CESifo WORKING PAPERS

9265
2021

August 2021

Generalization of the Deaton Theorem: Piecewise Linear Income Taxation and Participation Decisions

Robin Boadway, Katherine Cuff

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
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: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9265*

Generalization of the Deaton Theorem: Piecewise Linear Income Taxation and Participation Decisions

Abstract

Deaton (1979) showed that if preferences are weakly separable in goods and labour and quasi-homothetic in goods and the government imposes an optimal linear progressive tax, commodity taxes are redundant. Hellwig (2009) generalized the Deaton theorem by showing that the allocation obtained under differential commodity taxes and an arbitrary linear progressive income tax is Pareto-dominated by one with uniform commodity taxes and a reformed linear progressive income tax. We show that both the Deaton theorem and the Hellwig extension continue to apply if a) the government implements a piecewise linear progressive income tax and b) labour varies along both the intensive and extensive margins. Some extensions are considered.

JEL-Codes: H210, H230, H240.

Keywords: optimal income taxation, commodity taxation, piecewise linear income tax.

Robin Boadway
Queen's University / Canada
boadwayr@econ.queensu.ca

Katherine Cuff
McMaster University / Canada
cuffk@mcmaster.ca

1 Introduction

The choice of a mix of commodity and income taxes has been central to optimal tax analysis and has led to some of the most policy-relevant results in the literature. Of particular importance are the circumstances under which optimal commodity taxes should be uniform, in which case they can be subsumed in the income tax system so are theoretically redundant.[1] The study of the optimal structure of commodity taxes goes back at least as far as the classic paper by Ramsey (1927). He showed for a representative-agent setting that if revenue requirements are infinitesimal or preferences are quadratic, optimal commodity taxes should be chosen to reduce commodity demands in the same proportion. If preferences are additive and quasilinear in labour, optimal commodity tax rates should be inversely proportional to demand elasticities so will not generally be uniform. The modern analysis of the case for uniform commodity taxes began with Corlett and Hague (1953) who showed in a representative-agent model that higher commodity tax rates should be levied on goods that are most complementary with leisure. Sandmo (1976) showed in this context that commodity taxes should be uniform if preferences are weakly separable in goods and leisure and homothetic in goods.

The analysis of the structure of commodity taxes in a heterogeneous-agent setting began with Atkinson and Stiglitz (1976). They added multiple goods to the optimal income tax model of Mirrlees (1971) in which individuals differ by wage rates but have the same preferences over goods and labour or leisure. The Atkinson-Stiglitz theorem showed that if preferences are weakly separable in goods and leisure and if the government levies an optimal nonlinear income tax, commodity tax rates should be uniform. Although this is a powerful theorem, it only holds if the government imposes an optimal nonlinear income tax. Laroque (2005) and Kaplow (2006) generalized the Atkinson-Stiglitz theorem. They showed that if preferences are weakly separable in leisure and goods, the equilibrium allocation obtained when the government imposes an arbitrary nonlinear income tax and differential commodity taxes is Pareto dominated by a system of uniform, possibly zero, commodity taxes and a reformed nonlinear income tax. This result, along with the Atkinson-Stiglitz theorem, was enough to persuade the Mirrlees Review (2011) that VAT tax rates in the UK should be

[1] Even with uniform commodity taxes, one may want a mix of income and commodity taxes to mitigate against tax evasion (Boadway, Marchand and Pestieau, 1994).

uniform.[2]

Implementing an optimal nonlinear income tax is demanding, and the literature has often resorted to linear progressive taxation (Sheshinski, 1972; Atkinson, 1973). This is both much simpler to administer and qualitatively similar to the pattern of marginal tax rates obtained by simulations of optimal income taxation (Mirrlees, 1970; Tuomala, 1990). Deaton (1979) studied a model with possibly differential commodity taxes and a lump-sum transfer and showed that optimal commodity taxes are uniform—and therefore linear progressive taxation is optimal—if preferences are weakly separable in leisure and goods and quasi-homothetic in goods. Parallel to the Laroque and Kaplow extensions of the Atkinson-Stiglitz theorem, Hellwig (2009) generalized the Deaton theorem to show that the allocation obtained under an arbitrary linear progressive taxation and differential commodity taxation is Pareto-dominated by another allocation with uniform commodity taxes and a reformed linear progressive tax.[3]

Our purpose is to generalize the Deaton theorem and the Hellwig extension in two directions both of which add realism to the setting. First, we let labour supply decisions vary along both extensive and intensive margins, in effect combining the intensive-margin optimal tax approach of Mirrlees (1971) with the extensive-margin approach of Diamond (1980) and Saez (2002). This allows us to take into consideration the empirically important participation decision and its effect on optimal commodity and income taxation. Second, we assume that the income tax is piecewise linear thereby reflecting the form of income tax used in most countries. Our main focus will be on the circumstances under which uniform commodity taxation—or a uniform VAT— is optimal. The characterization of the optimal piecewise linear income tax is of secondary interest, although we shall derive the optimal income tax structure as part of our analysis.

Our main results are as follows. First, the Deaton theorem generalizes to the case of piecewise linear income taxation and extensive margin labour supply. If piecewise linear tax rates are set optimally and if preferences are weakly separable in goods and leisure and quasi-homothetic in goods, optimal commodity taxes are uniform. Second, the analogue of

[2]The Mirrlees Review recognized that preferences might not be separable, but they argued that deviations from separability were not likely to nullify the benefits of a uniform VAT.

[3]Hellwig (2009) used homogeneous preferences in his proof, but he subsequently noted in Hellwig (2010) that the proof also goes through with quasi-homothetic preferences.

the Hellwig extension also applies. If preferences satisfy the Deaton conditions, an allocation resulting from an arbitrary piecewise linear income tax and differential commodity taxes is Pareto-dominated by an allocation with uniform commodity taxes and a reformed piecewise linear progressive income tax.

In addition, we note a number of extensions to the main results. First, the Deaton theorem and the Hellwig extension continue to apply if the piecewise linear income tax has more than two tax brackets. Second, if preferences are quasi-homothetic in a subset of goods, optimal commodity tax rates will be uniform for goods in the subset, but not other goods. Third, we argue that the main results continue to hold if some taxpayers are at the kink of the budget constraint in income-consumption space. However, they will not apply if some taxpayers make no purchases of some goods because they are at a boundary. We also show that, contrary to Deaton (1979), the Deaton theorem does not hold when preferences are heterogeneous such that the slopes of Engel curves are the same for all individuals, but the intercepts are not.

2 A model with intensive and extensive margins and piecewise linear income taxation

We adopt the simplest assumptions necessary to derive our results. There are m goods, denoted by x_j for $j = 1, \cdots, m$ which are produced by a linear technology using only labour ℓ as an input. Individuals differ in two dimensions: their skill in producing output as reflected in their wage rate, and a fixed cost of participating in the labour market. The latter is in addition to the cost of supplying labour while working. There are n^i individuals of type-i where $i = 0, \cdots, N$ and $\sum_{i \geqslant 0} n^i = 1$.[4] Type-0's are unable to work and some workers of types $i \geqslant 1$ choose not to work. Thus, non-participants will include individuals of all types. Those type-i individuals who choose to work earn a wage rate of w^i per unit of hours worked, where $w^i > w^{i-1}$ and $w^0 = 0$. Individuals of a given wage-type also differ in their fixed utility cost of participating in the labour force denoted by α^i, where α^i is distributed on $[\underline{\alpha}^i, \overline{\alpha}^i]$ with density $F^i(\alpha^i)$ and cumulative distribution $f^i(\alpha^i)$. Note that

[4]We follow the convention of using superscripts to indicate individual types, including the income tax bracket of the individual's income, and subscripts to refer to goods and partial derivatives

the distribution of the costs of participation can differ among types, although this plays no essential role in our analysis.

Individuals make their labour market decisions in sequence. First they decide whether to participate in the labour market. Then, if they choose to participate, they decide how much labour ℓ to supply and how much of each of the m goods to consume. They are paid w^i for each unit of labour they supply so their labour income is $y = w^i\ell$, and they consume $(x_1, \cdots, x_m)$ financed by their after-tax or disposable income. They also bear the fixed cost of participation α^i. Those who do not participate earn no labour income and bear no cost of working. Non-participants decide on the amount of each good to consume which they finance by a transfer from the government. We assume that when the participation decision is taken, individuals foresee the consequences of their decision in the subsequent stage. The utility of type-i individuals who participate is given by $u(x_1, \cdots, x_m, \ell) - \alpha^i$, and the utility of non-participants of all types is given by $u(x_1, \cdots, x_m)$.

The government imposes a piecewise linear income tax with two tax brackets as well as a set of commodity taxes on the goods. The two income tax brackets are divided by the income level $\hat{y}$. Incomes in the first tax bracket, $y \leqslant \hat{y}$, are taxed at the rate t_1 while incomes above $\hat{y}$ are taxed at the rate t_2. All individual receive the same lump-sum transfer a whether they are working or not. Workers with labour incomes $y \leqslant \hat{y}$ pay $t_1 y - a$ in income tax, and those with incomes $y > \hat{y}$ pay $t_1\hat{y} + t_2(y - \hat{y}) - a$. The tax liability of non-participants is $-a$, which is the lump-sum transfer they receive. Disposable incomes of individuals in the three groups are:

$$d^0 = a, \quad d^{1i} = (1 - t_1)y^{1i} + a, \quad d^{2i} = (1 - t_2)y^{2i} + (t_2 - t_1)\hat{y} + a \tag{1}$$

where superscripts 0, $1i$ and $2i$ refer from now on to non-participants, workers of type i with incomes in the first tax bracket, and workers of type i with incomes in the second tax bracket.

Without loss of generality producer prices of the m goods are normalized to unity, while consumer prices are $q_j = 1 + \tau_j$, where τ_j is the commodity tax on good j. In our analysis, we begin with uniform commodity taxes and then study whether perturbing one of the taxes can improve welfare. Deaton's theorem will apply if no perturbation starting from uniformity increases social welfare. Since proportional commodity taxes can be replicated by adjustments to the income tax, we can normalize the initial commodity taxes on all goods to be zero. We then perturb the commodity tax on good x_k, τ_k. For notational simplicity, we assume that $\tau_k = \tau$. For all $j \neq k$, $\tau_j = 0$ in what follows. We investigate conditions under which $\tau = \tau_k = 0$ in the optimum. Overall, the government chooses the tax parameters $(a, t_1, t_2, \hat{y}, \tau)$. The choice of $\hat{y}$ does not affect our main results so we take it as given.[5]

2.1 Individual behaviour

Once the participation decision is taken, all individuals choose their consumption of the m goods, and participants choose how much labour to supply. We characterize the behaviour first of non-participants and then of participants in each of the two tax brackets. The outcome is a set of indirect utility functions in government policy variables and their properties. These then allow us to characterize the participation decision and to determine how it is affected by tax policy. For simplicity, we assume that incomes are increasing in wage rates. That implies that there will be a cut-off income level—and therefore wage level—such that all workers below that income level are in the first tax bracket while all above are in the second. We further assume that no workers choose the income level $\hat{y}$ separating tax brackets. We return to this assumption later.

2.1.1 Choice of consumption and labour supplies

We begin by characterizing the choices made by non-participants and participants after the participation decision has been taken. Subsequently, we characterize the participation decision.

Non-participants

Non-participants maximize their utility $u(x_1, \cdots, x_m)$ subject to the budget constraint $(1 + \tau)x_k + \sum_{j \neq k} x_j = d^0 = a$. This yields the set of uncompensated demands $\mathbf{x}^0 = \big(x_1^0(1 +$

[5] Deaton (1979) does not formally include a linear progressive income tax in his model. Instead, the government chooses a set of commodity tax rates and a lump-sum transfer. If optimal commodity taxes are uniform, optimal policy is equivalent to a linear progressive income tax. We have included a linear income tax explicitly in our approach since part of our focus is extending the Deaton result to a setting with piecewise linear progressive income taxation.

$\tau, a), \cdots, x_m^0(1+\tau, a)\big)$ and indirect utility $v^0(\tau, a)$ where v_a^0 is their marginal utility of income and $v_\tau^0 = -v_a^0 x_k^0$ by Roy's theorem.

Participants in the first tax bracket

For type-i individuals in the first tax bracket, the budget constraint is $(1+\tau)x_k + \sum_{j\neq k} x_j = d^{1i} = (1-t_1)w^i\ell^1 + a$. Maximizing $u(x_1, \cdots, x_m, \ell)$ subject to this constraint yields the set of uncompensated demands, $\mathbf{x}^1 = \big(x_1^1(1+\tau, (1-t_1)w^i, a), \cdots, x_m^1(1+\tau, (1-t_1)w^i, a)\big)$, and labour supply, $\ell^1\big(1+\tau, (1-t_1)w^i, a\big)$. Indirect utility is $v^1\big(\tau, (1-t_1)w^i, a\big)$ where v_a^{1i} is the marginal utility of income and the envelope theorem gives

$$v_\tau^{1i} = -v_a^{1i}x_k^{1i}, \quad \text{and} \quad v_{(1-t_1)w^i}^{1i} = v_a^{1i}\ell^{1i}. \tag{2}$$

Expenditure minimization yields the set of compensated demand and labour supply functions given by $\tilde{x}_1^1\big(1+\tau, (1-t_1)w^i, \bar{u}\big), \cdots, \tilde{x}_m^1\big(1+\tau, (1-t_1)w^i, \bar{u}\big)$ and $\tilde{\ell}^1\big(1+\tau, (1-t_1)w^i, \bar{u}\big)$. The expenditure function is $e^1\big(1+\tau, (1-t_1)w^i, \bar{u}\big) = (1+\tau)\tilde{x}_k^{1i} + \sum_{j\neq k}\tilde{x}_j^{1i} - (1-t_1)w^i\tilde{\ell}^{1i}$ and will be equal to a at the optimum. Applying Shephard's lemma gives $e_{(1+\tau)}^{1i} = \tilde{x}_k^{1i}$ and $e_{(1-t_1)w}^{1i} = -\tilde{\ell}^{1i}$. We can also derive the Slutsky equation for labour supply by noting that in the optimum compensated and uncompensated labour supplies are identical:

$$\tilde{\ell}^1\big(1+\tau, (1-t_1)w^i, \bar{u}\big) = \ell^1\big(1+\tau, (1-t_1)w^i, e(1+\tau, (1-t_1)w^i, \bar{u})\big).$$

Differentiating with respect to τ and $(1-t_1)w^i$, we obtain the Slutsky equations:

$$\ell_\tau^{1i} = \tilde{\ell}_\tau^{1i} - \ell_a^{1i}x_k^{1i} \qquad \text{and} \qquad \ell_{(1-t_1)w^i}^{1i} = \tilde{\ell}_{(1-t_1)w^i}^{1i} + \ell_a^{1i}\ell^{1i}. \tag{3}$$

