 > c_{10}$, $c_2 > c_{20}$. The budget constraint is $p_1 c_1 + p_2 c_2 = B$ where disposable income is $B = A - p(Y/w)$ reflecting that one unit of child care must be purchased for each hour of work. This generates a non-separability between leisure and child care through the budget constraint. Given pre-tax income Y and disposable income B , an individual chooses c_1 and c_2 in order to maximize utility subject to the budget constraint. This yields the indirect utility function,

$$(11) \quad V(Y, B) = (c_1 - c_{10})^{\beta_1} (c_2 - c_{20})^{\beta_2} - \frac{(Y/w)^{1+k}}{1+k}.$$

The demand functions are given by

$$(12) \quad c_1(p_1, p_2, B) = c_{10} + \beta_1 \frac{B - (p_1 c_{10} + p_2 c_{20})}{p_1},$$

⁸ This case occurs for instance when a high skill parent mimics a low skill non-parent. The issue of heterogeneity in needs and non-standard self-selection constraints of this kind is thoroughly discussed in Bastani et al. (2010).

$$(13) \quad c_1(p_1, p_2, B) = c_{20} + \beta_2 \frac{B - (p_1 c_{10} + p_2 c_{20})}{p_2}$$

The individual purchases the minimum amounts c_{10} and c_{20} of each good and spends the remaining income $B - (p_1 c_{10} + p_2 c_{20})$ on goods 1 and 2 in the proportions β_1 and β_2 , respectively.⁹ The expenditure elasticities, denoted η_1 and η_2 , are:

$$(14) \quad \eta_1 = \frac{\partial c_1}{\partial B} \frac{B}{c_1} = \beta_1 / \frac{p_1 c_1}{B} = \beta_1 / s_1,$$

$$(15) \quad \eta_2 = \frac{\partial c_2}{\partial B} \frac{B}{c_2} = \beta_2 / \frac{p_2 c_2}{B} = \beta_2 / s_2$$

where s_i is the fraction of an individual's income spent on good $i = 1, 2$. We will assume that good 1 represents 'basic needs' such as food and shelter and that good 2 represents 'other goods'. We therefore set $c_{10} > 0$ and $c_{20} = 0$. Because of the asymmetric basic needs assumption, $s_1 > \beta_1$ which implies $\eta_1 < 1$. For good 2, we instead have $s_2 < \beta_2$ and $\eta_2 > 1$. Thus good 2 has a larger income elasticity. Individuals with a higher disposable income spend a larger fraction of their income on good 2. As disposable income B rises, η_1 converges to 1 from below and η_2 converges to 1 from above.¹⁰

We now present the principles in calibration that we have chosen. We allow for two different skill levels (low and high skill) and two categories of agents (parents and non-parents). This yields a total of four wage types to be used in the numerical simulations. We approximate the wage distributions using percentiles. Each wage distribution is represented by the 33rd and 66th percentiles.

We calibrate the model to the UK using wage and cost of child care data from the Family Resources Survey (FRS). The wage distribution for our category 'parents' is computed using wages for women with at least one child in child care age (ages 0-4). The wage distribution for

⁹ The Cobb-Douglas case is obtained when $c_{10} = c_{20} = 0$.

¹⁰ To see this, one can refer to the expressions $\eta_1 = \beta_1 B / [\beta_1 B + (1 - \beta_1) p_1 c_{10}]$ and $\eta_2 = B / [1 - p_1 c_{10}]$. Note that in the Cobb-Douglas case, $s_1 = \beta_1$ and $s_2 = \beta_2$ implying $\eta_1 = \eta_2 = 0$.

individuals categorized as 'nonparents' is constructed using wages for the rest of the population. A measure of the wage rate was obtained by dividing total labour earnings by total hours worked. The cost of child care was obtained by computing the mean hourly child care across all modes of care and all users of child care in the sample. The mean hourly cost of child care was found to be 3.87 GBP which is around 50% of the wage rate of a low-skilled parent. The wage rates are reported in Table 1.

Table 1: Hourly wage rates for the low and high skill agent (GBP).

