
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Kanbur, Ravi; Paukkeri, Tuuli; Pirttilä, Jukka; Tuomala, Matti

Optimal taxation and public provision for poverty reduction

Published in:
INTERNATIONAL TAX AND PUBLIC FINANCE

DOI:
[10.1007/s10797-017-9443-6](https://doi.org/10.1007/s10797-017-9443-6)

Published: 01/02/2018

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Kanbur, R., Paukkeri, T., Pirttilä, J., & Tuomala, M. (2018). Optimal taxation and public provision for poverty reduction. *INTERNATIONAL TAX AND PUBLIC FINANCE*, 25(1), 64–98. <https://doi.org/10.1007/s10797-017-9443-6>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Optimal taxation and public provision for poverty reduction

Ravi Kanbur¹ · Tuuli Paukkeri^{2,3}  ·
Jukka Pirttilä^{4,5} · Matti Tuomala⁵

© The Author(s) 2017. This article is an open access publication

Abstract The existing literature on optimal taxation typically assumes there exists a capacity to implement complex tax schemes, which is not necessarily the case for many developing countries. We examine the determinants of optimal redistributive policies in the context of a developing country that can only implement linear tax policies due to administrative reasons. Further, the reduction of poverty is typically the expressed goal of such countries, and this feature is also taken into account in our model. We derive the optimality conditions for linear income taxation, commodity taxation, and public provision of private and public goods for the poverty minimization case and compare the results to those derived under a general welfarist objective function. We also study the implications of informality on optimal redistributive policies for such countries. The exercise reveals non-trivial differences in optimal tax rules under the different assumptions.

✉ Tuuli Paukkeri
tuuli.paukkeri@vatt.fi

Ravi Kanbur
sk145@cornell.edu

Jukka Pirttilä
jukka@wider.unu.edu

Matti Tuomala
matti.tuomala@uta.fi

- ¹ Cornell University, Ithaca, NY, USA
- ² VATT Institute for Economic Research, Helsinki, Finland
- ³ Aalto University, Helsinki, Finland
- ⁴ UNU-WIDER, Helsinki, Finland
- ⁵ University of Tampere, Tampere, Finland

Keywords Redistribution · Income taxation · Commodity taxation · Public good provision · Poverty

JEL Classification H21 · H40 · O12

1 Introduction

High levels of within-country inequality in many otherwise successful developing countries have become a key policy concern in global development debate. While some countries have very unequal inherent distributions (e.g., due to historical land ownership arrangements), in others the fruits of economic growth have been unequally shared. No matter what the underlying reason for the high inequality, often the only direct way for governments to affect the distribution of income is via redistributive tax and transfer systems. Clearly, public spending on social services also has an impact on the distribution of well-being, although some of the effects (such as skill-enhancing impacts from educational investment) only materialize over a longer time horizon.

Reflecting the desire to reduce poverty and inequality, redistributive transfer systems have, indeed, proliferated in many developing countries. Starting from Latin America, they are now spreading to low-income countries, including those in Sub-Saharan Africa.¹ In low-income countries, in particular, redistributive arrangements via transfers are still at an early stage, and they often consist of isolated, donor-driven programs. There is an urgent and well-recognized need to move away from scattered programs to more comprehensive tax-benefit systems.

This paper examines the optimal design of cash transfers, commodity taxes (or subsidies), the provision of public and private goods (such as education and housing), and financing them by a linear income tax. The paper also includes an analysis of optimal income taxation in the presence of an informal sector. The paper therefore provides an overview of many of the most relevant instruments for redistributive policies that are needed for a system-wide analysis of social protection. We build on the optimal income tax approach, which is extensively used in the developed country context², but much less applied for the design of redistributive systems in developing country circumstances. This approach, initiated by [Mirrlees \(1971\)](#), allows for a rigorous treatment of efficiency concerns (e.g., the potentially harmful effect of distortionary taxation on employment) and redistributive objectives. Achieving the government's redistributive objectives is constrained by limited information: the social planner cannot directly observe individuals' income-earning capacity, and therefore it needs to base its tax and transfer policies on observable variables, such as gross income. The most general formulations of optimal tax models apply nonlinear tax schedules, but in a developing country context, using fully nonlinear taxes is rarely feasible. In this paper, we therefore limit the analysis to redistributive linear income taxes, which combine a

¹ For a recent treatment and survey, see [Barrientos \(2013\)](#).

² See [IFS and Mirrlees \(2011\)](#) for an influential application of optimal tax theory to policy analysis for rich countries.

lump-sum transfer with a proportional income tax, and which can be implemented by withholding at source if necessary.

Linear income taxes are not very common in practice: less than 30 countries had flat tax rates for personal income in 2012, with some concentration in ex-Soviet Eastern Europe (Peichl 2014). It is noteworthy that even though flat taxes are not particularly common in low-income countries, in many instances in such countries the progressive income tax reaches only a small share of the population. This would indicate that despite the existence of a progressive income tax, these countries do not yet possess enough tax capacity to implement well-functioning progressive income taxes. This is one motivation for our interest of modeling optimal linear taxes. Peichl (2014) suggests that simplification benefits can be especially relevant for developing countries.³

In conventional optimal taxation models, the government's objective function is modeled as a social welfare function, which depends directly on individual utilities. We depart from this welfarist approach by presenting general non-welfarist tax rules, as in Kanbur et al. (2006), and, in particular, optimal tax and public good provision rules when the government is assumed to minimize poverty. We have chosen this approach as it resembles well the tone of much of the policy discussion in developing countries, including the Millennium Development Goals (MDGs) and the new Sustainable Development Goals (SDGs), where the objective is explicitly to reduce poverty rather than maximize well-being.⁴ Similarly, the discussion regarding cash transfer systems is often couched especially in terms of poverty alleviation. While we do not necessarily want to advocate poverty minimization over other social objectives, we regard examining its implications, and contrasting them with traditional welfaristic approaches, useful. Using non-welfarist objectives is, as such, nothing new in economics. In fact, as Sen (1985) has argued, one can be critical of utilitarianism for many reasons. Note also that the objective of poverty minimization is not at odds with the restriction of a linear tax scheme that we impose: a flat tax regime together with a lump-sum income transfer component can achieve similar amounts of redistribution toward the poor as a progressive tax system, if specified suitably (Keen et al. 2008; Peichl 2014). In all our analysis, we first present welfarist tax rules (which are mostly already available in the literature) to provide a benchmark to examine how applying poverty minimization as an objective changes the optimal tax and public service provision rules.

We also deal with some extensions to existing models, which are motivated by the developing country context, such as the case where public provision affects the individuals' income-earning capacity, thus capturing (albeit in a very stylized way) possibilities to affect their capabilities. An important feature to take into account in tax analysis of developing countries is the presence of a large informal sector, and we also examine the implications of this for optimal redistributive policies.

Our paper is related to various strands of earlier literature. First, Kanbur et al. (1994) and Pirttilä and Tuomala (2004) study optimal income tax and commodity tax rules, respectively, from the poverty alleviation point of view, but their papers build on the nonlinear tax approach which is not well suited to developing countries. Kanbur and

³ Note that it might be reasonable for some countries to move to a progressive income tax system as their tax capacity increases with development; the study of such dynamics is beyond the scope of this paper.

⁴ In fact, the first SDG is simply "End poverty in all of its forms everywhere."

Keen (1989) do consider linear income taxation together with poverty minimization, but they do not produce optimal tax rules but focus on a tax reform perspective, and provide tax rate simulations. Others have considered different departures from the welfarist standard. For example, Fleurbaey and Maniquet (2007) consider fairness as an objective of the tax-transfer system and its implications on optimal taxation. Roemer et al. (2003) employ a maximin type of social goal and characterize how well tax and transfer systems achieve the goal of equality of opportunity. Second, our work is related to new contributions in behavioral public finance, which address the situation where the behavioral biases of the individuals lead the social planner to adopt a different objective function than the individuals have; see Chetty (2015), Gerritsen (2016), Farhi and Gabaix (2015). A third strand of literature considers taxation and development more generally, such as Gordon and Li (2009), Keen (2009, 2012), Bird and Gendron (2007) and Besley and Persson (2013).⁵ This field, while clearly very relevant, has not concentrated much on the design of optimal redistributive systems. Finally, optimal linear income taxation has been studied from the standard welfarist perspective. We describe these models in Sect. 2.1. The most recent description of linear income tax models can be found in Piketty and Saez (2013). They also emphasize how linear tax rules, while analytically more feasible, provide the same intuition as the more complicated nonlinear models. The linear tax rules, they argue, are robust to alternative specifications⁶, and examining this forms part of our motivation: we study optimal linear tax policies, in our understanding for the first time, from the poverty minimization perspective.

The paper proceeds as follows. Section 2 examines optimal linear income taxation, while Sect. 3 turns to optimal provision rules for publicly provided private and public goods that are financed by such a linear income tax. Section 4 analyzes the combination of optimal linear income taxes and commodity taxation and asks under which conditions one should use differentiated commodity taxation if the government is interested in poverty minimization and also has optimal cash transfers at its disposal. The question of how optimal poverty-minimizing income tax policies are altered in the presence of an informal sector is examined in Sect. 5, whereas Sect. 6 presents a numerical illustration of optimal income taxation for poverty minimization. Finally, conclusions are provided in Sect. 7.

2 Linear income taxation

2.1 Optimal linear income taxation under the welfarist objective

In this section, we give an overview of some of the models and results for optimal linear income taxation as they have been presented in the literature. Many formulae for optimal taxation were developed in the 1970s and 1980s (see Dixit and Sandmo 1977;

⁵ Besley and Persson (2013) use a model with *groups* that can differ in their income-earning abilities. Their analysis focuses, however, on explaining how economic development and tax capacity are interrelated, and not on redistribution between *individuals*.

⁶ They also describe some implications of departures from the welfarist standard in the optimal nonlinear tax model.

Tuomala 1985 and the survey by Tuomala 1990), and they are still being used, whereas Piketty and Saez (2013) offer fresh expressions of the tax rules. Our exposition mainly follows that of Tuomala (1985), but Appendix 1 shows how the results relate to those in Piketty and Saez (2013).

