

# How should commodities be taxed? A counter-argument to the recommendation in the Mirrlees Review

By Spencer Bastani<sup>a</sup>, Sören Blomquist<sup>a</sup>, and Jukka Pirttilä<sup>b</sup>

<sup>a</sup>Uppsala Center for Fiscal Studies at the Department of Economics, Uppsala University

<sup>b</sup>UNU-WIDER and School of Economics, University of Tampere, 33014 Finland; e-mail: jukka.pirttila@uta.fi

## 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 article 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, commodity taxes should discourage consumption of goods with large expenditure elasticities. Our results imply that the optimal commodity tax system is dependent on the expenditure side of the government. For instance, if the government fully subsidizes the cost of the good needed for working, then commodity taxation is uniform under the standard separability assumption. Calibration exercises suggest that these results can be quantitatively important.

**JEL classifications:** H21, H42

## 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 *et al.* (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).<sup>1</sup> 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 persons. A high-skill person choosing the same taxable income as a low-skill person will have more leisure time 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 their 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.<sup>2</sup>

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.* (2010), 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 given in the report 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 to work (such as child care or elder care), commodity taxation of those goods that are separable from labour supply should not be uniform.

We adopt the framework that lies behind the Mirrlees Review; as a first approximation leisure is weakly separable from all commodities except one. We think of this exception as,

- 1 Some of the other key contributions are Christiansen (1984), Edwards *et al.* (1994) and Saez (2002).
- 2 Boadway and Pestieau (2003) discuss the scope of the Atkinson-Stiglitz theorem. One reason it would not hold is differences in needs between persons or differences in endowments, as in Cremer *et al.* (2001). Our argument regarding child care can be seen as related to these contributions, but the modelling and the mechanism here are novel. Another closely related paper is Boadway *et al.* (1994), who study the case where differentiated commodity taxation can be used to curb tax evasion in a model where commodity taxation cannot be evaded whereas income taxation can. We return to the linkages between their paper and ours later.

for example, child care. Elder care is also a relevant example.<sup>3</sup> Thus we study commodity taxes under the assumption that the utility function takes the form  $U[F(x_1, x_2, \dots, x_n), G(x_c, I)]$ . 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 elder care) is subject to a zero VAT tax rate, corresponding to the Mirrlees Review recommendation. In the other case child care (or elder care) is publicly 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 formally analysing optimal commodity taxation in the presence of child care use. In this article 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 to work, and individuals pay for this commodity themselves, taxes on all the other commodities should be differentiated. The consumption of goods with large income elasticities should be discouraged by the tax system. This recommendation is basically in accordance with the wisdom prevailing long before the Mirrlees analysis. However, if the commodity needed to work is publicly provided free of charge, then uniform commodity taxation on all other goods is optimal. Thus our article 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 practical: 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 UK (in accordance with the recommendations in the Mirrlees Review) implies an implicit subsidy rate on child care that is far too small. In fact, in our simulation analysis, providing child care free of charge turns out to be the optimal policy.

The article is organized as follows. In Section 2 we present the model of individual behaviour. Section 3 outlines the government's problem and presents our main theoretical

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

findings in the tagging and no-tagging cases. In Section 4 we present our numerical exercise. Section 5 offers a broader discussion of tagging and presents elder care as another example of a good needed to work. Section 6 offers concluding remarks.

## 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 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:  $U[F(x_1, x_2, \dots, x_n), G(\xi x_c, l)]$  where  $l$  denotes labour supply and  $\xi = 1$  for users and  $\xi = 0$  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:

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

The formulation is equivalent to assuming that leisure is weakly separable from all commodities except  $x_c$ . For simplicity we will focus on the case  $f(l) = l$ .<sup>4</sup>

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 elder 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  $\pi$ . 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  $\phi$  is the subsidy rate 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

4 This formulation 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 the tax administration do not share information regarding working hours. The functional form is chosen for analytical ease; it could also be a more complicated one, as in [Blomquist \*et al.\* \(2010\)](#), without changing any of the qualitative results.

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

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

We denote by  $x^b$  the vector of consumption goods and by  $q$  the vector of consumer prices. Individual optimization is decomposed into a two-step process. In the first stage, the individual chooses  $x^b$  to maximize  $U^b[F(x^b), l^b]$  subject to the budget constraint (1) where disposable income  $B^b$ , pre-tax income  $Y^b$  and consumer prices  $q$  all are treated as given. This maximization yields the conditional indirect utility function  $V^b[q, B^b, Y^b] = U^b\{F[x^b(q, B^b)], \frac{Y^b}{w^b}\}$  where  $x_i^b[q, B^b]$  is the demand function for good  $i$ . Note that this demand function does not directly depend on working hours (or  $Y$ ) due to the separability assumption. 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  $Y^b$  to maximize  $V^b[q, B^b, Y^b] = V^b[q, Y^b - T(Y^b) - (1 - \phi)(\pi/w^b)Y^b, Y^b]$ .

### 3. Government's problem

We now proceed with the government's optimization 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  $T$  directly, the government assigns pre-tax and after-tax income points  $(Y^b, A^b)$  for each agent (the tax schedule can implicitly be calculated as  $T(Y^b) = Y^b - A^b$ ). The set of self-selection constraints ensure that each agent weakly prefers the income point assigned to him rather than the income point assigned to any other agent. The number of agents of each type is normalized to unity.

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 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 might think that tagging would be relatively straightforward: benefits and taxes should be made contingent on the family having a child in day care age. Within the context of our highly stylized model, such tagging would be easy. However, in practice it might be difficult. Consider a tax system with separate taxation of spouses. Should a tagging scheme be such that all parents, fathers and mothers, face the same tax schedule? In a society where child care is mainly the responsibility of mothers, this would imply a potentially unwanted tax advantage for fathers compared to non-fathers. Another possibility is for mothers to face a separate tax schedule. In many countries such a tax treatment would be regarded as politically controversial. Yet another possibility is to allow families to choose which adult should face the favourable tax schedule. It is clearly not obvious how tagging should be implemented in practice.