	Parents	Non-Parents
w_1	7.56	8.53
w_2	13.29	13.97
Hourly cost of child care	3.87	

To perform simulations the values of all parameters in the utility function (10) must be specified. We set $k = 4$ implying a Frisch elasticity of $1/3$ which in the model is an upper bound on the intensive-margin compensated elasticity. This should be regarded as an intermediate value in light of the survey of Frisch and intensive-margin elasticities found in Chetty et al. (2012). The value is also consistent with the small intensive-margin labor supply elasticities for the subgroup of mothers with small children reported in Blundell and Shephard (2012). Furthermore, we set $\beta_1 = 0.1$, $\beta_2 = 0.9$ and $c_{10} = 3$, $c_{20} = 0$ which in the benchmark no-tagging model this yields budget shares for good 1 around $1/3$, elasticities for good 1 ranging between 0.34 to 0.48 and elasticities for good 2 ranging between 1.13 and 1.28 (depending on skill level). This is broadly consistent with empirical evidence. For instance, the UK consumption estimates reported in Table 2 suggest that zero-rated food and domestic energy, which arguably qualify as necessary basic needs goods, have income elasticities well below one. All other goods have expenditure elasticities beyond one, thus the model produces reasonable income elasticities.

The objective of the planner in the most general case is to maximize social welfare as defined by a weighted sum of individual utilities, but for simplicity we present results only from the Rawlsian case (where the welfare of the least well off household is maximized), and then discuss in the end the influence of this choice for the results. To construct an equivalent-variation type of welfare gain measure of policy reform we proceed as follows. We calculate the minimum amount of extra revenue Λ which needs to be injected into the government budget constraint in

the pre-reform equilibrium, to reach the social welfare level of the post-reform equilibrium. Finally Λ is divided by aggregate income in the pre-reform economy to obtain a welfare gain measure expressed in terms of percentage points of GDP.

Table 2: Income Elasticities of Demand for the United Kingdom, (Reproduced from Table J.9, p.525 in TAXUD/2010/DE/328).

Commodity	Income elasticity
Zero-rated food and drink	0.25
Standard-rated food and drink, restaurants, takeaways and alcohol	1.15
Leisure goods (inc. tobacco), and services (inc. hotels)	1.36
Domestic energy	0.17
Household goods and services	1.15
Personal goods and services (inc. adult clothing)	1.20
Private transport goods and services	1.02
Other zero-rated goods (children's clothing, public transport, books, etc.)	1.28

In the model we consider it is desirable to subsidize child care. We find that a 100% subsidy is optimal. In fact, an even larger subsidy would be desirable, but 100% is an upper bound because otherwise buyers and sellers of child care services could collude.¹¹ With a 100% subsidy rate we know that uniform commodity taxation is optimal. This is our first result.

If for some reason, child care cannot be subsidized to 100%, it is of interest to see how tax rates should be differentiated when child care is imperfectly subsidized. For this reason we let the subsidy rate on child care be 0%, in line with the Mirlees Review recommendation, and analyze differentiated taxation, under the assumption that the base line VAT is 20%.

In Table 3 we present the benchmark no-tagging allocation where child care is not subject to taxation and the tax structure on other goods is restricted to be uniform. In Table 4 we present the results where child care is not subject to taxation but the tax structure on other goods is allowed to be non-uniform. The reason for the focus on the no-tagging case is that it turns out that under tagging, there is little to be gained from commodity tax differentiation. We return to the discussion of why this is probably the case in the end of this Section.

¹¹ With a subsidy rate greater than 100%, both the provider and the customer would have an incentive to raise prices. The real price of the services would probably be difficult for the government to observe.

Table 3 (Benchmark Allocation)

Type	Y	c_1	c_2	$T'(Y)$
w_{1P}	0.000	4.066	9.593	49%
w_{1NP}	15.633	4.457	13.111	0%
w_{2P}	16.471	4.148	10.329	51%
w_{2NP}	30.194	5.236	20.120	0%
$\tau_2 = 0.20, \tau_1 = 0.20, \tau_c = 0$				

The results in Table 4 imply that the degree of tax differentiation is high: the VAT rate for ‘other goods’, good 2, should be set four times higher than the benchmark rate for necessities (which was assumed to be 20 per cent).¹²¹³ The monetary welfare gain following from tax differentiation equals 2.04 per cent of GDP. In 2010 figures it amounts to approximately 30 billion pounds. This means that the administrative costs of having two instead of one VAT rate would need to exceed this figure. With modern information technology, this seems to us as a rather high cost level.