The government collects a linear income tax τ , which it uses to finance a lump-sum transfer b , along with other exogenous public spending R . The individuals differ in their income-earning capacity (w^i), and z^i denotes individual labor income ($w^i L^i$, where L^i represents hours worked). Consumption equals $c^i = (1 - \tau)z^i + b$, where the superscript- i refers to individuals.⁷ There is a discrete distribution of N individuals, whose heterogeneous preferences over consumption and labor are captured by the utility function $u^i(c^i, z^i)$. The maximized (subject to the individual budget constraint) value of this utility function is captured by the indirect utility function, which is denoted by $V^i(1 - \tau, b)$, and we refer to the net-of-tax rate as $1 - \tau = a$. To simplify notation, subscript- a refers to the derivative with respect to the net-of-tax rate.

The government has redistributive objectives represented by a Bergson–Samuelson function $W(V^1, \dots, V^N)$ with $W' > 0$, $W'' < 0$. The government’s problem is to choose the tax rate τ and transfer b so as to maximize the social welfare function $\sum W(V^i(a, b))$ under the budget constraint $(1 - a) \sum z^i = Nb + R$.⁸ We denote the social marginal utility of income by $\beta^i = W_V V_b^i$.

All the mathematical details are presented in Appendix 1. There it is shown that the optimal tax rule is given by

$$\frac{\tau^*}{1 - \tau^*} = \frac{1}{\varepsilon} \left(1 - \frac{z(\beta)}{\bar{z}} \right), \tag{1}$$

where $\varepsilon = \frac{d\bar{z}}{d(1-\tau)} \frac{(1-\tau)}{\bar{z}}$ is the elasticity of total income with respect to the net-of-tax rate, \bar{z} is average income and $z(\beta) = \frac{\sum \beta^i z^i}{\sum \beta^i}$ welfare-weighted average income. Define $\Omega = \frac{z(\beta)}{\bar{z}}$, so that $I = 1 - \Omega$ is a normative measure of inequality or, equivalently, of the relative distortion arising from the second-best tax system. Clearly Ω should vary between zero and unity. One would expect it to be a decreasing function of τ (given the per capita revenue requirement $g = \frac{R}{N}$). There is a minimum feasible level of τ for any given positive g , and of course g must not be too large, or no equilibrium is possible. Hence any solution must also satisfy $\tau > \tau_{\min}$ if the tax system is to be progressive. That is, if the tax does not raise sufficient revenue to finance the non-transfer expenditure, R , the shortfall must be made up by imposing a poll tax ($b < 0$) on each individual. One would also expect the elasticity of labor supply with respect to the net-of-tax rate to be an increasing function of τ (it need not be).

We can rewrite (1) as $\tau^* = \frac{1 - \Omega}{1 - \Omega + \varepsilon}$ to illustrate the basic properties of the optimal tax rate. Because $\varepsilon \geq 0$ and $0 \leq \Omega < 1$, both the numerator and denominator are nonnegative. The optimal tax rate is thus between zero and one. The formula captures

⁷ We consider “income” here as the labor income of individuals, but considering that our model is intended especially for the poorer countries, agricultural income could as well be included in the concept of income. In Sect. 5 we discuss the implications of untaxed home consumption in agricultural production.

⁸ Summation is always over all individuals i , which is suppressed for simplification.

neatly the efficiency-equity trade-off. τ decreases with ε and Ω , and we have the following general results: (1) In the extreme case where $\Omega = 1$, i.e., the government does not value redistribution at all, $\tau = 0$ is optimal. We can call this case libertarian. According to the libertarian view, the level of disposable income is irrelevant (ruling out both basic income b , and other public expenditures, g , funded by the government). (2) If there is no inequality, then again $\Omega = 1$ and $\tau = 0$. There is no intervention by the government. The inherent inequality will be fully reflected in the disposable income. Furthermore, lump-sum taxation is optimal; $b = -g$ or $T = -b$. (3) We can call the case where $\Omega = 0$ as “Rawlsian” or maximin preferences. The government maximizes tax revenue (optimal $\tau = \frac{1}{1+\varepsilon}$) as it maximizes the basic income b (assuming the worst off individual has zero labor income). In fact, maximizing b can be regarded as a non-welfarist case, which is the focus in the next subsection.

2.2 Optimal linear income taxation under non-welfarist objectives

A non-welfarist government is one that follows a different set of preferences than those employed by individuals themselves (Kanbur et al. 2006). Thus, instead of maximizing a function of individual utilities, the government has other, paternalistic objectives that go beyond utilities. A special case taken up in more detail below is the objective of minimizing poverty in the society. To be as general as possible, let us define a “social evaluation function” (as in, e.g., Kanbur et al. 2006) as $S = \sum F(c^i, z^i)$, which the government maximizes instead of the social welfare function. $F(c^i, z^i)$ measures the social value of consumption c^i for a person with income z^i and can be related to $u(c^i, z^i)$ but is not restricted to it. Following Tuomala’s model as above, given the instruments available, linear income tax τ , lump-sum grant b and other expenditure R the government thus maximizes $\sum F(az^i + b, z^i)$ subject to the budget constraint $(1 - a) \sum z^i - Nb = R$. Define

$$\frac{\sum (F_c(z^i + az_a^i) + F_z z_a^i)}{\sum (F_c(1 + az_b^i) + F_z z_b^i)} \equiv \tilde{F}, \quad (2)$$

which reflects the relative impact of taxes and transfers on the social evaluation function. Using this definition, and following the same steps as in the previous section (see Appendix), the optimal tax rate becomes:

$$\frac{\tau^*}{1 - \tau^*} = \frac{1}{\varepsilon} \left(1 - \frac{\tilde{F}}{\bar{z}} \right). \quad (3)$$

The result resembles the welfarist tax rule in (1). In addition to labor supply considerations via the term $\frac{1}{\varepsilon}$, they both entail a term that measures the relative benefits of taxes and transfers, in the welfarist case via welfare-weighted income, in the non-welfarist case via \tilde{F} , the relative impact on the social evaluation function. Note that since under non-welfarism individuals are not necessarily at their utility optimum, the envelope condition does not apply and thus the behavioral responses z_a^i and z_b^i are not cancelled

out in \tilde{F} . That is, the impacts of tax changes on labor supply are not trivial under non-welfarism. The terms $\sum z_a^i (F_c a + F_z)$ in the numerator and $\sum z_b^i (F_c a + F_z)$ in the denominator of (2) capture these effects on the social evaluation function. If taxation had no behavioral impacts ($z_a^i = z_b^i = 0$), it would affect the value of the social evaluation function only by mechanically altering individual after-tax income. Note that in this case, $\tilde{F} = \frac{\sum F_c z^i}{\sum F_c}$ would be a more direct equivalent to $z(\beta) = \frac{\sum \beta^i z^i}{\sum \beta^i}$. The same equivalence would be achieved also when $F_c a + F_z = 0$, that is, the social marginal rate of substitution between income and consumption equals the private rate: $-\frac{F_z}{F_c} = a = -\frac{u_c^i}{u_l^i}$ (the latter is obtained from the individual's first-order condition). In these cases, \tilde{F} would be a purely redistributive term, albeit a non-welfaristic one. Paternalistic concerns additionally enter the optimal tax rule via labor supply changes, captured by the response of z . In this way, the tax rule in (3) can be decomposed, and this decomposition is similar in spirit to the corrective parts of the tax formulae in the new optimal tax literature with behavioral agents, such as Farhi and Gabaix (2015) and Gerritsen (2016).

The signs and magnitudes of F_c and F_z and thus of \tilde{F} depend on the specific objective of the government, that is, on the shape of F . Let us consider the specific case of poverty minimization below.

2.2.1 Special case: poverty minimization

Now let us derive the optimal linear tax results for a government whose objective is to minimize poverty in society. The instruments available to the government are the same, τ and b , and other exogenous expenditure is R . Note first that the revenue-maximizing tax rate is in fact equivalent to the tax rate obtained from a maximin objective function, since when the government only cares about the poverty (consumption) of the poorest individual, its only goal is to maximize redistribution to this individual, i.e., maximize tax revenue.

Let us first define the objective function of the government explicitly. Poverty is defined as deprivation of individual consumption c^i relative to some desired level \bar{c} and measured with a deprivation index $D(c^i, \bar{c})$, such that $D > 0 \forall c \in [0, \bar{c})$ and $D = 0$ otherwise, and $D_c < 0$, $D_{cc} \geq 0 \forall c \in [0, \bar{c})$, as in Pirttilä and Tuomala (2004). A typical example of such an index would be the P_α family of Foster–Greer–Thorbecke (FGT) poverty indices. We discuss the application of FGT indices in our model in Appendix 2. Note, however, that the choice of poverty index depends on the preferences of the government, whether they wish to minimize the total amount of deprivation in the society, or are for instance concerned especially about the incomes of the poorest of the poor. The social evaluation function $F(c^i, z^i)$ becomes $D(c^i, \bar{c})$ and the objective function is $\min P = \sum D(c^i, \bar{c})$. Now $F_c = D_c$ and $F_z = 0$, so

$$\tilde{F} = \tilde{D} = \frac{\sum D_c (z^i + a z_a^i)}{\sum D_c (1 + a z_b^i)}, \tag{4}$$

and the optimal tax rule becomes:

$$\frac{\tau^*}{1 - \tau^*} = \frac{1}{\varepsilon} \left(1 - \frac{\tilde{D}}{\bar{z}} \right). \quad (5)$$

Since now $F_z = 0$, the result is closer to (1) than (3) was, although part of the labor supply impacts still remain. Here \tilde{D} describes the relative efficiency of taxes and transfers in reducing deprivation. Both the numerator and denominator of \tilde{D} depend on D_c , so the difference in the relative efficiency of the two depends on z_a^i and z_b^i . The more people react to taxes (relative to transfers) by earning less, the higher is \tilde{D} and the lower should the tax rate be. In (1), the higher is the social value of income, the higher is $z(\beta)$ and the lower should the tax rate be.