Another example of a good needed to work is elder care. For instance, a daughter taking care of an elderly parent might need elder care for her parent in order to work. In this case the ideal way to implement tagging in practice 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, a tax system which seems difficult to operationalize. In sum, it is important to note that there often is a difference between a fully optimal tagging and the type of tagging

schemes that are 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 four 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 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, the (2NP, 1NP) constraint is  $V^{2,NP}(q, B^{2,NP}, Y^{2,NP}) \geq \widehat{V}^{2,NP}(q, B^{1,NP}, Y^{1,NP})$  and the (2P, 1P) constraint  $V^{2,P}(q, B^{2,P}, Y^{2,P}) \geq \widehat{V}^{2,P}(q, \widehat{B}^{1,P}, Y^{1,P})$ . Here,  $\widehat{V}$  is used to denote the indirect utility of a mimicker and  $\widehat{B}$  is the amount of income the mimicker has available for private consumption. Note that  $\widehat{B}$  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 do not work the same number of hours.<sup>5</sup> For example, in the case of the (2P, 1P) self-selection constraint, the mimicking high-ability parent has a disposable income of  $B^{1P} = A^{1P} - (1 - \phi)(\pi/w^{2P})Y^{1P}$  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.

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

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

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

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

$$\sum_b (Y^b - A^b) + \sum_{b=1P, 2P} \sum_i t_i x_i^b(q, A^b - (\pi/w^b)Y^b) + \sum_{b=1NP, 2NP} \sum_i t_i x_i^b(q, A^b) = R \quad (3)$$

where  $R$  is an exogenous revenue requirement of the public sector. The Lagrange multipliers of the constraints (2NP, 1NP), (2P, 1P), and eq. (3) are denoted by  $\lambda^{NP}$ ,  $\lambda^P$ , and  $\gamma$ , respectively. The first-order conditions with respect to the commodity tax and the after-tax income are presented in Appendix 1. These are standard and similar to those in

5 In an alternative interpretation of the Mirrlees model where work intensity (instead of working hours) would be unobservable, the mimicker and a true low-ability type would not necessarily need different amounts of child care. This alternative interpretation would, however, also affect other optimal income tax models whose results depend on the relationship between commodity use and leisure.

Edwards *et al.* (1994). Using the first-order conditions the optimal commodity tax rule can be derived:

$$\sum_b \sum_i t_i \frac{\partial s_k^b}{\partial q_i} = \lambda^{*P} (x_k^{1P} - \hat{x}_k^{2P}) + \lambda^{*NP} (x_k^{1NP} - \hat{x}_k^{2NP}); k = 1, \dots, n \quad (4)$$

where  $\lambda^{*P} = (\lambda^P / \gamma) \hat{V}_B^{2P}$  and  $\lambda^{*NP} = (\lambda^{NP} / \gamma) \hat{V}_B^{2NP}$  are both positive and  $s_k^b$  denotes individual  $b$ 's compensated demand for good  $k$ . The left-hand side of eq. (4) is the aggregate compensated change (weighted by the commodity tax rates) 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 at the right of eq. (4) is zero. Now contrast the demand for good  $k$  between a mimicking parent of type 2,  $\hat{x}_k^{2P} [q, A^{1,P} - (\pi/w^{2,P})Y^{1,P}]$ , and a true type 1 parent,  $x_k^{1P} [q, A^{1,P} - (\pi/w^{1,P})Y^{1,P}]$ . 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, that is,  $x_k^{1P} < \hat{x}_k^{2P}$ . This tends to imply that the consumption of good  $k$  should be discouraged by the tax system. The extent of discouragement should be greater, other things being equal, the higher is the expenditure elasticity. Of course, the tax on a good for which the mimicker's demand is higher could still be low if other terms in the optimal tax formula eq. (4) drive the tax rate in the opposite direction. This could happen, for instance, if the compensated own-price elasticity of this good would be very high in comparison to other goods, rendering taxes on it highly distortive.<sup>6</sup> Thus, if there is one commodity that needs to be consumed to work, then even if the utility from leisure is separable from other commodities, non-uniform commodity taxation should in general be used. The tax rule involves a term that suggests that consumption of goods with higher income elasticities should be discouraged by the tax system. This also implies that the Atkinson-Stiglitz result does not hold.

### 3.2 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, and with two-dimensional heterogeneity, in principle, a full set of incentive constraints needs to be considered.

6 In our numerical example in Section 4, it turns out that the good which is according to the formula discouraged by the optimal tax system is indeed subject to the higher tax rate. The fact that one cannot, on theoretical grounds, immediately draw the conclusion that consumption of goods that are discouraged by the optimal tax system also should be subject to higher commodity tax rates, is related to the discussion in Boddaway *et al.* (1994). These authors study a completely different issue (tax evasion), but some of the mechanisms in the tax formulae have the same logic. In their paper, mimickers evade more taxes on income, and hence a higher tax rate on goods that the mimickers buy more can be a useful instrument to deter mimicking, but in the end the overall optimal tax formula depends also on other factors, and straightforward conclusions cannot be drawn.



We confine ourselves to cases where the type-2 non-parent is always the top income earner whom no one wants to mimic. We also suppose that the type-1 parent is the lowest income earner, and this type does not want to mimic others. This leaves us with seven possible self-selection constraints, which include the possibility that non-parents may want to mimic parents, and vice versa. For example, one needs to consider the constraint that type 1 non-parent does not want to mimic type 2 parent, that is, (1NP, 2P):  $V^{1,NP}(q, B^{1,NP}, Y^{1,NP}) \geq \widehat{V}^{1,NP}(q, \widehat{B}^{2,P}, Y^{2,P})$ .

