Optimal Commodity Taxation When Households Earn Multiple Incomes

Kevin Spiritus*

4th December 2023

Please click here for the latest version

Abstract

I characterize the optimal linear commodity taxes when households differ in multiple characteristics and earn multiple incomes, in presence of an optimal non-linear tax schedule on the taxpayers' labour incomes. The government should tax a commodity more heavily if, conditional on labour income, more deserving taxpayers consume larger quantities of that commodity. Furthermore, the government wishes to tax commodities at different rates to the extent that doing so reduces the distortions caused by the labour income tax. This is the case when households with different incomes have different consumption preferences, or when households with different labour supplies also have different consumption patterns.

Keywords: optimal commodity taxation, multidimensional taxation **JEL Classification:** H21, H24

^{*}Department of Economics, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands, e-mail: spiritus@ese.eur.nl. I am grateful grateful to David Agrawal and two anonymous referees for their insightful and constructive remarks. I also thank Bart Capéau, Aart Gerritsen, Bas Jacobs and Dirk Schindler for valuable comments. Declarations of interest: none.

1 Introduction

The optimal taxation of commodities remains a contentious issue. International organizations and academic reviews recommend that VAT rates are uniform (for example, OECD, 2020, p.42 and Mirrlees et al., 2011, ch.9). Yet most governments do apply reduced VAT rates. By doing so, they aim to increase the purchasing power of low-income groups, to stimulate the consumption of merit goods such as culture, and to redistribute towards specific groups, for example, through reduced rates for female hygiene products.¹

The recommendation that commodities are taxed at uniform rates is rooted in a seminal paper by Atkinson and Stiglitz [1976]. They show that when preferences are weakly separable between leisure and consumption and taxpayers differ only in their labour productivities, commodities should indeed be taxed at uniform rates. The reason is that with such a one-dimensional population, if all households have equal preferences, individual consumption patterns contain no information about the households' types beyond what is already revealed by the households' labour incomes. Given then that any redistribution attained by differentiated commodity taxes can also be attained by a tax on labour income with equal distortions of labour supply but without distortions of consumption decisions, it is best to not use differentiated commodity taxes for redistribution. Only when there are clear complementarities between labour supply and the consumption of specific goods, most notably childcare and transportation to work, can differentiated commodity taxes help reduce the distortions caused by a progressive tax on labour income (Gordon and Kopczuk, 2014 and Pirttilä and Suoniemi, 2014).

A notable shortcoming of the model of Atkinson and Stiglitz [1976] is that taxpayers differ only in their labour productivities. Saez [2002] studies the desirability of differentiated linear commodity taxes when taxpayers also differ in their tastes. Studying linear commodity taxes besides a non-linear income tax is useful due to the possibility of arbitrage between taxpayers, and because linear commodity taxes can be levied at the point of sale rather than on the individual level. Saez [2002] shows that starting from a situation without commodity taxes, introducing a small linear tax on one commodity increases social welfare if, on average, higher-income taxpayers

¹See, for example, OECD [2020] and Terricabras [2016]. According to OECD [2020], other goods enjoy reduced VAT rates because they are locally supplied and their production is labour-intensive. In this paper, I do not consider this motive.

have a relatively higher taste for consuming this commodity.²

Ferey et al. [2022] derive optimal tax conditions in the setting studied by Saez [2002], still for the case where households earn a single labour income. The relation between their conditions for the linear optimum, and the desirability conditions found by Saez [2002], remains unclear. Saez [2002, p.229] leaves this as an "extremely useful" task for future research.

In this paper, I derive formulas for the optimal linear taxation of commodities in a broad generalization of the model of Atkinson and Stiglitz [1976]. By doing so, I identify the sufficient statistics that allow quantifying the optimal commodity tax rates. Extending Saez [2002], I allow for taxpayers to differ in multiple characteristics such as tastes, productivities, gender, etc. Furthermore, I allow for taxpayers to earn multiple labour incomes. Allowing multiple labour incomes in the household allows studying, for example, the effect of different complementarities of primary and secondary labour supply with the consumption of certain goods (for example, childcare). I briefly characterize the optimal non-linear schedule for the multidimensional labour income tax, and then focus on the optimal linear tax rates on commodities.

Main Results My results are as follows. If the tax schedule on the labour incomes is optimal, then one should only deviate from uniform commodity taxes if doing so either contributes to the distributional objectives of the government in ways that cannot be accomplished through the tax on labour income, or if it reduces the distortions caused by the taxation of labour income.

A first reason to deviate from uniform commodity taxes is when, on average and conditional on labour income, the covariance between the social marginal utilities and the consumption of a particular good differs from zero. This correlation could be due to differences in preferences affecting the marginal utility of consumption, or it could be influenced by societal judgments incorporated in the welfare function and the cardinalization of utility functions. For instance, the government might find a strong taste for a certain good to be socially beneficial, or it might aim to compensate those who consume specific goods to offset other injustices.³ The larger the covari-

 $^{^{2}}$ The word 'tastes' here is to be interpreted as a characteristic of individual preferences. Differences in consumption are thus driven by differences in consumption preferences rather than by differences in disposable income.

³A crucial assumption is that the government cannot perfectly observe the type of the taxpayer, or cannot directly condition taxes on the type. If the government had this capability, Kaplow [2008] argues that it would be more effective to adjust income taxes based on the taxpayer's type rather than implementing differentiated commodity taxes.

ance between the social marginal utilities and the consumption of a specific good among individuals with the same income, the greater the distortion that the government should tolerate from taxing that good. This motive for deviating from uniform commodity taxes is absent if taxpayers differ only in their labour productivities.

A second reason to deviate from uniform commodity taxes arises when, on average and conditional on labour income, there is a difference between the cross-sectional variation and the individual variation in the consumption of some good as one changes one of the labour incomes. I show that this can be the case for two reasons. First, around the labour incomes under consideration, taxpayers with different incomes may have different consumption preferences. Second, again around the labour incomes under consideration, taxpayers who earn the same disposable incomes but supply different quantities of labour, may consume different quantities of the good. This could be due to complementarities between leisure and consumption of that good. If there is a difference between the cross-sectional and individual variations for either of these reasons, then consumption patterns reveal information about taxpayer types, enabling more efficient redistribution than through labour tax alone. The larger this difference, the larger the acceptable distortion from taxing that good. Moreover, the optimal distortion is larger if the marginal excess burden of taxing labour income is larger, because shifting taxation from labour to consumption yields greater benefits.

A third reason to deviate from uniform commodity taxes is more technical. If on average and conditional on labour income, households for whom the individual variation of the consumption of some good with one of the labour incomes is larger also face a larger marginal excess burden of taxing that labour income, then shifting the burden of taxation from that labour income to the good under consideration increases the efficiency of the tax system. As noted by Saez [2002], such covariance would be difficult to measure empirically. This third term is absent in models where households differ only in their labour productivities.

Saez [2002] outlines three conditions under which, in presence a single labour income, optimal commodity taxes would be uniform. However, he does not show how these conditions relate to a condition that characterizes the optimum. In this paper, I do characterize the optimum in the presence of multiple optimally taxed labour incomes. I show how the various marginal costs and benefits of perturbing a tax rate, as identified by Saez [2002], should be balanced against each other. When all three conditions identified by Saez [2002] are met, some terms remain in the optimal tax expression, but the resulting expression indeed implies uniform commodity taxation, even in presence of multiple labour incomes.

Relationship to Existing Literature This paper makes several contributions to the literature. Saez [2002] studies the linear taxation of commodities when taxpayers differ in their labour productivities and their tastes, but they earn only one labour income. He only finds desirability conditions for non-uniform commodity taxes and does not characterize the optimal tax rates. Jacobs and Boadway [2014] do derive optimal linear commodity taxes, for the case where households differ only in their labour productivities and they earn only one labour income. Because they consider a population that differs in one dimension only, they find less reasons than Saez [2002] to levy differentiated commodity taxes. Kaplow [2008] studies various ways in which preference heterogeneity leads to correlations between the social marginal utilities and the consumption of particular goods. He does not show how these correlations fit into an optimal tax expression. Ferey et al. [2022] derive optimal linear commodity taxes for the case studied by Saez [2002], without showing how their optimality conditions relate to the desirability conditions found by Saez [2002]. I extend the literature by deriving the optimal linear commodity tax rates for the general case, allowing households to differ in characteristics beyond their labour productivities and preferences, and to earn multiple labour incomes. I demonstrate how the three conditions for uniform commodity taxes, as identified by Saez [2002], correspond to the terms in the optimal tax expression.

Road Map We proceed as follows. I introduce the model in Section 2. In Section 3, I derive properties of taxpayer behaviour, and I introduce sufficient statistics conditional on labour income. I present the optimal tax schedules and their relation to the findings of Saez [2002] in Section 4, before concluding in Section 5.