Since the form of the result is similar in the welfarist and the poverty minimization cases, the analysis could be also seen as a special case of the argument in Saez and Stantcheva (2016), who derived generalized social welfare weights and express the tax formulae in terms of those.⁹ Here, the generalized social welfare weight would thus be derived from a poverty minimization objective. It could be close to a suitably defined welfarist criterion, and clearly it would be exactly the same only if the welfarist criterion would correspond to the chosen poverty minimization objective.

We can also rewrite \tilde{D} , using $a = 1 - \tau$, as: $\frac{\sum D_c(z^i + (1-\tau)\frac{\partial z^i}{\partial(1-\tau)})}{\sum D_c(1+(1-\tau)z_b^i)} = \frac{\sum D_c(1 + \frac{(1-\tau)}{z^i} \frac{\partial z^i}{\partial(1-\tau)})z^i}{\sum D_c(1+(1-\tau)z_b^i)}$. Thus the \tilde{D} in the optimal tax result (5) entails a further consideration that depends on labor supply responses. It combines paternalistic preferences—how much poverty is reduced—with the behavioral responses to a tax system—how much labor income increases when the take-home pay goes up. The latter effect tends to lower the optimal tax rate to induce the poor to work more. Kanbur et al. (1994) find a similar result in their nonlinear poverty-minimizing tax model. Here, however, we are restricted to lower the tax on everyone instead of only the poorest individuals.

To summarize, the non-welfarist tax rules differ from the welfarist ones, depending on the definition of non-welfarism in question (the F_c and F_z terms). However, when we take poverty minimization as the specific case of non-welfarism, the tax rules are quite similar to welfarist ones. The basic difference is that equity is not considered in welfare terms but in terms of poverty reduction effectiveness. A more notable difference arises from efficiency considerations. With linear taxation, taking into account labor supply responses means that everybody's tax rate is affected, instead of just the target group's. If we want to induce the poor to work more to reduce their poverty, we need to lower everyone's tax rate. The welfarist linear tax rule does not take this into account. It is not, however, possible to state that under poverty minimization tax rates are optimally lower than under welfare maximization, since we cannot directly compare the welfare and deprivation terms. However, there is an additional efficiency consideration involved under poverty minimization. Nonlinear tax rules of course make it possible to target lower tax rates on the poorer individuals, but in a developing country context with

⁹ We are grateful to a referee for this point.

lower administrative capacity this is not necessarily possible, and such considerations affect everyone’s tax rate.

3 Public good provision with linear income taxes

3.1 Optimal public provision under the welfarist objective

Let us first extend the welfarist model of linear taxation to include the provision of pure public goods. The government offers a universal pure public good G , which enters individual utilities in addition to the consumption of private goods. The government’s objective function is now $\sum W(V^i(a, b, G))$, whereas the budget constraint becomes $(1 - a) \sum z^i - Nb - N\pi G = R$ where π is the producer price of the public good. The consumer price of private consumption is normalized to 1. Let us now define the marginal willingness to pay for the public good by the expression $\sigma = \frac{V_G}{V_b}$ and $\sigma^* = \frac{\sum \beta^i \sigma^i}{\sum \beta^i}$ as the welfare-weighted average marginal rate of substitution between public good and income for individual i . The rule for public provision can then be written as

$$\pi = \sigma^* - \tau (\sigma^* \bar{z}_b - \bar{z}_G). \tag{6}$$

This public good provision rule is a version of a modified Samuelson rule. It equates the relative cost of providing the public good to the welfare-weighted sum of marginal rates of substitution (MRS). It also includes a revenue term, which takes into account the impacts of public good provision and income transfers on labor supply and thus tax revenue.

Consider first the case when labor supply does not depend on public good provision and there are no income effects, i.e., $\bar{z}_G = \bar{z}_b = 0$. Then we are left with a more familiar rule that welfare-weighted aggregate MRS must equal the cost of the public good. When we add income effects so that $\bar{z}_b < 0$, and since σ^* is positive, then because of the second term in (6), the financing costs of the public good are reduced. Likewise, if labor supply and public provision are positively related, the financing costs of the public good are reduced.

3.2 Optimal provision of public goods under poverty minimization

Now consider a non-welfarist government interested in minimizing poverty. The public good G which it offers enters the deprivation index separately from other, private consumption x : $D(x, G, \bar{x}, \bar{G})$. The government still offers a lump-sum cash transfer b as well and finances its expenses with the linear income tax τ .

Again alternative formulations of the public good provision rule can be written. The first is

$$\pi = D^* - \tau (D^* \bar{z}_b - \bar{z}_G), \tag{7}$$

which can be compared with Eq. (6). Here, $D^* = \frac{\sum D_G + \sum D_x a z_G^i}{\sum D_x (1 + a z_b^i)}$ captures the efficiency of the public good in reducing deprivation relative to the income trans-

fer (because $D_G, D_x < 0, D^* > 0$). Again, if $\bar{z}_G = \bar{z}_b = 0$, the equation reduces to $\pi = D^* = \frac{\sum D_G}{\sum D_x}$. This rule highlights a considerable difference to the standard modified Samuelson rules, reflecting instead of a welfare-based MRS the direct poverty reduction impact of the public good. With $\bar{z}_G \neq 0$ and $\bar{z}_b \neq 0$, D^* also depends on the indirect impacts of the public good via labor supply on consumption. As previously, the right-hand side includes a tax revenue term. Using the same example as in the context of (6), if $\bar{z}_G = 0$ and $\bar{z}_b < 0$, the price π of the public good would be higher than its relative efficiency in eliminating deprivation.

Here we have allowed the government to be directly interested in the consumption of some pure public good. But if the government is solely interested in reducing income poverty, it might not include such goods in the deprivation measure.¹⁰ However, suppose that individual welfare does not directly depend on the public good provided but the public good can have a productivity increasing impact. An example could be publicly provided education services that affect individuals' productivity via the wage rate. We therefore suppose that the direct impact of the public good on deprivation cancels out (i.e., $D_G = 0$), whereas the wage rate becomes an increasing function of G , i.e., $w'(G) > 0$ (denoting $z = w(G)L$). This means that the expression for D^* is rewritten as

$$D^* = \frac{\sum D_x a \left(w \frac{\partial L}{\partial G} + w' L \right)}{\sum D_x \left(1 + a w \frac{\partial L}{\partial b} \right)}. \tag{8}$$

This means that even if labor supply would not react to changes in public good provision, such provision would still be potentially desirable through its impact on the wage rate. In this way, public good provision can be interpreted as increasing the capability of the individuals to earn a living wage, which serves as a poverty reducing tool, and which can in some cases be a more effective way to reduce poverty rather than direct cash transfers. The optimality depends on the relative strength of $w'(G) > 0$ versus the direct impact of the transfers.

An alternative provision rule for the public good, which results from extending the Piketty–Saez approach, in the usual case where it also enters individuals' utility function is

$$\frac{\int \left(D_G + D_x (1 - \tau) \frac{\partial z^i}{\partial G} \right) dv(i)}{\int D_x dv(i)} = \pi - \tau \frac{dZ}{dG}. \tag{9}$$

In the numerator of the left-hand side, the first term is the direct deprivation effect of G and the second term captures the indirect deprivation effect, operating via the labor supply impacts of the public good, which affect the level of private consumption, x . These impacts are scaled by the poverty alleviation impact of private consumption itself (the impact of a cash transfer). The right-hand side reflects the costs of public good provision: besides the direct cost of the good there is an indirect tax revenue effect

¹⁰ See also Appendix 2 for multidimensional considerations in poverty measurement.

operating through labor supply. The condition is directly comparable to the welfarist rule, given in (39) in the Appendix, because even though the welfarist case relies on utilities, in the FOC for G no envelope condition is evoked. The only difference between Eqs. (39) and (9) is that the utility and welfare weight terms are exchanged for deprivation terms.

Consider finally the provision of a quasi-private good, such that in addition to the publicly provided amount, individuals can purchase (“top-up”) the good themselves as well. The good is denoted by s and its total amount consists of private purchases h and public provision G : $s = G + h$. In addition to good s , individuals consume other private goods, denoted by x . The individual budget constraint is thus $c^i = x^i + ph^i = (1 - \tau)z^i + \tau Z(1 - \tau) - R - \pi G$, where p is the consumer price of private purchases of the quasi-private good. The producer price of education in the private sector (p) or in the public sector (π) can be equal, or one sector could have access to cheaper technology. Deprivation is determined in terms of consumption of x and s , so the objective function is $\min P = \int D(x^i, s^i, \bar{x}, \bar{s}) \, d\nu(i)$. In this case, the provision rule is

$$\frac{\int \left[D_x \left((1 - \tau) \frac{\partial z^i}{\partial s} \frac{\partial s}{\partial G} - p \frac{\partial h^i}{\partial G} \right) + D_s \frac{\partial s^i}{\partial G} \right] \, d\nu(i)}{\int D_x \, d\nu(i)} = \pi - \tau \frac{dZ}{dG}. \tag{10}$$

The result is analogous to the pure public good result in (9), with the difference that now the impact G has on poverty depends on whether public provision fully crowds out private purchases of the good (i.e., $\frac{dh}{dG} = -1 \Leftrightarrow \frac{ds}{dG} = 0$) or not (i.e., $\frac{dh}{dG} = 0 \Leftrightarrow \frac{ds}{dG} = 1$). If there is full crowding out, an increase in public provision of G that is fully funded via a corresponding increase in the tax rate has no impact on the consumption of s and consequently no impact on poverty. If there is no crowding out, however, the FOC becomes

$$\frac{\int \left[D_x \left((1 - \tau) \frac{\partial z^i}{\partial s} \right) + D_s \right] \, d\nu(i)}{\int D_x \, d\nu(i)} = \pi - \tau \frac{dZ}{dG}, \tag{11}$$

which is the same as in the case of a pure public good in Eq. (9).