Participants in the second tax bracket

The budget constraint of type-i individuals in the second tax bracket is $(1+\tau)x_k + \sum_{j\neq k} x_j = d^{2i} = (1-t_2)w^i\ell^2 + (t_2-t_1)\hat{y} + a$. Maximizing $u\big(x_1, \cdots, x_m, \ell^2\big)$ subject to this constraint yields the set of uncompensated demands, $\mathbf{x}^2 = \big(x_1^2(1+\tau, (1-t_2)w^i, (t_2-t_1)\hat{y} + a), \cdots, x_m^2(1+\tau, (1-t_2)w^i, (t_2-t_1)\hat{y}+a)\big)$, labour supply, $\ell^2\big(1+\tau, (1-t_2)w^i, (t_2-t_1)\hat{y}+a\big)$, and indirect utility, $v^2\big(\tau, (1-t_2)w^i, (t_2-t_1)\hat{y}+a\big)$, where the envelope theorem gives

$$v_\tau^{2i} = -v_a^{2i}x_k^{2i}, \qquad v_{(1-t_2)w^i}^{2i} = v_a^{2i}\ell^{2i}, \qquad v_a^{2i} = v_{(t_2-t_1)\hat{y}}^{2i}. \tag{4}$$

Expenditure minimization by a type-i individual yields the set of compensated demands, $\tilde{x}_1^2\big(1+\tau,(1-t_2)w^i,\bar{u}\big),\cdots,\tilde{x}_m^2\big(1+\tau,(1-t_2)w^i,\bar{u}\big)$, labour supply, $\tilde{\ell}^2\big(1+\tau,(1-t_2)w^i,\bar{u}\big)$, and the expenditure function, $e^2\big(1+\tau,(1-t_2)w^i,(t_2-t_1)\hat{y},\bar{u}\big)=(1+\tau)\tilde{x}_k^{2i}+\sum_{j\neq k}\tilde{x}_j^{2i}-(1-t_2)w^i\tilde{\ell}^{2i}-(t_2-t_1)\hat{y}$. Applying Shephard's lemma yields

$$e^{2i}_{(1+\tau)}=\tilde{x}_k^{2i},\qquad e^{2i}_{(1-t_2)w^i}=-\tilde{\ell}^{2i},\qquad e^{2i}_{(t_2-t_1)\hat{y}}=-1. \tag{5}$$

At the optimum, $e^2\big(1+\tau,(1-t_2)w^i,(t_2-t_1)\hat{y},\bar{u}\big)=a$. To derive the relevant Slutsky equations, note that in the optimum we have:

$$\tilde{\ell}^2\big(1+\tau,(1-t_2)w^i,\bar{u}\big)=\ell^2\big(1+\tau,(1-t_2)w^i,(t_2-t_1)\hat{y}+e(1+\tau,(1-t_2)w^i,(t_2-t_1)\hat{y},\bar{u})\big).$$

Differentiating with respect to τ, $(1-t_2)w^i$ and $(t_2-t_1)\hat{y}$, and using (5) gives:

$$\ell^{2i}_{\tau}=\tilde{\ell}^{2i}_{\tau}+\ell^{2i}_a x_k^{2i},\quad \ell^{2i}_{(1-t_2)w^i}=\tilde{\ell}^{2i}_{(1-t_2)w^i}+\ell^{2i}_a\ell^{2i},\quad \ell^{2i}_{(t_2-t_1)\hat{y}}+\ell^{2i}_e e_{(t_2-t_1)\hat{y}}=0. \tag{6}$$

2.1.2 Participation decisions

As mentioned, to determine how individuals of different wage rates allocate themselves to the first and second income tax bracket, we assume that income y^i is increasing in skill-type regardless of income bracket. A sufficient condition for this to be the case is that optimal labour supply is increasing in skill-type or, equivalently, that the substitution effect of an increase in the wage rate dominates the income effect. Otherwise, we require that the elasticity of labour supply not to be too negative.[6] For a given piecewise linear income tax schedule, it is possible that y^{1i} is greater or less than y^{2i}, but in either case this assumption ensures that the individual of skill-type $i+1$ earns more, that is, $y^{1i+1}>y^{1i}$ and $y^{2i+1}>y^{2i}$. Consequently, if an individual earns y^i which is less (more) than some cut-off $\hat{y}$ all individuals with a wage less (greater) than w^i will also earn less (more) than $\hat{y}$. We use $\hat{\imath}$ to denote a fictional skill level associated with this cut-off income and assume that no individual earns exactly $\hat{y}$.

All workers of skill-type $i<\hat{\imath}$ choose to earn income in the first income tax bracket, $y^i<\hat{y}$, and all those of skill-type $i>\hat{\imath}$ choose to earn income in the second income tax

[6] Differentiating income, $y=w\ell$, with respect to w yields $d(w\ell)/dw=\ell+w\ell_w=\ell(1+w\ell_w/\ell)$ which is positive provided $w\ell_w/\ell>-1$.

bracket, $y^i > \hat{y}$, that is,

$$v^1\big(\tau,(1-t_1)w^i\big) > v^2\big(\tau,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big) \quad \forall i < \hat{\imath}$$
$$v^1\big(\tau,(1-t_1)w^i\big) < v^2\big(\tau,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big) \quad \forall i > \hat{\imath}$$

We now characterize the participation decisions for individuals in the two tax brackets.

Individuals of skill-type $i < \hat{\imath}$

These individuals will participate if $v^1\big(\tau,(1-t_1)w^i,a\big) - \alpha^i \geqslant v^0(\tau,a)$. For the marginal type$-i$ participant in the first tax bracket, the cost of participation is:

$$\hat{\alpha}^1\big(\tau,(1-t_1)w^i,a\big) = v^1\big(\tau,(1-t_1)w^i,a\big) - v^0(\tau,a)$$

where

$$\hat{\alpha}^{1i}_\tau = v^{1i}_\tau - v^0_\tau = -v^{1i}_a x^{1i}_k + v^0_a x^0_k, \quad \hat{\alpha}^{1i}_{(1-t_1)w^i} = v^{1i}_{(1-t_1)w^i} = v^{1i}_a \ell^{1i}, \quad \hat{\alpha}^{1i}_a = v^{1i}_a - v^0_a.$$

The number of type-i participants in the first tax bracket will then be given by $h^1(\tau,(1-t_1)w^i,a) = n^i F^i\big(\hat{\alpha}^1(\tau,(1-t_1)w^i,a)\big)$ where

$$h^{1i}_\tau = (v^{1i}_\tau - v^0_\tau)n^i f^i(\hat{\alpha}^{1i}) = (-v^{1i}_a x^{1i}_k + v^0_a x^0_k)n^i f^i(\hat{\alpha}^{1i}),$$
$$h^{1i}_{(1-t_1)w^i} = v^{1i}_{(1-t_1)w^i} n^i f^i(\hat{\alpha}^{1i}) = v^{1i}_a \ell^{1i} n^i f^i(\hat{\alpha}^{1i}), \quad h^{1i}_a = (v^{1i}_a - v^0_a)n^i f^i(\hat{\alpha}^{1i}). \tag{7}$$

Income effects on participation arise in this model because the value of an additional dollar of disposable income depends on whether an individual is participating or not in the labour market. Given income effects, we have the following relationship

$$v^0_a \ell^{1i} n^i f^i(\hat{\alpha}^{1i}) = h^{1i}_{(1-t_1)w^i} - \ell^{1i} h^{1i}_a \tag{8}$$

where the left-hand side of (8) can be interpreted as the compensated change in participation from a change in the after-tax wage rate. Assuming no income effects on participation implies that $h_a = 0$ or $v^{1i}_a = v^0_a$ and consequently the compensated change in participation is equal to the uncompensated change. With income effects, this is no longer the case and what matters for efficiency is the compensated change in participation with respect to the after-tax wage. For later use, we can rewrite (8) in terms of elasticities, where η is the income elasticity of participation and $\tilde{\sigma}$ and σ are the compensated and uncompensated elasticities of participation defined with respect to the after-tax wage

$$\tilde{\sigma}^{1i} = \sigma^{1i} - \eta^{1i}\frac{(1-t_1)w^i \ell^{1i}}{a}. \tag{9}$$

Individuals of skill-type $i > \hat{\imath}$

Similarly, these individuals will participate provided $v^2\big(\tau,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big)-\alpha^i \geqslant v^0(\tau,a)$. The marginal type$-i$ participant in the second tax bracket satisfies

$$\hat{\alpha}^2\big(\tau,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big) = v^2\big(\tau,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big) - v^0(\tau,a)$$

where

$$\hat{\alpha}^{2i}_\tau = v^{2i}_\tau - v^0_\tau = -v^{2i}_a x^i_k + v^0_a x^0_k, \qquad \hat{\alpha}^{2i}_{(1-t_2)w^i} = v^{2i}_{(1-t_2)w^i} = v^{2i}_a \ell^{2i},$$
$$\hat{\alpha}^{2i}_{(t_2-t_1)\hat{y}} = v^{2i}_{(t_2-t_1)\hat{y}} = v^{2i}_a, \qquad \hat{\alpha}^{2i}_a = v^{2i}_a - v^0_a.$$

The number of type-i participants in the second tax bracket will be $h^2\big(\tau,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big) = n^i F^i\big(\hat{\alpha}^2(\tau,(1-t_2)w^i,(t_2-t_1)\hat{y}+a)\big)$ where

$$h^{2i}_\tau = (v^{2i}_\tau - v^0_\tau)n^i f^i(\hat{\alpha}^{2i}) = (-v^{2i}_a x^{2i}_k + v^0_a x^0_k)n^i f^i(\hat{\alpha}^{2i}),$$
$$h^{2i}_{(1-t_2)w^i} = v^{2i}_{(1-t_2)w^i} n^i f^i(\hat{\alpha}^{2i}) = v^{2i}_a \ell^{2i} n^i f^i(\hat{\alpha}^{2i}),$$
$$h^{2i}_{(t_2-t_1)\hat{y}} = v^{2i}_{(t_2-t_1)\hat{y}} n^i f^i(\hat{\alpha}^{2i}) = v^{2i}_a n^i f^i(\hat{\alpha}^{2i}), \quad h^{2i}_a = (v^{2i}_a - v^0_a)n^i f^i(\hat{\alpha}^{2i}). \tag{10}$$

Using (10), we can again obtain an expression for the compensated change in participation with respect to the after-tax wage and the corresponding elasticities as

$$v^0_a \ell^{2i} n^i f^i(\hat{\alpha}^{2i}) = h^{2i}_{(1-t_2)w^i} - \ell^{2i} h^{2i}_a \quad \Rightarrow \quad \tilde{\sigma}^{2i} = \sigma^{2i} - \eta^{2i}\frac{(1-t_2)w^i\ell^{2i}}{a}. \tag{11}$$

For later use, we also note the following relationship:

$$v^0_a n^i f^i(\hat{\alpha}^{2i}) = h^{2i}_{(t_2-t_1)\hat{y}} - h^{2i}_a$$

where we can interpret the left-hand side as the compensated change in participation with a change in $(t_2-t_1)\hat{y}$ and define the following compensated elasticity

$$\tilde{\chi}^{2i} = \chi^{2i} - \eta^{2i}\frac{(t_2-t_1)\hat{y}}{a} \tag{12}$$

where χ is the elasticity of participation with respect to $(t_2-t_1)\hat{y}$.[7]

The number of non-participants is $h^0 = n^0 + \sum_{i<\hat{\imath}}(n^i - h^{1i}) + \sum_{i>\hat{\imath}}(n^i - h^{2i}) = 1 - \sum_{i<\hat{\imath}} h^{1i} - \sum_{i>\hat{\imath}} h^{2i}$. This includes all type-0 individuals as well as those of types $i \geqslant 1$ who choose not to participate.

[7]With a linear progressive income tax system, there is only a single income bracket and $\tilde{\chi} = \chi = 0$.

2.2 The government problem

The government maximizes an additive social welfare function in individual utilities subject to a budget constraint with no public expenditures for simplicity. Our methodology is to evaluate the social welfare effects of a change in τ starting at $\tau = 0$ and assuming the government is choosing income tax parameters a, t_1 and t_2 optimally.[8] Let $W\big(v^0(\tau, a)\big)$ be the social utility of a non-participating individual, while for participants of type i in the first tax bracket, social utility is $W\big(v^{1i}(\cdot) - \alpha^i\big)$ and in the second tax bracket it is $W\big(v^{2i}(\cdot) - \alpha^i\big)$. We assume that the same social utility function applies to all participants and non-participants, and is concave.[9]

Social welfare is the sum of social utilities and can be written:

$$h^0 W\big(v^0(\tau, a)\big) + \sum_{i<\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{1i}} W\Big(v^i\big(\tau, (1-t_1)w^i, a\big) - \alpha^i\Big) dF^i(\alpha^i)$$

$$+ \sum_{i>\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{2i}} W\Big(v^2\big(\tau, (1-t_2)w^i, (t_2-t_1)\hat{y} + a\big) - \alpha^i\Big) dF^i(\alpha^i) \tag{13}$$

where the first term includes non-participants and the last two terms includes workers in the first and second tax brackets respectively. Income and commodity tax revenues must equal the lump-sum transfer, so the government's budget constraint is:

$$a = t_1 \sum_{i<\hat{\imath}} h^1\big(\tau, (1-t_1)w^i, a\big) w^i \ell^1\big(1+\tau, (1-t_1)w^i, a\big)$$

$$+(t_1 - t_2) \sum_{i>\hat{\imath}} h^2\big(\tau, (1-t_2)w^i, (t_2-t_1)\hat{y}, a\big)\hat{y} \tag{14}$$

$$+t_2 \sum_{i>\hat{\imath}} h^2\big(\tau, (1-t_2)w^i, (t_2-t_1)\hat{y}, a\big) w^i \ell^2\big(1+\tau, (1-t_2)w^i, (t_2-t_1)\hat{y} + a\big)$$

$$+\tau\Bigg(h^0 x_k^0(1+\tau, a) + \sum_{i<\hat{\imath}} h^{1i} x_k^1\big(1+\tau, (1-t_1)w^i, a\big) + \sum_{i>\hat{\imath}} h^{2i} x_k^2\big(1+\tau, (1-t_2)w^i, (t_2-t_1)\hat{y}+a\big)\Bigg).$$

[8] We could also derive an optimality condition on the break-point $\hat{y}$ as in Apps, Van Long and Rees (2014). While this would be useful for fully characterizing the optimal piecewise linear income tax, we do not need this condition for deriving our results on extending the Deaton theorem.