Table 4 (Differentiated Commodity Tax Optimum)

Type	Y	c_1	c_2	$T'(Y)$
w_{1P}	0.000	4.617	9.563	49%
w_{1NP}	14.520	5.042	12.073	-17%
w_{2P}	14.520	4.689	9.989	51%
w_{2NP}	30.105	6.072	18.162	-45%
$\Lambda / GDP = 2.04\% \quad \tau_2 = 0.83, \tau_1 = 0.20, \tau_c = 0$				

In the above results the proportion of parents in the economy is set at 15%, which we consider a reasonable benchmark. The analysis above was based on a Rawlsian social welfare function. We have also examined the welfare gains under various weighted Utilitarian social welfare functions and found that the welfare gains from non-uniform commodity taxation are still sizable. We have also examined models with more than two wage levels. In some of these specifications, the welfare gains in fact turned out to be the greatest for weighted Utilitarianism, not for the Rawlsian case.

¹² Note that the last column in the tables refers to the marginal income tax rate, not the marginal effective income tax rate. For top earners the marginal effective income tax rate is zero. With positive commodity taxes, this can imply a negative marginal income tax rate.

¹³ It should also be noted that agent 1NP and 2P are pooled in the differentiated tax optimum. This should not be surprising given that these two agents have very similar wage rates once the child care expenses of the 2P-agents are taken into account.

In Table 5 we allow for differentiated commodity tax rates but now let the tax/subsidy on child care be optimized. As mentioned above, a 100% subsidy on child care is optimal. Thus the numerical results confirm proposition 2, namely that if the good needed in order to work is provided for free by the government, uniform taxation of other goods is optimal. Of course, the fact that τ_1 and τ_2 take on the value 0.20 is due to our normalization which mirrors the prevailing baseline VAT rate in the UK. The important thing to note from Table 5 is that the tax structure is uniform (the same social optimum could be achieved by setting $\tau_1 = \tau_2 = 0$ and properly adjusting the income tax schedule). Note that the welfare gain from the child care subsidy is very substantial and exceeds 4% of GDP. Thus an important lesson from the model is that the recommendations of the Mirrlees review, quoted in the introduction, that child care should be subject to “a low (perhaps zero) VAT rate” is incomplete. In fact, we find that child care should be publicly provided.

Table 5 (Full Optimum, Public Provision)

Type	Y	c_1	c_2	$T'(Y)$
w_{1P}	5.865	4.167	10.506	91%
w_{1NP}	12.457	4.317	11.854	49%
w_{2P}	19.871	4.464	13.172	65%
w_{2NP}	30.194	5.077	18.692	0%
$\Lambda / GDP = 4.29\% \quad \tau_2 = 0.20, \tau_1 = 0.20, \tau_c = -1.00$				

Tagging versus public provision

As mentioned above, with tagging there is in our model little to be gained from non-uniform tax rates. The reason for this is, within our modelling framework, that in the tagging optimum, there is only one self-selection constraint which could be mitigated by the differentiated tax scheme, namely the constraint linking the high skill parent and the low skill parent, within the tagged group of parents. Moreover, in this tagging optimum the low-skill parent works very little, and therefore his or her need for child care services is limited, implying that the disposable income of a true low-skilled individual and a mimicker are almost the same. This leaves little scope for beneficial tax differentiation. In the no-tagging regime, however, there is also the self-selection constraint linking parents with non-parents. It would seem, therefore, that the quantitatively significant benefits of commodity tax differentiation are limited to the case without tagging

within our model. But one must remember that this requires perfect tagging, and such policy is rarely obtained in practice.

Public provision of child care and tagging can under ideal conditions be close substitutes. However, in reality public provision and tagging can be quite different. Perhaps most importantly is that subsidized child care is self-targeting. The intended beneficiaries of subsidized child care are mainly secondary earners (women) with children in child care ages. In principle it would be possible to have a separate tax schedule for this group. However, we do not observe such tagging. Even in countries where there in principle is separate taxation of spouses, the transfer system is based on the joint income of the household. For example, in the UK the working tax credit is a tag applying to the household, not the individual. The tag applies to both the primary and secondary earner implying that the tag is far from perfect.