To summarize, the welfarist public provision rule, when public goods are financed with linear income taxes and supplemented with lump-sum transfers, differs from the standard modified Samuelson rule. It equates a welfare-weighted sum of MRS to the marginal cost where tax revenue impacts are taken into account. Indirect effects of public provision (through labor supply decisions and thus private consumption) are incorporated. The poverty-minimizing public provision rule, however, replaces the welfare-weighted sum of MRS with the relative marginal returns to deprivation reduction. Here the “MRS” term measures how well public good is translated to reduced poverty (incorporating indirect effects as well), relative to private consumption. Finally, when the public good has positive effects on productivity, its provision can be desirable even if it would not have any direct impact on poverty.

4 Commodity taxation with linear income taxes

4.1 Optimal commodity taxation with linear income tax under the welfarist objective

This section considers the possibility that the government also uses commodity taxation (subsidies) to influence consumers' welfare. We follow the modeling of [Diamond \(1975\)](#). Unlike the analysis above, there are J consumer goods x_j instead of just two. Working with many goods is used to be able to more clearly describe the conditions under which uniform commodity taxation occurs at the optimum. The government levies a tax t_j on the consumption of good x_j , so that its consumer price is $q_j = p_j + t_j$, where p_j represents the producer price (a commodity subsidy would be reflected by $t_j < 0$). Let q denote the vector of all consumer prices. In addition, the government can use a lump-sum transfer, b . Note that in this exposition, leisure is the untaxed numeraire commodity. Alternatively, one could also imply a linear tax on labor supply as above and treat one of the consumption goods as the untaxed numeraire. However, choosing leisure as the numeraire makes the exposition easier. Thus, the consumer's budget constraint is $\sum_j q_j x_j^i = z^i + b$.

The government maximizes $\sum_i W(V^i(b, q))$ subject to its budget constraint $\sum_i \sum_j t_j x_j^i - Nb = R$. It is useful to define, following [Diamond \(1975\)](#),

$$\gamma^i = \beta^i + \lambda \sum_j t_j \frac{\partial x_j^i}{\partial b} \quad (12)$$

as the net social marginal utility of income for person i . This notion takes into account the direct marginal social gain, β^i , and the tax revenue impact arising from commodity demand changes. The rule for optimal commodity taxation for good k is shown to be

$$\frac{1}{N} \sum_i \sum_j t_j \frac{\partial \bar{x}_k^i}{\partial q_j} = \frac{1}{\lambda} \text{cov}(\gamma^i, x_k^i). \quad (13)$$

The left-hand side of the rule is the aggregate compensated change (weighted by commodity taxes) of good k when commodity prices are changed. The right-hand side refers to the covariance of the net marginal social welfare of income and consumption of the good in question. The rule says that the consumption of those goods whose demand is the greatest for people with low net social marginal value of income (presumably, the rich) should be discouraged by the tax system. Likewise the consumption of goods such as necessities should be encouraged by the tax system.

The key policy question is whether or when uniform commodity taxes are optimal, or, in other words, when would a linear income tax combined with an optimal demogrant be sufficient to reach the society's distributional goals at the smallest cost. [Deaton \(1979\)](#) shows that weakly separable consumption and leisure and linear Engel curves are sufficient conditions for the optimality of uniform commodity taxes. These requirements are quite stringent and unlikely to hold in practice; however, the economic importance they imply is unclear. If implementing differentiated commodity

taxation entails significant administrative costs, they may easily outweigh the potential benefits of distributional goals and that is why economists have typically been quite skeptical about non-uniform commodity taxation when applied to practical tax policy.

4.2 Optimal commodity taxation with linear income tax under poverty minimization

Poverty could be measured in many ways when there are multiple commodities: the government may care about overall consumption, the consumption of some of the goods (those that are in the basket used to measure poverty) or then it cares about both the overall consumption and the relative share of different kinds of consumption goods (such as merit goods). We discuss these measurement issues in Appendix 2, but here we examine the simplest set-up where deprivation only depends on disposable income, $c^i = z^i + b$. Using the consumer’s budget constraint, this is equal to the overall consumption level, $\sum_j q_j x_j^i$.

The government thus minimizes the sum of the poverty index $D(\sum_j q_j x_j^i, \bar{c})$, and the budget constraint is the same as before. It is again useful to define

$$\gamma_P^i = D_c \sum_j q_j \frac{\partial x_j^i}{\partial b} + \lambda \sum_j t_j \frac{\partial x_j^i}{\partial b} \tag{14}$$

as the net poverty impact of additional income for person i . This notion takes into account the direct impact on poverty and the tax revenue impact arising from commodity demand changes.

As shown in Appendix 1 section “Commodity taxation”, this leads to an optimal tax rule as below:

$$\frac{1}{N} \sum_i \sum_j t_j \frac{\partial \bar{x}_k^i}{\partial q_j} = -\frac{1}{\lambda} \left[\frac{1}{N} \sum_i D_c x_k^i + \frac{1}{N} \sum_i \sum_j D_c q_j \frac{\partial \bar{x}_k^i}{\partial q_j} \right] + \frac{1}{\lambda} \text{cov}(\gamma_P^i, x_k^i). \tag{15}$$

In this formulation, the left-hand side is the same as in the welfarist case and it reflects the aggregate compensated change in the demand of good k . The first two terms in the square brackets at the right-hand side capture the impacts of tax changes on poverty: the first term is the direct impact of the price change (keeping consumption unaffected) on measured poverty, whereas the second depends on the behavioral shift in consumption. Multiplied by the minus sign, the former term implies that the consumption of the good should be encouraged, whereas if demand decreases when the prices increase, the latter term actually serves to discourage consumption. The last term on the right reflects the same principles as the covariance rule in Eq. (13), the correlation of the net poverty impact of income and the consumption of the good in question. That is, the covariance part of the tax rule moves the tax rule in the direction of favoring goods that have high poverty reduction impact on the poor (i.e., that the poor consume more).

The key lesson to note from the optimal commodity tax rule in the poverty minimization case is that the conventional conditions for uniform commodity tax to be optimal are not valid anymore. The reason is that even if demand was separable from labor supply, the first term on the right still remains in the rule, and its magnitude clearly varies depending on the quantity of good consumed. Thus, income transfers are not sufficient to alleviate poverty when the government aims to minimize poverty that depends on disposable income. The intuition is very simple: commodity tax changes have a direct effect on the purchasing power of the consumer, and these depend on the amount consumed. The extent of encouraging the consumption of the goods is the greater, the larger is their share of consumption among the consumption bundles of the poor. The result resembles that of [Pirttilä and Tuomala \(2004\)](#), meaning that the intuition from optimal nonlinear income taxation under poverty minimization carries over to linear income taxation. A formal proof is provided in [Appendix 1](#).

In sum, the rule for optimal commodity taxation is changed when we shift from welfare maximization to poverty minimization. The welfarist rule reflects a fairly straightforward trade-off between efficiency (tax revenue) and equity (distributional impacts). The poverty-minimizing commodity tax rule brings new terms; the interrelations of which are not easy to disentangle. It, however, also takes into account efficiency considerations (tax revenue through indirect labor supply effects) and equity (direct impact of the taxed good on poverty and indirect impact via labor supply effects). Most importantly, the conventional wisdom of when uniform commodity taxation is sufficient fails to hold in the poverty minimization case. Thus, observed commodity subsidies in developing countries, such as fuel or food subsidies, can be considered optimal given the preference for poverty minimization.¹¹ In practice, it would be wise to limit the number of differentiated commodity tax rates to a few essential categories such as fuel and food, in order to keep the administrative complexity at a minimum.

5 Poverty minimization in the presence of an informal sector

An important issue for a developing country attempting to collect taxes is the issue of a large informal sector. If part of tax revenue is lost due to tax evasion in the informal sector, which is likely to be the case in the less developed economies, then the income transfer is reduced and redistributive targets may not be met. In this section, we discuss the implications of informality for optimal redistributive policies for a government wishing to minimize poverty.¹² The results can thus be contrasted to those obtained in previous sections.

¹¹ [Keen \(2014\)](#) uses a tax reform approach and examines how much more effective transfers need to be than differentiated commodity subsidies in reaching the poor to achieve the same poverty reduction with lower government outlays.

¹² Such a society might also reflect poor administrative power and corruption in the tax collecting authority. Notice, however, that considering only the “leakage” of tax revenue in the model would only reduce the extent of poverty reduction achieved with taxation by lowering the income transfer for everyone. The poverty reduction efficiency of taxation would thus be lowered, but there would be no differential effects across individuals.

Following Kanbur (2015) and Kanbur and Keen (2014), informal operators can be categorized as those who should comply with regulations but illegally choose not to, and those who legally remain outside regulation, e.g., due to the smaller size of operations (either naturally or by adjusting size as a response to regulation). For our purposes, however, it is enough to lump these categories into one “informal sector,” where it is possible to avoid taxes at least to some extent. It is also possible for workers to work in both sectors, such that part of total income is declared for taxation and part is evaded (consider, e.g., supplementing official employment income with street vending). Note also that especially in the case of agriculture, evasion can also consist of home production. In this case, the reason for “informality” would be the small size of the producing entity, such that they are naturally not liable for taxes. Production for own consumption is, however, still relevant for the well-being and measured poverty of the family.

In this application, we follow the approach pioneered in Besley and Persson (2013). They work with a model that fits into the description above, where part of the tax base evades taxes. We thus take informality as given, and do not consider whether informality is “natural,” illegal or a response to taxation. Furthermore, this intensive margin model (what extent of income is earned in the informal sector), they argue, yields essentially similar results as an extensive margin model (whether to participate in the formal job market).