In addition to the constraints mentioned above, the following can bind: (2NP,2P), (2NP,1P), (2P,1NP), and (1NP,1P). Suppose first that income levels are ordered so that  $Y^{1,P} < Y^{2,P} < Y^{1,NP} < Y^{2,NP}$  then it is likely that the (1NP,2P) constraint is one of the binding ones. If, alternatively, income levels are ordered so that  $Y^{1,P} < Y^{1,NP} < Y^{2,P} < Y^{2,NP}$ , then the (2P, 1NP) constraint is more likely to bind:  $V^{2,P}(q, B^{2,P}, Y^{2,P}) \geq \widehat{V}^{2,P}(q, \widehat{B}^{1,NP}, Y^{1,NP})$ .

Due to the large number of possible binding self-selection constraints, assume first for expositional reasons that income levels are ordered as  $Y^{1,P} < Y^{2,P} < Y^{1,NP} < Y^{2,NP}$  and the binding constraints are the same as in the tagging case (2NP,1NP), (2P,1P), and the new constraint (1NP, 2P). Let  $\delta$  be the Lagrange multiplier associated with the last constraint. The Lagrangean incorporating all three self-selection constraints is presented in Appendix 1. The rule for commodity taxation is in this case

$$\sum_b \sum_i t_i \frac{\partial s_k^b}{\partial q_i} = \lambda^{*P}(x_k^{1P} - \widehat{x}_k^{2P}) + \lambda^{*NP}(x_k^{1NP} - \widehat{x}_k^{2NP}) + \delta^*(x_k^{2P} - \widehat{x}_k^{1NP}), \quad k = 1, \dots, n \quad (5)$$

where  $\delta^* = (\delta/\gamma)\widehat{V}_B^{1NP} > 0$ . Similar arguments as above imply that  $x_k^{1NP} = \widehat{x}_k^{2NP}$  whereas  $x_k^{1P} < \widehat{x}_k^{2P}$  (if  $k$  is a normal good). These two terms together mean that the consumption of the good in question is likely to be discouraged (again supposing that other terms affecting the relative level of taxes across goods are kept the same). Now compare  $x_k^{2P}$  [ $q, A^{2,P} - (\pi/w^{2,P})Y^{2,P}$ ] and  $\widehat{x}_k^{1,NP}$  [ $q, A^{2,P}$ ]. Since the mimicker does not have to buy child care, the money left to buy good 1 is greater for him ( $B^{2P} < \widehat{B}^{1NP}$ ). If good  $k$  is normal, this means that  $x_k^{2P} < \widehat{x}_k^{1NP}$ . This term also works towards discouraging the consumption of good  $k$ . In this case, a positive effective tax helps 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 be discouraged more heavily relative to other goods.

The commodity tax rule would give rise to similar recommendations even if, in addition, the self-selection constraints (2NP,2P), (2NP,1P), and (1NP, 1P) would bind. In all these cases, the mimicker is a non-parent and does not need to buy child care and his or her money left to buy other goods is greater than that of the mimicked person.

If instead the constraint (2P, 1NP) is among the binding ones, then the commodity tax could look like the following (again letting  $\delta$  be the Lagrange multiplier associated with this constraint):

$$\sum_b \sum_i t_i \frac{\partial s_k^b}{\partial q_i} = \lambda^{*P}(x_k^{1P} - \widehat{x}_k^{2P}) + \lambda^{*NP}(x_k^{1NP} - \widehat{x}_k^{2NP}) + \delta^\#(x_k^{1NP} - \widehat{x}_k^{2P}), \quad k = 1, \dots, n \quad (6)$$

where  $\delta^\# = (\delta/\gamma)\widehat{V}_B^{2P} > 0$ . Consider the third term on the right-hand side of eq. (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, for example, those of the form (1NP, 2P), where a non-parent mimics a parent and those 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 for work and agents pay for this good themselves, uniform taxation is not optimal, contrasting the Mirrlees Review recommendation.

*Proposition 2* If the individual optimisation problem satisfies Assumption 1, and tagging is feasible and used, the consumption of commodities with higher income elasticities should be discouraged by the tax system. If tagging is not used or is infeasible, depending on which self-selection constraints bind in the government's optimum, the consumption of commodities with higher income elasticities should either be encouraged or discouraged by the tax system.

Proposition 2 highlights the fact, that our argument for differentiated commodity taxation differs from the traditional case, that is, the argument stating that when leisure is not weakly separable from goods, those goods whose demand is increasing in the amount of leisure should be taxed more heavily. The rationale for the standard argument for differentiated commodity taxation is that mimickers have more leisure time. Our argument is different since we assume a preference structure where commodity demand is independent of leisure. Instead, in our setting with tagging, mimickers have a higher disposable income than those mimicked, which is an argument working towards discouraging the consumption of goods whose income elasticity of demand is larger.

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

Underlying Propositions 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 persons, including mimickers;  $A^b = B^b$  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:

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

This constraint replaces the original budget constraint in the government optimization 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 eq. (5) continues to hold.<sup>7</sup> This leads to the following result:

*Proposition 3* If the individual optimisation 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: whether uniform commodity taxation is optimal 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 of goods used in conjunction with labour supply (as in Scandinavia) than countries with a more limited public provision (such as the UK).

### 3.4 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 individual's indifference curve in the  $(Y, B)$ -space at the individual optimum. The second stage of an individual's maximisation problem presented at the end of Section 2 implies that  $V^b[q, B^b, Y^b]$  is maximized with respect to  $Y^b$  where  $B^b = Y^b - T(Y^b) - (\pi/w^b)Y^b$ . The first-order condition for this maximisation yields the following (implicit) expression for the marginal tax rate:

$$T'(Y^b) = 1 + \frac{V_Y^b}{V_B^b} - \frac{\pi}{w^b} \quad (8)$$

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 Appendix 1 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  $\pi/w^b$  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 which 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.