In reality there are also other groups than parents of young children who consume services linked to working hours. One prominent example is adults who take care of their elderly parents. Elderly care, and care of the functionally impaired, have strong similarities to child care. In many countries, it is the case that elderly persons are cared for by a near relative, like a daughter, daughter-in-law, son, or a (younger) spouse. As a concrete example, if a woman is responsible for the care of her elderly father, and this care requires, say, 8 hours a day, free elderly care would affect this woman's budget constraint in the same way as if she had a child and received free child care. To put things into perspective, in 2005 the Swedish government spent 4.6 percent of GDP on elderly care and 2.1 percent on child care. According to a computation in Blomquist et al. (2010) approximately one third of Swedish women in ages 50–65 are affected by subsidies to elderly care. In 2005, the employment rate of this group was around 70%. In comparison, the average for OECD Europe is 33 percent and in the US the corresponding figure is 56 percent. Thus, subsidies to elderly care, and care of the functionally impaired might well be one important reason why the labor force participation for women in Sweden in ages 50–65, with the exception of Iceland, is the highest in the OECD area.¹⁴ Thus the connection between hours of work and expenditures on elderly care is likely to be strong.¹⁵

¹⁴ Blomquist et al. (2010) report that there were around 290,000 persons who received some form of public elderly care, or care for the functionally impaired in Sweden in 2005. In the absence of publicly provided elderly care, each of these persons would have to be cared for by a close relative, like a daughter. This means that approximately one third of Swedish women in ages 50–65 might be classified as needing to purchase elderly care in order to work. Even though many elderly have the financial means to buy care themselves, they still might be cared for by a daughter (or son) since it might be financially more advantageous for the daughter or son to care of the elderly parent than to let him/her buy care for himself/herself if this would increase the future inheritance (see Bernheim et al. 1985 on this point).

¹⁵ For instance, Bonsang (2007) studies the extent to which adult children spend time caring for their elderly parents and finds a strong negative correlation between market work and time spent caring for an elderly parent.

In similarity to child care, public provision of elderly care is self-targeting as well, and we believe it would be very hard to construct the tax system so that the group of individuals that take care of an elderly person can be tagged perfectly via the income tax system. It is hard to capture the imperfections of tagging with a theoretical model, and we have considered two polar cases in terms of the sophistication of tagging schemes, but we leave open the possibility that what we have labelled as “no tagging” in fact might lie closer to real tax systems than the case with perfect tagging. Thus, it is possible that with realistic tagging schemes the welfare gains of tax differentiation still might be sizable.¹⁶

Finally, the differences in the income elasticities implied by our parameter assumptions are moderate, and we have only analysed two different tax rates. With more consumption categories, within which there would be also greater differences in income elasticities, one would probably end up with more variation to the optimal VAT rates. Then it could be the case that commodity tax differentiation becomes welfare improving also under (perfect) tagging. On balance, therefore, the range of differentiation implied by a more complete analysis could also be substantial.

5. Concluding remarks

A recommendation in the recent Mirrlees review is that commodity taxes should be uniform, with the exception that child care should not be taxed. This recommendation builds on results presented in a background chapter by Crawford, Keen and Smith. Although these authors find that leisure is not strictly separable from commodities, as a rough approximation it is. The Mirrlees review also recognizes that for many parents, child care is needed in order to be able to work. In this paper we study the implications of the preference structure presented in the Mirrlees review. However, the conclusions about the tax structure that we reach are quite different.

We find that when leisure is weakly separable from all commodities, except one commodity, a commodity that is needed in order to work, then if individuals pay for this commodity themselves, taxes on all the other commodities should be differentiated. There should be higher taxes on goods with large income elasticities. This recommendation is basically in accordance with the wisdom prevailing before the Mirrlees analysis. However, if the commodity needed in order to work is publicly provided, then uniform commodity taxation on all other goods is optimal.

¹⁶ In fact, it is the infeasibility of perfect tagging which motivates the need to use distortionary taxation in the first place.