[9] Assuming identical social utility functions is not innocuous as it gives social weight to the cost of working. This assumption does not, however, affect our main results. We could also assume social utility applies only to $v(\cdot)$ and obtain the same results.

The government maximizes (13) subject to (14). Denote the Lagrangian expression for this problem as $\mathcal{L}(a, t_1, t_2, \tau)$. The first-order conditions on a, t_1 and t_2 evaluated at $\tau = 0$ are as follows, where λ is the Lagrangian multiplier on (14):

$$\mathcal{L}_a\big|_{\tau=0} \equiv h^0 W_0' v_a^0 + \sum_{i<\hat{\imath}} n^i v_a^{1i} \int_{\underline{\alpha}^i}^{\hat{\alpha}^{1i}} W_{1i}' dF^i(\alpha^i) + \sum_{i>\hat{\imath}} n^i v_a^{2i} \int_{\underline{\alpha}^i}^{\hat{\alpha}^{2i}} W_{2i}' dF^i(\alpha^i) + \lambda\Bigg(t_1 \sum_{i<\hat{\imath}} h_a^{1i} w^i \ell^{1i}$$

$$+ t_1 \sum_{i>\hat{\imath}} h_a^{2i} \hat{y} + t_2 \sum_{i>\hat{\imath}} h_a^{2i}(w^i \ell^{2i} - \hat{y}) + t_1 \sum_{i<\hat{\imath}} h^{1i} w^i \ell_a^{1i} + t_2 \sum_{i>\hat{\imath}} h^{2i} w^i \ell_a^{2i} - 1 \Bigg) = 0, \qquad (15)$$

$$\mathcal{L}_{t_1}\big|_{\tau=0} \equiv -\sum_{i<\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{1i}} W_{1i}' v_{(1-t_1)w^i}^{1i} w^i dF^i(\alpha^i) - \sum_{i>\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{2i}} W_{2i}' v_a^{2i} \hat{y} dF^i(\alpha^i)$$

$$+ \lambda\Bigg(\sum_{i<\hat{\imath}} h^{1i} w^i \ell^{1i} + \sum_{i>\hat{\imath}} h^{2i} \hat{y} - t_1 \sum_{i<\hat{\imath}} h_{(1-t_1)w^i}^{1i} w^i \ell^{1i} w^i - t_1 \sum_{i<\hat{\imath}} h^{1i} w^i \ell_{(1-t_1)w^i}^{1i} w^i$$

$$- t_1 \sum_{i>\hat{\imath}} h_{(t_2-t_1)\hat{y}}^{2i} \hat{y}\hat{y} - t_2 \sum_{i>\hat{\imath}} h_{(t_2-t_1)\hat{y}}^{2i} (w^i \ell^{2i} - \hat{y})\hat{y} - t_2 \sum_{i>\hat{\imath}} h^{2i} w^i \ell_a^{2i} \hat{y} \Bigg) = 0, \qquad (16)$$

$$\mathcal{L}_{t_2}\big|_{\tau=0} \equiv \sum_{i>\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{2i}} W_{2i}' \big(v_{(1-t_2)w^i}^{2i}(-w^i) + v_a^{2i} \hat{y} \big) dF^i(\alpha^i) + \lambda\Bigg(\sum_{i>\hat{\imath}} h^{2i}(w^i \ell^{2i} - \hat{y})$$

$$+ t_1 \sum_{i>\hat{\imath}} \Big(h_{(1-t_2)w^i}^{2i} \hat{y}(-w^i) + h_{(t_2-t_1)\hat{y}}^{2i} \hat{y}\hat{y} \Big) - t_2 \sum_{i>\hat{\imath}} h_{(1-t_2)w^i}^{2i} (w^i \ell^{2i} - \hat{y}) w^i$$

$$+ t_2 \sum_{i>\hat{\imath}} h_{(t_2-t_1)\hat{y}}^{2i} (w^i \ell^{2i} - \hat{y})\hat{y} + t_2 \sum_{i>\hat{\imath}} h^{2i} w^i \big(\ell_a^{2i} \hat{y} - w^i \ell_{(1-t_2)w^i}^{2i} \big) \Bigg) = 0. \qquad (17)$$

Define the *average social marginal value of an additional unit of income* to each type as

$$\beta^0 = W'(v^0) v_a^0, \quad \beta^{1i} = \frac{n^i v_a^{1i}}{h^{1i}} \int_{\underline{\alpha}^i}^{\hat{\alpha}^{1i}} W'(v^{1i} - \alpha^i) dF^i(\alpha^i) \quad \text{for} \quad i < \hat{\imath},$$

$$\beta^{2i} = \frac{n^i v_a^{2i}}{h^{2i}} \int_{\underline{\alpha}^i}^{\hat{\alpha}^{2i}} W'(v^{2i} - \alpha^i) dF^i(\alpha^i) \quad \text{for} \quad i > \hat{\imath}$$

and the *average net social marginal values of an additional unit of income* (when $\tau = 0$) as

$$b^0 = \frac{\beta^0}{\lambda}, \quad b^{1i} = \frac{\beta^{1i}}{\lambda} + \frac{t_1 w^i \ell^{1i} h_a^{1i}}{h^{1i}} + t_1 w^i \ell_a^{1i} \quad \text{for} \quad i < \hat{\imath},$$

$$b^{2i} = \frac{\beta^{2i}}{\lambda} + \frac{\big(t_2(w^i \ell^{2i} - \hat{y}) + t_1 \hat{y} \big) h_a^{2i}}{h^{2i}} + t_2 w^i \ell_a^{2i} \quad \text{for} \quad i > \hat{\imath}. \qquad (18)$$

Using these definitions and the fact that the population is normalized to unity, we can rewrite the first-order condition on a given by (15) as $h^0 b^0 + \sum_{i<\hat{\imath}} h^{1i} b^{1i} + \sum_{i>\hat{\imath}} h^{2i} b^{2i} = 1$, or

$$h^0(1-b^0) = -\sum_{i<\hat{\imath}} h^{1i}(1-b^{1i}) - \sum_{i>\hat{\imath}} h^{2i}(1-b^{2i}). \tag{19}$$

Using the definitions (18), the envelope expressions (2) and (4), the derivatives of the participation functions (7) and (10), and the Slutsky equations (3) and (6), we can rewrite the first-order conditions on t_1 in (16) and t_2 in (17) evaluated at $\tau = 0$ as:

$$\begin{aligned} \mathcal{L}_{t_1}\big|_{\tau=0} &= \sum_{i<\hat{\imath}} h^{1i}(1-b^{1i})w^i\ell^{1i} - t_1 \sum_{i<\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{1i}) w^i \ell^{1i} w^i \ell^{1i} - t_1 \sum_{i<\hat{\imath}} h^{1i} w^i w^i \tilde{\ell}^{1i}_{(1-t_1)w^i} \\ &+ \sum_{i>\hat{\imath}} h^{2i}(1-b^{2i})\hat{y} - t_2 \sum_{i>\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y})\hat{y} - t_1 \sum_{i>\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{2i})\hat{y}\hat{y} = 0, \end{aligned} \tag{20}$$

$$\begin{aligned} \mathcal{L}_{t_2}\big|_{\tau=0} &\equiv \sum_{i>\hat{\imath}} h^{2i}(1-b^{2i})(w^i\ell^{2i}-\hat{y}) - t_1 \sum_{i>\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{2i})\hat{y}(w^i\ell^{2i}-\hat{y}) \\ &-t_2 \sum_{i>\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y})(w^i\ell^{2i}-\hat{y}) - t_2 \sum_{i>\hat{\imath}} h^{2i} w^i w^i \tilde{\ell}^{2i}_{(1-t_2)w^i} = 0. \end{aligned} \tag{21}$$

Next, take the derivative of the government's Lagrangian expression with respect to τ at $\tau = 0$:

$$\begin{aligned} \mathcal{L}_\tau\big|_{\tau=0} &\equiv h^0 W_0' v_\tau^0 + \sum_{i<\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{1i}} W_{1i}' v_\tau^{1i} dF^i(\alpha^i) + \sum_{i>\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{2i}} W_{2i}' v_\tau^{2i} dF^i(\alpha^i) + \lambda\Big(h^0 x_k^0 + \sum_{i<\hat{\imath}} h^{1i} x_k^{1i} \\ &+ \sum_{i>\hat{\imath}} h^{2i} x_k^{2i} + t_1 \sum_{i<\hat{\imath}} h_\tau^{1i} w^i \ell^{1i} + t_1 \sum_{i>\hat{\imath}} h_\tau^{2i}\hat{y} + t_2 \sum_{i>\hat{\imath}} h_\tau^{2i}(w^i\ell^{2i}-\hat{y}) + t_1 \sum_{i<\hat{\imath}} h^{1i} w^i \ell_\tau^{1i} + t_2 \sum_{i>\hat{\imath}} h^{2i} w^i \ell_\tau^{2i}\Big) \end{aligned}$$

which can be rewritten using (2)–(10), (15) and (18) as:

$$\begin{aligned} \frac{1}{\lambda}\mathcal{L}_\tau\big|_{\tau=0} &\equiv \sum_{i<\hat{\imath}} h^{1i}\left(1-b^{1i}\right)(x_k^{1i}-x_k^0) + \sum_{i>\hat{\imath}} h^{2i}\left(1-b^{2i}\right)(x_k^{2i}-x_k^0) \\ &-t_1 \sum_{i<\hat{\imath}} v_a^0 (x_k^{1i}-x_k^0) n^i f^i(\hat{\alpha}^{1i}) w^i \ell^{1i} - t_1 \sum_{i>\hat{\imath}} v_a^0 (x_k^{2i}-x_k^0) n^i f^i(\hat{\alpha}^{2i})\hat{y} \\ &-t_2 \sum_{i>\hat{\imath}} v_a^0 (x_k^{2i}-x_k^0) n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y}) + t_1 \sum_{i<\hat{\imath}} h^{1i} w^i \tilde{\ell}_\tau^{1i} + t_2 \sum_{i>\hat{\imath}} h^{2i} w^i \tilde{\ell}_\tau^{2i}. \end{aligned} \tag{22}$$

In the following section, we assume as in Deaton (1979) that preferences are weakly separable in labour and goods and quasi-homothetic in the latter. We show that the expression in (22) equals zero when the government sets a, t_1 and t_2 optimally so (19), (20) and (21) are satisfied.

3 Generalization of the Deaton theorem to piecewise linear taxation and extensive margin

Suppose preferences are separable in goods and labour, so

$$u(x_1, \cdots, x_m, \ell) = u\big(\phi(x_1, \cdots, x_m), \ell\big).$$

Assume that the sub-utility function in goods $\phi(\cdot)$ is quasi-homothetic (that is, homothetic to some point in commodity space, not necessarily the origin). We refer to these as Deaton preferences in what follows. Following Deaton (1979), quasi-homotheticity implies that we can write the indirect sub-utility function as:[10]

$$\phi\big(x_1(1+\tau, d), \cdots, x_m(1+\tau, d)\big) = \mu(1+\tau) + \psi(1+\tau)d \tag{23}$$

where, recall, consumer prices are $q_k = 1+\tau$ and $q_j = 1$ for $j \neq k$ and we have suppressed the latter prices as arguments of $\mu(\cdot)$ and $\psi(\cdot)$ for ease of notation. Disposable incomes d are given by (1) above.

By Roy's theorem, uncompensated demands for good k can be written:

$$x_k = -\frac{\mu_\tau(1+\tau) + \psi_\tau(1+\tau)d}{\psi(1+\tau)} \equiv \rho_k(1+\tau) + \gamma_k(1+\tau)d, \tag{24}$$

where μ_τ and ψ_τ are the derivatives of $\mu(\cdot)$ and $\psi(\cdot)$ with respect to τ, or equivalently $1+\tau$, the consumer price of good k. For ease of notation, in (24) we define $\rho_k(1+\tau) = -\mu_\tau(1+\tau)/\psi(1+\tau)$ and $\gamma_k(1+\tau) = -\psi_\tau(1+\tau)/\psi(1+\tau)$. Note that the Engel curve for x_k is linear with the same slope $\gamma_k(1+\tau)$ for all participating and non-participating individuals.

We can rewrite (22) for the case of Deaton preferences by developing expressions for $x_k^{1i} - x_k^0$, $x_k^{2i} - x_k^0$, $\tilde{\ell}_\tau^{1i}$ and $\tilde{\ell}_\tau^{2i}$. Substituting the definitions of disposable income for non-participants and participants in (1) into (24), we obtain Lemma 1.

[10] In fact, Deaton writes the indirect sub-utility function as $\mu^i(q_1, \cdots, q_m) + \psi(q_1, \cdots, q_m)d$, where $\mu^i(\cdot)$ can vary among households. This would imply that although the slopes of all Engel curves are the same among individuals, their intercepts are not. Equivalently, preferences for different individuals are quasi-homothetic to different origins. As we show later, the Deaton theorem does not actually apply with these heterogeneous preferences.

Lemma 1 *With Deaton preferences, the following expressions hold:*

$$x_k^{1i} - x_k^0 = \gamma_k (1-t_1) w^i \ell^{1i} \qquad i < \hat{\imath},$$

$$x_k^{2i} - x_k^0 = \gamma_k \big((1-t_2) w^i \ell^{2i} + (t_2 - t_1)\hat{y} \big) \qquad i > \hat{\imath}.$$

To derive expressions for the compensated labour supply derivatives $\tilde{\ell}_\tau$ that appear in (22), note first that applying Young's theorem to the expenditure function $e^1\big(1+\tau, (1-t_1)w^i, \bar{u}\big) = (1+\tau)\tilde{x}_k^{1i} + \sum_{j \neq k} \tilde{x}_j^{1i} - (1-t_1) w^i \tilde{\ell}^{1i}$ yields

$$\frac{\partial^2 e^1(\cdot)}{\partial(1+\tau)\partial\big((1-t_1)w^i\big)} = \frac{\partial^2 e^1(\cdot)}{\partial\big((1-t_1)w^i\big)\partial(1+\tau)} \quad \Longrightarrow \quad \frac{\partial \tilde{x}_k^{1i}}{\partial\big((1-t_1)w^i\big)} = -\tilde{\ell}_\tau^{1i}. \tag{25}$$

Second, in equilibrium the following must be satisfied for workers of types $i < \hat{\imath}$:

$$e^1(1+\tau, (1-t_1)w^i, \bar{u}) = a,$$

$$\tilde{x}_k^1\big(1+\tau, (1-t_1)w^i, \bar{u}\big) = x_k^1\big(1+\tau, (1-t_1)w^i, e(1+\tau, (1-t_1)w^i, \bar{u})\big),$$

$$\tilde{\ell}^1\big(1+\tau, (1-t_1)w^i, \bar{u}\big) = \ell^1\big(1+\tau, (1-t_1)w^i, e(1+\tau, (1-t_1)w^i, \bar{u})\big).$$

Substituting the expression for disposable income d^{1i} from (1) into (24), then using the above relationships to substitute out uncompensated demand, uncompensated labour supply and a in the resulting expression, and differentiating with respect to the after-tax wage $(1-t_1)w^i$ yields:

$$\frac{\partial \tilde{x}_k^{1i}}{\partial\big((1-t_1)w^i\big)} = \gamma_k (1-t_1) w^i \tilde{\ell}^{1i}_{(1-t_1)w^i}. \tag{26}$$

Combining (25) and (26), and following an identical procedure for workers of types $i > \hat{\imath}$, we obtain Lemma 2.