7 Instead, as emphasized by Blomquist *et al.* (2010), the presence of public provision affects the optimal structure of labour income 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 elder care). We perform four sets of simulations. The first, which we take as our benchmark, considers the polar case where there is uniform taxation on all goods except child care, which has a zero VAT. This is the tax system recommended in the Mirrlees Review. In our theoretical analysis we showed that under the preference structure considered it is welfare enhancing to use differentiated commodity taxation on other goods in the case where child care faces a zero VAT. We perform simulations to see how large these gains can be. We also perform simulations for our second polar case with child care provided free of charge and uniform taxes on all other commodities. Although we did not study the full optimum in the theory section, in our simulations we consider this case. We find that the polar case with child care provided free of charge is the fully optimal tax system.

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, not to provide a full-fledged numerical analysis of optimal commodity taxation.<sup>8</sup>

Our theoretical analysis has highlighted three factors of relevance for our tax differentiation result. First, as indicated by eq. (4), the degree of tax differentiation depends on the differences in the relative demand of other goods than child care between true low-ability types and mimickers. This difference, in turn, hinges on how sizeable a fraction of total consumption child care costs represent. According to the OECD Family Database and the report ‘Doing Better for Families’ (2011), child care costs in the UK for an average worker are estimated to lie around 25%, as a fraction of net family income, which is amongst the highest in the OECD. Hence child care costs are a quantitatively important portion out 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. 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. Third, unless income tax schedules can be tagged, there might be non-standard self-selection constraints that occur when the mimicker has a lower disposable income than the agent being mimicked.<sup>9</sup> 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

8 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).

9 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).

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

$$U(x_1, x_2, l) = (x_1 - x_{10})^{\beta_1} (x_2 - x_{20})^{\beta_2} - \frac{l^\varepsilon}{\varepsilon} \quad (9)$$

where  $\beta_1 + \beta_2 = 1$ ,  $\beta_i > 0$ ,  $i = 1, 2$ , and  $x_1 > x_{10}$ ,  $x_2 > x_{20}$ . The budget constraint is  $q_1 x_1 + q_2 x_2 = B$  where disposable income is  $B = A - \pi(Y/w)$  for parents, reflecting that one unit of child care must be purchased for each hour of work, and  $B = A$  for non-parents. Given pre-tax income  $Y$  and disposable income  $B$ , an individual chooses  $x_1$  and  $x_2$  to maximize utility subject to the budget constraint. This yields commodity demands

$$x_1(q_1, q_2, B) = x_{10} + \beta_1 \frac{B - (q_1 x_{10} + q_2 x_{20})}{q_1} \quad (10)$$

$$x_2(q_1, q_2, B) = x_{20} + \beta_2 \frac{B - (q_1 x_{10} + q_2 x_{20})}{q_2} \quad (11)$$

and when then optimized demands are inserted back into the utility function, the indirect utility function takes the form

$$V(q_1, q_2, B, Y) = [x_1(q_1, q_2, B) - x_{10}]^{\beta_1} [x_2(q_1, q_2, B) - x_{20}]^{\beta_2} - \frac{(Y/w)^\varepsilon}{\varepsilon} \quad (12)$$

The individual purchases the minimum amounts  $x_{10}$  and  $x_{20}$  of each good and spends the remaining income  $B - (q_1 x_1 + q_2 x_2)$  on goods 1 and 2 in the proportions  $\beta_1$  and  $\beta_2$ , respectively.<sup>10</sup> The expenditure elasticities, denoted  $\eta_1$  and  $\eta_2$ , are:

$$\eta_1 = \frac{\partial x_1}{\partial B} \frac{B}{x_1} = \beta_1 \frac{q_1 x_1}{B_1} = \frac{\beta_1}{s_1}, \quad (13)$$

$$\eta_2 = \frac{\partial x_2}{\partial B} \frac{B}{x_2} = \beta_2 \frac{q_2 x_2}{B_2} = \frac{\beta_2}{s_2}, \quad (14)$$

where  $s_i$  is the fraction of an individuals' income spent on good  $i = 1, 2$ . We assume that good 1 represents 'basic needs' such as food and shelter and that good 2 represents 'other goods'. We therefore set  $x_{10} > 0$  and  $x_{20} = 0$ . 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,  $\eta_1$  converges to 1 from below and  $\eta_2$  converges to 1 from above.<sup>11</sup>

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

11 To see this, one can refer to the expressions  $\eta_1 = \beta_1 B / [\beta_1 B + (1 - \beta_1) q_1 x_{10}]$  and  $\eta_2 = B / [1 - q_1 x_{10}]$ . Note that in the Cobb-Douglas case,  $s_1 = \beta_1$  and  $s_2 = \beta_2$  implying  $\eta_1 = \eta_2 = 1$ .

**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	

Source: Own calculations based on the UK Family Resources Survey.

We now present the principles in the calibration that we have chosen. We allow for two different skill levels (low and high) 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 individuals categorized as ‘non-parents’ 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 price 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.<sup>12</sup> The wage rates are reported in Table 1.

To perform simulations the values of all parameters in the utility function (10) must be specified. We set  $\varepsilon = 4$  implying a labour supply 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 of labour supply elasticities, surveyed in Chetty *et al.* (2013). The value is also consistent with the small intensive-margin labour 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  $x_{10} = 3$ ,  $x_{20} = 0$  which in the benchmark no-tagging model 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 for 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

12 In our static model it could perhaps be considered more appropriate to calculate these costs as a percentage of life-time income, suggesting that the importance of child care costs would decline. On the other hand, taking a life-time perspective, a greater proportion of people would be classified as parents. Therefore, adopting a life-cycle perspective would not necessarily diminish the role of child care costs in our model.

**Table 2** Income elasticities of demand for the UK

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

Source: Reproduced from Table J.9, p. 525 in TAXUD/2010/DE/328.