2 The model

2.1 Households

Households differ in labour productivities $\boldsymbol{w} \equiv (w^1, \ldots, w^L) \in \mathcal{W}$, where $L \geq 1$ and $\mathcal{W} \subset \mathbb{R}^L_+$ is closed and convex. For each productivity dimension $l = 1, \ldots, L$, households supply labour ℓ^l , yielding labour income $z^l = w^l \ell^l$. Households pay labour income taxes $T(\boldsymbol{z})$, where the tax schedule can depend on all labour incomes $\boldsymbol{z} \equiv$ (z^1, \ldots, z^L) simultaneously in complicated ways.

Households spend their disposable income on *numéraire* good c and commodities $\boldsymbol{x} \equiv (x^1, \ldots, x^J)$. Producer prices are normalized to one, and commodities are taxed at linear rates $\boldsymbol{t} \equiv (t^1, \ldots, t^J)$. The budget constraint for a household is:

$$c + \sum_{j=1}^{J} x^{j} (1+t^{j}) = \sum_{l=1}^{L} z^{l} - T(\boldsymbol{z}).$$
(1)

Besides their labour productivities, households can differ in other dimensions. We parametrise the additional characteristics using the vector $\boldsymbol{\theta} \equiv (\theta^1, \dots, \theta^K) \in \Theta \subset \mathbb{R}_+^K$. We denote the type vectors as $\boldsymbol{\omega} \equiv (\boldsymbol{w}, \boldsymbol{\theta}) \in \Omega \equiv \mathcal{W} \times \Theta$. We denote the joint cumulative distribution of the types as $G^{\boldsymbol{\omega}}(\boldsymbol{\omega})$ with corresponding twice continuously differentiable density function $g^{\boldsymbol{\omega}}(\boldsymbol{\omega})$. We do not make any assumptions about the joint distribution of the types. The only exception is that we assume there are no deterministic relationships between the productivity parameters \boldsymbol{w} and the additional characteristics $\boldsymbol{\theta}$.

Each household makes its decisions as one unit. They choose consumptions c and x, and labour incomes z, taking into account their budget constraint (1), to maximize a thrice continuously differentiable, weakly concave utility function:

$$\max_{c,\boldsymbol{x},\boldsymbol{z}} u\left(c,\boldsymbol{x},\frac{z^1}{w^1},\ldots,\frac{z^L}{w^L},\boldsymbol{\omega}\right),\tag{2}$$

with partial derivatives $u_c > 0$, $u_{x^j} > 0$ and $u_{\ell^l} < 0$, and with indifference sets that are strictly convex in $(c, \boldsymbol{x}, \boldsymbol{z})$ for all utility levels and types. We assume that for all households, the second-order conditions hold for optimization problem (1)–(2), and that the household's optimization problem admits a single global maximum. These assumptions, together with the assumption that the tax schedule $\boldsymbol{z} \mapsto T(\boldsymbol{z})$ is thrice continuously differentiable, prevent that small tax perturbations lead to "jumps" in the choices of the households.⁴ For a household of type $\boldsymbol{\omega}$ and for tax policies T and \boldsymbol{t} , we denote indirect utility as $v(\boldsymbol{\omega}, T, \boldsymbol{t})$.

⁴See Spiritus et al. [2022] for a proof.

2.2 Government

The government sets its tax policies to maximize an additive social welfare function subject to a budget constraint. The social welfare function is:⁵

$$\max_{T, t} \iint_{\Omega} W\left(v(\boldsymbol{\omega}, T, t), \boldsymbol{\omega}\right) \mathrm{d}G^{\boldsymbol{\omega}}(\boldsymbol{\omega}), \tag{3}$$

where $(v, \omega) \mapsto W(v, \omega)$ is a twice continuously differentiable, increasing, and weakly concave transformation of household utility that captures the government's normative preferences. Note that the weight given to a household can depend on its type ω .⁶ The government's budget constraint is:

$$\iint_{\Omega} \left(T(\boldsymbol{z}) + \sum_{j=1}^{J} t^{j} x^{j} \right) \mathrm{d}G^{\boldsymbol{\omega}}(\boldsymbol{\omega}) \ge E, \tag{4}$$

with E an exogenously given revenue requirement.

3 Taxpayer behaviour

The first-order conditions for household optimization problem (1)–(2) for labour incomes l = 1, ..., L and commodities j = 1, ..., J are, respectively:

$$\frac{u_{\ell^l}}{u_c} = -w^l (1 - T_{z^l}) \text{ and } \frac{u_{x^j}}{u_c} = 1 + t^j.$$
(5)

For each household type $\boldsymbol{\omega}$, we denote the optimal labour incomes as $\boldsymbol{Z}(\boldsymbol{\omega}) \equiv (Z^1(\boldsymbol{\omega}), \dots, Z^L(\boldsymbol{\omega}))$, the optimal consumption of the numéraire good as $C(\boldsymbol{\omega})$, and the optimal consumption of the other commodities as $\boldsymbol{X}(\boldsymbol{\omega}) \equiv (X^1(\boldsymbol{\omega}), \dots, X^J(\boldsymbol{\omega}))$. We denote the income space as the set of incomes that are chosen by any household. In other words, we represent the income space as $\boldsymbol{\mathcal{Z}} \equiv \{\boldsymbol{z} \in \mathbb{R}^L_+ | \exists \boldsymbol{\omega} \in \boldsymbol{\mathscr{W}} : \boldsymbol{Z}(\boldsymbol{\omega}) = \boldsymbol{z}\}$.

With the assumptions made in Subsection 2.1, the households' responses to small perturbations of the tax policies are well defined. Let ρ denote a lump sum perturb-

 $^{{}^{5} \}iint_{\Omega}$ denotes a multidimensional integral over the set Ω .

⁶It remains an open question how the government should compare taxpayers that have different preferences. One possible approach is the construction of equivalent incomes that cardinalize the households' utilities in a way that incorporates the planner's normative judgements. See Fleurbaey and Maniquet [2018] for a review of principles of fairness in optimal taxation theory. In this paper, we accept the social preferences as given.

ation, and let τ^{l} denote a compensated perturbation of the *l*-th marginal tax rate, such that the perturbed tax schedule is:

$$\boldsymbol{z} \mapsto T(\boldsymbol{z}) - \rho + \sum_{l=1}^{L} (z^l - Z^l(\boldsymbol{\omega})) \tau^l.$$
 (6)

We denote the behavioural responses to an increase in the lump sum income for any commodity j as $\partial X^{j}(\boldsymbol{\omega})/\partial \rho$, and for any labour income l as $\partial Z^{l}(\boldsymbol{\omega})/\partial \rho$. We denote the behavioural responses to a compensated perturbation of the m-th marginal income tax rate as $\partial X^{j*}(\boldsymbol{\omega})/\partial \tau^{m}$ and $\partial Z^{l*}(\boldsymbol{\omega})/\partial \tau^{m}$. Lastly, we denote the behavioural responses to a change of the linear tax rate on the *i*-th commodity as $\partial X^{j}(\boldsymbol{\omega})/\partial t^{i}$ and $\partial Z^{l}(\boldsymbol{\omega})/\partial t^{i}$.⁷

Standard derivations yield Slutsky decompositions (omitting function arguments, and with asterisks denoting compensated effects):

$$\forall j, m : \frac{\partial X^m}{\partial t^j} = \frac{\partial X^{m*}}{\partial t^j} - \frac{\partial X^m}{\partial \rho} X^j, \tag{7}$$

and:

$$\forall j, l: \frac{\partial Z^l}{\partial t^j} = \frac{\partial Z^{l*}}{\partial t^j} - \frac{\partial Z^l}{\partial \rho} X^j, \tag{8}$$

and Slutsky symmetries:

$$\forall l, m, j, k : \frac{\partial Z^{l*}}{\partial \tau^m} = \frac{\partial Z^{m*}}{\partial \tau^l}, \quad \frac{\partial X^{j*}}{\partial \tau^l} = \frac{\partial Z^{l*}}{\partial t^j} \quad \text{and} \quad \frac{\partial X^{k*}}{\partial t^j} = \frac{\partial X^{j*}}{\partial t^k}. \tag{9}$$

Following Christiansen [1984], Saez [2002] and Jacobs and Boadway [2014], we split the decision process of the households into two phases. In the second phase, households take their labour incomes and thus their disposable incomes as given, and they decide how to spend their money on consumption goods. In the first phase, households choose their labour incomes, taking the ensuing outcome of the second phase as given. Conditional on the combination of labour incomes \boldsymbol{z} , we denote the corresponding optimal consumption of the *j*-th good as $X^{cj}(\boldsymbol{\omega}, \boldsymbol{z})$.