Lemma 2 *With Deaton preferences, the following expressions hold:*

$$\tilde{\ell}_\tau^{1i} = -\gamma_k (1-t_1) w^i \tilde{\ell}^{1i}_{(1-t_1)w^i} \qquad i < \hat{\imath},$$

$$\tilde{\ell}_\tau^{2i} = -\gamma_k (1-t_2) w^i \tilde{\ell}^{2i}_{(1-t_2)w^i} \qquad i > \hat{\imath}.$$

Substituting the expressions from Lemmas 1 and 2 into (22) and rearranging, we obtain Lemma 3.

Lemma 3 *With Deaton preferences, the following expression holds:*

$$\frac{1}{\lambda}\mathcal{L}_\tau\big|_{\tau=0} \equiv$$

$$\gamma_k(1-t_1)\Bigg(\sum_{i<\hat{\imath}} h^{1i}\left(1-b^{1i}\right)w^i\ell^{1i} - t_1\sum_{i<\hat{\imath}} v_a^0 w^i\ell^{1i}n^i f^i(\hat{\alpha}^{1i})w^i\ell^{1i} + \sum_{i>\hat{\imath}} h^{2i}\left(1-b^{2i}\right)\hat{y}$$

$$-t_1\sum_{i<\hat{\imath}} h^{1i}w^iw^i\tilde{\ell}^{1i}_{(1-t_1)w^i} - t_2\sum_{i>\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y})\hat{y} - t_1\sum_{i>\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{2i})\hat{y}\hat{y}\Bigg)$$

$$+\gamma_k(1-t_2)\Bigg(\sum_{i>\hat{\imath}} h^{2i}\left(1-b^{1i}\right)(w^i\ell^{2i}-\hat{y}) - t_2\sum_{i>\hat{\imath}} h^{2i}w^iw^i\tilde{\ell}^{2i}_{(1-t_2)w^i}$$

$$-t_2\sum_{i>\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y})(w^i\ell^{2i}-\hat{y}) - t_1\sum_{i>\hat{\imath}} v_a^0 n^i f^i(\hat{\alpha}^{2i})\hat{y}(w^i\ell^{2i}-\hat{y})\Bigg).$$

Proposition 1 follows directly by substituting the first-order conditions on t_1 and t_2, given by (20) and (21), into the expression in Lemma 3.

Proposition 1 *With Deaton preferences and the optimal piecewise linear income tax system in place,*

$$\frac{1}{\lambda}\mathcal{L}_\tau\big|_{\tau=0} = 0.$$

Proposition 1 applies for an incremental change in the tax of good k starting with all commodity taxes set to zero. Since the tax on any other good could have been changed instead, the proposition implies that changing the tax on any good starting from zero commodity taxes will not change social welfare. Therefore, if preferences are weakly separable in goods and labour and quasi-homothetic in goods as in Deaton (1979), and if the government sets the piecewise linear income tax optimally, commodity taxes are redundant. That is, the Deaton theorem generalizes to the case with piecewise linear income taxation and extensive labour supply decisions.

It is worth noting that quasi-homothetic preferences have two features that capture redistributive concerns. First, since the slopes of Engel curves can vary among goods, these preferences are compatible with income elasticities of demand differing over goods. That is, some goods can be necessities in the sense that their demand only rises moderately with income, while others are luxuries whose demand increases proportionately more with income. Second, a common way to characterize quasi-homothetic preferences in goods is by the following sub-utility function:

$$\phi(x_1 - \delta_1, \cdots, x_m - \delta_m)$$

where $\phi(\cdot)$ is homothetic in its arguments. The arguments are demands in excess of some minimal required amounts, δ_j, which can differ by goods. One might expect that the commodity tax system would favor goods with low income income elasticities of demand (necessities) and with high required amounts. The remarkable thing about the Deaton theorem is that piecewise linear income taxes—which do not differentiate among goods—suffice to achieve redistribution objectives as long as preferences over goods are quasi-homothetic.

3.1 Optimal Income Tax Rates with Uniform Commodity Taxation

Given that the Deaton theorem generalizes to piecewise linear income taxation and extensive labour supply decisions, the optimal tax system includes only an income tax. We now characterize the optimal income tax rates, given uniform commodity taxation.

Consider first the case of the optimal linear progressive income tax schedule as in Deaton (1979) but extended to allow for participation decisions. Using the above analysis, this is equivalent to having a single income tax bracket where all individuals receive a and all participants face a marginal tax rate t_1. Setting $t_1 = t$ and suppressing the subscript for the first income bracket, the optimal linear progressive tax rate t (when $\tau = 0$) satisfies the following first-order condition:

$$\sum_{i \geqslant 1} h^i (1 - b^i) w^i \ell^i - t \left(\sum_{i \geqslant 1} h^i_{(1-t)w^i} w^i w^i \ell^i - \sum_{i \geqslant 1} \ell^i h^i_a w^i w^i \ell^i + \sum_{i \geqslant 1} h^i w^i \tilde{\ell}^i_{(1-t)w^i} w^i \right) = 0$$

where

$$\beta^i = \frac{n^i v^i_a \int_{\underline{\alpha}^i}^{\hat{\alpha}^i} W'(v^i - \alpha^i) dF^i(\alpha^i)}{h^i}, \quad b^i = \frac{\beta^i}{\lambda} + \frac{t_1 w^i \ell^i h^i_a}{h^i} + t w^i \ell^i_a.$$

Defining $\tilde{\epsilon}^i$ as the compensated labour supply elasticity with respect to the after-tax wage rate and using the definition of the compensated participation elasticity given by (9), the first-order condition on the optimal linear income tax can be rewritten as

$$\frac{t}{1-t} = \frac{\sum_{i \geqslant 1}(1-b^i)h^i w^i \ell^i}{\sum_{i \geqslant 1}\left(\tilde{\epsilon}^i + \tilde{\sigma}^i\right) h^i w^i \ell^i}.$$

Since $w^0 = 0$ and the average value of b^i over all $i \geq 0$ is one (from the first-order condition on a), the above can be rewritten in terms of the covariance of b^i and total income for all types $i \geq 0$ as

$$\frac{t}{1-t} = \frac{-\text{cov}(b^i, h^i w^i \ell^i)}{\sum_{i \geqslant 1}\left(\tilde{\epsilon}^i + \tilde{\sigma}^i\right) h^i w^i \ell^i}. \tag{27}$$

The optimal linear income tax rule given by (27) takes the standard form where the numerator captures equity considerations and the denominator captures efficiency considerations. With only an extensive margin labour supply decision, $\epsilon^i = 0$ and assuming no income effects on participation (as in Saez, 2005), (27) reduces to what the optimal participation tax rate would be when taxes cannot be conditioned on skill type.[11] Conversely, with only an intensive margin labour supply decision, the participation elasticity is zero and $h^i = n^i$. Eq. (27) reduces to the optimal linear income tax rule with a discrete number of types. In this case, we could rewrite (27) using income shares defined as

$$s^i = \frac{n^i w^i \ell^i}{\sum_{i \geq 0} n^i w^i \ell^i} \quad \text{where} \quad \sum_{i \geq 0} s^i = 1 \quad \text{and} \quad s^0 = 0$$

to obtain a similar expression as in Piketty and Saez (2012):

$$\frac{t}{1-t} = \frac{-\text{cov}(b^i, s^i)}{\sum_{i \geqslant 1} \tilde{\epsilon}^i s^i}.$$

The denominator is the income share-weighted compensated elasticities of labour supply, and can be interpreted as the elasticity of total income with respect to the net-of-tax price, $1-t$.

Consider now the optimal piecewise linear income tax rates. The first-order condition on t_1 can be written as

$$\frac{t_1}{1-t_1} = \frac{\sum_{i<\hat{\imath}} h^{1i}(1-b^{1i}) w^i \ell^{1i} + \sum_{i>\hat{\imath}} h^{2i}(1-b^{2i})\hat{y}}{\sum_{i<\hat{\imath}}\left(\tilde{\epsilon}^{1i} + \tilde{\sigma}^{1i}\right) h^{1i} w^i \ell^{1i}} + \frac{\sum_{i>\hat{\imath}}\left(\tilde{\chi}^{2i} - \tilde{\sigma}^{2i}\frac{t_2}{(1-t_2)}\right) h^{2i}\hat{y}}{\sum_{i<\hat{\imath}}\left(\tilde{\epsilon}^{1i} + \tilde{\sigma}^{1i}\right) h^{1i} w^i \ell^{1i}} \tag{28}$$

[11] Saez (2005) assumes income taxes can be conditioned on skill-type. His expressions for the optimal participation taxes can be obtained here by assuming marginal tax rates are conditional on skill-type, that is t^i.

and the first-order condition on t_2 can be written as

$$\frac{t_2}{1-t_2} = \frac{\sum_{i>\hat{\imath}} h^{2i}(1-b^{2i})(w^i\ell^{2i}-\hat{y})}{\sum_{i>\hat{\imath}}(\tilde{\epsilon}^{2i}h^{2i}w^i\ell^{2i}+\tilde{\sigma}^{2i}h^{2i}(w^i\ell^{2i}-\hat{y}))} + \frac{\sum_{i>\hat{\imath}} \tilde{\chi}^{2i}h^{2i}(w^i\ell^{2i}-\hat{y})}{\sum_{i>\hat{\imath}}(\tilde{\epsilon}^{2i}h^{2i}w^i\ell^{2i}+\tilde{\sigma}^{2i}h^{2i}(w^i\ell^{2i}-\hat{y}))}. \tag{29}$$

Each optimal tax rate consists of two terms. The first term is the ratio of an equity to an efficiency term resulting from a change in the marginal tax rate and has the standard interpretation. The efficiency effect from a change in the marginal tax rate in the denominator depends on both the compensated labour supply and compensated participation elasticities. The equity effect from a change in the marginal tax rate reflects the fact that an increase in t_1 impacts all participants while an increase in t_2 only affects participants in the second income bracket.

The second term on the right-hand is new and is an additional efficiency term that arises because the participation decision of those in the second income bracket is also affected by the income tax differential, $(t_2-t_1)\hat{y}$.[12] Consider first the common expression, $\tilde{\chi}^{2i}h^{2i}\hat{y}$, that appears in the numerator of this term in both (28) and (29), but of opposite sign. This expression reflects the effect a change in h^{2i} (arising from a change in $(t_2-t_1)\hat{y}$) has on the net revenue $h^{2i}(t_2-t_1)\hat{y}$. Of course, the government is levying t_2 on the income earned above $\hat{y}$ by participants in the second bracket. Therefore a change in h^{2i} affects this revenue as captured in the second expression in the numerator of this last term in (28) and the first expression in the numerator of this last term in (29).

4 Generalization of the Hellwig extension

Deaton's result is restrictive in the sense that it only applies when the optimal linear income tax is in place. If income taxes are non-optimal, commodity taxes should in general be non-uniform. For example, Boadway and Song (2016) show in a two-good setting with a linear progressive income tax that if preferences satisfy the Deaton conditions, the commodity tax rate should be higher on the good with the lowest income elasticity of demand if the marginal income tax rate is lower than optimal, and vice versa. That assumes that the income tax

[12] Eqs. (28) and (29) correspond with marginal tax rates obtained by Jacquet, Lehmann and Van der Linden (2013) for the case of optimal nonlinear income taxes when labour supply varies along both intensive and extensive margins. Efficiency terms include both labour supply elasticities and participation elasticities in their analysis as well.

rate is taken as given. Hellwig (2009) demonstrates how Deaton's result can apply when the linear progressive tax system is not optimal as long as the government can reform the tax parameters t and a. His analysis shows that with quasi-homothetic preferences in goods, a differential commodity tax system combined with an arbitrary, not necessarily optimal, linear progressive income tax is Pareto-dominated by a uniform commodity tax system (including with zero tax rates) and a suitably reformed linear progressive tax. In this section, we show that Hellwig's result generalizes to our setting with piecewise linear income taxation and both extensive- and intensive-margin labour supply variation. We illustrate this in the simple setting used above with m goods and a discrete number of skill-types.

Suppose there is an arbitrary piecewise linear income tax and a set of differential commodity taxes in place. The parameters of the income tax are a, t_1 and t_2 as above, and the break point between the first and second tax brackets is $\hat{y}$. Let consumer prices be $q_j = (1+\tau_j)p_j$ for $j = 1, \cdots, m$ where producer prices are p_j and define $\mathbf{q} = (q_1, \cdots, q_m)$ and $\mathbf{p} = (p_1, \cdots, p_m)$ as the vector of consumer and producer prices, respectively.[13] Differential commodity taxes are initially in place, so τ_j varies by good. As above, there are three types of individuals: non-participants, workers in first tax bracket, and workers in second tax bracket. Their respective budget constraints are:

$$\sum_{j=1}^{m} q_j x_j^0 = a = d^0, \tag{30}$$

$$\sum_{j=1}^{m} q_j x_j^{1i} = (1-t_1)w^i \ell^{1i} + a = d^{1i} \qquad i < \hat{\imath}, \tag{31}$$

$$\sum_{j=1}^{m} q_j x_j^{2i} = (1-t_2)w^i \ell^{2i} + (t_2 - t_1)\hat{y} + a = d^{2i} \qquad i > \hat{\imath}. \tag{32}$$

The common utility function is separable and takes the form $u\big(\phi(x_1, \cdots, x_m), \ell\big)$.

Denote the consumption vector chosen by the three types in the initial tax setting (status quo s) as $\mathbf{x}^0(s)$, $\mathbf{x}^{1i}(s)$ and $\mathbf{x}^{2i}(s)$, and disposable incomes of participants as $d^{1i}(s)$ and $d^{2i}(s)$. We can define the value of the subutility functions of the three types in their optima as follows:

$$\omega^0(s) \equiv \phi\big(\mathbf{x}^0(s)\big), \qquad \omega^{1i}(s) \equiv \phi\big(\mathbf{x}^{1i}(s)\big), \qquad \omega^{2i}(s) \equiv \phi\big(\mathbf{x}^{2i}(s)\big) \tag{33}$$

[13] In the previous section, we normalized $p_j = 1$ for all $j = 1, ...m$ for simplicity. This was without loss of generality. Proposition 1 holds under any arbitrary vector of producer prices for the m commodities.