Consider the case of income taxation. We can incorporate informality into the model by noting that people can shelter part e of their labor income from taxation. The extent of evasion is assumed to increase when the tax rate goes up, and thus $\frac{\partial e}{\partial a} < 0$. Income taxes are only paid from income $z^i - e^i$. It is noteworthy that for a government wishing to minimize income poverty, this is in fact beneficial: disposable incomes rise. The more this effect is concentrated among the poor who enter the deprivation index, the better. Individual consumption is now $z^i - \tau(z^i - e^i) + b = e^i + a(z^i - e^i) + b$. On the other hand, tax collections are reduced: the budget constraint becomes $(1 - a) \sum (z^i - e^i) = Nb + R$. Our formulation follows that of Besley and Persson (2013), but we simplify it in order to explicitly consider the problem of optimal taxation, whereas they focus on the issue of investments in the state’s fiscal capacity (we abstract from this issue here and take evasion as given).¹³ The framework, however, nicely captures the essential trade-offs a government faces when there is tax evasion.

The government now minimizes the Lagrangian $L = \sum D(e^i + a(z^i - e^i) + b, \bar{c}) + \lambda((1 - a) \sum (z^i - e^i) - Nb - R)$. The first-order condition with respect to the net-of-tax rate is:

$$\begin{aligned} & \sum D_c \left(\frac{\partial e^i}{\partial a} + z^i - e^i + a \left(\frac{\partial z^i}{\partial a} - \frac{\partial e^i}{\partial a} \right) \right) \\ & = \lambda \left(\sum (z^i - e^i) - (1 - a) \sum \left(\frac{\partial z^i}{\partial a} - \frac{\partial e^i}{\partial a} \right) \right), \end{aligned} \tag{16}$$

¹³ Another difference is that in their original formulation, people face costs of evasion. When the tax rate goes up, the relative attractiveness of tax evasion increases, producing the same kind of effect ($\frac{\partial e}{\partial a} < 0$) we assume directly here for brevity. (These costs could be related to, e.g., Allingham–Sandmo-type risk of being caught and facing sanctions.) Also Slemrod’s (1990) review suggests that higher tax rates tend to increase the supply of labor to the informal sector.

whereas, under the assumption that there are no income effects in evasion, the first-order condition with respect to b stays the same. From here, we can derive a rule for the optimal tax following the same steps as in Sect. 2.2:

$$\frac{\tau^*}{1 - \tau^*} = \frac{1}{\varepsilon^e} \left(1 - \frac{\tilde{D}^e}{\bar{z}^e} \right), \quad (17)$$

where now ε^e is a tax elasticity of the net-of-evasion tax base $\bar{z}^e = \bar{z} - \bar{e}$ and \tilde{D}^e represents the relative impact of taxes and transfers on the deprivation index (see Appendix 1 for further detail). The rule represents a trade-off between poverty reduction and efficiency, both of which are now altered by evasion. There is a pressure toward lower tax rates, as now distortions of taxation are increased by evasion behavior, so $\varepsilon^e > \varepsilon$. Contrary to this effect, \tilde{D}^e is reduced compared to \tilde{D} because reducing taxes (increasing a) is now a less useful instrument for poverty reduction, as part of the taxes have been evaded. As $\frac{\partial \varepsilon^e}{\partial a} < 0$, people pay more taxes when tax rates are reduced, and therefore poverty in fact increases. \tilde{D}^e thus works to increase tax rates.

Therefore, an interesting trade-off arises: informality increases the cost of raising taxes, but it also means that higher taxes are less harmful as those in the informal sector do not need to pay them (and they are still entitled to the lump-sum transfer).¹⁴ These countervailing forces have not been noted by the literature before. The presence of informality therefore seems to give rise to tax policy rules that are far from trivial. Future work could also look more deeply into the issue of the tax mix in the presence of informality. If income tax is more easily evaded than commodity taxation, as Boadway et al. (1994) suggest, this could give rise to policies that focus taxation and redistribution on commodity taxes and subsidies, instead of income taxes and lump-sum transfers. Slemrod and Gillitzer (2014) have also suggested focusing on a “tax systems approach” and including, among other things, evasion behavior into optimal taxation analysis to obtain more useful prescriptions for actual tax policy. This topic certainly deserves a more detailed analysis.

6 A numerical illustration

To further illustrate the differences of tax rates under poverty minimization and welfareism, we provide a simple numerical simulation. Here we concentrate on the special case where there are no income effects on labor supply and the elasticity of labor supply with respect to the net-of-tax wage rate is constant. If ε denotes this elasticity, the quasi-linear indirect utility function is given by $v(w(1 - \tau), b) = b + \frac{[w(1 - \tau)]^{1 + \varepsilon}}{1 + \varepsilon}$, so that ε is constant. Like most work on optimal nonlinear and linear income taxation, we use the lognormal distribution $\ln(n, m\sigma^2)$ to describe the distribution of productivities with support $[0, \infty)$ and parameters m and σ (see Aitchison and Brown 1957). The first parameter, m , is the log of the median wage. The second parameter, the variance

¹⁴ The idea that those in the informal sector can still receive transfers matches well with reality: many of the cash transfer systems reach those with little or no connection to the formal sector.

of log wage σ^2 , is itself an inequality measure. As is well known, the lognormal distribution fits reasonably well over a large part of the income range but diverges markedly at both tails. The Pareto distribution in turn fits well at the upper tail. We also use the two-parameter version of the Champernowne distribution (known also as the Fisk distribution). This distribution approaches asymptotically a form of Pareto distribution for large values of wages but it also has an interior maximum. In our simulations, the revenue requirement is set to zero; thus, the system is purely redistributive.

To illustrate the poverty-minimizing tax formula in (3), we also need to specify a measure of poverty. Typically, poverty indices consist of computing some average measure of deprivation by setting individual needs as defined above at the agreed upon poverty line \bar{c} . For this purpose, we take a poverty index of the form developed by Foster et al. (1984). They have proposed defining a poverty index as the average of these poverty gaps across individuals raised to some power α . When $\alpha = 1$, it is just the proportion of units below the poverty line multiplied by the average poverty gap. (See Appendix 2 for more details.) We consider the cases where either 30 or 40% of the population lie below the poverty line.

The results from the simulation of the optimal tax when the government minimizes the poverty gap for the lognormal case are presented in Table 1. Results are shown for two different values of labor supply elasticity ε , two different values regarding income dispersion σ , and two values of the share of population below the poverty line $F(\bar{w})$. The tax rates are high, above 60%, for all the combinations of parameter values.¹⁵

Comparing these results to the welfarist case is not straightforward, as those depend on the chosen welfare function. We adopt a constant relative inequality aversion form of the welfare function: the contribution to social welfare of the i th individual is $\frac{w_i^{1-\eta}}{1-\eta}$, where η is the constant relative inequality aversion coefficient. Hence, the social marginal value of income to an individual with wage rate w is proportional to $w^{-\eta}$. Using the property of the lognormal distribution $\ln(E(w^s)) = sm + s^2\frac{\sigma^2}{2}$, we can calculate the optimal tax rate from the following formula: $\frac{\tau}{1-\tau} = \frac{1}{\varepsilon}[1 - e^{-\eta(1+\varepsilon)\sigma^2}]$. Or, using the property of the lognormal distribution that $\ln(1 + cv^2) = \sigma^2$, where cv is the coefficient of variation, we can rewrite $\tau = \frac{1}{1+\varepsilon/[1+cv^2]^{-\eta(1+\varepsilon)}}$.

A wide range of values for the inequality aversion parameter η have been employed in the literature, varying typically from 0.5 to 2. Note that, as discussed in Sect. 2.1, as $\eta \rightarrow \infty$, social preferences approach “maximin” preferences, where the optimal tax rate is the same as the revenue-maximizing tax rate, $\tau = \frac{1}{1+\varepsilon}$, which does not depend on the original income distribution. Naturally, if there is no regard for inequality in the society, $\eta = 0$ and $\tau = 0$. Table 2 displays the welfaristic tax simulation results for two different values of labor supply elasticity ε , for two different values of income dispersion σ , and for five different values of inequality aversion η .

The simulation results illustrate clearly that at conventional inequality aversion levels, optimal welfaristic tax rates lie well below the poverty-minimizing rates. Only as inequality aversion becomes extremely high do the welfaristic rates approach the

¹⁵ The results are very similar using the Champernowne distribution (with income dispersion parameters chosen so that inequality is similar in both cases), which is not very surprising as the distributions only differ at the top of the income schedule. These results are available upon request.

Table 1 Simulated tax rates for poverty minimization under different values of ε , σ , and $F(\bar{w})$

ε	$\sigma = 0.7$		$\sigma = 1.0$	
	$F(\bar{w}) = 0.3$	$F(\bar{w}) = 0.4$	$F(\bar{w}) = 0.3$	$F(\bar{w}) = 0.4$
0.25	79	77	79	78
0.5	65	63	66	64

Table 2 Simulated tax rates in the welfaristic case under different values of ε , σ , and η

ε	$\sigma = 0.7$			
	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta \rightarrow \infty$
0.25	43	58	69	80
0.5	31	44	56	67
ε	$\sigma = 1.0$			
	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta \rightarrow \infty$
0.25	52	65	74	80
0.5	38	51	61	67

poverty-minimizing ones. With poverty minimization as the social objective, optimal tax rates are close to the revenue-maximizing “maximin” rate.

Another point of comparison could be the welfaristic linear tax simulations of [Stern \(1976\)](#). His calculations differ from ours as he incorporates income effects and a non-constant elasticity of labor supply with respect to the tax rate.¹⁶ With the elasticity of substitution between consumption and leisure at 0.5 and income dispersion described by $\sigma = 0.39$, as concern for inequality rises from low to medium and high, he finds tax rates rising from 19 to 43 and 48%. The extreme “maximin” result is 80%. These tax rates are also clearly lower than the poverty-minimizing rates, except at very extreme values of inequality aversion.

These numerical examples and [Stern’s \(1976\)](#) results tend to suggest that the tax rates for the poverty minimization case are likely to be higher than for many welfarist examples. The results compare to [Kanbur et al. \(1994\)](#), who also found that the (non-linear) marginal tax rates on the poor are fairly high under the poverty minimization objective. Both their and our results are interesting from the point of view that the analytical formulae for the optimal tax rate include a term that, ceteris paribus, encourages labor supply, but in computational results its influence is offset, most likely, by the need to minimize the poverty gap. The higher the poverty rate, the higher the lump-sum grant financed by these taxes needs to be, in order to raise more people out of poverty.