⁷Strictly, we should write supplies Z and demands C and X as functions of the type and the perturbation parameters. Since we always evaluate the effects of perturbations starting from the situation where $\rho = 0$ and $\tau = 0$, we simplify notations by writing supplies and demands as functions of the type alone.

4 Optimal policies

Saez [2001] was the first to characterize the optimal labour income tax schedule in terms of sufficient statistics. Werquin et al. [2015] did the same for the case with multiple labour incomes. The optimal taxation of multiple labour incomes is complicated, and its interpretation is beyond the scope of the present paper. I simply state the optimality conditions for the tax on the labour incomes in Subsection 4.1, and take these optimality conditions as given in Subsection 4.2 to characterize the optimal linear commodity taxes. In Subsection 4.3, I state the conditions for uniform commodity taxation identified by Saez [2002] in the notations of the present paper, and I relate those conditions to the equation that characterizes the optimum.

4.1 Optimal tax on labour incomes

Before stating the optimality condition for the tax schedule on the labour incomes, I introduce some notations. We denote the marginal excess burden of a tax on the l-th labour income as:

$$\forall \boldsymbol{\omega} : \mathcal{W}^{l}(\boldsymbol{\omega}) \equiv -\sum_{m=1}^{L} T_{z_{m}} \frac{\partial Z^{m*}(\boldsymbol{\omega})}{\partial \tau^{l}} - \sum_{j=1}^{J} t^{j} \frac{\partial X^{j*}(\boldsymbol{\omega})}{\partial \tau^{l}}.$$
 (10)

The marginal excess burden of a tax on the l-th labour income indicates the compensated loss in tax revenues caused by an increase in the l-th marginal tax rate.

We denote the social marginal utility of income as (with λ denoting the government's budget Lagrange multiplier):⁸

$$b(\boldsymbol{\omega}) \equiv rac{W_v u_c}{\lambda},$$

and the corresponding net social marginal valuation of income:

$$\forall \boldsymbol{\omega} : \beta(\boldsymbol{\omega}) \equiv b(\boldsymbol{\omega}) + B_{\rho}(\boldsymbol{\omega}). \tag{11}$$

Here, $B_{\rho}(\boldsymbol{\omega})$ captures the effect on tax revenues caused by behavioural responses to

 $^{^8 \}mathrm{See}$ Diamond [1975].

a marginal change in ρ :

$$B_{\rho}(\boldsymbol{\omega}) \equiv \sum_{l=1}^{L} T_{z^{l}} \frac{\partial Z^{l}(\boldsymbol{\omega})}{\partial \rho} + \sum_{j=1}^{J} t^{j} \frac{\partial X^{j}(\boldsymbol{\omega})}{\partial \rho}.$$
 (12)

It is the marginal propensity to pay taxes out of additional income.

The derivations and the interpretation of the optimal tax schedule on the labour incomes are beyond the scope of this paper. Werquin et al. [2015] and Spiritus et al. [2022] show that, given our assumptions, the optimal tax schedule is characterized by:⁹

$$\forall z : \sum_{l=1}^{L} \frac{\partial}{\partial z^{l}} \left(\overline{\mathcal{W}^{l}}(\boldsymbol{z}) g^{z}(\boldsymbol{z}) \right) = -(1 - \overline{\beta}(z)) g^{z}(\boldsymbol{z}), \tag{13}$$

where for any function $\boldsymbol{\omega} \mapsto h(\boldsymbol{\omega})$, the notation $\boldsymbol{z} \mapsto \overline{h}(\boldsymbol{z})$ denotes the average of $h(\boldsymbol{\omega})$ over all types $\boldsymbol{\omega}$ who choose labour income $\boldsymbol{Z}(\boldsymbol{\omega}) = \boldsymbol{z}$; $\boldsymbol{z} \mapsto g^{z}(\boldsymbol{z})$ denotes the probability density function for the labour incomes; and $\boldsymbol{z} \mapsto G^{z}(\boldsymbol{z})$ denotes the cumulative distribution function.

At each point \boldsymbol{z} on the boundary of the income space $\boldsymbol{\mathcal{Z}}$, the following boundary condition must hold:¹⁰

$$\sum_{l=1}^{L} \overline{\mathcal{W}^{l}}(\boldsymbol{z}) e^{l}(\boldsymbol{z}) = 0, \qquad (14)$$

where $e^{l}(\boldsymbol{z})$ is the *l*-th component of the unit vector normal to the boundary of the income space at the point \boldsymbol{z} . Applying the Divergence Theorem to condition (13) and substituting the boundary conditions (14) shows the traditional condition that the net social marginal valuations of income should average to one:

$$\underbrace{\iint}_{\Omega} \beta(\boldsymbol{\omega}) \mathrm{d} G^{\boldsymbol{\omega}}(\boldsymbol{\omega}) = 1.$$
(15)

⁹Integrating (13) for the case with one labour income, L = 1, shows that it is equivalent to Saez' [2001] ABC-style characterization for the optimal tax on labour income.

¹⁰See Werquin et al. [2015] and Spiritus et al. [2022] for proofs. For the case with one labour income, L = 1, condition (14) is equivalent to the traditional transversality conditions, which demand that the marginal tax rates equal zero at the bottom and at the top of the income distribution. See, for example, Mirrlees [1971] and Saez [2001].

4.2 Optimal commodity taxes

The linear tax rate on the *j*-th commodity is optimal if perturbing it leaves social welfare unaffected. Demanding that the sum of all effects of a perturbation of t^{j} on social welfare equals zero, the government's first-order condition for the optimal linear tax on the *j*-th commodity is:

$$\iint_{\Omega} \left\{ \left(1 - \frac{W'u_c}{\lambda} \right) X^j + \sum_{l=1}^L T_{z^l} \frac{\partial Z^l}{\partial t^j} + \sum_{k=1}^J t^k \frac{\partial X^k}{\partial t^j} \right\} \mathrm{d}G^{\boldsymbol{\omega}}(\boldsymbol{\omega}) = 0, \qquad (16)$$

where we omit the function arguments in the integrand to simplify notations.

Substitute Slutsky decompositions (7)-(8) and definition (11) of the net social marginal utilities:

$$\iint_{\Omega} \left\{ (1-\beta) X^{j} + \sum_{l=1}^{L} T_{z^{l}} \frac{\partial Z^{l*}}{\partial t^{j}} + \sum_{k=1}^{J} t^{k} \frac{\partial X^{k*}}{\partial t^{j}} \right\} \mathrm{d}G^{\boldsymbol{\omega}}(\boldsymbol{\omega}) = 0.$$
(17)

Condition (17) characterizes the optimal linear taxes on the commodities, even when the tax schedule on the labour incomes is not optimal. We are interested in the optimal tax rates on the commodities when the tax schedule on the labour incomes is optimal, thus optimal tax conditions (13) and boundary conditions (14) are fulfilled.¹¹

I will start by introducing the theorem that characterizes the optimal commodity taxes. Following that, in the following subsections, I will provide intuitions for the most important terms. Lastly, I will restate the conditions for uniform commodity taxation in the optimum, as identified by Saez [2002], using this paper's notations, and I will show how they correspond to the terms in the optimal tax expression.

$$\frac{\operatorname{cov}\left(\beta, X^{j}\right)}{\iint_{\Omega} X^{j*} \mathrm{d}G^{\boldsymbol{\omega}}(\boldsymbol{\omega})} = \frac{\iint_{\Omega} \left\{ \sum_{l=1}^{L} T_{z^{l}} \frac{\partial Z^{i*}}{\partial t^{j}} + \sum_{k=1}^{J} t^{k} \frac{\partial X^{**}}{\partial t^{j}} \right\} \mathrm{d}G^{\boldsymbol{\omega}}(\boldsymbol{\omega})}{\iint_{\Omega} X^{j*} \mathrm{d}G^{\boldsymbol{\omega}}(\boldsymbol{\omega})}.$$
(18)

¹¹If the marginal tax rates on labour income are not optimal, one can use the derivations of Atkinson and Stiglitz [1980] to study the properties of condition (17). For example, assuming the population average of the net social marginal valuations of income equals one, so (15) is satisfied, we find:

We thus find that in the optimum, the distributional characteristic of each good should be equal to the normalized efficiency losses caused by taxing that good.