We prove the following proposition.

Proposition 2 *Assume utility takes the form* $u\big(\phi(x_1, \cdots, x_m), \ell\big)$ *with subutility* $\phi(\mathbf{x})$ *satisfying the Deaton conditions. Let* $\big(\mathbf{x}^h(s), d^h(s)\big)$ *be the allocation of a type-h individual for* $h = 0, 1i, 2i$ *under the arbitrary piecewise linear progressive income tax system* (t_1, t_2, a) *with differential commodity taxes* $\tau_1, \cdots, \tau_m$*. If consumer prices are not proportional to producer prices, the allocation* $\big(\mathbf{x}^h(s), d^h(s)\big)$ *is strictly dominated by another allocation that can be implemented by the linear progressive income tax* $(\hat{t}_1, \hat{t}_2, \hat{a})$ *with consumer prices equal to producer prices* $\mathbf{q} = \mathbf{p}$*.*

Proof: Following Hellwig (2009), we proceed to prove the Proposition 2 in two steps.

Step 1

If consumer prices are not proportional to producer prices, there exists another feasible allocation of consumption $\mathbf{x}^h$ for the three types yielding the same subutility level for each type $\omega^h(s)$ that requires fewer resources. Consider the following problem for a type-h individual, given $d^h(s)$:

$$\min_{x_1, \cdots, x_m} \sum_{j=1}^{m} p_j x_j \qquad \text{s.t.} \qquad \phi(\mathbf{x}) = \omega^h(s) \tag{34}$$

where $\omega^h(s)$ satisfies (33). The solution gives the allocation $\big(\hat{\mathbf{x}}^h, d^h(s)\big)$ with the same utility as in the initial situation above. For each type h, the above problem implies:

$$\sum_{j=1}^{m} p_j \hat{x}_j^h < \sum_{j=1}^{m} p_j x_j^h(s) \tag{35}$$

That is, less resources are required to produce $\hat{\mathbf{x}}^h$ than the initial allocation. The same applies for all three types of individuals, non-participants ($h = 0$), workers in the first tax bracket ($h = 1i$) and workers in the second tax bracket ($h = 2i$).

Step 2

We next show that for each individual h, the allocation $\big(\hat{\mathbf{x}}^h, d^h(s)\big)$ can be implemented by a linear progressive income tax $(\hat{t}_1, \hat{t}_2, \hat{a})$ with consumer prices equal to producer prices when quasi-homothetic preferences apply. (Note that in this allocation, individuals choose the same income or labour supply but different commodity bundles.)

By the Deaton conditions as shown in (23) and (24), the consumption allocation obtained from the type-h individual's problem (34), given $d^h(s)$, satisfies:

$$\hat{x}_j^0 = \rho_j(\mathbf{p}) + \gamma_j(\mathbf{p})a \qquad j = 1, \cdots, m, \tag{36}$$

$$\hat{x}_j^{1i} = \rho_j(\mathbf{p}) + \gamma_j(\mathbf{p})\big((1-t_1)w^i\ell^{1i} + a\big) \qquad j = 1, \cdots, m, \tag{37}$$

$$\hat{x}_j^{2i} = \rho_j(\mathbf{p}) + \gamma_j(\mathbf{p})\big((1-t_2)w^i\ell^{2i} + (t_2 - t_1)\hat{y} + a\big) \qquad j = 1, \cdots, m. \tag{38}$$

Multiplying each of these by p_j and summing over all goods, we obtain:

$$\sum_{j=1}^{m} p_k\hat{x}_j^0 = \sum_{j=1}^{m} p_j\rho_j(\mathbf{p}) + \sum_{j=1}^{m} p_j\gamma_j(\mathbf{p})a, \tag{39}$$

$$\sum_{j=1}^{m} p_j\hat{x}_j^{1i} = \sum_{j=1}^{m} p_j\rho_j(\mathbf{p}) + \sum_{j=1}^{m} p_j\gamma_j(\mathbf{p})a + \sum_{j=1}^{m} p_j\gamma_j(\mathbf{p})(1-t_1)w^i\ell^{1i}, \tag{40}$$

$$\sum_{j=1}^{m} p_j\hat{x}_j^{2i} = \sum_{j=1}^{m} p_j\rho_j(\mathbf{p}) + \sum_{j=1}^{m} p_j\gamma_j(\mathbf{p})a + \sum_{j=1}^{m} p_j\gamma_j(\mathbf{p})\big((1-t_2)w^i\ell^{2i} + (t_2-t_1)\hat{y}\big). \tag{41}$$

Consider an alternative tax system $(\hat{t}_1, \hat{t}_2, \hat{a})$, and suppose it is related to the original tax system as follows:

$$(1-\hat{t}_1) = \sum_{j=1}^{m} p_j\gamma_j(\mathbf{p})(1-t_1), \quad (1-\hat{t}_2) = \sum_{j=1}^{m} p_j\gamma_j(\mathbf{p})(1-t_2), \quad \hat{a} = \sum_{j=1}^{m} p_j\big(\rho_j(\mathbf{p}) + \gamma_j(\mathbf{p})a\big). \tag{42}$$

The first two expressions imply:

$$\hat{t}_2 - \hat{t}_1 = \sum_{j=1}^{m} p_j\gamma_j(t_2 - t_1). \tag{43}$$

Substituting (42) and (43) into (39)–(41), we obtain:

$$\sum_{j=1}^{m} p_j\hat{x}_j^0 = \hat{a}, \tag{44}$$

$$\sum_{j=1}^{m} p_j\hat{x}_j^{1i} = (1-\hat{t}_1)w^i\ell^{1i} + \hat{a}, \tag{45}$$

$$\sum_{j=1}^{m} p_j\hat{x}_j^{2i} = (1-\hat{t}_2)w^i\ell^{2i} + (\hat{t}_2 - \hat{t}_1)\hat{y} + \hat{a}. \tag{46}$$

These are the same budget constraints as in (30)–(32) with the new tax system and no commodity taxes. It shows that if type-h individuals are faced with the new tax system and consumer prices are equal to producer prices, they can choose $\hat{\mathbf{x}}^h$ and the original disposable income level $d^h(s)$. We have already shown in Step 1 that if the individuals choose this allocation fewer resources are required and they are equally well-off as in the original tax system. This means that the lump-sum component $\hat{a}$ can be increased to use up the extra resources so everyone can be made better off. ■

5 Extensions and limitations

Our results have been developed for the general case where labour supply is variable along both the intensive and extensive margins and the government implements a piecewise linear progressive tax. Obviously, the results will also apply in the special cases where a) labour supply varies along the intensive margin only and piecewise linear income tax is in place, b) labour supply varies along the extensive margin only and piecewise linear income taxation is used, and c) labour varies along both margins and a linear progressive income tax applies. In each of these cases, both the Deaton theorem and the Hellwig generalization—which were derived for the case where labour varies along the intensive margin and a linear progressive income tax is used—apply.

In our model, we adopted a number of simplifying assumptions for analytical convenience. In this section, we consider the consequences of relaxing some of those assumptions. In some cases, our main propositions remain intact, while for others, they do not. However, in the latter cases, the Deaton theorem does not apply either.

5.1 Multiple tax brackets

It is straightforward to show that our results continue to apply if we increase the number of tax brackets above two, as long as our other assumptions remain satisfied. If the optimal multi-bracket linear progressive income tax is in place and preferences are weakly separable in goods and labour and quasi-homothetic in goods, commodity taxes should be uniform (or zero). Similarly, Hellwig's generalization continues to apply. An arbitrary multi-bracket linear income tax combined with differential commodity taxes is Pareto-dominated by uniform commodity taxes and a suitably reformed multi-bracket linear income tax.

These results are relevant from a policy perspective. Most countries do implement a multi-bracket linear progressive income tax. To the extent that preferences do not diverge too much from quasi-homotheticity, uniform commodity taxes cause limited inefficiency and save the costs of administering a differential commodity tax system.

5.2 Quasi-homotheticity in a subset of goods

Deaton extended his analysis to the case where utility is quasi-homothetic in a subset of goods as well as being weakly separable. He showed that with optimal commodity taxes and a lump-sum transfer in place, commodity taxes should be uniform within that subset, but will generally be differential for goods outside that subset. We show in Appendix A that the analogue of this is also true in the case of an optimal piecewise linear income tax system and with both extensive and intensive labour supply decisions. In particular, the optimal tax system will include a piecewise linear income tax, uniform commodity taxes for goods in the quasi-homothetic subset, and differential commodity taxes for other goods.

To illustrate this, assume utility is quasi-homothetic and weakly separable in all goods except good m, so utility can be written as $u(\phi(x_1, \cdots, x_{m-1}), x_m, \ell)$. Further, assume that good m is taxed at rate τ_m, so the government chooses $(a, t_1, t_2, \tau_m, \tau)$ where τ is the tax rate on an arbitrary good k from the set $(x_1, \cdots, x_{m-1})$ as above. We can follow the same steps as the previous analysis to prove that the government would want to impose a uniform commodity tax on all goods except good m as shown in Proposition 3 in Appendix A. This analysis could be extended so that there is an arbitrary set of commodities in the sub-utility function ϕ and provided the government optimally chooses both the piecewise linear income tax system and the commodity taxes on all other goods the Deaton theorem holds.

The Hellwig extension also extends to the case where quasi-homotheticity applies only to a subset of goods. Suppose utility is weakly separable and quasi-homothetic in a subset of goods $k \in Q$, so can be written $u\big(\phi(\mathbf{x}_Q), \mathbf{x}_N, \ell\big)$ where $\mathbf{x}_Q$ is the vector of goods in the subset Q and $\mathbf{x}_N$ is the subset of the rest of the goods. We show in Appendix B that the allocation obtained under an arbitrary piecewise linear income tax and differential commodity taxes is Pareto-dominated by an allocation when commodity taxes on goods $\mathbf{x}_Q$ are zero and both the income tax parameters and the commodity taxes in $\mathbf{x}_N$ are suitably reformed.

5.3 Heterogeneous preferences

Deaton (1979, p. 357) stated his theorem in the following simplified form: if Engel curves for all goods are linear and the slope of each Engel curve is the same for all individuals, commodity taxes should be uniform if an optimal linear progressive income tax is in place. Equivalently, he argued that his proof of the optimality of uniform commodity taxes required only that preferences for individual h yield an indirect sub-utility function of the following form: $v(\mathbf{q}, d^h) = \mu^h(\mathbf{q}) + \psi(\mathbf{q})d^h$, where $\mathbf{q}$ is the vector of consumer prices and d^h is disposable income. This differs from (23) since the term $\mu^h(\mathbf{q})$ can differ among individual types. This indirect sub-utility function yields demand curves for good k, analogous to (24) above:

$$x_k^h = -\frac{\mu_{q_k}^h(\mathbf{q}) + \psi_{q_k}(\mathbf{q})d^h}{\psi(\mathbf{q})} \equiv \rho_k^h(\mathbf{q}) + \gamma_k(\mathbf{q})d^h. \tag{47}$$

This implies that Engel curves have the same slope for all individuals, $\gamma_k(\mathbf{q})$, but different intercepts, $\rho_k^h(\mathbf{q})$. Using these demands for x_k^h in our above analysis, we find that neither the Deaton theorem nor Hellwig's extension apply either with linear progressive or piecewise linear progressive income taxes. In other words, assuming linear Engel curves for all goods with the same slopes across individuals is not sufficient for optimal commodity taxes to be uniform when income taxes are set optimally. However, linear Engel curves for all goods with the same slopes across individuals combined with common intercepts, as quasi-homotheticity implies, is sufficient. That is, all individuals must choose consumption allocations along the same Engel curves.

This can be seen most easily in the proof of the Hellwig extension above. The alternative tax system $(\hat{t}_1, \hat{t}_2, \hat{a})$ satisfied (42) for all individual types. With heterogeneous preferences, the expression for $\hat{a}$ for a person of type h must satisfy

$$\hat{a}^h = \sum_{j=1}^{m} p_j \big(\rho_j^h(\mathbf{p}) + \gamma_j(\mathbf{p})a\big).$$

Since this differs among types, the alternative tax system would require changes in the lump-sum transfer a to differ by household type which is not feasible. Thus, the Hellwig generalization would not apply in this case.

In fact, there is an algebraic error in Deaton's proof as we show in Appendix C. Optimality of uniform commodity taxes when $\mu^h(\mathbf{q})$ varies by household does not hold in general.

5.4 Nonlinear budget sets and corner solutions

In our analysis we assumed that all individuals chose an interior solution for their commodity demands and labour supply. However, the use of Deaton preferences along with piecewise linear income taxation can lead to corner solutions and multiple equilibria. There are various possibilities.

Bunching at the income tax break point

If the piecewise linear income tax leads to a strictly concave budget constraint, some individuals may choose incomes at the kink point in which case their indifference curves may not be tangential to the budget constraint. They may nonetheless be at an interior in their commodity demands. For those taxpayers whose incomes are at the kink point and whose indifference curves are not tangential to the budget line in income-consumption (disposable income) space, the envelope theorem does not apply for changes in the marginal tax rate. This complicates our analysis of the Deaton theorem considerably. Nonetheless, we can infer from studying the Hellwig generalization that uniform commodity taxes should apply at the optimum when utility functions exhibit Deaton preferences.

To see this, note that the Hellwig generalization continues to apply when some taxpayers are at the kink point on their budget lines. The proof of the Hellwig generalization is based on a revealed preference argument and does not require differentiability in income. That is, Step 1 in the proof above applies for whatever incomes taxpayers choose in the initial allocation. Step 2 then shows that the incomes chosen in the initial allocation can still be supported after commodity taxes are abolished and the income tax suitably reformed. Revealed preference then implies that all persons can be made better off by the reform, given the saving in resources the move to zero commodity taxes implies.

The fact that the Hellwig generalization applies means that any tax system with differential commodity taxes cannot be optimal. This implies indirectly that the Deaton theorem must be satisfied.

Concave budget sets

When the budget constraint is strictly convex, bunching in the interior of the budget constraint will not occur, but there is a remote possibility that some worker-type will be indifferent to being on either segment of the budget constraint (that is, either tax bracket). As well, taxpayers may change their incomes discretely in response to a tax change by moving from one tax bracket to another. (This possibility cannot arise if labour only varies along the extensive margin since then labour incomes are fixed.) Our proof of the Deaton theorem in this case is complicated by the non-convexity of the taxpayer's problem.[14] However, as in the bunching case, the Hellwig extension remains valid when the budget constraint is convex (i.e., the budget set is concave). Whatever income taxpayers choose is taken as given in the proof of the Hellwig extension. That being the case, differential commodity taxation cannot be optimal when Deaton preferences apply and the government uses an optimal piecewise linear income tax.