¹⁶ Our simplifying assumptions allow us to provide tax rates with respect to the three parameters in [Table 2](#).

7 Conclusion

This paper examined optimal linear income taxation, public provision of public and private goods and the optimal combination of linear income tax and commodity taxes when the government's aim is to minimize poverty. The linear tax environment was chosen because such taxes are more easily implementable in a developing country context and since optimal linear tax rules are seen to provide similar intuition as the more complex nonlinear tax formulas.

The results show that the linear income tax includes additional components that work toward lowering the marginal tax rate. This result arises from the goal to boost earnings to reduce income poverty. Unlike in the optimal nonlinear income tax framework, this lower marginal tax affects all taxpayers in the society. However, the numerical simulations offered suggest that this mechanism is offset by the distributive concerns and in practice the optimal tax rates for poverty minimization appear high. Public good provision in the optimal tax framework under poverty minimization was shown to depend on the relative efficiency of public provision versus income transfers in generating poverty reductions. One particular avenue where public provision is useful is via its potentially beneficial impact on individuals' earnings capacity. Thus, public provision can be desirable even if its direct welfare effects were non-existent.

Perhaps more importantly, poverty minimization as an objective changes completely the conditions under which uniform commodity taxation is optimal. When the government's objective is to minimize poverty that depends on disposable income, uniform commodity taxation is unlikely to be ever optimal: this is because the commodity tax changes have first-order effects on consumers' budget via the direct impact on the cost of living, and this direct effect depends on the relative importance of different goods in the overall consumption bundle. Separability in demand coupled with linear Engel curves is not sufficient to guarantee optimality of uniform commodity taxes. In reality, the administrative difficulties of implementing commodity taxation with many tax rates must, of course, be taken into account, as well.

We also examined the implications of the presence of an informal sector for optimal tax and transfer policies. The results revealed that when the government is concerned about income poverty, the presence of the informal sector is, on the one hand, useful, as it reduces the poverty-increasing effect of higher taxes but, on the other hand, it is also costly since it is likely to increase the elasticity of the tax base. Examining the implications of informality on the role of other instruments of government policies is an important avenue for future work.

Another strand of follow-up work should address the question of complementary policies for redistribution, such as minimum wages. It should be borne in mind that different policies impose different requirements on administrative capacity,¹⁷ and

¹⁷ For example, [Lee and Saez \(2012\)](#) show how a minimum wage policy can usefully complement an optimal nonlinear income tax and transfer policy under welfarist objectives. However, imposing minimum wage regulation implies that the government needs to be either able to observe individual wage rates, or has sufficient institutional strength to rely on whistleblowers to denounce non-complying employers, in order to enforce the legislation.

examining which poverty reduction instruments become available only as the societies advance on their development path is an interesting avenue for further work.

Acknowledgements We are grateful to two anonymous referees and seminar audiences at the UNU-WIDER Conference on ‘Inequality: Measurement, Trends, Impacts and Policies’, the Annual Meeting of the Finnish Economic Association, the HECER Development Economics seminar and the IIPF conference in Dublin for useful comments. This research originates from the UNU-WIDER project *The economics and politics of taxation and social protection*. Funding from the Academy of Finland (Grant No. 268082) is gratefully acknowledged.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1: Mathematical appendix

Linear income taxation

Welfarism

Consider first the welfarist case. Using λ to denote the multiplier associated with the budget constraint, the government’s Lagrangian is $L = \sum W(V^i(a, b)) + \lambda((1 - a) \sum z^i - Nb - R)$. The social marginal utility of income is $\beta^i = W_V V_b^i$. Using Roy’s theorem, $V_a^i = V_b^i z^i$, we have $W_V V_a^i = \beta^i z^i$. The first-order conditions with respect to a and b , respectively, are then:

$$\sum \beta^i z^i = \lambda \left(\sum z^i - (1 - a) \sum z_a^i \right) \quad (18)$$

$$\sum \beta^i = \lambda \left(N - (1 - a) \sum z_b^i \right). \quad (19)$$

Divide (18) by (19) to get:

$$\frac{\sum \beta^i z^i}{\sum \beta^i} = \frac{\sum z^i - (1 - a) \sum z_a^i}{N - (1 - a) \sum z_b^i}. \quad (20)$$

Denote average income $\bar{z} = \frac{\sum z^i}{N}$ and welfare-weighted average income $z(\beta) = \frac{\sum \beta^i z^i}{\sum \beta^i}$ to get:

$$z(\beta) = \frac{\bar{z} - (1 - a)\bar{z}_a}{1 - (1 - a)\bar{z}_b}. \quad (21)$$

Multiply the government’s revenue constraint by $\frac{1}{N}$ and define $g = \frac{R}{N}$ to get $(1 - a)\bar{z} - b = g$, and totally differentiate, keeping g constant:

$$\frac{db}{da} \Big|_{g\text{const}} = \frac{\bar{z} - (1 - a)\bar{z}_a}{-1 + (1 - a)\bar{z}_b} = -z(\beta). \tag{22}$$

The fact that $z(\beta) = -\frac{db}{da} \Big|_{g\text{const}}$ tells us that welfare-weighted labor supply should be equal to the constant-revenue effect of tax rate changes in b .

By totally differentiating average labor income \bar{z} and using (22), we have

$$\frac{d\bar{z}}{da} \Big|_{g\text{const}} = \bar{z}_a + \bar{z}_b \frac{db}{da} \Big|_{g\text{const}} = \bar{z}_a - \bar{z}_b z(\beta). \tag{23}$$

When we impose g as a constant we have to give up one of our degrees of freedom. Now the interpretation of $\frac{d\bar{z}}{da} \Big|_{g\text{const}}$ is then the effect on labor supply when a is changed, as is b , in order to keep tax revenue constant. Using (23) we can write (21):

$$z(\beta) - \bar{z} = -(1 - a) \frac{d\bar{z}}{da} \Big|_{g\text{const}} = -\tau \frac{d\bar{z}(1 - \tau)\bar{z}}{d(1 - \tau)(1 - \tau)\bar{z}}, \tag{24}$$

from which we get the optimal tax rate of Eq. (1).

We now derive the results in the form of the **Piketty and Saez (2013)** model. In their model, there is a continuum of individuals, whose distribution is $v(i)$ (population size is normalized to one). Individuals maximize their utility $u^i((1 - \tau)z^i + b, z^i)$, and their FOC implicitly defines the Marshallian earnings function $z_u^i(1 - \tau, b)$. Using this, aggregate earnings are $Z_u(1 - \tau, b)$. The government’s budget constraint $b + R = \tau Z_u(1 - \tau, b)$ implicitly defines b as a function of τ , and consequently Z_u can also be defined solely as a function of τ : $Z(1 - \tau) = Z_u(1 - \tau, b(\tau))$. Z has elasticity $\varepsilon = \frac{1 - \tau}{Z} \frac{dZ}{d(1 - \tau)}$.

To start, note that if the government only cared about maximizing tax revenue $\tau Z(1 - \tau)$, it would set τ such that $\frac{\partial(\tau Z(1 - \tau))}{\partial \tau} = 0$: $Z(1 - \tau) - \tau \frac{dZ}{d(1 - \tau)} = 0$. Using $\frac{\tau}{Z} \frac{dZ}{d(1 - \tau)} = \frac{\tau}{1 - \tau} \varepsilon$, this gives

$$\begin{aligned} \frac{\tau^*}{1 - \tau^*} &= \frac{1}{\varepsilon} \\ \Leftrightarrow \tau^* &= \frac{1}{1 + \varepsilon}. \end{aligned} \tag{25}$$

When the government is concerned about social welfare, its problem is to $\max \text{SWF} = \int \omega^i W(u^i((1 - \tau)z^i + \tau Z(1 - \tau) - R, z^i)) dv(i)$, where use has been made of the individual consumption $c^i = (1 - \tau)z^i + b = (1 - \tau)z^i + \tau Z(1 - \tau) - R$. Here ω is a Pareto weight and W is an increasing and concave transformation of utilities. The FOC $\frac{\partial \text{SWF}}{\partial \tau} = 0$ is:

$$\int \omega^i W_u \left[u_c^i \left(-z^i + (1 - \tau) \frac{\partial z^i}{\partial \tau} + Z + \tau \frac{dZ}{d\tau} \right) + u_z^i \frac{\partial z^i}{\partial \tau} \right] dv(i) = 0,$$

which, using the individual’s envelope condition, becomes:

$$\int \omega^i W_u u_c^i \left(-z^i + Z - \tau \frac{dZ}{d(1 - \tau)} \right) dv(i) = 0.$$

Taking $Z - \tau \frac{dZ}{d(1-\tau)}$ out of the integrand and leaving it to the left-hand side, we have on the right-hand side $\frac{\int \omega^i W_u u_c^i z^i dv(i)}{\int \omega^i W_u u_c^i dv(i)}$. Piketty and Saez define $\beta^i = \frac{\omega^i W_u u_c^i}{\int \omega^i W_u u_c^i dv(i)}$ as a normalized social marginal welfare weight for individual i , so that the term can be simplified to:

$$Z - \tau \frac{dZ}{d(1-\tau)} = \int \beta^i z^i dv(i).$$

Using the definition of aggregate elasticity of earnings and defining $\bar{\beta} = \frac{\int \beta^i z^i dv(i)}{Z}$ as the average normalized social marginal welfare weight, weighted by labor incomes z^i (it can also be interpreted as the ratio of the average income weighted by individual welfare weights β^i to the average income Z), we can rewrite this as $1 - \frac{\tau}{1-\tau} \varepsilon = \bar{\beta}$, which gives the optimal social welfare-maximizing tax rate:

$$\frac{\tau^*}{1-\tau^*} = \frac{1}{\varepsilon} (1 - \bar{\beta}). \tag{26}$$

According to Piketty and Saez, $\bar{\beta}$ “measures where social welfare weights are concentrated on average over the distribution of earnings.” The welfare-maximizing tax rate is thus decreasing in both the average marginal welfare weight and the tax elasticity of aggregate earnings. A higher $\bar{\beta}$ reflects a lower taste for redistribution, and thus a lower desire to tax for redistributive reasons.