**Table 3** Benchmark allocation: Mirrlees Review recommendation

Type	Y	$c_1$	$c_2$	$T'(Y)$	EMTR	Utility
$w_{1P}$	0.000	4.068	9.608	49%	57%	7.713
$w_{1NP}$	12.709	4.238	11.144	36%	46%	7.713
$w_{2P}$	0.000	4.068	9.608	71%	76%	7.713
$w_{2NP}$	30.194	4.969	17.725	-20%	0%	8.775
$\tau_2 = 0.20, \tau_1 = 0.20, \tau_c = 0$						

Source: Own calculations.

equivalent-variation type of welfare gain measure of policy reform we proceed as follows. We calculate the minimum amount of extra revenue  $\Lambda$  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  $\Lambda$  is divided by aggregate income in the pre-reform economy to obtain a welfare gain measure expressed in terms of percentage points of GDP.

In our first analysis we consider the polar case where the tax rate on child care is set to 0%, in line with the Mirrlees Review recommendation, and analyse 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.

In the tables we present the optimal allocation together with the marginal tax rates, the effective marginal tax rates (EMTR), and the commodity tax rates.<sup>13</sup> The last column

13 Note that we have defined the marginal tax rate as one minus the MRS, following Stiglitz (1982). Thus, even though type  $w_{1P}$  and  $w_{2P}$  in Table 3 are pooled at zero income, their marginal tax rates are different since the slopes of their indifference curve are different at this point. The same type of reasoning explains why type  $w_{1NP}$  and  $w_{2P}$  in Table 4 have different marginal tax rates even though they are pooled together.

**Table 4** Differentiated commodity tax optimum

Type	Y	$c_1$	$c_2$	$T'(Y)$	EMTR	Utility
$w_{1P}$	0.000	4.617	9.563	49%	71%	8.006
$w_{1NP}$	14.520	5.042	12.073	-40%	19%	8.006
$w_{2P}$	14.520	4.689	9.989	47%	69%	8.006
$w_{2NP}$	30.105	6.072	18.162	-74%	0%	9.816
$\Lambda/GDP = 2.23\%$	$\tau_2 = 0.83, \tau_1 = 0.20, \tau_c = 0$					

Source: Own calculations.

of each table shows the utility levels. Note that for top earners the marginal effective income tax rate is always zero. With positive commodity taxes, this can imply a negative marginal income tax rate.<sup>14</sup>

The results in Table 4 imply that the optimal 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%).<sup>15</sup> One should note, however, that the exact magnitude of the tax differential depends on what is assumed of the tax rates for other commodities. The monetary welfare gain following from tax differentiation equals 2.23% of GDP, and moving from uniform to differentiated commodity taxation also entails a Pareto improvement. In 2010 figures the overall welfare gain amounts to approximately 30 billion GBP. This means in the context of this simple model that the administrative costs of having two instead of one VAT rate would need to exceed this figure for uniform taxation to be preferred. With modern information technology, this seems to us as a rather high cost level.

In the foregoing 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 max-min 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 smaller but still sizeable.<sup>16</sup>

In Table 5 we show the simulation results for our second polar case where child care is provided free of charge and there is a uniform tax. The fact that we in the simulations have set the uniform commodity tax rate to 0.2 simply mirrors the prevailing baseline VAT rate in the UK. We could equally well set the commodity tax rate to zero, the allocation would still be the same. We see that the welfare gain as compared to the benchmark case is almost 5%.

We have also performed simulations where we search for the full tax optimum, that is, the tax/subsidy on child care and the other commodity tax rates are optimized simultaneously. We do this under the constraint that the subsidy on child care cannot be larger than

14 Note that the fact that the first three utility levels are the same is a consequence of the government maximizing a max-min social welfare function.

15 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.

16 In Appendix 2 we present simulations regarding a model with parents only but with more skill levels and there the use of weighted utilitarianism produces significant welfare gains.



**Table 5** Public provision (full optimum)

Type	$Y$	$c_1$	$c_2$	$T'(Y)$	EMTR	Utility
$w_{1P}$	5.865	4.167	10.506	90%	91% (40%)	8.343
$w_{1NP}$	12.457	4.317	11.854	39%	49%	8.378
$w_{2P}$	19.871	4.464	13.172	58%	65% (36%)	9.323
$w_{2NP}$	30.194	5.077	18.692	-20%	0%	9.551
$\Lambda/GDP = 4.78\%$	$\tau_2 = 0.20, \tau_1 = 0.20, \tau_c = -1.00$					

Source: Own calculations. Distortionary component of effective marginal tax rate in parenthesis.

100%; this bound is imposed because otherwise buyers and sellers of child care services could collude.<sup>17</sup> We find that given our parameterization of the model economy, a 100% subsidy is optimal.<sup>18</sup> In fact, an even larger subsidy would be desirable, but 100% is an upper bound. This implies that the polar case shown in Table 5 is also the full optimum. The intuition follows that in Blomquist *et al.* (2010): child care subsidies that are paid back via the income tax not only cover the child care costs but deter mimicking as child care services are less valuable for mimickers.

It is worth emphasizing that the welfare gain from the child care subsidy is very substantial, amounting to almost 5% of GDP. Thus an important lesson from our numerical exercise 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’ implies a too small subsidization of child care. In fact, we find that child care should be publicly provided. Finally, note that in the next to last column of Table 5 we have indicated the so-called distortionary component of the effective marginal tax rate in parenthesis. These numbers have been obtained by subtracting the non-distortionary component  $\pi/w$ , as defined by Blomquist *et al.* (2010), from the effective marginal tax rate. It reflects that a substantial part of the effective marginal tax rate facing parents, is corrective, forcing agents to internalize the cost of child care incurred in the government budget constraint when child care is publicly provided.

An interesting further result is that in the uniform commodity tax case of Table 3, neither parent works, whereas one of them works in the partial optimum of Table 4 and both work in the full optimum of Table 5. Finally note that moving from the differentiated commodity tax optimum (Table 4) to the full optimum (Table 5), we also attain a Pareto improvement.