Corner solutions in commodity demands

Deaton preferences implies linear Engel curves which can lead to corner solutions for some commodity demands at low income levels. Boadway and Song (2016) considered the case where low-income individuals choose a corner solution in a luxury good (more precisely are constrained by not being able to buy a negative quantity). They show that the Deaton theorem no longer applies in this case, so our generalization will not apply either. It would be welfare-improving to differentiate commodity tax rates, although the form of differentiation is not clearcut. Lowering the tax rate on necessity goods improves equity but has an ambiguous effect on efficiency.

[14] Apps, Van Long and Rees (2014) address the problems of bunching at kink points and strictly convex budget constraints by simulation techniques. They consider separately the case of strictly concave and convex budget constraints and simulate the properties of the optimal piecewise linear income tax. They have only one consumption good so their focus is solely on optimal income taxation.

Corner solution in labour supplies

We have allowed for individuals to choose zero labour income rather than a discrete income by choosing not to participate. Zero labour incomes can also be chosen in an intensive-margin setting. Boadway, Cuff and Marchand (2000) show that this can occur when preferences are quasi-linear in leisure so that leisure or labour is chosen as a residual. The Hellwig result applies in this case as long as the demands for goods are in the interior, so uniform commodity taxes are optimal and the Deaton theorem is satisfied.

6 Concluding remarks

The Deaton theorem along with the Atkinson-Stiglitz theorem are among the most important policy-relevant results to have come out of the optimal income tax literature. Their relevance have been enhanced by the Hellwig extension to the Deaton theorem and the Laroque (2005) and Kaplow (2006) extensions to the Atkinson-Stiglitz theorem. Our purpose has been to increase the policy relevance of the Deaton theorem further by generalizing it in two realistic directions. One is to allow individuals to vary their labour supply choices along the extensive margin as well as along the intensive margin by incorporating a participation decision of the sort used by Diamond (1980) and Saez (2002) into a Mirrleesian optimal income tax setting. The second is to generalize the linear into tax system to a piecewise linear income tax with multiple tax brackets. Both of these extensions are empirically relevant. For example, governments typically deploy piecewise linear income taxes.

Our main result is that the Deaton theorem and the Hellwig extension continue to hold with extensive margin labour supply and piecewise linear income taxation. That is, if preferences are weakly separable in goods and leisure, and quasi-homothetic in goods, optimal commodity taxes should be uniform. We have also considered various extensions and and complications. Following Deaton, we have shown that a restrictive version of the Deaton theorem and Hellwig's extension applies when the Deaton conditions hold for only to a subset of goods. We have also argued that uniformity of commodity taxes should continue to apply if some taxpayers choose income at the kink point of a concave budget constraint, or if some are indifferent between two segments of a convex budget constraint. At the same time, we have shown that the Deaton theorem does not hold if some individuals choose a corner solution in some goods, or if preferences are heterogeneous in the way Deaton considered.

As mentioned earlier, quasi-homothetic preferences are relatively general in the sense that they allow for different income elasticities of demand and different minimal required levels of consumption among goods. Despite that, linear or piecewise linear progressive income taxes are sufficient to achieve the government's redistributive objectives without any need to use differential commodity taxes. Moreover, the Hellwig extension to the Deaton theorem applies as well. It is never efficient to use differential commodity taxes to achieve redistributive objectives, for example, by favouring necessities at the expense of luxury goods. The elimination of differential commodity taxes combined with a suitable reform of piecewise linear income tax can be Pareto-improving. That is the relevant policy message of our results.

References

Apps, Patricia, Ngo Van Long and Ray Rees (2014), 'Optimal Piecewise Linear Income Taxation,' *Journal of Public Economic Theory* 16(4), 523–45.

Atkinson, Anthony B. (1973), 'How Progressive Should Income Tax Be,' in M. Parkin and A.R. Nobay (eds.), *Essays in Modern Economics* (London: Longman), 90–109.

Atkinson, Anthony B. and Joseph E. Stiglitz (1976), 'The Design of Tax Structure: Direct vs. Indirect Taxation,' *Journal of Public Economics* 6, 55–75.

Boadway, Robin, Maurice Marchand, and Pierre Pestieau (1994), 'Towards a Theory of the Direct-Indirect Tax Mix,' *Journal of Public Economics* 55, 71–88.

Boadway, Robin and Zhen Song (2016), 'Indirect Taxes for Redistribution: Should Necessity Goods be Favored?,' *Research in Economics* 70, 64–88.

Corlett, W. J., and D. C. Hague (1953), 'Complementarity and the Excess Burden of Taxation,' *Review of Economic Studies* 21, 21–30.

Deaton, Angus (1979),'Optimally Uniform Commodity Taxes,' *Economics Letters* 2, 357–61.

Diamond, Peter A. (1980), 'Income Taxation with Fixed Hours of Work,' *Journal of Public Economics* 13, 101–10.

Hellwig, Martin F. (2009), 'A Note on Deaton's Theorem on the Undesirability of Nonuniform Excise Taxation.' *Economic Letters* 105, 186–88.

Hellwig, Martin F. (2010), 'A Generalization of the Atkinson-Stiglitz (1976) Theorem on the Undesirability of Nonuniform Excise Taxation,' *Economic Letters* 108, 156–58.

Jacquet, Laurence, Etienne Lehmann and Bruno Van der Linden (2013), 'The Optimal Marginal Tax Rates with both Extensive and Intensive Responses,' *Journal of Economic Theory* 148, 1770–805.

Kaplow, Louis (2006), 'On the Desirability of Commodity Taxation even when Income Taxation is not Optimal,' *Journal of Public Economics* 90, 1235–50.

Laroque, Guy (2005), 'Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof,' *Economics Letters* 87, 141–44.

Mirrlees, James A. (1971), 'An Exploration in the Theory of Optimum Income Taxation.' *Review of Economic Studies* 38, 175–208.

Mirrlees, James, Stuart Adam, Timothy Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba (2011), *Tax by Design: The Mirrlees Review* (Oxford: Oxford University Press).

Piketty, Thomas and Emmanuel Saez and Thomas Piketty (2013), 'Optimal Labor Income Taxation,' in Auerbach, A., R. Chetty, M. Feldstein and E.Saez (eds.), *Handbook of Public Economics*, Volume 5 (Amsterdam: North-Holland), 391–474.

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

Saez, Emmanuel (2002), 'Optimal Income Transfer Programs: Intensive vs Extensive Labor Supply Responses,' *Quarterly Journal of Economics* 117,1039–73.

Sandmo, Agnar (1976), 'Optimal Taxation: An Introduction to the Literature,' *Journal of Public Economics* 6, 37–54.

Sheshinski, Eytan (1972), 'The Optimal Linear Income Tax,' *Review of Economic Studies* 39, 297–302.

Slemrod, Joel, Shlomo Yitzhaki, Joram Mayshar, and Michael Lundholm (1994), 'The Optimal Two-Bracket Linear Income Tax,' *Journal of Public Economics* 53, 269–90.

Tuomala, Matti (1990), *Optimal Income Tax and Redistribution* (Oxford: Clarendon Press).

Appendix A

Deaton theorem when utility is quasi-homothetic in a subset of goods

Assume utility is quasi-homothetic in all goods except good m. Let there be a commodity tax τ_m on good m and commodity tax τ on good $k \neq m$. The commodity taxes on all other goods are assumed to be zero. We can show that $\tau = 0$ in the optimum. The approach is similar to the proof of Proposition 1. Using the first-order conditions on t_1, t_2, a and τ_m evaluated at $\tau = 0$, we show that $\partial\mathcal{L}/\partial\tau$ also evaluated at $\tau = 0$ is zero.

Note first that since government policy variables now include τ_m, the uncompensated and compensated demands for goods by all persons and supply of labour for participants with incomes in the two brackets, as well as expenditure functions now include $1+\tau_m$ as an argument. Similarly, indirect utility functions and participation functions include τ_m as an argument. The envelope conditions and Slutsky equations apply with respect to changes in τ_m as well.

Social welfare is now given by

$$h^0 W\big(v^0(\tau,\tau_m,a)\big) + \sum_{i<\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{1i}} W\Big(v^i\big(\tau,\tau_m,(1-t_1)w^i,a\big) - \alpha^i\Big)dF^i(\alpha^i)$$

$$+\sum_{i>\hat{\imath}} n^i \int_{\underline{\alpha}^i}^{\hat{\alpha}^{2i}} W\Big(v^2\big(\tau,\tau_m,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big) - \alpha^i\Big)dF^i(\alpha^i)$$

and the government's budget constraint is:

$$a = t_1 \sum_{i<\hat{\imath}} h^1\big(\tau,\tau_m,(1-t_1)w^i,a\big)w^i\ell^1\big(1+\tau,1+\tau_m,(1-t_1)w^i,a\big)$$

$$+(t_1-t_2)\sum_{i>\hat{\imath}} h^2\big(\tau,\tau_m,(1-t_2)w^i,(t_2-t_1)\hat{y},a\big)\hat{y}$$

$$+t_2\sum_{i>\hat{\imath}} h^2\big(\tau,\tau_m,(1-t_2)w^i,(t_2-t_1)\hat{y},a\big)w^i\ell^2\big(1+\tau,1+\tau_m,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big)$$

$$+\tau\Big(h^0 x_k^0(1+\tau,1+\tau_m,a)+\sum_{i<\hat{\imath}} h^{1i}x_k^1\big(1+\tau,1+\tau_m,(1-t_1)w^i,a\big)$$
$$+\sum_{i>\hat{\imath}} h^{2i}x_k^2\big(1+\tau,1+\tau_m,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big)\Big)$$
$$+\tau_m\Big(h^0 x_m^0(1+\tau,1+\tau_m,a)+\sum_{i<\hat{\imath}} h^{1i}x_m^1\big(1+\tau,1+\tau_m,(1-t_1)w^i,a\big)$$
$$+\sum_{i>\hat{\imath}} h^{2i}x_m^2\big(1+\tau,1+\tau_m,(1-t_2)w^i,(t_2-t_1)\hat{y}+a\big)\Big).$$

The government maximizes social welfare subject to the budget constraint. Denote the Lagrangian expression for this problem as $\mathcal{L}(a,t_1,t_2,\tau_m,\tau)$. The first-order condition on a evaluated at $\tau=0$ is:

$$\mathcal{L}_a\Big|_{\tau=0} \equiv h^0 W_0' v_a^0+\sum_{i<\hat{\imath}} n^i v_a^{1i}\int_{\underline{\alpha}^i}^{\hat{\alpha}^{1i}} W_{1i}' dF^i(\alpha^i)+\sum_{i>\hat{\imath}} n^i v_a^{2i}\int_{\underline{\alpha}^i}^{\hat{\alpha}^{2i}} W_{2i}' dF^i(\alpha^i)+\lambda\Bigg(t_1\sum_{i<\hat{\imath}} h_a^{1i} w^i \ell^{1i}$$
$$+t_1\sum_{i>\hat{\imath}} h_a^{2i}\hat{y}+t_2\sum_{i>\hat{\imath}} h_a^{2i}(w^i\ell^{2i}-\hat{y})+t_1\sum_{i<\hat{\imath}} h^{1i}w^i\ell_a^{1i}+t_2\sum_{i>\hat{\imath}} h^{2i}w^i\ell_a^{2i} \tag{48}$$
$$+\tau_m\Big(h^0\frac{\partial x_m^0}{\partial a}+\sum_{i<\hat{\imath}} h^{1i}\frac{\partial x_m^{1i}}{\partial a}+\sum_{i>\hat{\imath}} h^{2i}\frac{\partial x_m^{2i}}{\partial a}+\sum_{i<\hat{\imath}} h_a^{1i}(x_m^{1i}-x_m^0)+\sum_{i>\hat{\imath}} h_a^{2i}(x_m^{2i}-x_m^0)\Big)-1\Bigg)=0.$$

Define the average net social marginal values of an additional unit of income when $\tau=0$:

$$b^0=\frac{\beta^0}{\lambda}+\tau_m\frac{\partial x_m^0}{\partial a},$$
$$b^{1i}=\frac{\beta^{1i}}{\lambda}+\frac{(t_1w^i\ell^{1i}+\tau_m(x_m^{1i}-x_m^0))h_a^{1i}}{h^{1i}}+t_1w^i\ell_a^{1i}+\tau_m\frac{\partial x_m^{1i}}{\partial a}\quad\text{for}\quad i<\hat{\imath}, \tag{49}$$
$$b^{2i}=\frac{\beta^{2i}}{\lambda}+\frac{\big(t_2(w^i\ell^{2i}-\hat{y})+t_1\hat{y}+\tau_m(x_m^{2i}-x_m^0)\big)h_a^{2i}}{h^{2i}}+t_2w^i\ell_a^{2i}+\tau_m\frac{\partial x_m^{2i}}{\partial a}\quad\text{for}\quad i>\hat{\imath}.$$

where β^0, β^{1i} and β^{2i} are defined as before. Using these definitions and the fact that the total population is normalized to unity, the first-order condition on a in (48) becomes:

$$h^0(1-b^0)=-\sum_{i<\hat{\imath}} h^{1i}(1-b^{1i})-\sum_{i>\hat{\imath}} h^{2i}(1-b^{2i}). \tag{50}$$