Piketty and Saez also note that (26) can be written in the form of $\tau^* = \frac{-\text{cov}(\beta^i, \frac{z^i}{Z})}{-\text{cov}(\beta^i, \frac{z^i}{Z}) + \varepsilon}$.

If higher incomes are valued less (lower β), then the covariances are negative and the tax rate is positive. This is a similar formulation as in [Dixit and Sandmo \(1977\)](#), Eq. (20), where $\tau^* = -\frac{1}{\lambda} \frac{-\text{cov}(z^i, \mu^i)}{\frac{\partial \bar{z}}{\partial (1-\tau)}|_{\text{comp}}}$ (here λ represents the government’s budget constraint

Lagrange multiplier and μ^i the individual’s marginal utility of income, s.t. $u_c = \mu^i$). Here the numerator reflects the equity element and the denominator the efficiency component, similar as in (26).

Non-welfarism

In the non-welfarist case, the Lagrangian function is $L = \sum F(a z^i + b, z^i) + \lambda((1-a) \sum z^i - Nb - R)$. The first-order conditions with respect to a and b are:

$$\sum (F_c(z^i + a z_a^i) + F_z z_a^i) = \lambda (\sum z^i - (1-a) \sum z_a^i) \tag{27}$$

$$\sum (F_c(1 + a z_b^i) + F_z z_b^i) = \lambda (N - (1-a) \sum z_b^i). \tag{28}$$

Dividing the first equation with the second and dividing through the right-hand side with N , we get:

$$\frac{\sum (F_c(z^i + az_a^i) + F_z z_a^i)}{\sum (F_c(1 + az_b^i) + F_z z_b^i)} = \frac{\bar{z} - (1 - a)\bar{z}_a}{1 - (1 - a)\bar{z}_b}, \tag{29}$$

which gives Eq. (3). Minimizing a deprivation index D is a special case of this, such that $F_c = D_c$ and $F_z = 0$. Otherwise the derivation of (5) is analogous to the above.

Let us next derive the poverty-minimizing tax rule following the formulation of Piketty and Saez. Given the government’s instruments, consumption is $c^i = (1 - \tau)z^i + b = (1 - \tau)z^i + \tau Z(1 - \tau) - R$. The poverty minimization objective in the continuous case thus reads:

$$\begin{aligned} \min P &= \int D(c^i, \bar{c}) \, d\nu(i) \\ &= \int D\left((1 - \tau)z^i + \tau Z(1 - \tau) - R, \bar{c}\right) \, d\nu(i). \end{aligned} \tag{30}$$

The optimal tax rate is found from the government’s FOC, $\frac{\partial P}{\partial \tau} = 0$:

$$\begin{aligned} \int D_c \left(-z^i + (1 - \tau) \frac{\partial z^i}{\partial \tau} + Z + \tau \frac{dZ}{d\tau} \right) \, d\nu(i) &= 0 \\ \Leftrightarrow \int D_c \left(-z^i - (1 - \tau) \frac{\partial z^i}{\partial(1 - \tau)} + Z - \tau \frac{dZ}{d(1 - \tau)} \right) \, d\nu(i) &= 0. \end{aligned} \tag{31}$$

Define a “normalized marginal deprivation weight” as $\beta^i = \frac{D_c}{\int D_c d\nu(j)}$. Using this definition, $\left(Z - \tau \frac{dZ}{d(1 - \tau)} \right) \int D_c \, d\nu(i) = \int D_c \left(z^i + (1 - \tau) \frac{\partial z^i}{\partial(1 - \tau)} \right) \, d\nu(i)$ can be written as:

$$Z - \tau \frac{dZ}{d(1 - \tau)} = \int \beta^i \left(z^i + (1 - \tau) \frac{\partial z^i}{\partial(1 - \tau)} \right) \, d\nu(i). \tag{32}$$

Using the definition of the elasticity of individual labor earnings $\varepsilon^i = \frac{1 - \tau}{z^i} \frac{\partial z^i}{\partial(1 - \tau)}$, we have $(1 - \tau) \frac{\partial z^i}{\partial(1 - \tau)} = z^i \varepsilon^i$ and using elasticity of aggregate earnings $\varepsilon = \frac{1 - \tau}{Z} \frac{dZ}{d(1 - \tau)}$, we have $Z - \tau \frac{dZ}{d(1 - \tau)} = 1 - \frac{\tau}{1 - \tau} \varepsilon$ and we can rewrite the above as:

$$Z \left(1 - \frac{\tau}{1 - \tau} \varepsilon \right) = \int \beta^i \left(z^i + z^i \varepsilon^i \right) \, d\nu(i). \tag{33}$$

This leads to the poverty-minimizing rule of

$$\frac{\tau^*}{1 - \tau^*} = \frac{1}{\varepsilon} (1 - \bar{\beta} - \bar{\beta}^\varepsilon), \tag{34}$$

where analogously to Piketty–Saez, $\bar{\beta} = \frac{\int \beta^i z^i dv(i)}{Z} \left(= \frac{\int D_c z^i dv(i)}{Z \int D_c dv(j)} \right)$ is an average normalized deprivation weight, weighted by labor incomes (or, analogously, average labor income weighted by individual deprivation weights). In addition, we have defined $\bar{\beta}^\varepsilon = \frac{\int \beta^i z^i \varepsilon^i dv(i)}{Z} \left(= \frac{\int D_c z^i \varepsilon^i dv(i)}{Z \int D_c dv(j)} \right)$, which describes average labor incomes weighted by their corresponding individual elasticities and deprivation weights. This can be interpreted as a combined deprivation and efficiency effect.

As in the welfarist setting, the more elastic average earnings are to taxation, the lower is the optimal tax rate (a regular efficiency effect). The optimal poverty-minimizing tax rate is decreasing in the average deprivation weight $\bar{\beta}$, as a higher taste for redistribution toward the materially deprived implies a lower $\bar{\beta}$ and thus higher taxation for redistributive purposes. The effect is analogous to the welfarist tax rate, of course with slightly different definitions for $\bar{\beta}$.

The new term $\bar{\beta}^\varepsilon$ can be interpreted as a combined deprivation weight and efficiency effect. The elasticity term implicit in $\bar{\beta}^\varepsilon$ takes into account the incentive effects of taxation on working and works to reduce τ^* . To avoid discouraging the poor from working, their tax rates should be lower. But because the tax instrument is forced to be linear, tax rates are then lowered for everyone, as we found in the Tuomala model in Eq. (5). The value of $\bar{\beta}^\varepsilon$ depends on the relationship of the individual earnings elasticities and income: if the elasticity is the same across income levels, there is just a level effect moving from $\bar{\beta}$ to $\bar{\beta}^\varepsilon$; however, if the elasticity were higher for more deprived individuals, for example, $\bar{\beta}^\varepsilon$ would most likely be higher than under a flat elasticity. This works toward a lower tax rate in order to avoid discouraging the poorest from working. However, whether $\bar{\beta}^\varepsilon$ is high or low does not depend only on the shape of the elasticity but also on the shape of the deprivation weights, which also affect $\bar{\beta}$.

Finally, the third way for expressing the optimal tax rule in the case of poverty minimization is one following the [Dixit and Sandmo \(1977\)](#) formulation and it can be written as

$$\tau^* = -\frac{1}{\lambda} \frac{\text{cov}(D_c, z^i) + \frac{1}{N} \sum D_c a \tilde{z}_a^i + \text{cov}(D_c a \tilde{z}_b^i, z^i)}{\frac{1}{N} \sum \tilde{z}_a^i}. \quad (35)$$

In this expression, the denominator is the same as in Eq. (20) of [Dixit and Sandmo \(1977\)](#) presented before, that is, the average derivative of compensated labor supply with respect to the net-of-tax rate. In the numerator, the first term measures the strength of the association between income and poverty impact: when the association between overall poverty and small income is strong (this would be the case with the squared poverty gap), the tax should be high so that it will finance a sizable lump-sum transfer. If the association is weaker (as with the headcount rate), the tax rate is optimally smaller. The second and the third terms in the numerator are new. They measure the indirect effects from changes in the tax rate on labor supply. Here \tilde{z} is the compensated (Hicksian) labor supply. The greater is the reduction in the labor supply following an increase in the tax rate (it is the compensated change as the tax increase is linked with a simultaneous increase in the lump-sum transfer), the smaller should the tax rate be in order to avoid increases in deprivation arising from lower earned income. The last two terms in the numerator are closely linked with a formulation $D_c(1 - \tau) \frac{\partial \tilde{z}}{\partial q} |_{\text{comp}}$,

where the idea is that the last covariance term serves as a corrective device for the mean impact of taxes on labor supply (similarly as in the denominator in the original Dixit–Sandmo formulation).

Public good provision

Welfarism

The Lagrangian is $L = \sum W(V^i(a, b, G)) + \lambda((1 - a) \sum z^i - Nb - N\pi G - R)$. Maximizing the Lagrangian with respect to b and G gives:

$$\sum \beta^i = \lambda \left(N - (1 - a) \sum z_b^i \right) \tag{36}$$

$$\sum W_V V_G^i = \lambda \left(N\pi - (1 - a) \sum z_G^i \right). \tag{37}$$

Dividing (37) by (36) we obtain

$$\frac{\sum \beta^i \sigma^i}{\sum \beta^i} = \frac{\pi - (1 - a)\bar{z}_G}{1 - (1 - a)\bar{z}_b}, \tag{38}$$

where we define $\sigma^* = \frac{\sum \beta^i \sigma^i}{\sum \beta^i}$ to be the welfare-weighted average marginal rate of substitution between public good and income for individual i . Rewriting this rule gives Eq. (6) in the main text.