Whilst our analysis dealt with a simplified set-up with two ability types, we have also examined the robustness of the results with respect to the number of skill levels and the wage calibration strategy. This analysis, presented in Appendix 2, focusses only on parents but allows for a different number of skill types. The results of our exercise reveal that the welfare gains that we find in the two-type case do not vanish when the number of types is increased. The results also imply that in a model with parents only, moving from uniform commodity taxation with a zero tax on child care to public provision of child care

17 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.

18 In principle in this specification without tagging, subsidies on child care could tighten some self-selection constraints linking non-parents to parents, rendering less than 100% subsidy optimal.

would entail a Pareto improvement. Our conclusion is therefore that the theoretical mechanism that we have focused on in this article can indeed be quantitatively significant. In a model with non-parents as well, the extent of optimal tax differentiation could be lower as it is optimal from the non-parents' point of view to have uniform commodity taxation.

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 could 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. Discussion

### 5.1 Tagging versus public provision

As mentioned, with tagging there is in our model little to be gained from non-uniform tax rates. The reason for this is that within our modelling framework, 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 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. One must remember that this requires perfect tagging, and such policy is rarely obtained in practice.

If the objective of the government is to support families with small children, public provision of child care and tagging (of parents with small children) can under certain conditions be close substitutes. However, in reality public provision and tagging can be quite different. Perhaps most important 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 in principle there 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. But the benefits of public provision can also be shared within the households. Therefore, both instruments are imperfect in reaching out only those who need child-care services to work.

### 5.2 Elder care as another example

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. Elder 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, eight hours a day, free elder care would affect this woman's budget constraint in a similar 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% of GDP on elderly care and 2.1% 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 elder care. In 2005, the employment rate of this group was around 70%. In comparison, the average for OECD Europe is 33% and in the USA the corresponding figure is 56%. Thus, subsidies to elder care and care of the functionally impaired might well be one important reason the labour force participation for women in Sweden in ages 50–65, with the exception of Iceland, is the highest in the OECD area.<sup>19</sup> Thus the connection between hours of work and expenditures on elder care is likely to be strong.<sup>20</sup>

In similarity to child care, public provision of elder 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 sizeable.<sup>21</sup>

## 6. 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 \*et al.\* \(2010\)](#). 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 to be able to work. In this article 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.

- 19 [Blomquist \*et al.\* \(2010\)](#) report that in Sweden in 2005, there were around 290,000 persons who received some form of public elder care or care for the functionally impaired. In the absence of publicly provided elder 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 elder care 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 adult child to care for the elderly parent than to let him or her buy care for him- or herself if this would increase the future inheritance (see [Bernheim \*et al.\* 1985](#) on this point).
- 20 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.
- 21 In fact, it is the infeasibility of perfect tagging (i.e. personalized lump-sum taxes) that motivates the need to use distortionary taxation in the first place.

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

It is interesting to note that in the traditional case for differentiated commodity taxes, that is, 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 the demand of those goods whose demand increase in income should be discouraged by commodity taxes.

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, 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 of goods used in conjunction with labour supply (as in Scandinavia) than in countries with a more limited public provision (such as the UK).

Our computational exercise, whilst 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 that we were only able to examine perfect tagging, and such a 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 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.

## Supplementary material

[Supplementary material](#) is available online at the OUP website.

## Acknowledgements

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). Comments from thoughtful referees led to considerable insights for which we are thankful. Mike Brewer and Andrew Shephard gave valuable advice regarding UK data sources.

## Funding

Riksbankens Jubileumsfond (S2007-1304:1-E, RS10-1276:1) and the Jan Wallander and Tom Hedelius Foundation (W2012-0438:1).

## 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 Stiglitz, J. (1976) The design of tax structure: direct vs indirect taxation, *Journal of Public Economics* **6**, 55–75.
- Bastani, S. (2013), Using the discrete model to derive optimal income tax rates, Uppsala Center for Fiscal Studies Working Paper No 2013:11, Department of Economics, Uppsala University.
- Bastani, S., Blomquist, S., and Micheletto L. (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 Summers L. (1985) The strategic bequest motive, *Journal of Political Economy*, **93**: 1045–76.
- Blomquist, S., Christiansen, V., and Micheletto, L. (2010) Public provision of private goods and nondistortionary marginal tax rates, *American Economic Journal: Economic Policy*, **2**, 1–27.
- Blundell, R. and Shephard, A. (2012) Employment, hours of work and the optimal taxation of low income families, *Review of Economic Studies*, **79**, 481–510.
- Boadway, R., Marchand, M., and Pestieau, P. (1994) Towards a theory of the direct-indirect tax mix, *Journal of Public Economics*, **55**, 71–88.
- Boadway, R. and Pestieau, P. (2003) Indirect taxation and redistribution: the scope of the Atkinson-Stiglitz theorem, in R. Arnott, B. Greenwald, R. Kanbur, and B. Nalebuff (eds) *Economics for an Imperfect World: Essays in Honor of Joseph Stiglitz*, MIT Press, Cambridge, MA.
- Bonsang, E. (2007) How do middle-aged children allocate time and money transfers to their older parents in Europe?, *Empirica*, **34**, 171–88.
- Browning, M. and Meghir, C. (1991) The effects of male and female labour supply on commodity demands, *Econometrica*, **59**, 925–51.
- Chetty, R., Guren, A., Manoli, D. S., and Weber, A. (2013) Does indivisible labour explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities, *NBER Macroeconomics Annual 2012*, **27**, 1–56.
- Christiansen, V. (1984) Which commodity taxes should supplement the income tax? *Journal of Public Economics*, **24**, 195–220.
- Crawford, I., Keen, M., and Smith, S. (2010), Value added tax and excises, in Institute for Fiscal Studies and James Mirrlees (eds) *Dimensions of Tax Design: The Mirrlees Review*, Oxford University Press, Oxford.
- Cremer, H., Pestieau, P. and Rochet, J.-C. (2001) Direct vs indirect taxation: the design of the tax structure revisited, *International Economic Review*, **42**, 781–99.
- Edwards, S., Keen, M. and Tuomala, M. (1994) Income tax, commodity tax and public good provision, *FinanzArchiv*, **51**, 472–87.
- Mirrlees, J. (1976) Optimal tax theory: a synthesis, *Journal of Public Economics*, **6**, 327–58.
- Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, B., Chote, R., Gammie, M., Johnson, P., Myles, G., and Poterba, J. (2011) *Tax by Design: The Mirrlees Review*, Oxford University Press, Oxford.
- OECD (2011) *Doing Better for Families*, OECD, Paris.
- Saez, E. (2002) The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes, *Journal of Public Economics*, **83**, 217–30.
- Stiglitz, J. E. (1982) Self-selection and Pareto efficient taxation, *Journal of Public Economics*, **17**, 213–40.