It is interesting to note that in the traditional case for differentiated commodity taxes, i.e., when leisure is not weakly separable from goods, then those goods whose demand is increasing in the amount of leisure should be taxed more heavily. This is because mimickers have more leisure time. With the preference structure we study here demand is independent of leisure. Instead mimickers have more disposable income than those mimicked and to deter mimicking those goods whose demand increase in income should be taxed more heavily.

We also want to emphasize that, under weak separability, when the commodity needed in order to work is publicly provided; uniform taxation of other commodities is optimal, irrespective of the availability of tagging. In our view, this is a novel point: the optimal structure of commodity taxation depends on the expenditure side of the government and, in particular, on the extent of public provision. This could mean that the case for uniform commodity taxation is stronger in countries with extensive public provision (as in Scandinavia) than countries with a more limited public provision (such as the UK).

Our computational exercise, while clearly a first pass regarding the issue, suggest that the real-world importance of tax differentiation could be substantial if tagging is not used in the income tax system. We also pointed out, however, that we were only able to examine perfect tagging, and such policy is hard to achieve in the real world. In the end, the choice regarding whether the commodity tax should be uniform or not is an empirical matter. Correct decisions on the optimal mixed tax system would require a more complete simulation with a large number of agents and a sufficiently rich structure for commodity demand. Such analysis is clearly urgently needed in optimal tax research also more generally speaking, not only in connection with the present model. We leave these areas for further research.

References

- Akerlof, G. A. (1978) 'The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning', *American Economic Review* 68, 8-19.
- Atkinson, A. and J. Stiglitz (1976) 'The Design of Tax Structure: Direct vs Indirect Taxation', *Journal of Public Economics* 6, 55-75.
- Bastani, S., Blomquist, S. and L. Micheletto (2010) 'Public Provision of Private Goods, Tagging and Optimal Income Taxation with Heterogeneity in Needs,' CESifo Working Paper 3275, CESifo Group Munich.
- Bernheim, B., Shleifer, A. and L. Summers (1985) 'The Strategic Bequest Motive.' *Journal of Political Economy*, 93(6): 1045-76.

- Blomquist, S., V. Christiansen and L. Micheletto (2010) ‘Public Provision of Private Goods and Nondistortionary Marginal Tax Rates’ , *American Economic Journal: Economic Policy* 2(2), 1-27.
- Blundell, R. and A. Shephard (2012) ‘Employment, Hours of Work and the Optimal Taxation of Low Income Families’ , *Review of Economic Studies* 79(2), 481-510
- Bonsang, E. (2007) ‘How Do Middle-Aged Children Allocate Time and Money Transfers to Their Older Parents in Europe?’ *Empirica*, 34(2): 171–88.
- Browning, M. and C. Meghir (1991) ‘The Effects of Male and Female Labour Supply on Commodity Demands’ , *Econometrica* 59, 925-951.
- Chetty, R., A. Guren, D. S. Manoli, and A. Weber (2011) ‘Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities,’ NBER Working Papers 16729, National Bureau of Economic Research, Inc
- Christiansen, V. (1984) ‘Which commodity taxes should supplement the income tax?’ *Journal of Public Economics* 24, 195-220.
- Crawford, I., M. Keen, and S. Smith (2011), “Value Added Tax and Excises”, chapter four in *Dimensions of Tax Design: The Mirrlees Review*, Edited by Institute for Fiscal Studies (IFS) and James Mirrlees, Oxford University Press, Oxford.
- Edwards, S., M. Keen and M. Tuomala (1994) ‘Income Tax, Commodity Tax and Public Good Provision’ , *FinanzArchiv* 51, 472-487.
- Mirrlees, J. (1976) ‘Optimal Tax Theory: A Synthesis’ , *Journal of Public Economics* 6, 327-358.
- Mirrlees, J., S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba (2011) *Tax by Design: the Mirrlees Review*. Oxford University Press.
- Saez, E. (2002) The Desirability of Commodity Taxation Under Non-Linear Income Taxation and Heterogeneous Tastes, *Journal of Public Economics*, 83, 217-230.
- European Commission (2011) ‘A retrospective evaluation of elements of the EU VAT system’, TAXUD/2010/DE/328