4.2.1 Condition for optimal commodity taxes

Theorem 1. Suppose households differ in one or more unobservable characteristics, they earn one or more labour incomes, and they make their decisions as one unit. Suppose the tax schedule applied to the labour incomes is optimal. Differentiated commodity taxes then may offer distributional benefits over and above those offered by the taxes on labour income. Moreover, differentiated commodity taxes may help reduce the distortions that are caused by the tax on the labour incomes. If these benefits are present, then the optimal commodity taxes are differentiated. These benefits must then be balanced against the costs of distorting the household's consumption choices. The following condition characterizes the optimal linear commodity taxes, for all commodities x^j :

$$\iint_{\mathcal{Z}} \sum_{k=1}^{J} t^{k} \overline{\frac{\partial X^{ck*}}{\partial t^{j}}} dG^{z}(\boldsymbol{z}) = \iint_{\mathcal{Z}} \operatorname{cov}(b, X^{j} | \boldsymbol{z}) dG^{z}(\boldsymbol{z})$$

$$- \sum_{l=1}^{L} \iint_{\mathcal{Z}} \overline{\mathcal{W}^{l}} \left(\frac{\partial \overline{X^{j}}}{\partial z^{l}} - \frac{\overline{\partial X^{cj}}}{\partial z^{l}} \right) dG^{z}(\boldsymbol{z})$$

$$+ \sum_{l=1}^{L} \iint_{\mathcal{Z}} \operatorname{cov} \left(\mathcal{W}^{l}, \frac{\partial X^{cj}}{\partial z^{l}} \middle| \boldsymbol{z} \right) dG^{z}(\boldsymbol{z})$$

$$+ \iint_{\mathcal{Z}} \operatorname{cov}(B_{\rho}, X^{j} | \boldsymbol{z}) dG^{z}(\boldsymbol{z}).$$

$$(19)$$

Proof. See Appendix A.

I will discuss the different terms of Theorem 1 in the following Subsections. The key distinction between Conditions (17) and (19) lies in the assumption of the optimality of the tax schedule on labour incomes in the latter. Commodity taxes, to some extent, act as a surrogate for a tax on labour income. They modify the purchasing power associated with any given labour income, leading to distributional effects and influencing labour supply decisions. To the extent that commodity taxes mirror a tax on labour income, their associated costs and benefits are already balanced at the margin due to the optimality of the labour income tax schedule. Therefore, Condition (19) balances the costs and benefits of commodity taxes only to the extent that they cannot be duplicated by a tax on labour incomes.

If the right-hand side of (19) differs from zero, then in the optimum, at least one

of the commodity tax rates must differ from zero. If we recall that the *numéraire* good c remains untaxed, it follows that if the right-hand side of (19) differs from zero, then it is desirable to levy differentiated taxes on the commodities.

4.2.2 Marginal excess burden conditional on labour income

First, let us study the left-hand side of Condition (19). It captures the compensated revenue costs associated with a minor increase in the tax rate for the j-th commodity, assuming that households' work incomes (and therefore their disposable incomes) remain constant. This part of Equation (19) indicates that the optimal marginal excess burden induced by taxing the good under consideration, conditional on labour income, is larger as the right-hand side of (19) increases.

Condition (19) presents a challenge because it does not provide a straightforward formula for determining the optimal tax rates for each good. The main issue is that the right-hand side of Condition (19) is influenced by the tax rates in complex ways. If we overlook this complexity, we could consider the right-hand side and the behavioural elasticities on the left-hand side of Condition (19) as fixed, and view the optimal tax conditions for each good as a system of linear equations. To see this, note that the marginal tax rates can be moved outside of the integral:

$$\iint_{\mathcal{Z}} \sum_{k=1}^{J} t^{k} \overline{\frac{\partial X^{ck*}}{\partial t^{j}}} \mathrm{d} G^{\mathsf{z}}(\boldsymbol{z}) = \sum_{k=1}^{J} t^{k} \iint_{\mathcal{Z}} \overline{\frac{\partial X^{ck*}}{\partial t^{j}}} \mathrm{d} G^{\mathsf{z}}(\boldsymbol{z}).$$

If the compensated cross-prices elasticities are zero (so $\partial X^{ck*}/\partial t^j = 0$ for $k \neq j$), then a commodity with a lower own-price elasticity will have a higher tax rate. Since $\partial X^{ck*}/\partial t^k$ is negative, it then follows that the tax rate t^k must have the opposite sign of the right-hand side of (19). However, when compensated cross-price elasticities are not zero, it seems there is no simple equation for expressing individual optimal tax rates in terms of sufficient statistics.^{12,13} Therefore, moving forward, we will

¹²Such formulations do exist for non-linear commodity taxes when households earn one labour income and have homogeneous preferences. See, for example, Mirrlees [1976], Gerritsen et al. [2020] and Ferey et al. [2022]. When it comes to linear commodity taxes, it is common practice to present the optimal tax conditions in the form of Theorem 1, with marginal excess burdens on the left-hand side. See, for example, Atkinson and Stiglitz [1980] and Jacobs and Boadway [2014].

¹³To gain further insights, we can modify the left-hand side of the optimal-tax expression (19) by multiplying it by t^j and summing over j. This results in the expression $\iint_{\mathcal{Z}} \sum_{j,k=1}^{J} t^k \overline{(\partial X^{ck*}/\partial t^j)} t^j dG^z(\boldsymbol{z})$. Given that the Slutsky matrix is negative semi-definite, the

primarily focus on discussing optimal marginal excess burdens rather than individual commodity tax rates.

4.2.3 Distributional characteristics conditional on labour income

We now turn to the first term on the right-hand side of (19). This term shows the covariance between the social marginal utilities and the consumption of the commodity under consideration, among households who earn certain levels of income from labour. It tells us whether, conditional on the labour incomes, the government considers taxpayers who consume more of the good under consideration to be more deserving of additional resources. This first term is zero in models in which taxpayers only differ in their labour productivities, such as the models of Atkinson and Stiglitz [1976] and Jacobs and Boadway [2014].

Let us consider an example.¹⁴ Suppose each household has two labour productivities w^1 and w^2 and they earn labour incomes z^1 and z^2 . They derive utility from consuming a *numéraire* good c and they obtain disutility $v(\ell^1, \ell^2)$ from supplying labour. Furthermore, some households consume a good x. The utility function is:

$$U(c, x, z^{1}, z^{2}, w^{1}, w^{2}, \gamma) = \frac{c^{1-\sigma}}{1-\rho} + \gamma \frac{x^{1-\theta}}{1-\theta} - v\left(\frac{z^{1}}{w^{1}}, \frac{z^{2}}{w^{2}}\right).$$

We assume that the parameter $\gamma \geq 0$ varies among households and is independent of their labour productivity. This parameter represents a physiological need to consume good x, which could be something like a necessary medical treatment. The parameters $\sigma > 0$ and $\theta > 1$ are the same for all households. They indicate the concavities of the households' consumption utilities from *numéraire* good c and good x, respectively. Because $\theta > 1$, the utility from consuming good x is always negative but increasing in x for a person who needs it ($\gamma > 0$). Now, consider two households with the same

resulting expression is necessarily negative. Consequently, if we substitute all instances of X^j on the right-hand side of (19) with the commodity tax revenue $\sum_{j=1}^{J} t^j X^j$, the resulting expression must also be negative. If the efficiency terms on the second and third lines of (19) are zero, it implies that $\iint_{\mathcal{Z}} \operatorname{cov}(\beta, \sum_{j=1}^{J} t^j X^j | \mathbf{z}) dG^{\mathbf{z}}(\mathbf{z})$ must be negative. This means that individuals with lower net social marginal valuations of income should, on average, pay more commodity taxes. However, the implication for individual tax rates remain ambiguous.

¹⁴Kaplow [2008] illustrates many ways in which preferences can be heterogeneous and how they affect the evolution of the marginal social utilities with the consumed quantities. We study only one example because I merely aim to illustrate the intuition behind the first term on the right-hand side of (19). A contribution of the present paper is that I show how the correlations between the social marginal utilities and the quantities consumed fit into the optimal tax expressions.

labour incomes z^1 and z^2 . However, one households needs to consume x ($\gamma > 0$) and the other does not ($\gamma = 0$). The household that needs to consume x will never reach the same level of utility as the household that does not need to consume x, no matter how much of x they consume. The first-order conditions for the consumption of cand x yield:

$$X = \left(\frac{\gamma C^{\sigma}}{1 + t_x}\right)^{\frac{1}{\theta}}.$$
(20)

Among households with the same incomes, those with a greater need for good x consume more of x and less of the *numéraire* good c. Because of the concave preferences with respect to the *numéraire* good, households with a greater need for good x have higher marginal utilities of the *numéraire*. Since these households achieve lower utility levels, the derivative W' of the concave social welfare function is also larger. Therefore, we find that the covariance between consumption of x and the social marginal utilities of income, conditional on income, is positive $(\operatorname{cov}(b, x | z) > 0)$.¹⁵ This means that taxpayers consuming more of commodity x are more deserving of redistribution. According to Theorem 1, the positive covariance between the social marginal utilities and consumption of commodity x implies there is a reason to subsidize commodity x in the optimum.