### Appendix 1. Some mathematical details

#### 1. The case with tagging

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

$$\begin{aligned}
 W = & \rho^{1,P} V^{1,P}(q, A^{1,P} - (\pi/w^{1,P}) Y^{1,P}, Y^{1,P}) + \rho^{2,P} V^{2,P}(q, A^{2,P} - (\pi/w^{2,P}) Y^{2,P}, Y^{2,P}) \\
 & + \rho^{1,NP} V^{1,NP}(q, A^{1,NP}, Y^{1,NP}) + \rho^{2,NP} V^{2,NP}(q, A^{2,NP}, Y^{2,NP}) \\
 & + \lambda^P \left[ V^{2,P}(q, B^{2,P}, Y^{2,P}) - \widehat{V}^{2,P}(q, \widehat{B}^{1,P}, Y^{1,P}) \right] \\
 & + \lambda^{NP} \left[ V^{2,NP}(q, B^{2,NP}, Y^{2,NP}) - \widehat{V}^{2,NP}(q, B^{1,NP}, Y^{1,NP}) \right] \\
 & + \gamma \left[ \sum_b (Y^b - A^b) + \sum_{b=1P,2P} \sum_i t_i x_i^b(q, A^b - (\pi/w^b) Y^b) \right. \\
 & \left. + \sum_{b=1NP,2NP} \sum_i t_i x_i^b(q, A^b) - R \right]
 \end{aligned} \tag{A.1}$$

The first-order conditions are

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

$$Y^{2NP} : (\rho^{2,NP} + \lambda^{NP}) V_Y^{2NP} = -\gamma \tag{A.3}$$

$$A^{1NP} : \rho^{1,NP} V_B^{1NP} - \lambda^{NP} \widehat{V}_B^{2NP} = \gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{1,NP}}{\partial B} \right) \tag{A.4}$$

$$Y^{1NP} : \rho^{1,NP} V_Y^{1NP} - \lambda^{NP} \widehat{V}_Y^{2NP} = -\gamma \tag{A.5}$$

$$A^{2P} : (\rho^{2,P} + \lambda^P) V_B^{2P} = \gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{2,P}}{\partial B} \right) \tag{A.6}$$

$$Y^{2P} : (\rho^{2,P} + \lambda^P) V_Y^{2P} - (\rho^{2,P} + \lambda^P) V_B^{2P} \frac{\pi}{w^{2,P}} = -\gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{2,P}}{\partial B} \frac{\pi}{w^{2,P}} \right) \tag{A.7}$$

$$A^{1P} : \rho^{1,P} V_B^{1P} = \lambda^P \widehat{V}_B^{2P} + \gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{1,P}}{\partial B} \right) \tag{A.8}$$

$$Y^{1P} : \rho^{1,P} V_Y^{1P} - \lambda^P \widehat{V}_Y^{2P} - \rho^{1,P} V_B^{1P} \frac{\pi}{w^{1,P}} + \lambda^P \widehat{V}_B^{2P} \frac{\pi}{w^{2,P}} = -\gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{1,P}}{\partial B} \frac{\pi}{w^{1,P}} \right) \tag{A.9}$$

$$q_k : \sum_b \rho^b V_{q_k}^b + \lambda^P (V_{q_k}^{2P} - \widehat{V}_{q_k}^{2P}) + \lambda^{NP} (V_{q_k}^{2NP} - \widehat{V}_{q_k}^{2NP}) + \gamma \sum_b x_k^b + \sum_b \sum_i \gamma t_i \frac{\partial x_i^b}{\partial q_k} = 0 \tag{A.10}$$

The commodity tax rule can be obtained by substituting from the Roy's identity  $\frac{\partial V^h}{\partial q_k} = -x_k^h \frac{\partial V^h}{\partial B}$  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 (A.10):

$$\begin{aligned}
 & -x_k^{2NP}(\rho^{2,NP} + \lambda^{NP})V_B^{2NP} - x_k^{2P}(\rho^{2,P} + \lambda^P)V_B^{2P} - x_k^{1NP}\rho^{1,NP}V_B^{1NP} - x_k^{1P}\rho^{1,P}V_B^{1P} \\
 & + \lambda^{NP}x_k^{2NP}\widehat{V}_B^{2NP} + \lambda^P x_k^{2P}\widehat{V}_B^{2P} + \gamma \sum_b x_k^b + \gamma \sum_b \sum_i t_i \frac{\partial s_i^b}{\partial q_k} - \gamma \sum_b \sum_i t_i x_k^b \frac{\partial x_i^b}{\partial B} = 0
 \end{aligned} \tag{A.11}$$

Substituting from eqs (A.2), (A.4), (A.6), and (A.8), cancelling terms, and using the Slutsky symmetry  $\frac{\partial s_i^b}{\partial q_k} = \frac{\partial s_k^b}{\partial q_i}$  leads to the expression in eq. (4) in the main text.