Appendix

The case with tagging

The Lagrangean in the case of tagging and no public provision is:

$$\begin{aligned}
 (A1) \quad L = & \rho^{1,P} V^{1,P}(q, A^{1,P} - (\pi/w^{1P})Y^{1,P}, Y^{1,P}) + \rho^{2,P} V^{2,P}(q, A^{2,P} - (\pi/w^{2P})Y^{2,P}, Y^{2,P}) \\
 & + \rho^{1,NP} V^{1,NP}(q, A^{1,NP}, Y^{1,NP}) + \rho^{2,NP} V^{2,NP}(q, A^{2,NP}, Y^{2,NP}) \\
 & + \lambda^P \left[V^{2,P}(q, B^{2,P}, Y^{2,P}) - \hat{V}^{2,P}(q, \hat{B}^{1,P}, Y^{1,P}) \right] \\
 & + \lambda^{NP} \left[V^{2,NP}(q, B^{2,NP}, Y^{2,NP}) - \hat{V}^{2,NP}(q, \hat{B}^{1,NP}, Y^{1,NP}) \right] \\
 & + \gamma \left[\sum_h (Y^h - A^h) + \sum_h \sum_i t_i x_i^h (q, A^h - (\pi/w^h)Y^h, Y^h) - R \right]
 \end{aligned}$$

The first-order conditions are

$$(A2) \quad A^{2,NP} : (\rho^{2,NP} + \lambda^{NP}) V_B^{2,NP} = \gamma \left(1 - \sum_i t_i \frac{\partial x_i^{2NP}}{\partial B} \right)$$

$$(A3) \quad Y^{2,NP} : (\rho^{2,NP} + \lambda^{NP}) V_Y^{2,NP} = -\gamma \left(1 + \sum_i t_i \frac{\partial x_i^{2NP}}{\partial Y} \right)$$

$$(A4) \quad A^{1,NP} : \rho^{1,NP} V_B^{1,NP} = \lambda^{NP} \hat{V}_B^{2,NP} + \gamma \left(1 - \sum_i t_i \frac{\partial x_i^{1NP}}{\partial B} \right)$$

$$(A5) \quad Y^{1,NP} : \rho^{1,NP} V_Y^{1,NP} - \lambda^{NP} \hat{V}_Y^{2,NP} = -\gamma \left(1 + \sum_i t_i \frac{\partial x_i^{1NP}}{\partial Y} \right)$$

$$(A6) \quad A^{2,P} : (\rho^{2,P} + \lambda^P) V_B^{2,P} = \gamma \left(1 - \sum_i t_i \frac{\partial x_i^{2P}}{\partial B} \right)$$

$$(A7) \quad Y^{2,P} : (\rho^{2,P} + \lambda^P) V_Y^{2,P} - (1 + \lambda^P) V_B^{2,P} \frac{\pi}{w^{2P}} = -\gamma \left(1 + \sum_i t_i \frac{\partial x_i^{2P}}{\partial Y} - \sum_i t_i \frac{\partial x_i^{2P}}{\partial B} \frac{\pi}{w^{2P}} \right)$$

$$(A8) \quad A^{1,P} : \rho^{1,P} V_B^{1,P} = \lambda^{NP} \hat{V}_B^{2,P} + \gamma \left(1 - \sum_i t_i \frac{\partial x_i^{1P}}{\partial B} \right)$$

$$(A9) \quad Y^{1,P} : \rho^{1,P} V_Y^{1,P} - \lambda^P \hat{V}_Y^{2,P} - V_B^{1,P} \frac{\pi}{w^{1P}} + \lambda^P \hat{V}_B^{2,P} \frac{\pi}{w^{2P}} = -\gamma \left(1 + \sum_i t_i \frac{\partial x_i^{1P}}{\partial Y} - \sum_i t_i \frac{\partial x_i^{1P}}{\partial B} \frac{\pi}{w^{1P}} \right)$$

$$(A10) \quad q_k : \sum_h \rho^h V_{q_k}^h + \lambda^P (V_{q_k}^{2P} - \hat{V}_{q_k}^{2P}) + \lambda^{NP} (V_{q_k}^{2,NP} - \hat{V}_{q_k}^{2,NP}) + \gamma \sum_h x_k^h + \sum_h \sum_i \gamma_i \frac{\partial x_i^h}{\partial q_k} = 0$$