4.2.4 Cross-sectional variation and individual variation of consumption

The second term on the right-hand side of (19) captures, for each labour income, the difference between the cross-sectional variation and the individual variation in the consumption of the good under consideration. To see this, we can follow the exposition of Saez [2002]. Recall that the quantity $\overline{X^j}(z)$ denotes the average consumption of the *j*-th commodity among all households who earn labour incomes z. If we slightly increase the *l*-th labour income such that the income bundle increases by a quantity dz^l , then the term $\overline{X^j}(z + dz^l)$ represents the average of X^j over all households who choose to earn labour incomes $z + dz^l$. The quantities $\overline{X^j}(z)$ and $\overline{X^j}(z + dz^l)$ thus concern different households. Consequently, the term $\partial \overline{X^j}/\partial z^l(z)$ represents the variation in consumption X^j as we compare different households who choose income bundles that differ only by a slight change in the *l*-th labour income. In other words, the term concerns the *cross-sectional* variation of X^j around z. Note that these households also have different disposable incomes.

¹⁵See Kaplow [2008] for more formal proofs.

The quantity $\partial X^{cj}(\boldsymbol{\omega}, \boldsymbol{z})/\partial z^l$, on the other hand, concerns the change in consumption of a single household of type $\boldsymbol{\omega}$ if it is forced to earn a slightly higher *l*-th labour income z^l , thus earning a slightly higher disposable income. It thus concerns the *individual* variation in the consumption of commodity x^j with labour income z^l . The term $\overline{\partial X^{cj}/\partial z^l}(\boldsymbol{z})$ averages this variation over all households at income bundle \boldsymbol{z} .

To gain further insight into the second term on the right-hand side of (19), assume for a moment that for any vector $\boldsymbol{\theta}$ of additional characteristics, a one-to-one mapping exists between the labour productivities \boldsymbol{w} and the bundle of labour incomes \boldsymbol{z} . In other words, there exists a vector-valued function $\boldsymbol{z} \mapsto \boldsymbol{W}(\boldsymbol{z}, \boldsymbol{\theta})$, which for each vector of additional characteristics $\boldsymbol{\theta}$ indicates for each income bundle \boldsymbol{z} , the labour productivities of the households that choose it. Under this assumption, we derive the following identity in Appendix B, for any bundle of incomes \boldsymbol{z} :

$$\frac{\partial \overline{X^k}}{\partial z^l}(\boldsymbol{z}) - \frac{\partial \overline{X^{ck}}}{\partial z^l}(\boldsymbol{z}) = \sum_m \frac{\partial \overline{X^{ck}}}{\partial w^m} \frac{\partial W^m}{\partial z^l}(\boldsymbol{z}) + \operatorname{cov}\left(X^k, \left.\frac{\partial \ln g^{\boldsymbol{\theta}|\boldsymbol{z}}}{\partial z^l}\right| z\right), \quad (21)$$

where $\boldsymbol{\theta} \mapsto g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{z})$ indicates the density function of the additional characteristics $\boldsymbol{\theta}$ conditional on labour incomes \boldsymbol{z} .

Let us study the first term on the right-hand side of (21).¹⁶ The quantity $\partial X^{ck}(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{z})/\partial w^m$ tells us for given labour incomes \boldsymbol{z} how consumption of good x^k changes as the *m*-th labour productivity changes. There are two reasons why x^k might change. First, changing the *m*-th labour productivity while keeping constant the incomes \boldsymbol{z} , comes down to comparing households with different labour supplies but equal incomes. The first term on the right-hand side of (21) then tells us if households who supply different quantities of labour but earn equal disposable incomes, consume different amounts of the good under consideration. This reason for differential commodity taxes was first identified by Christiansen [1984]. A second reason why x^k might change as the *m*-th labour productivity changes, is that there is a deterministic relation between the taxpayers' consumption preferences and the labour productivity. This reason to differentially tax commodities was first identified by Mirrlees [1976].

To gain further intuition for the first term on the right-hand side of (21), let us

¹⁶This term was also identified by Jacobs and Boadway [2014] for the case where all taxpayers have the same preferences.

consider an example. Suppose each household has two labour productivities w^1 and w^2 , one for each spouse, and they earn labour incomes z^1 and z^2 . They derive utility from consuming a *numéraire* good c and they obtain disutility from supplying labour. Furthermore, households consume a good x. The utility function is:

$$U(c, x, \boldsymbol{z}, \boldsymbol{w}, \sigma) = c + \left(x^{\frac{\sigma-1}{\sigma}} - \left(\frac{z^1}{w^1}\right)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - v^2\left(\frac{z^2}{w^2}\right),$$

where σ indicates the complementarity between the consumption of x and the first labour supply, and the function v^2 represents the disutility from the second labour supply. Assume that σ is independent from the labour productivities. Standard derivations yield the conditional demand for x:

$$X^{c}(\boldsymbol{w},\sigma,\boldsymbol{z}) = \left(\frac{1-T_{z^{1}}}{1+t_{x}}\right)^{\sigma} \frac{z^{1}}{(w^{1})^{1-\sigma}},$$

with partial derivatives:

$$\frac{\partial X^c}{\partial w^1} = (\sigma - 1) \left(\frac{1 - T_{z^1}}{1 + t_x}\right)^{\sigma} \frac{z^1}{(w^1)^{2-\sigma}} \text{ and } \frac{\partial X^c}{\partial w^2} = 0$$

It follows that consumption of good x is complementary with the leisure of the first spouse if $\sigma > 1$, in which case Theorem 1 implies that good x should be taxed at higher rates.

Let us now study the second term on the right-hand side of (21). The term $\partial \ln g^{\theta|z}(\theta, z) / \partial z^l$ differs from zero if the conditional density $g^{\theta|z}(\theta|z)$ of the additional characteristics is different for different labour incomes z^l . If this is the case, then the additional characteristics θ contain information about the value of the income z^l , even if no deterministic relation exists between them.

Assuming that the vector $\boldsymbol{\theta}$ is composed of taste parameters for the different consumption goods, the term $\partial \ln g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}, \boldsymbol{z})/\partial z^l$ will differ from zero if, on average around the income bundle under consideration, households with different labour incomes z^l have relatively different tastes. The values of $\partial \ln g^{\boldsymbol{\theta}|\boldsymbol{z}}/\partial z^l$ represent the amount of information that the taste parameters $\boldsymbol{\theta}$ reveal about the households' incomes.

Of course, the government cannot directly tax the unobservable parameters $\boldsymbol{\theta}$. However, if the consumption of commodity x^k is correlated with $\partial \ln g^{\boldsymbol{\theta}|\boldsymbol{z}}/\partial z^l$, then also the consumption of this commodity reveals information about the value of labour income z^l . It then becomes optimal for the government to levy a differential tax on commodity x^k .

Keep in mind that the conditional densities $g^{\theta|z}(\theta|z)$ must add up to one for all types who pool at income bundle z.¹⁷ Therefore, if $\partial \ln g^{\theta|z}(\theta, z)/\partial z^l$ is positive for some values of θ , it must be negative for some others. If on average $\partial \ln g^{\theta|z}(\theta, z)/\partial z^l$ is negative for households with low consumption of commodity x^k and positive for those with high consumption of commodity x^k , then the covariance in (21) is positive. This implies that taxing that commodity serves as an indirect way to tax households with labour income z^l .

To further clarify, consider the following example. Assume that households differ in three dimensions, their labour productivities w^1 and w^2 , and their *taste* γ for the consumption of good x. Households obtain utility from *numéraire* good c and from good x, and disutility from obtaining labour incomes z^1 and z^2 . They maximize a utility function:

$$U(c, x, w^{1}, w^{2}, \gamma, z^{1}, z^{2}) = c + \gamma \frac{x^{1-\sigma}}{1-\sigma} - v^{1} \left(\frac{z^{1}}{w^{1}}\right) - v^{2} \left(\frac{z^{2}}{w^{2}}\right),$$

where the functions v^1 and v^2 represent the disutilities of labour supply. Standard derivations yield the demand for x conditional on the labour incomes:

$$x^{c}(w^{1}, w^{2}, \gamma, z^{1}, z^{2}) = \left(\frac{\gamma}{1+t_{x}}\right)^{1/\sigma}$$

If there is a deterministic relation between the taste parameters γ and either of the productivities w^1 and w^2 , then the first term on the right-hand side of (21) will not be zero. Let us exclude that possibility. Given the separability of preferences between leisure and consumption, one might then think that good x should not be taxed. However, suppose that the taste parameter γ is positively correlated with the first labour income z^1 . For lower γ values, the density $g(\gamma|z^1, z^2)$ will decrease in z^1 for most z^1 values, making $\partial g(\gamma|z^1, z^2)/\partial z^1 < 0$. For higher γ values, the density $g(\gamma|z^1, z^2)/\partial z^1 > 0$. Since consumption of commodity x increases with the taste parameter γ conditional on the incomes, we find that conditional consumption of x is positively correlated with

¹⁷Put differently, $\iint_{\Theta} g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{z}) \mathrm{d}\boldsymbol{\theta} = 1.$

 $\partial g(\gamma | z^1, z^2) / \partial z^1$. This creates a reason for a positive tax on that commodity. This rationale for implementing differentiated commodity taxes was first uncovered by Saez [2002].