Using (49) and (50) as well as the analogues to the envelope expressions (2) and (4), the derivatives of the participation functions (7) and (10), and the Slutsky equations (3) and (6), we can write the first-order conditions on t_1, t_2 and τ_m in the government problem evaluated at $\tau=0$ as:

$$\mathcal{L}_{t_1}\Big|_{\tau=0}\equiv$$

$$\sum_{i<\hat{\imath}} h^{1i}(1-b^{1i})w^i\ell^{1i} + \sum_{i>\hat{\imath}} h^{2i}(1-b^{2i})\hat{y} - t_1 \sum_{i<\hat{\imath}} h^{1i}w^iw^i\tilde{\ell}^{1i}_{(1-t_1)w^i} - \tau_m \sum_{i<\hat{\imath}} h^{1i}\frac{\partial \tilde{x}^{1i}_m}{\partial(1-t_1)w^i}w^i$$
$$-t_1 \sum_{i<\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{1i})w^i\ell^{1i}w^i\ell^{1i} - t_2 \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y})\hat{y} - t_1 \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})\hat{y}\hat{y} \quad (51)$$
$$-\tau_m \sum_{i<\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{1i})(x^{1i}_m - x^0_m)w^i\ell^{1i} - \tau_m \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(x^{2i}_m - x^0_m)\hat{y} = 0,$$

$$\mathcal{L}_{t_2}\big|_{\tau=0} \equiv \sum_{i>\hat{\imath}} h^{2i}(1-b^{2i})(w^i\ell^{2i}-\hat{y}) - t_2 \sum_{i>\hat{\imath}} h^{2i}w^iw^i\tilde{\ell}^{2i}_{(1-t_2)w^i} - \tau_m \sum_{i>\hat{\imath}} h^{2i}\frac{\partial \tilde{x}^{2i}_m}{\partial(1-t_2)w^i}w^i$$
$$-t_1 \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})\hat{y}(w^i\ell^{2i}-\hat{y}) - t_2 \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y})(w^i\ell^{2i}-\hat{y}) \quad (52)$$
$$-\tau_m \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(x^{2i}_m - x^0_m)(w^i\ell^{2i}-\hat{y}) = 0,$$

$$\mathcal{L}_{\tau_m}\big|_{\tau=0} \equiv \sum_{i<\hat{\imath}} h^{1i}(1-b^{1i})(x^{1i}_m - x^0_m) + \sum_{i<\hat{\imath}} h^{2i}(1-b^{2i})(x^{2i}_m - x^0_m)$$
$$-t_1 \sum_{i<\hat{\imath}} v^0_a(x^{1i}_m - x^0_m)n^i f^i(\hat{\alpha}^{1i})w^i\ell^{1i} + t_1 \sum_{i<\hat{\imath}} h^{1i}w^i\tilde{\ell}^{1i}_{\tau_m}$$
$$-t_1 \sum_{i>\hat{\imath}} v^0_a(x^{2i}_m - x^0_m)n^i f^i(\hat{\alpha}^{2i})\hat{y} - t_2 \sum_{i>\hat{\imath}} v^0_a(x^{2i}_m - x^0_m)n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y}) + t_2 \sum_{i>\hat{\imath}} h^{2i}w^i\tilde{\ell}^{2i}_{\tau_m}$$
$$+\tau_m\left(h^0\frac{\partial \tilde{x}^0_m}{\partial(1+\tau_m)} + \sum_{i<\hat{\imath}} h^{1i}\frac{\partial \tilde{x}^{1i}_m}{\partial(1+\tau_m)} + \sum_{i>\hat{\imath}} h^{2i}\frac{\partial \tilde{x}^{2i}_m}{\partial(1+\tau_m)}\right) \quad (53)$$
$$-\tau_m\left(\sum_{i<\hat{\imath}} v^0_a(x^{1i}_m - x^0_m)n^i f^i(\hat{\alpha}^{1i})(x^{1i}_m - x^0_m) + \sum_{i>\hat{\imath}} v^0_a(x^{2i}_m - x^0_m)n^i f^i(\hat{\alpha}^{2i})(x^{2i}_m - x^0_m)\right) = 0.$$

Similarly, take the derivative of the government's Lagrangian expression with respect to τ at $\tau = 0$ and using the same procedure rewrite it as:

$$\frac{1}{\lambda}\mathcal{L}_{\tau}\big|_{\tau=0} \equiv \sum_{i<\hat{\imath}} h^{1i}\left(1-b^{1i}\right)(x^{1i}_k - x^0_k) + \sum_{i>\hat{\imath}} h^{2i}\left(1-b^{2i}\right)(x^{2i}_k - x^0_k)$$
$$+t_1 \sum_{i<\hat{\imath}} h^{1i}w^i\tilde{\ell}^{1i}_{\tau} + t_2 \sum_{i>\hat{\imath}} h^{2i}w^i\tilde{\ell}^{2i}_{\tau} + \tau_m h^0\frac{\partial \tilde{x}^0_m}{\partial(1+\tau)} + \tau_m \sum_{i<\hat{\imath}} h^{1i}\frac{\partial \tilde{x}^{1i}_m}{\partial(1+\tau)} + \tau_m \sum_{i>\hat{\imath}} h^{2i}\frac{\partial \tilde{x}^{2i}_m}{\partial(1+\tau)}$$
$$-t_1 \sum_{i<\hat{\imath}} v^0_a(x^{1i}_k - x^0_k)n^i f^i(\hat{\alpha}^{1i})w^i\ell^{1i} - t_1 \sum_{i>\hat{\imath}} v^0_a(x^{2i}_k - x^0_k)n^i f^i(\hat{\alpha}^{2i})\hat{y} \quad (54)$$
$$-t_2 \sum_{i>\hat{\imath}} v^0_a(x^{2i}_k - x^0_k)n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y})$$
$$-\tau_m \sum_{i<\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{1i})(x^{1i}_m - x^0_m)(x^{1i}_k - x^0_k) - \tau_m \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(x^{2i}_m - x^0_m)(x^{2i}_k - x^0_k).$$

Assume as in Deaton (1979) that preferences can be written as $u\big(\phi(x_1, \cdots, x_{m-1}), x_m, \ell\big)$, so goods' demands will be given by:

$$x_k^0 = \rho_k(1+\tau) + \gamma_k(1+\tau)\big(a - (1+\tau_m)x_m^0\big), \tag{55}$$

$$x_k^{1i} = \rho_k(1+\tau) + \gamma_k(1+\tau)\big((1-t_1)w^i\ell^{1i} + a - (1+\tau_m)x_m^{1i}\big), \tag{56}$$

$$x_k^{2i} = \rho_k(1+\tau) + \gamma_k(1+\tau)\big((1-t_1)w^i\ell^{2i} + (t_2-t_1)\hat{y} + a - (1+\tau_m)x_m^{2i}\big). \tag{57}$$

Using (55)–(57), we obtain Lemma 4.

Lemma 4 *With Deaton preferences in all goods $j = 1, \cdots, m-1$, the following expressions hold:*

$$x_k^{1i} - x_k^0 = \gamma_k(1-t_1)w^i\ell^{1i} - \gamma_k(1+\tau_m)(x_m^{1i} - x_m^0) \qquad i < \hat{\imath},$$

$$x_k^{2i} - x_k^0 = \gamma_k\big((1-t_2)w^i\ell^{2i} + (t_2-t_1)\hat{y}\big) - \gamma_k(1+\tau_m)(x_m^{2i} - x_m^0) \qquad i > \hat{\imath}.$$

In equilibrium the following must be satisfied

$$e^0(1+\tau, 1+\tau_m, \bar{u}) = a,$$

$$\tilde{x}_k^0\big(1+\tau, 1+\tau_m, \bar{u}\big) = x_k^0\big(1+\tau, 1+\tau_m, e(1+\tau, 1+\tau_m, \bar{u})\big).$$

Using these relationships to substitute out x_k^0, x_m^0 and a in (55), differentiating with respect to $1+\tau_m$, and using $e^0_{1+\tau_m} = \tilde{x}_m^0$, we obtain:

$$\frac{\partial \tilde{x}_k^0}{\partial(1+\tau_m)} = \gamma_k\left(-(1+\tau_m)\frac{\partial \tilde{x}_m^0}{\partial(1+\tau_m)}\right) = \frac{\partial \tilde{x}_m^0}{\partial(1+\tau)} \tag{58}$$

where the last equality follows from the symmetry of the substitution effect (Young's theorem). Following the same procedure for $i < \hat{\imath}$ and $i > \hat{\imath}$ yields

$$\frac{\partial \tilde{x}_k^{1i}}{\partial(1+\tau_m)} = \gamma_k(1-t_1)w^i\tilde{\ell}_{\tau_m}^{1i} - \gamma_k(1+\tau_m)\frac{\partial \tilde{x}_m^{1i}}{\partial(1+\tau_m)} = \frac{\partial \tilde{x}_m^{1i}}{\partial(1+\tau)}, \tag{59}$$

$$\frac{\partial \tilde{x}_k^{2i}}{\partial(1+\tau_m)} = \gamma_k(1-t_2)w^i\tilde{\ell}_{\tau_m}^{2i} - \gamma_k(1+\tau_m)\frac{\partial \tilde{x}_m^{2i}}{\partial(1+\tau_m)} = \frac{\partial \tilde{x}_m^{2i}}{\partial(1+\tau)}. \tag{60}$$

Since

$$\tilde{\ell}_{\tau_m}^{1i} = -\frac{\partial \tilde{x}_m^{1i}}{\partial((1-t_1)w^i)} \quad \text{and} \quad \tilde{\ell}_{\tau_m}^{2i} = -\frac{\partial \tilde{x}_m^{2i}}{\partial((1-t_2)w^i)}$$

by the symmetry of the substitution effect, Lemma 5 follows.

Lemma 5 *With Deaton preferences in goods $j = 1, \cdots, m-1$, the following relations hold:*

$$\frac{\partial \tilde{x}_m^0}{\partial(1+\tau)} = -\gamma_k(1+\tau_m)\frac{\partial \tilde{x}_m^0}{\partial(1+\tau_m)},$$

$$\frac{\partial \tilde{x}_m^{1i}}{\partial(1+\tau)} = -\gamma_k(1-t_1)w^i\frac{\partial \tilde{x}_m^{1i}}{\partial((1-t_1)w^i)} - \gamma_k(1+\tau_m)\frac{\partial \tilde{x}_m^{1i}}{\partial 1+\tau_m},$$

$$\frac{\partial \tilde{x}_m^{2i}}{\partial(1+\tau)} = -\gamma_k(1-t_2)w^i\frac{\partial \tilde{x}_m^{2i}}{\partial((1-t_2)w^i)} - \gamma_k(1+\tau_m)\frac{\partial \tilde{x}_m^{2i}}{\partial(1+\tau_m)}.$$

Next, in equilibrium the following must be satisfied for workers of types $i < \hat{\imath}$:

$$e^1\big(1+\tau, 1+\tau_m, (1-t_1)w^i, \bar{u}\big) = a,$$

$$\tilde{x}_k^1\big(1+\tau, 1+\tau_m, (1-t_1)w^i, \bar{u}\big) = x_k^1\big(1+\tau, 1+\tau_m, (1-t_1)w^i, e(1+\tau, (1-t_1)w^i, \bar{u})\big),$$

$$\tilde{\ell}^1\big(1+\tau, 1+\tau_m, (1-t_1)w^i, \bar{u}\big) = \ell^1\big(1+\tau, 1+\tau_m, (1-t_1)w^i, e(1+\tau, (1-t_1)w^i, \bar{u})\big).$$

Using these relationships to substitute out x_k^{1i}, x_m^{1i}, ℓ^{1i} and a in (56) and differentiating with respect to the after-tax wage $(1-t_1)w^i$ yields:

$$\begin{aligned}\frac{\partial \tilde{x}_k^{1i}}{\partial((1-t_1)w^i)} &= \gamma_k(1-t_1)w^i\tilde{\ell}^{1i}_{(1-t_1)w^i} - \gamma_k(1+\tau_m)\frac{\partial \tilde{x}_m^{1i}}{\partial((1-t_1)w^i)} \\ &= \gamma_k(1-t_1)w^i\tilde{\ell}^{1i}_{(1-t_1)w^i} + \gamma_k(1+\tau_m)\tilde{\ell}^{1i}_{\tau_m}\end{aligned}$$

where the second equality follows from the symmetry if the substitution effect. Following an identical procedure for workers of types $i > \hat{\imath}$, we obtain Lemma 6.

Lemma 6 *With Deaton preferences in all goods $j = 1, \cdots, m-1$, the following expressions hold:*

$$\tilde{\ell}_\tau^{1i} = -\gamma_k(1-t_1)w^i\tilde{\ell}^{1i}_{(1-t_1)w^i} - \gamma_k(1+\tau_m)\tilde{\ell}^{1i}_{\tau_m} \qquad i < \hat{\imath},$$

$$\tilde{\ell}_\tau^{2i} = -\gamma_k(1-t_2)w^i\tilde{\ell}^{2i}_{(1-t_2)w^i} - \gamma_k(1+\tau_m)\tilde{\ell}^{2i}_{\tau_m} \qquad i > \hat{\imath}.$$

Substituting the expressions in Lemmas 4, 5, and 6 into (54), we obtain the following

$$\frac{1}{\lambda}\mathcal{L}_\tau\big|_{\tau=0} \equiv \gamma_k(1-t_1)\Bigg(\sum_{i<\hat{\imath}} h^{1i}\left(1-b^{1i}\right)w^i\ell^{1i} - t_1\sum_{i\leqslant\hat{\imath}} v_a^0 w^i\ell^{1i}n^i f^i(\hat{\alpha}^{1i})w^i\ell^{1i}$$

$$
\begin{aligned}
&+\sum_{i>\hat{\imath}} h^{2i}\left(1-b^{2i}\right)\hat{y} - t_1 \sum_{i\leqslant\hat{\imath}} h^{1i} w^i w^i \tilde{\ell}^{1i}_{(1-t_1)w^i} - t_2 \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(w^i\ell_{2i}-\hat{y})\hat{y} \\
&-t_1 \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})\hat{y}\hat{y} - \tau_m \sum_{i<\hat{\imath}} h^{1i}\frac{\partial \tilde{x}^{1i}_m}{\partial(1-t_1)w^i} w^i \\
&-\tau_m \sum_{i<\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{1i})(x^{1i}_m - x^0_m) w^i \ell^{1i} - \tau_m \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(x^{2i}_m - x^0_m)\hat{y} \Bigg) \\
&+\gamma_k(1-t_2)\Bigg(\sum_{i\geqslant\hat{\imath}+1} h^{2i}\left(1-b^{1i}\right)(w^i\ell^{2i}-\hat{y}) - t_2\sum_{i>\hat{\imath}} h^{2i} w^i w^i \tilde{\ell}^{2i}_{(1-t_2)w^i} \qquad (61)\\
&-t_2 \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y})(w^i\ell^{2i}-\hat{y}) - t_1 \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})\hat{y}(w^i\ell^{2i}-\hat{y}) \\
&-\tau_m \sum_{i>\hat{\imath}} h^{2i}\frac{\partial \tilde{x}^{2i}_m}{\partial(1-t_2)w^i} w^i - \tau_m \sum_{i>\hat{\imath}} v^0_a n^i f^i(\hat{\alpha}^{2i})(x^{2i}_m - x^0_m)(w^i\ell^{2i}-\hat{y})\Bigg) \\
&-\gamma_k(1+\tau_m)\Bigg(\sum_{i<\hat{\imath}} h^{1i}(1-b^{1i})(x^{1i}_m - x^0_m) + \sum_{i<\hat{\imath}} h^{2i}(1-b^{2i})(x^{2i}_m - x^0_m) \\
&-t_1 \sum_{i<\hat{\imath}} v^0_a (x^{1i}_m - x^0_m) n^i f^i(\hat{\alpha}^{1i}) w^i \ell^{1i} + t_1 \sum_{i<\hat{\imath}} h^{1i} w^i \tilde{\ell}^{1i}_{\tau_m} \\
&-t_1 \sum_{i>\hat{\imath}} v^0_a (x^{2i}_m - x^0_m) n^i f^i(\hat{\alpha}^{2i})\hat{y} - t_2 \sum_{i>\hat{\imath}} v^0_a (x^{2i}_m - x^0_m) n^i f^i(\hat{\alpha}^{2i})(w^i\ell^{2i}-\hat{y}) + t_2 \sum_{i>\hat{\imath}} h^{2i} w^i \tilde{\ell}^{2i}_{\tau_m} \\
&+\tau_m\left(h^0 \frac{\partial \tilde{x}^0_m}{\partial(1+\tau_m)} + \sum_{i<\hat{\imath}} h^{1i}\frac{\partial \tilde{x}^{1i}_m}{\partial(1+\tau_m)} + \sum_{i>\hat{\imath}} h^{2i}\frac{\partial \tilde{x}^{2i}_m}{\partial(1+\tau_m)}\right) \\
&-\tau_m\left(\sum_{i<\hat{\imath}} v^0_a (x^{1i}_m - x^0_m) n^i f^i(\hat{\alpha}^{1i})(x^{1i}_m - x^0_m) + \sum_{i>\hat{\imath}} v^0_a (x^{2i}_m - x^0_m) n^i f^i(\hat{\alpha}^{2i})(x^{2i}_m - x^0_m)\right)\Bigg).
\end{aligned}
$$

By substituting the first-order conditions on t_1, t_2 and τ_m given by (51), (52), and (53) respectively into (61), Proposition 3 follows directly.