The commodity tax rule can be substituting from the Roy's identity $\frac{\partial V}{\partial q_k} = -x_k^h \frac{\partial V}{\partial B}$ and the

Slutsky equation $\frac{\partial x_i^h}{\partial q_k} = \frac{\partial s_i^h}{\partial q_k} - x_k^h \frac{\partial x_i^h}{\partial B}$ in (A10):

$$(A11) \quad \begin{aligned} & -x_k^{2NP} (\rho^{2,NP} + \lambda^{NP}) V_B^{2NP} - x_k^{2P} (\rho^{2,P} + \lambda^P) V_B^{2P} - x_k^{1NP} \rho^{1,NP} V_B^{1NP} - x_k^{1P} \rho^{1,P} V_B^{1P} \\ & + \lambda^{NP} \hat{x}_k^{2NP} \hat{V}_B^{2NP} + \lambda^P \hat{x}_k^{2P} \hat{V}_B^{2P} + \gamma \sum_h x_k^h + \gamma \sum_h \sum_i t_i \frac{\partial s_i^h}{\partial q_k} - \gamma \sum_h \sum_i t_i x_k^h \frac{\partial x_i^h}{\partial B} = 0 \end{aligned}$$

Substituting from (A2), (A4), (A6) and (A8), cancelling terms, and using the Slutsky symmetry

$\frac{\partial s_i^h}{\partial q_k} = \frac{\partial s_k^h}{\partial q_i}$ leads to the expression in equation (6) in the main text.

The case without tagging

Suppose first that the additional self-selection constraint that needs to be taken care of is (1NP, 2P). Then the Lagrangean, again without public provision, is rewritten as

$$(A12) \quad \begin{aligned} W = & \rho^{1,P} V^{1,P}(q, A^{1,P} - (\pi/w^{1P})Y^{1,P}, Y^{1,P}) + \rho^{2,P} V^{2,P}(q, A^{2,P} - (\pi/w^{2P})Y^{2,P}, Y^{2,P}) \\ & + \rho^{1,NP} V^{1,NP}(q, A^{1,NP}, Y^{1,NP}) + \rho^{2,NP} V^{2,NP}(q, A^{2,NP}, Y^{2,NP}) \\ & + \lambda^P \left[V^{2,P}(q, B^{2,P}, Y^{2,P}) - \hat{V}^{2,P}(q, \hat{B}^{1,P}, Y^{1,P}) \right] \\ & + \lambda^{NP} \left[V^{2,NP}(q, B^{2,NP}, Y^{2,NP}) - \hat{V}^{2,NP}(q, \hat{B}^{1NP}, Y^{1NP}) \right] \\ & + \delta \left[V^{1,NP}(q, B^{1,NP}, Y^{1,NP}) - \hat{V}^{1,NP}(q, \hat{B}^{2P}, Y^{2P}) \right] \\ & + \gamma \left[\sum_h (Y^h - A^h) + \sum_h \sum_i t_i x_i^h (q, A^h - (\pi/w^h)Y^h, Y^h) - R \right] \end{aligned}$$

And the first-order conditions (now for brevity only for after-tax income and the commodity tax) are:

$$(A13) \quad A^{2NP} : (\rho^{2,NP} + \lambda^{NP}) V_B^{2NP} = \gamma \left(1 - \sum_i t_i \frac{\partial x_i^{2NP}}{\partial B} \right)$$

$$(A14) \quad A^{1NP} : \rho^{1,NP} V_B^{1NP} = \lambda^{NP} \hat{V}_B^{2NP} - \delta V_B^{1NP} + \gamma \left(1 - \sum_i t_i \frac{\partial x_i^{1NP}}{\partial B} \right)$$

$$(A15) \quad A^{2P} : (\rho^{2,P} + \lambda^P)V_B^{2P} = \delta \hat{V}_B^{1NP} + \gamma \left(1 - \sum_i t_i \frac{\partial x_i^{2P}}{\partial B} \right)$$

$$(A16) \quad A^{1P} : \rho^{1,P}V_B^{1P} = \lambda^{NP}\hat{V}_B^{2P} + \gamma \left(1 - \sum_i t_i \frac{\partial x_i^{1P}}{\partial B} \right)$$

$$(A17) \quad q_k : \sum_h \rho^h V_{q_k}^h + \lambda^P (V_{q_k}^{2P} - \hat{V}_{q_k}^{2P}) + \lambda^{NP} (V_{q_k}^{2NP} - \hat{V}_{q_k}^{2NP}) + \delta (V_{q_k}^{1NP} - \hat{V}_{q_k}^{1NP}) \\ + \gamma \sum_h x_k^h + \sum_h \sum_i \gamma t_i \frac{\partial x_i^h}{\partial q_k} = 0$$

Using exactly the same procedure as in the case with no tagging leads to the rule in equation (8) in the main text.