We have discussed two potential reasons why the second term on the right-hand side of (19) might not be zero. If there is a difference between the cross-sectional variation and the individual variation in the consumption of the good under consideration, then the consumption pattern reveals information about the individual types beyond what is revealed by the labour incomes alone. Saez [2002] shows that if this is the case, differentiated commodity taxes help mitigate the distortions caused by the tax on labour income. The second term on the right-hand side of (19) lends further support to this intuition: the optimal marginal excess burden resulting from taxing a specific good is proportional to the marginal excess burdens of taxing the labour incomes. Indeed, the larger are the marginal excess burdens from taxing the labour incomes, the larger are the potential benefits from shifting from labour incomes taxes to differentiated consumption taxes.

4.2.5 Additional terms

The third term on the right-hand side of (19) captures the possibility that conditional on the labour incomes, household for whom a higher labour income leads to a higher consumption of the good under consideration, have higher compensated responses to taxes on labour income. By shifting the tax burden from that labour income to the considered good, the government can achieve the same redistribution benefits with fewer efficiency losses.

The fourth term on the right-hand side of (19) contains a covariance between the income effects on government revenues and the consumption level of the considered good. This term is present because the social marginal utilities considered in the characterisation of the optimum should be net of income effects.

Saez [2002] also obtains similar terms and notes that these third and fourth terms are difficult to measure empirically. These terms do not appear in models where households only differ in their labour productivities.

4.3 Relation to Saez [2002]

Saez [2002] identifies three conditions under which uniform commodity taxation is

optimal. He starts from a situation where all commodity tax rates are zero and examines the welfare effects of perturbing one of the tax rates. If the welfare effects are zero, then deviating from uniformity is not optimal. In his investigations, Saez [2002] assumes there is only one labour income.

I will now explore how the conditions found by Saez [2002] relate to my statement of the full optimality condition in Theorem 1. His conditions are as follows, in the notations of the present paper:

1. conditional on each income level z, the marginal social utilities b and the consumption of the considered good x^{j} are uncorrelated:

$$\operatorname{cov}(b, X^j | z) = 0;$$

2. conditional on each income level z, behavioural reponses $\partial Z^*(\boldsymbol{\omega})/\partial \tau$ and $\partial Z(\boldsymbol{\omega})/\partial \rho$ are independent of consumption patterns X^j and $\partial X^{cj}/\partial z$:¹⁸

$$\operatorname{cov}\left(\frac{\partial Z(\boldsymbol{\omega})}{\partial \rho}, X^{j} \mid \boldsymbol{z}\right) = 0 \text{ and } \operatorname{cov}\left(\frac{\partial Z^{*}(\boldsymbol{\omega})}{\partial \tau^{l}}, \frac{\partial X^{cj}}{\partial z} \mid \boldsymbol{z}\right) = 0$$

3. for any income level, the cross-sectional variation in the consumption of the considered good is equal to the individual variation, as one varies labour income:

$$\frac{\mathrm{d}\overline{X^j}}{\mathrm{d}z} - \frac{\overline{\partial X^{cj}}}{\partial z} = 0.$$

One difference between Theorem 1 and the conditions found by Saez [2002], is the number of labour incomes. We see that the different terms in Theorem 1 are conditional on the entire bundle of incomes \boldsymbol{z} , and when we consider the effects of variations in one of the incomes \boldsymbol{z}^l , we sum those effects over all incomes.

Under the presumption that Assumptions 1–3 found by Saez [2002] hold for all labour incomes, substitute these three assumptions into (19) and substitute definitions (10) and (12) for the marginal excess burdens and the marginal propensities to pay taxes out of additional income. What then remains is the following condition for the

¹⁸There is some ambiguity in Saez's [2002] statement of his second assumption. He also appears to assume that $\operatorname{cov}\left(\partial Z(\boldsymbol{\omega})/\partial \rho, \partial X^{cj}/\partial z|z\right) = \operatorname{cov}\left(\partial Z^*(\boldsymbol{\omega})/\partial \tau, X^j|z\right) = 0$. These assumptions, however, are not necessary, neither for our nor for his results.

optimum:

$$\iint_{\mathcal{Z}} \sum_{k=1}^{J} t^{k} \overline{\frac{\partial X^{ck*}}{\partial t^{j}}} \mathrm{d}G^{\mathrm{z}}(\boldsymbol{z}) = -\sum_{j=1}^{J} t^{j} \sum_{l=1}^{L} \iint_{\mathcal{Z}} \mathrm{cov}\left(\frac{\partial X^{j*}(\boldsymbol{\omega})}{\partial \tau^{l}}, \frac{\partial X^{cj}}{\partial z^{l}} \middle| \boldsymbol{z}\right) \mathrm{d}G^{\mathrm{z}}(\boldsymbol{z}) + \sum_{j=1}^{J} t^{j} \iint_{\mathcal{Z}} \mathrm{cov}\left(\frac{\partial X^{j}(\boldsymbol{\omega})}{\partial \rho}, X^{j} \middle| \boldsymbol{z}\right)) \mathrm{d}G^{\mathrm{z}}(\boldsymbol{z}).$$

This is a linear equation in the commodity tax rates, with solution t = 0. We thus find that Saez's [2002] assumptions remain sufficient for uniform taxation to be optimal in the case with multiple optimally taxed labour incomes.

5 Conclusions

I characterized the optimal linear taxation of commodities alongside multiple optimally taxed labour incomes, taking into account that individuals may differ in multiple unobserved characteristics. There are several possible reasons to deviate from uniform tax rates on commodities. A first possible reason is that conditional on labour income, taxpayers who consume more of a particular good are more deserving in the eyes of the government. This can be the case when a strong taste for a particular good is considered socially desirable or if the government wishes to compensate types who typically consume certain goods (for example, to compensate for other injustices). The acceptable size of the distortion from taxing a particular good, conditional on labour income, is then larger if the covariance conditional on income between the social marginal utilities and the consumption of that good is larger.

A second possible reason to differentiate commodity taxes is that on average and conditional on income, the cross-sectional variation of the consumption of some commodity differs from the individual variation as one varies one of the labour incomes. This can happen for two reasons. The first is that around the considered incomes, taxpayers with different incomes have different unobserved characteristics besides their labour productivities, and these characteristics correlate with their consumption patterns. The second reason why the individual variation may differ from the cross-sectional variation, is that around the considered incomes, individuals who have the same disposable incomes but have supply different amounts of labour tend to consume different quantities of the good in question. In each case, shifting the burden of taxation from the labour incomes to differentiated commodity taxes increases the efficiency of the tax system. The acceptable size of the distortion of taxing a particular good then depends on the size of the difference between the cross-sectional and the individual variations of the consumption of that good along each labour income, and the size of the marginal excess burdens of the taxes on those labour incomes.

The formulas in this paper identify which statistics are sufficient to characterize the optimal commodity taxes. Most of these statistics are not yet available in the empirical literature. Moreover, it remains unclear how the social welfare function can capture common ethical intuitions that may lead to differentiated commodity taxes in the optimum. For these reasons, numerical simulations of the optimal commodity tax rates would currently be based on arbitrary parameters. It would be very useful if further research could quantify the needed sufficient statistics for a numerical exercise. Further useful extensions would include moving beyond the unitary decision framework for the households. Such an extension, however, would first require advances in our knowledge on the optimal taxation of labour income when households do not take their decisions as one unit.

References

- Anthony B. Atkinson and Joseph E. Stiglitz. The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, 6(1-2):55–75, 1976.
- Anthony B. Atkinson and Joseph E. Stiglitz. Lectures on public economics. McGraw-Hill College, 1980.
- Vidar Christiansen. Which commodity taxes should supplement the income tax? Journal of Public Economics, 24(2):195–220, 1984.
- Peter A Diamond. A many-person ramsey tax rule. Journal of Public Economics, 4 (4):335–342, 1975.
- Antoine Ferey, Benjamin B Lockwood, Dmitry Taubinsky, et al. Sufficient statistics for nonlinear tax systems with general across-income heterogeneity. Technical report, CESifo, 2022.
- Marc Fleurbaey and François Maniquet. Optimal income taxation theory and principles of fairness. *Journal of Economic Literature*, 56(3):1029–79, 2018.