Proposition 3 *With Deaton preferences in a subset of goods $j = 1, \cdots, m-1$ and both the optimal piecewise linear income tax system and optimal commodity tax on good m in place, it follows that*

$$
\frac{1}{\lambda}\mathcal{L}_\tau\big|_{\tau=0} = 0.
$$

Therefore, if preferences are weakly separable in goods and labour and quasi-homothetic in a subset of goods as in Deaton (1979), and if the government sets the piecewise linear income tax and commodity taxes on all other goods optimally, commodity taxes on the subset of goods are redundant.

Appendix B

Hellwig extension when utility is quasi-homothetic in a subset of goods

The utility function is $u\big(\phi(\mathbf{x}_Q), \mathbf{x}_N, \ell\big)$ with sub-utility $\phi(\mathbf{x}_Q)$ quasi-homothetic. Write the budget constraints for the three types of individuals (30)–(32) as:

$$\sum_{k\in Q} q_k x_k^0 = a - \sum_{j\in N} q_j x_j^0,$$

$$\sum_{k\in Q} q_k x_k^{1i} = (1-t_1)w^i \ell^{1i} + a - \sum_{j\in N} q_j x_j^{1i},$$

$$\sum_{k\in Q} q_k x_k^{2i} = (1-t_2)w^i \ell^{2i} + (t_2 - t_1)\hat{y} + a - \sum_{j\in N} q_j x_j^{2i}.$$

Let $\big(\mathbf{x}_Q^h(s), \mathbf{x}_N^h(s), d^h(s)\big)$ be the allocation for individuals $h = 0, 1i, 2i$ when income tax parameters are t_1, t_2 and a and differential commodity tax rates are $\tau_1, \cdots, \tau_m$. Let the value of the sub-utilities of goods in Q in the initial allocation be given by the analogue of (33):

$$\omega^0(s) \equiv \phi\big(\mathbf{x}_Q^0(s)\big), \qquad \omega^{1i}(s) \equiv \phi\big(\mathbf{x}_Q^{1i}(s)\big), \qquad \omega^{2i}(s) \equiv \phi\big(\mathbf{x}_Q^{2i}(s)\big).$$

We show that the allocation under this initial tax system is Pareto-dominated by one with income tax parameters $\hat{t}_1$, $\hat{t}_2$ and $\hat{a}$, commodity taxes $\hat{\tau}_j$ for goods $j \in N$ and zero for those in set Q. We proceed as above in two steps.

Step 1

Consider the following problem for a type-h individual, given $d^h(s), \mathbf{x}_N^h(s)$:

$$\min_{\mathbf{x}_Q} \sum_{k\in Q} p_k x_k \qquad \text{s.t.} \qquad \phi(\mathbf{x}_Q) = \omega^h(s)$$

where $\omega^h(s)$ is defined above. The solution gives the allocation $\left(\hat{\mathbf{x}}_Q^h, \mathbf{x}_N^h(s), d^h(s)\right)$ with the same utility as in the initial situation above. For each type h, the above problem implies:

$$\sum_{k\in Q} p_k \hat{x}_k^h < \sum_{k\in Q} p_k x_k^h(s).$$

That is, less resources are required to produce the subset of goods $\hat{\mathbf{x}}_Q^h$ than the initial allocation. The same applies for all three types of individuals, non-participants ($h = 0$), workers in the first tax bracket ($h = 1i$) and workers in the second tax bracket ($h = 2i$).

Step 2

We next show that for each individual h, the allocation $\left(\hat{\mathbf{x}}_Q^h, \mathbf{x}_N^h(s), d^h(s)\right)$ can be implemented by a piecewise linear progressive income tax $(\hat{t}_1, \hat{t}_2, \hat{a})$ and commodity taxes $\hat{\tau}_j$ for goods $j \in N$, with consumer prices equal to producer prices for goods $k \in Q$ when quasi-homothetic preferences apply to the latter.

By the Deaton conditions, the consumption allocations for goods $k \in Q$ obtained from the individuals' expenditure minimization problem given $d^h(s)$ and $\mathbf{x}_N^h(s)$ are

$$\hat{x}_k^0 = \rho_k(\mathbf{p}) + \gamma_k(\mathbf{p})\Big(a - \sum_{j\in N} q_j x_j^0(s)\Big) \qquad k \in Q,$$

$$\hat{x}_k^{1i} = \rho_k(\mathbf{p}) + \gamma_k(\mathbf{p})\Big((1-t_1)w^i\ell^{1i} + a - \sum_{j\in N} q_j x_j^{1i}(s)\Big) \qquad k \in Q,$$

$$\hat{x}_k^{2i} = \rho_k(\mathbf{p}) + \gamma_k(\mathbf{p})\Big((1-t_2)w^i\ell^{2i} + (t_2 - t_1)\hat{y} + a - \sum_{j\in N} q_j x_j^{2i}(s)\Big) \qquad k \in Q.$$

Multiplying each of these by p_k and summing over all goods in the set Q, we obtain:

$$\sum_{k\in Q} p_k \hat{x}_k^0 = \sum_{k\in Q} p_k \rho_k(\mathbf{p}) + \sum_{k\in Q} p_k \gamma_k(\mathbf{p})\Big(a - \sum_{j\in N} q_j x_j^0(s)\Big),$$

$$\sum_{k\in Q} p_k \hat{x}_k^{1i} = \sum_{k\in Q} p_k \rho_k(\mathbf{p}) + \sum_{k\in Q} p_k \gamma_k(\mathbf{p})\Big((1-t_1)w^i\ell^{1i} + a - \sum_{j\in N} q_j x_j^{1i}(s)\Big),$$

$$\sum_{k\in Q} p_k \hat{x}_k^{2i} = \sum_{k\in Q} p_k \rho_k(\mathbf{p}) + \sum_{k\in Q} p_k \gamma_k(\mathbf{p})\Big((1-t_2)w^i\ell^{2i} + (t_2 - t_1)\hat{y} + a - \sum_{j\in N} q_j x_j^{2i}(s)\Big).$$

Consider an alternative tax system $\left(\hat{t}_1, \hat{t}_2, \hat{a}, \hat{\mathbf{q}}_N\right)$, and suppose it is related to the original tax system as follows:

$$(1-\hat{t}_1) = \sum_{k\in Q} p_k \gamma_k(\mathbf{p})(1-t_1), \quad (1-\hat{t}_2) = \sum_{k\in Q} p_k \gamma_k(\mathbf{p})(1-t_2),$$

$$\hat{a} = \sum_{k \in Q} p_k \big(\rho_k(\mathbf{p}) + \gamma_k(\mathbf{p}) a\big), \quad \hat{q}_j = \sum_{k \in Q} p_k \gamma_k(\mathbf{p}) q_j \quad j \in N.$$

Substituting these into the above budget constraints, we obtain:

$$\sum_{k \in Q} p_k \hat{x}_k^0 = \hat{a} - \sum_{j \in N} \hat{q}_j x_j^0(s),$$

$$\sum_{k \in Q} p_k \hat{x}_k^{1i} = (1 - \hat{t}_1) w^i \ell^{1i} + \hat{a} - \sum_{j \in N} \hat{q}_j x_j^{1i}(s),$$

$$\sum_{k \in Q} p_k \hat{x}_k^{2i} = (1 - \hat{t}_2) w^i \ell^{2i} + (\hat{t}_2 - \hat{t}_1) \hat{y} + \hat{a} - \sum_{j \in N} \hat{q}_j x_j^{2i}(s).$$

These are the same budget constraints as in the initial situation with the new tax system and no commodity taxes on the subset of goods $k \in Q$. It shows that if type-h individuals are faced with the new tax system and consumer prices are equal to producer prices for $k \in Q$, they can choose $\hat{\mathbf{x}}_Q^h$ and the original levels of $d^h(s)$ and $\mathbf{x}_N^h(s)$. We have shown in Step 1 that if the individuals choose this allocation fewer resources are required and they are equally well-off as in the original tax system. This means that the lump-sum component $\hat{a}$ can be increased to use up the extra resources so everyone can be made better off.

Appendix C

The Deaton theorem does not hold with heterogeneous preferences

Assume following Deaton (1979) that there are a discrete number of individuals $h = 1, \cdots, H$ who differ in their wage rates. The government imposes a set of commodity taxes $t_1, \cdots, t_m$ on the m goods and gives a transfer of a to all persons. With uniform commodity taxes, this is equivalent to a linear progressive income tax.

Deaton defines the average demand for good k, $\overline{x}_k$, and the equity-weighted demand for good k, x_k^*, as follows:

$$\overline{x}_k = \frac{1}{H} \sum_h x_k^h, \qquad x_k^* = \sum_h \frac{\lambda^h}{H \overline{\lambda}} x_k^h \tag{62}$$

where

$$\lambda^h = \theta^h + \frac{\overline{\theta} r^h}{1 - \overline{r}}.$$

Here, θ^h is the marginal social utility of a dollar to person h, so it is equivalent to our β^h, and $\overline{\theta}$ is the average of the θ^h's. The term r^h is the marginal government revenue resulting from giving person h a dollar, and $\overline{r}$ is its average. So, λ^h is like our b^h, the net social marginal utility of a dollar to individual h.

Deaton shows that the optimal commodity tax system satisfies

$$\sum_{j=1}^{m} \overline{s}_{kj} t_j = -(\overline{x}_k - x_k^*) \qquad \forall\ k$$

where $\overline{s}_{kj}$ is the average of the substitution effects s_{kj}^h over all individuals h. If taxes are to be uniform, so $t_k = \tau p_k$, we require

$$\tau \overline{s_{k0} p_0} = \overline{x}_k - x_k^*. \tag{63}$$

Following (47), with heterogeneous preferences as assumed by Deaton, the demand for good k by individual h is:

$$x_k^h = -\frac{\mu_{q_k}^h(\mathbf{q}) + \psi_{q_k}(\mathbf{q}) d^h}{\psi(\mathbf{q})} = \rho_k^h(\mathbf{q}) + \gamma_k(\mathbf{q}) d^h.$$

Deaton shows that the left-hand side of (63) is proportional to $\gamma_k(\mathbf{q})$, and he claims that the right-hand side is also proportional to $\gamma_k(\mathbf{q})$ and independent of $\rho_k^h(\mathbf{q})$. In particular, he claims that:

$$\overline{x}_k - x_k^* = \gamma_k(\mathbf{q})(\overline{d} - d^*) \tag{64}$$

where $\overline{d}$ is average disposable income d^h, and d^* is the equity weighted average of d^h. However, (64) does not apply with heterogeneous preferences. To see this, use the definitions in (62) to give:

$$\overline{x}_k = \frac{1}{H} \sum_h \rho_k^h(\mathbf{q}) + \frac{\gamma_k(\mathbf{q})}{H} \sum_h d^h = \overline{\rho}_k(\mathbf{q}) + \gamma_k(\mathbf{q}) \overline{d},$$

$$x_k^* = \sum_h \frac{\lambda^h}{H\overline{\lambda}} \rho_k^h(\mathbf{q}) + \sum_h \frac{\lambda^h}{H\overline{\lambda}} \gamma_k(\mathbf{q}) d^h = \rho_k^*(\mathbf{q}) + \gamma_k(\mathbf{q}) d^*.$$

So, we obtain:

$$\overline{x}_k - x_k^* = \overline{\rho}_k(\mathbf{q}) - \rho_k^*(\mathbf{q}) + \gamma_k(\mathbf{q})(\overline{d} - d^*). \tag{65}$$

This is only equivalent to (64) if $\overline{\rho}_k(\mathbf{q}) = \rho_k^*(\mathbf{q})$, or:

$$\frac{1}{H} \sum_h \rho_k^h(\mathbf{q}) = \sum_h \frac{\lambda^h}{H\overline{\lambda}} \rho_k^h(\mathbf{q}).$$

This will not generally be the case, so the condition for uniform commodity taxation is not satisfied.

A similar condition can be obtained in our model when like Deaton we assume preferences over commodities are perfectly correlated with skill and take the Deaton form with differing intercepts for the Engel curves. Under these assumptions, we find that uniform commodity taxation are optimal provided:

$$\sum_{i\geq 0}\rho_\tau^i = n_0 b_0 \rho_\tau^0 + \sum_{i<\hat{\imath}}(n_i - h^{1i})b^{0i}\rho_\tau^i + \sum_{i>\hat{\imath}}(n_i - h^{2i})b^{0i}\rho_\tau^i + \sum_{i<\hat{\imath}} h^{1i}b^{1i}\rho_\tau^i + \sum_{i>\hat{\imath}} h^{2i}b^{2i}\rho_\tau^i$$

where the left-hand side expression is the average of ρ_τ^i given the population size is unity. The right-hand side expression is the average of $b\rho_\tau$ where from the optimal choice of the lump-sum transfer a the government will ensure the average value of b is unity. (That is, $H = 1$, $\lambda = b$ and $\bar{b} = 1$.) A sufficient condition for the above to be satisfied is if ρ_τ^i is the same for all individuals.