Consider next the opposite case where the additional self-selection constraint is (2P, 1NP). All what changes is that the corresponding constraint in the Lagrangean is rewritten as $\delta [V^{2,P}(q, B^{2,P}, Y^{1,1NP}) - \hat{V}^{2,P}(q, \hat{B}^{1NP}, Y^{2P})]$. In the commodity tax rule, all other terms remain unchanged and in the last term the mimicker and the true type person change places.

Effective marginal tax rates

The total tax burden of the household is $\tau(Y^h) = T(Y^h) + \sum_i t_i x_i^h$. The effective marginal tax rate is

$$(A18) \quad \tau'(Y^h) = T'(Y^h) + \sum_i t_i \left[\frac{\partial x_i^h}{\partial Y} + \left(1 - T' - \frac{\pi}{w^h} \right) \frac{\partial x_i^h}{\partial B} \right]$$

Using the expression for the marginal income tax rate in (3), this equation can be rewritten as

$$(A19) \quad \tau'(Y^h) = \frac{V_Y^h}{V_B^h} + 1 - \frac{\pi}{w^h} + \sum_i t_i \left[\frac{\partial x_i^h}{\partial Y} - \frac{V_Y^h}{V_B^h} \frac{\partial x_i^h}{\partial B} \right] = \frac{V_Y^h}{V_B^h} \left(1 - \sum_i t_i \frac{\partial x_i^h}{\partial B} \right) + 1 - \frac{\pi}{w^h} + \sum_i t_i \frac{\partial x_i^h}{\partial Y}$$

For brevity we concentrate here for the expressions of effective marginal tax rate (EMTR) in the case of tagging. To obtain the EMTR for parents of type 2 for the case of tagging and no public provision, first substitute for $(1 + \lambda^P)V_B^2$ from (A5) in (A6) to get

$$(A20) \quad (\rho^{2,P} + \lambda^P)V_Y^{2P} = \gamma \frac{\pi}{w^{2P}} - \gamma \left(1 + \sum_i t_i \frac{\partial x_i^{2P}}{\partial Y} \right)$$

Dividing this by (A5) gives us

$$(A21) \quad \frac{V_Y^{2P}}{V_B^{2P}} = \frac{\frac{\pi}{w^{2P}} - \left(1 + \sum_i t_i \frac{\partial x_1^{2P}}{\partial Y}\right)}{1 - \sum_i t_i \frac{\partial x_1^{2P}}{\partial B}}$$

Combining this with right-hand side of (A19) indicates that the EMTR at the top for the parents in the case of tagging and no public provision should be zero. For non-parents, it is zero obviously, too.

However, when there is public provision, the expression in (A19) is as in the standard model as

$$(A22) \quad \tau'(Y^h) = \frac{V_Y^h}{V_B^h} \left(1 - \sum_i t_i \frac{\partial x_1^h}{\partial B}\right) + 1 + \sum_i t_i \frac{\partial x_1^h}{\partial Y}$$

The first-order conditions for after-tax income are unaltered. However, the first-order conditions for Y is, for example for the type 2 parent in the case of tagging as,

$$(A23) \quad Y^{2P} : (\rho^{2,P} + \lambda^P) V_Y^{2P} + \gamma \left(1 + \sum_i t_i \frac{\partial x_i^{2P}}{\partial Y} - \frac{\pi}{w^{2P}}\right) = 0,$$

where the last term appears because of the presence of the cost of child care in the government's budget constraint. Dividing this by (A6) and combining with (A22) yields $\tau'(Y^{2P}) = \pi/w^{2P} > 0$.

Similar corrective terms appear in the tax rules for other types as well.