### 2. 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

$$\begin{aligned}
 W = & \rho^{1,P}V^{1,P}(q, A^{1,P} - (\pi/w^{1,P})Y^{1,P}, Y^{1,P}) + \rho^{2,P}V^{2,P}(q, A^{2,P} - (\pi/w^{2,P})Y^{2,P}, Y^{2,P}) \\
 & + \rho^{1,NP}V^{1,NP}(q, A^{1,NP}, Y^{1,NP}) + \rho^{2,NP}V^{2,NP}(q, A^{2,NP}, Y^{2,NP}) \\
 & + \lambda^P \left[ V^{2,P}(q, B^{2,P}, Y^{2,P}) - \widehat{V}^{2,P}(q, \widehat{B}^{1,P}, Y^{1,P}) \right] \\
 & + \lambda^{NP} \left[ V^{2,NP}(q, B^{2,NP}, Y^{2,NP}) - \widehat{V}^{2,NP}(q, B^{1,NP}, Y^{1,NP}) \right] \\
 & + \delta \left[ V^{1,NP}(q, B^{1,NP}, Y^{1,NP}) - \widehat{V}^{1,NP}(q, \widehat{B}^{2,P}, Y^{2,P}) \right] \\
 & + \gamma \left[ \sum_b (Y^b - A^b) + \sum_{b=1P,2P} \sum_i t_i x_i^b(q, A^b - (\pi/w^b)Y^b) \right. \\
 & \left. + \sum_{b=1NP,2NP} \sum_i t_i x_i^b(q, A^b) - R \right]
 \end{aligned} \tag{A.12}$$

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

$$A^{2NP} : (\rho^{2,NP} + \lambda^{NP})V_B^{2NP} = \gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{2NP}}{\partial B} \right) \tag{A.13}$$

$$A^{1NP} : \rho^{1,NP}V_B^{1NP} = \lambda^{NP}\widehat{V}_B^{2NP} - \delta V_B^{1NP} + \gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{1NP}}{\partial B} \right) \tag{A.14}$$

$$A^{2P} : (\rho^{2,P} + \lambda^P)V_B^{2P} = \delta \widehat{V}_B^{1NP} + \gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{2P}}{\partial B} \right) \tag{A.15}$$

$$A^{1P} : \rho^{1,P}V_B^{1P} = \lambda^{NP}\widehat{V}_B^{2P} + \gamma \left( 1 - \sum_i t_i \frac{\partial x_i^{1P}}{\partial B} \right) \tag{A.16}$$



$$\begin{aligned}
 q_k : \sum_b \rho^b V_{q_k}^b + \lambda^P (V_{q_k}^{2P} - \widehat{V}_{q_k}^{2P}) + \lambda^{NP} (V_{q_k}^{2NP} - \widehat{V}_{q_k}^{2NP}) + \delta (V_{q_k}^{1NP} - \widehat{V}_{q_k}^{1NP}) \\
 + \gamma \sum_b x_k^b + \sum_b \sum_i \gamma t_i \frac{\partial x_i^b}{\partial q_k} = 0
 \end{aligned}
 \tag{A.17}$$

Using the same procedure as in the case with no tagging leads to the rule in eq. (5) 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  $\delta[V^{2,P}(q, B^{2,P}, Y^{1,NP}) - \widehat{V}^{2,P}(q, \widehat{B}^{1NP}, Y^{2P})]$ . In the commodity tax rule, all other terms remain unchanged, and in the last term the mimicker and the true type person change places.

### 3. Effective marginal tax rates

The total tax burden of the household is  $\tau(Y^b) = T(Y^b) + \sum_i t_i x_i^b$ . Recognizing that a marginal increase in  $Y^b$  induces a change in disposable income of  $1 - T' - \frac{\pi}{w^b}$ , the effective marginal tax rate is

$$\tau'(Y^b) = T'(Y^b) + \sum_i t_i \left(1 - T' - \frac{\pi}{w^b}\right) \frac{\partial x_i^b}{\partial B}
 \tag{A.18}$$

Using the expression for the marginal income tax rate in eq. (8), this equation can be rewritten as

$$\tau'(Y^b) = \frac{V_Y^b}{V_B^b} \left(1 - \sum_i t_i \frac{\partial x_i^b}{\partial B}\right) + 1 - \frac{\pi}{w^b}
 \tag{A.19}$$

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  $(\rho^{2,P} + \lambda^P)$  from eq. (A.6) in eq. (A.7) to get

$$(\rho^{2,P} + \lambda^P) V_Y^{2P} = \gamma \frac{\pi}{w^{2,P}} - \gamma
 \tag{A.20}$$

Dividing this by eq. (A.6) gives us

$$\frac{V_Y^{2P}}{V_B^{2P}} = \frac{\frac{\pi}{w^{2,P}} - 1}{1 - \sum_i t_i \frac{\partial x_i^{2P}}{\partial B}}
 \tag{A.21}$$

Combining this with right-hand side of eq. (A.19) 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 of type 2, it is zero obviously, too.

However, when there is public provision, the expression in eq. (A.19) is as in the standard model as

$$\tau'(Y^b) = \frac{V_Y^b}{V_B^b} \left(1 - \sum_i t_i \frac{\partial x_i^b}{\partial B}\right) + 1
 \tag{A.22}$$

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

$$Y^{2P} : (\rho^{2,P} + \lambda^P) V_Y^{2P} + \gamma \left( 1 - \frac{\pi}{w^{2,P}} \right) = 0 \quad (\text{A.23})$$

where the last term appears because of the presence of the cost of child care in the government's budget constraint. Dividing this by eq. (A.6) and combining with eq. (A.22) yields  $\tau'(Y^{2P}) = \pi/w^{2P} > 0$ . Similar corrective terms appear in the tax rules for other types as well.