- Aart Gerritsen, Bas Jacobs, Alexandra V. Rusu, and Kevin Spiritus. Optimal taxation of capital income with heterogeneous rates of return. CESifo Working Paper No. 8395, 2020.
- Roger H Gordon and Wojciech Kopczuk. The choice of the personal income tax base. Journal of Public Economics, 118:97–110, 2014.
- Bas Jacobs and Robin Boadway. Optimal linear commodity taxation under optimal non-linear income taxation. *Journal of Public Economics*, 117:201–210, 2014.
- Louis Kaplow. Optimal policy with heterogeneous preferences. The BE Journal of Economic Analysis & Policy, 8(1), 2008.
- James Mirrlees. Optimal tax theory : A synthesis. Journal of Public Economics, 6 (4):327–358, 1976.
- James Mirrlees, Stuart Adam, Tim Besley, Richard Blundell, Steve Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba. Tax By Design: The Mirrlees Review. Oxford University Press, Sep 2011.
- James A Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208, 1971.
- OECD. Consumption Tax Trends. 2020.
- Jukka Pirttilä and Ilpo Suoniemi. Public provision, commodity demand, and hours of work: An empirical analysis. The Scandinavian Journal of Economics, 116(4): 1044–1067, 2014.
- Emmanuel Saez. Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1):205–29, January 2001.
- Emmanuel Saez. The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics*, 83(2):217–230, 2002.
- Kevin Spiritus, Etienne Lehmann, Sander Renes, and Floris Zoutman. Optimal taxation with multiple incomes and types. *CEPR Discussion Paper*, (DP16797), 2022.
- Josep-Maria Terricabras. Question for written answer E-006958-16 to the Commission. Rule 130. European Parliament parliamentary questions, September 2016. URL https://www.europarl.europa.eu/doceo/document/

E-8-2016-006958_EN.html.

Nicolas Werquin, Aleh Tsyvinski, and Mikhail Golosov. A variational approach to the analysis of tax systems. Technical report, Society for Economic Dynamics, 2015.

A Proof of Theorem 1

Proof. To account for the conditions for the optimal tax schedule on the labour incomes, we reformulate optimality condition (17) in the income space. Let $G^{\boldsymbol{\omega}|\boldsymbol{z}}(\boldsymbol{\omega}|\boldsymbol{z})$ denote the cumulative distribution function of the characteristics $\boldsymbol{\omega}$ conditional on labour incomes \boldsymbol{z} , with corresponding density function $g^{\boldsymbol{\omega}|\boldsymbol{z}}(\boldsymbol{\omega}|\boldsymbol{z})$. We can then change the integration domain of (17), to find for any commodity j:

$$\iint_{\mathcal{Z}} \iint_{\{\boldsymbol{\omega}|\boldsymbol{Z}(\boldsymbol{\omega})=\boldsymbol{z}\}} \left\{ (1-\beta) X^{j} + \sum_{l=1}^{L} T_{z^{l}} \frac{\partial Z^{l*}}{\partial t^{j}} + \sum_{k=1}^{J} t^{k} \frac{\partial X^{k*}}{\partial t^{j}} \right\} \mathrm{d}G^{\boldsymbol{\omega}|\boldsymbol{z}}(\boldsymbol{\omega}|\boldsymbol{z}) \mathrm{d}G^{\boldsymbol{z}}(\boldsymbol{z}) = 0,$$
(22)

or shorter:

$$\iint_{\mathcal{Z}} \overline{(1-\beta)X^{j}} dG^{z}(\boldsymbol{z}) + \iint_{\mathcal{Z}} \left\{ \sum_{l=1}^{L} T_{z^{l}} \overline{\frac{\partial Z^{l*}}{\partial t^{j}}} + \sum_{k=1}^{J} t^{k} \overline{\frac{\partial X^{k*}}{\partial t^{j}}} \right\} dG^{z}(\boldsymbol{z}) = 0.$$
(23)

We now rewrite the first integral in (23) as a covariance conditional on labour income:

$$\iint_{\mathcal{Z}} \overline{(1-\beta)X^{j}} \mathrm{d}G^{\mathbf{z}}(\boldsymbol{z}) = -\iint_{\mathcal{Z}} \mathrm{cov}(\beta, X^{j}|\boldsymbol{z}) \mathrm{d}G^{\mathbf{z}}(\boldsymbol{z}) + \iint_{\mathcal{Z}} (1-\overline{\beta})\overline{X^{j}} \mathrm{d}G^{\mathbf{z}}(\boldsymbol{z}).$$
(24)

Substitute optimality condition (13) and then apply partial integration taking into

account boundary conditions (14):¹⁹

$$\iint_{\mathcal{Z}} \overline{(1-\beta)X^{j}} dG^{z}(\boldsymbol{z}) = -\iint_{\mathcal{Z}} \operatorname{cov}(\beta, X^{j}|\boldsymbol{z}) dG^{z}(\boldsymbol{z}) - \iint_{\mathcal{Z}} \sum_{l=1}^{L} \frac{\partial}{\partial z^{l}} \left(\overline{\mathcal{W}^{l}} g^{z}(\boldsymbol{z}) \right) \overline{X^{j}} d\boldsymbol{z} \qquad (25)$$

$$= -\iint_{\mathcal{Z}} \operatorname{cov}(\beta, X^{j} | \boldsymbol{z}) \mathrm{d}G^{\mathrm{z}}(\boldsymbol{z}) + \iint_{\mathcal{Z}} \sum_{l=1}^{L} \left(\overline{\mathcal{W}^{l}} \frac{\partial \overline{X^{j}}}{\partial z^{l}} \right) \mathrm{d}G^{\mathrm{z}}(\boldsymbol{z}).$$
(26)

Substitute the latter result into optimality condition (23) and rearrange:²⁰

$$\iint_{\mathcal{Z}} \overline{\left\{ \sum_{l=1}^{L} \left(\mathcal{W}^{l} \frac{\partial X^{cj}}{\partial z^{l}} + T_{z^{l}} \frac{\partial Z^{l*}}{\partial t^{j}} \right) + \sum_{k=1}^{J} t^{k} \frac{\partial X^{k*}}{\partial t^{j}} \right\}} dG^{z}(\boldsymbol{z})$$

$$= \sum_{l=1}^{L} \iint_{\mathcal{Z}} \operatorname{cov} \left(\mathcal{W}_{l}, \frac{\partial X^{cj}}{\partial z^{l}} \middle| \boldsymbol{z} \right) dG^{z}(\boldsymbol{z})$$

$$- \sum_{l=1}^{L} \iint_{\mathcal{Z}} \overline{\mathcal{W}^{l}} \left(\frac{\partial \overline{X^{j}}}{\partial z^{l}} - \frac{\partial \overline{X^{cj}}}{\partial z^{l}} \right) dG^{z}(\boldsymbol{z})$$

$$+ \iint_{\mathcal{Z}} \operatorname{cov}(\beta, X^{j} | \boldsymbol{z}) dG^{z}(\boldsymbol{z}).$$
(27)

Finally, we rewrite the left-hand side of the latter equation. Substitute definition

 $^{^{19}}$ The partial integration that is used in going from (25) to (26) is an application of the Divergence Theorem, a central result in multivariable calculus which states that for a sufficiently regular vector-valued function $\mathbf{z} \mapsto \mathbf{f}(\mathbf{z})$ one has: $\iint_{\mathcal{Z}} \sum_{l=1}^{L} \partial f^{l}(\mathbf{z}) / \partial z^{l} d\mathbf{z} = \iint_{\partial \mathcal{Z}} \sum_{l=1}^{L} f^{l}(\mathbf{z}) e^{l}(\mathbf{z}) d\mathbf{z}$, with $e^{l}(\mathbf{z})$ the *l*-th component of the unit vector normal to the boundary $\partial \mathcal{Z}$ of \mathcal{Z} at \mathbf{z} . With one dimension, L = 1, the latter is the Fundamental Theorem of Calculus, which states that $\int_{z}^{\overline{z}} (df(z)/dz) dz =$ $f(\overline{z}) - f(\underline{z}). \text{ I apply the Divergence Theorem to } f^{l}(z) \equiv \overline{\mathcal{W}^{l}}(z)\overline{X^{j}}(z)g^{z}(z) \text{ and use boundary conditions (14) to find } \iint_{\mathcal{Z}} \sum_{l=1}^{L} \partial \left(\overline{\mathcal{W}^{l}}(z)\overline{X^{j}}(z)g^{z}(z) \right) / \partial z^{l} \mathrm{d}z = 0.$ ²⁰To find (27), first add $\sum_{l=1}^{L} \iint_{\mathcal{Z}} \overline{\mathcal{W}^{l}}(\partial X^{cj*}/\partial z^{l}) \mathrm{d}G^{z}(z)$ on both sides of (23).

(10) of the excess burden and rearrange:

$$\iint_{\mathcal{Z}} \overline{\left\{ \sum_{l=1}^{L} \left(\mathcal{W}^{l} \frac{\partial X^{cj}}{\partial z^{l}} + T_{z^{l}} \frac{\partial Z^{l*}}{\partial t^{j}} \right) + \sum_{k=1}^{J} t^{k} \frac{\partial X^{k*}}{\partial t^{j}} \right\}} dG^{z}(\boldsymbol{z}) \\
= \iint_{\mathcal{Z}} \overline{\left\{ \sum_{l=1}^{L} T_{z^{l}} \left(\frac{\partial Z^{l*}}{\partial t^{j}} - \sum_{m=1}^{L} \frac{\partial Z^{l*}(\boldsymbol{\omega})}{\partial \tau^{m}} \frac{\partial X^{cj}}{\partial z^{m}} \right) \right\}} dG^{z}(\boldsymbol{z}) \\
+ \iint_{\mathcal{Z}} \overline{\left\{ \sum_{k=1}^{J} t^{k} \left(\frac{\partial X^{k*}}{\partial t^{j}} - \sum_{l=1}^{L} \frac{\partial X^{k*}(\boldsymbol{\omega})}{\partial \tau^{l}} \frac{\partial X^{cj}}{\partial z^{l}} \right) \right\}} dG^{z}(\boldsymbol{z}). \tag{28}$$

Apply Slutsky symmetries (9):

$$\iint_{\mathcal{Z}} \left\{ \sum_{l=1}^{L} \left(\mathcal{W}^{l} \frac{\partial X^{cj}}{\partial z^{l}} + T_{z^{l}} \frac{\partial Z^{l*}}{\partial t^{j}} \right) + \sum_{k=1}^{J} t^{k} \frac{\partial X^{k*}}{\partial t^{j}} \right\} dG^{z}(\boldsymbol{z}) \\
= \iint_{\mathcal{Z}} \left\{ \sum_{l=1}^{L} T_{z^{l}} \left(\frac{\partial X^{j*}}{\partial \tau^{l}} - \sum_{m=1}^{L} \frac{\partial Z^{m*}}{\partial \tau^{l}} \frac{\partial X^{cj}}{\partial z^{m}} \right) \right\} dG^{z}(\boldsymbol{z}) \\
+ \iint_{\mathcal{Z}} \left\{ \sum_{k=1}^{J} t^{k} \left(\frac{\partial X^{j*}}{\partial t^{k}} - \sum_{l=1}^{L} \frac{\partial Z^{l*}}{\partial t^{k}} \frac{\partial X^{cj}}{\partial z^{l}} \right) \right\} dG^{z}(\boldsymbol{z}). \tag{29}$$

Now note that the compensated response of the consumption of the k-th consumption good to a reform in the j-th commodity tax can be decomposed as follows:²¹

$$\frac{\partial X^{k*}(\boldsymbol{\omega})}{\partial t^j} = \frac{\partial X^{ck*}(\boldsymbol{\omega}, \boldsymbol{Z}(\boldsymbol{\omega}))}{\partial t^j} + \sum_{m=1}^L \frac{\partial X^{ck}(\boldsymbol{\omega}, \boldsymbol{Z}(\boldsymbol{\omega}))}{\partial z^m} \frac{\partial Z^{m*}(\boldsymbol{\omega})}{\partial t^j}.$$
 (30)

For a given type and keeping constant the chosen combination of labour incomes, a compensated perturbation of the tax on labour income does not directly affect the consumption choices. We thus find:

$$\frac{\partial X^{j*}(\boldsymbol{\omega})}{\partial \tau^l} = \sum_{m=1}^{L} \frac{\partial X^{cj}(\boldsymbol{\omega}, \boldsymbol{Z}(\boldsymbol{\omega}))}{\partial z^m} \frac{\partial Z^{m*}(\boldsymbol{\omega})}{\partial \tau^l}.$$
(31)

²¹Note that altering one of the labour incomes while fixing the labour productivities, comes down to altering one of the labour supplies. Since individuals are optimizing, a perturbation of labour supply does not alter individual utility, so we have $\partial X^{ck*}(\boldsymbol{\omega}, \boldsymbol{Z}(\boldsymbol{\omega}))/\partial z^m = \partial X^{ck}(\boldsymbol{\omega}, \boldsymbol{Z}(\boldsymbol{\omega}))/\partial z^m$.

Substitute properties (30)–(31) on the right-hand side of (29) to simplify:

$$\iint_{\mathcal{Z}} \left\{ \sum_{l=1}^{L} \left(\mathcal{W}^{l} \frac{\partial X^{cj}}{\partial z^{l}} + T_{z^{l}} \frac{\partial Z^{l*}}{\partial t^{j}} \right) + \sum_{k=1}^{J} t^{k} \frac{\partial X^{k*}}{\partial t^{j}} \right\} \mathrm{d}G^{\mathrm{z}}(\boldsymbol{z})$$
$$= \iint_{\mathcal{Z}} \sum_{k=1}^{J} t^{k} \frac{\overline{\partial X^{ck*}}}{\partial t^{j}} \mathrm{d}G^{\mathrm{z}}(\boldsymbol{z}).$$
(32)

Substitute the latter result into condition (27) to find the characterization of the optimal tax rates in Theorem 1. $\hfill \Box$

B Difference between cross-sectional and individual variation of X^j

Proof. Recall that the quantity $\overline{X^k}(\boldsymbol{z})$ represents the average of consumption X^k of the k-th commodity over all households who choose to earn income bundle z. We then obtain:

$$\frac{\partial X^k}{\partial z^l}(\boldsymbol{z}) = \frac{\partial}{\partial z^l} \iint_{\{\boldsymbol{\omega} \in \Omega | \boldsymbol{Z}(\boldsymbol{\omega}) = \boldsymbol{z}\}} X^k(\boldsymbol{\omega}) \mathrm{d} G^{\boldsymbol{\omega}|_{\mathbf{Z}}}(\boldsymbol{\omega}|\boldsymbol{z}).$$

Recall that we assume that for any vector $\boldsymbol{\theta}$ of additional characteristics, a one-to-one relation exists between labour productivities \boldsymbol{w} and the corresponding labour incomes $\boldsymbol{Z}(\boldsymbol{w}, \boldsymbol{\theta})$. The inverse of this relation is denoted as $\boldsymbol{z} \mapsto \boldsymbol{W}(\boldsymbol{z}, \boldsymbol{\theta})$. Use the fact that $X^k(\boldsymbol{W}(\boldsymbol{z}, \boldsymbol{\theta}), \boldsymbol{\theta}) = X^{ck}(\boldsymbol{W}(\boldsymbol{z}, \boldsymbol{\theta}), \boldsymbol{\theta}, \boldsymbol{z})$ to obtain:

$$\frac{\partial \overline{X^k}}{\partial z^l}(\boldsymbol{z}) = \frac{\partial}{\partial z^l} \iint_{\Theta} X^{ck}(\boldsymbol{W}(\boldsymbol{z},\boldsymbol{\theta}),\boldsymbol{\theta},\boldsymbol{z}) g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{z}) \mathrm{d}\boldsymbol{\theta}.$$

Bring the partial derivative inside of the integral:

$$\begin{split} \frac{\partial \overline{X^k}}{\partial z^l}(\boldsymbol{z}) = & \overline{\frac{\partial X^{ck}}{\partial z^l}}(z) + \iint_{\Theta} \sum_m \frac{\partial X^{ck}(\boldsymbol{W}(\boldsymbol{z}, \boldsymbol{\theta}), \boldsymbol{\theta}, \boldsymbol{z})}{\partial w^m} \frac{\partial W^m(\boldsymbol{z}, \boldsymbol{\theta})}{\partial z^l} g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{z}) \mathrm{d}\boldsymbol{\theta} \\ &+ \iint_{\Theta} X^k(\boldsymbol{W}(\boldsymbol{z}, \boldsymbol{\theta}), \boldsymbol{\theta}) \frac{\partial g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{z})}{\partial z^l} \mathrm{d}\boldsymbol{\theta}. \end{split}$$

Finally, recall that $\iint_{\Theta} g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{z}) d\boldsymbol{\theta} = 1$. We then find $\iint_{\Theta} \partial g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{z}) / \partial z^{l} d\boldsymbol{\theta} = \partial \left(\iint_{\Theta} g^{\boldsymbol{\theta}|\boldsymbol{z}}(\boldsymbol{\theta}|\boldsymbol{z}) d\boldsymbol{\theta}\right) / \partial z^{l} = 0$, so:

$$\frac{\partial \overline{X^k}}{\partial z^l}(\boldsymbol{z}) - \frac{\partial \overline{X^{ck}}}{\partial z^l}(z) = \sum_m \overline{\frac{\partial \overline{X^{ck}}}{\partial w^m}} \frac{\partial W^m}{\partial z^l}(\boldsymbol{z}) + \operatorname{cov}\left(X^k, \frac{\partial \ln g^{\boldsymbol{\theta}|\boldsymbol{z}}}{\partial z^l}\right) |z|.$$

The latter equation corresponds directly to (21).