Extending the Piketty and Saez approach to include public provision, the government’s goal function is

$$SWF = \int \omega^i W \left(u^i \left((1 - \tau) z^i + \tau Z((1 - \tau), G) - R - \pi G, G, z^i \right) \right) dv(i).$$

The FOC for τ is as before, and the FOC for public good provision G is

$$\int \omega^i W_u \left(u_G^i + u_x^i \left((1 - \tau) \frac{\partial z^i}{\partial G} + \tau \frac{dZ}{dG} - \pi \right) \right) dv(i) = 0,$$

which produces the following public good provision rule:

$$\frac{\int \omega^i W_u \left(u_G^i + u_x^i (1 - \tau) \frac{\partial z^i}{\partial G} \right) dv(i)}{\int \omega^i W_u u_x^i dv(i)} = \pi - \tau \frac{dZ}{dG}. \tag{39}$$

The left-hand side relates the welfare gains of public good provision (a direct (u_G) and indirect effect ($u_x(1 - \tau) \frac{\partial z^i}{\partial G}$ via labor supply reactions)) to the welfare gains of directly increasing consumption (cash transfers) and the right-hand side relates the

costs of providing the public good (both its price and the effect it has on tax revenue) to the costs of directly increasing consumption (equal to 1 in this model).¹⁸

Poverty minimization

Using Tuomala’s model, and the deprivation index $D(x, G, \bar{x}, \bar{G})$ defined over consumption of the public good G and other private consumption x , we can divide the government’s first-order condition for G (analogous to Eq. 37) with that of b (analogous to Eq. 36) to get the following relationship:

$$D^* = \frac{\pi - (1 - a)\bar{z}_G}{1 - (1 - a)\bar{z}_b}, \tag{40}$$

where $D^* = \frac{\sum D_G + \sum D_x a z^i_G}{\sum D_x (1 + a z^i_b)}$. This can be rewritten to get Eq. (7).

In the Piketty–Saez type of model, individual private consumption is $x = (1 - \tau)z^i + b = (1 - \tau)z^i + \tau Z((1 - \tau), G) - R - \pi G$. The government’s problem is then:

$$\begin{aligned} \min P &= \int D(x^i, G, \bar{x}, \bar{G}) \, d\nu(i) \\ &= \int D((1 - \tau)z^i + \tau Z((1 - \tau), G) - R - \pi G, G, \bar{x}, \bar{G}) \, d\nu(i). \end{aligned} \tag{41}$$

The first-order condition for optimal tax τ is unchanged, and the FOC for public good provision is $\int [D_G + D_x ((1 - \tau)\frac{\partial z^i}{\partial G} + \tau \frac{dZ}{dG} - \pi)] \, d\nu(i) = 0$, which gives the public provision rule of (9).

The poverty minimization problem in the case of provision of a quasi-private good is

$$\begin{aligned} \min P &= \int D(x^i, s^i, \bar{x}, \bar{s}) \, d\nu(i) \\ &= \int D((1 - \tau)z^i + \tau Z((1 - \tau), G) - R - \pi G - ph^i, s^i, \bar{x}, \bar{s}) \, d\nu(i). \end{aligned} \tag{42}$$

The FOC for public good provision G is $\int [D_x ((1 - \tau)\frac{\partial z^i}{\partial s} \frac{\partial s}{\partial G} + \tau \frac{dZ}{dG} - \pi - p \frac{\partial h^i}{\partial G}) + D_s \frac{\partial s^i}{\partial G}] \, d\nu(i) = 0$, which gives the public provision rule (10).

¹⁸ In Eq. (39), we could define a normalized marginal social welfare weight, similar as before, $\beta^i = \frac{\omega^i W_{uu} u^i_x}{\int \omega^j W_{uu} u^j_x \, d\nu(j)}$ to get $\frac{\int \omega^j W_{uu} u^j_G \, d\nu(j)}{\int \omega^j W_{uu} u^j_x \, d\nu(j)} + \int \beta^i (1 - \tau) \frac{\partial z^i}{\partial G} \, d\nu(i) = \pi - \tau \frac{dZ}{dG}$.

Commodity taxation

Welfarism

The Lagrangian of the government’s optimization problem is the following:

$$L = \sum_i W \left(V^i(b, q) \right) + \lambda \left(\sum_i \sum_j t_j x_j^i - Nb - R \right). \tag{43}$$

The first-order conditions with respect to b and q_k are:

$$\sum_i \beta^i + \lambda \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b} - \lambda N = 0 \tag{44}$$

$$- \sum_i \beta^i x_k^i + \lambda \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial q_k} + \lambda \sum_i x_k^i = 0, \tag{45}$$

where Roy’s identity has been used in (45), i.e., $\frac{\partial V^i}{\partial q_k} = -\frac{\partial V^i}{\partial b} x_k^i$. Using the definition of γ^i , this means that (44) can be rewritten as

$$\frac{\sum_i \gamma^i}{N} = \lambda, \tag{46}$$

implying that the average net social marginal utility of income must equal the shadow price of budget revenues at the optimum. Next use the definition of γ^i and the Slutsky equation for the commodity demand

$$\frac{\partial x_j^i}{\partial q_k} = \frac{\partial \tilde{x}_j^i}{\partial q_k} - x_k^i \frac{\partial x_j^i}{\partial b},$$

where \tilde{x}_j^i denotes the compensated (Hicksian) demand for good x_j^i , in (45), to get

$$\sum_i \sum_j t_j \frac{\partial \tilde{x}_j^i}{\partial q_k} = \frac{1}{\lambda} \sum_i \left(\gamma^i - \lambda \right) x_k^i. \tag{47}$$

The covariance between γ^i and the demand of the good x_k can be written as (using (46))

$$\text{cov} \left(\gamma^i, x_k^i \right) = \frac{\sum_i \gamma^i x_k^i}{N} - \frac{\sum_i \gamma^i}{N} \frac{\sum_i x_k^i}{N} = \frac{\sum_i \gamma^i x_k^i}{N} - \lambda \frac{\sum_i x_k^i}{N}.$$

Using Slutsky symmetry, Eq. (47) can therefore be written as a covariance rule (13).

Poverty minimization

The deprivation index to be minimized is $D\left(\sum_j q_j x_j^i, \bar{c}\right)$. The first-order conditions with respect to b and q_k are:

$$\sum_i D_c \sum_j q_j \frac{\partial x_j^i}{\partial b} + \lambda \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b} - \lambda N = 0 \tag{48}$$

$$\sum_i D_c x_k^i + \sum_i D_c \sum_j q_j \frac{\partial x_j^i}{\partial q_k} + \lambda \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial q_k} + \lambda \sum_i x_k^i = 0. \tag{49}$$

Using the Slutsky equation in Eq. (49) and dividing by N leads to

$$\begin{aligned} & \frac{1}{N} \sum_i D_c x_k^i + \frac{1}{N} \sum_i D_c \sum_j q_j \left(\frac{\partial \bar{x}_j^i}{\partial q_k} - x_k^i \frac{\partial x_j^i}{\partial b} \right) \\ & + \frac{\lambda}{N} \sum_i \sum_j t_j \left(\frac{\partial \bar{x}_j^i}{\partial q_k} - x_k^i \frac{\partial x_j^i}{\partial b} \right) + \frac{\lambda}{N} \sum_i x_k^i = 0. \end{aligned} \tag{50}$$

Multiplying Eq. (48) by $\frac{\sum_i x_k^i}{N^2}$ and adding it with Eq. (50) gives

$$\begin{aligned} & \frac{1}{N} \sum_i D_c x_k^i + \frac{1}{N} \sum_i D_c \sum_j q_j \frac{\partial \bar{x}_j^i}{\partial q_k} - \frac{1}{N} \sum_i \sum_j D_c q_j x_k^i \frac{\partial x_j^i}{\partial b} \\ & + \frac{1}{N} \frac{\sum_i D_c}{N} \sum_j q_j \frac{\partial x_j^i}{\partial b} \sum_i x_k^i + \frac{\lambda}{N} \sum_i \sum_j t_j \frac{\partial \bar{x}_j^i}{\partial q_k} \\ & - \frac{\lambda}{N} \sum_i \sum_j t_j x_k^i \frac{\partial x_j^i}{\partial b} + \frac{1}{N} \frac{\lambda}{N} \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b} \sum_i x_k^i = 0. \end{aligned} \tag{51}$$

Noticing that the covariance of γ_p^i and x_k^i can be written as $\frac{1}{N} \sum_i \sum_j D_c q_j x_k^i \frac{\partial x_j^i}{\partial b} + \frac{\lambda}{N} \sum_i \sum_j t_j x_k^i \frac{\partial x_j^i}{\partial b} - \frac{1}{N} \frac{\sum_i D_c}{N} \sum_j q_j \frac{\partial x_j^i}{\partial b} \sum_i x_k^i - \frac{1}{N} \frac{\lambda}{N} \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b} \sum_i x_k^i$, the rule above can be written as Eq. (15) in the main text.

Non-optimality of uniform commodity taxation

We demonstrate formally how uniform commodity taxation is not optimal in the case of poverty minimization. To see this, rewrite first the FOC with respect to b (Eq. 48) as

$$\frac{1 - \frac{1}{N} \sum_i \sum_j t_j \frac{\partial x_j^i}{\partial b}}{\frac{1}{N} \sum_i D_c \sum_j q_j \frac{\partial x_j^i}{\partial b}} = \frac{1}{\lambda}. \tag{52}$$

Next, rewriting the FOC for q_k (Eq. 50) yields

$$\begin{aligned} \frac{1}{N} \sum_i \sum_j t_j \frac{\partial \tilde{x}_k^i}{\partial q_j} &= -\frac{1}{\lambda} \frac{1}{N} \sum_i D_c x_k^i - \frac{1}{\lambda} \frac{1}{N} \sum_i D_c \sum_j q_j \frac{\partial \tilde{x}_k^i}{\partial q_j} \\ &+ \frac{1}{\lambda} \frac{1}{N} \sum_i D_c \sum_j q_j \frac{\partial x_j^i}{\partial b} x_k^i + \frac{1}{N} \sum_i \left(\sum_j t_j \frac{\partial x_j^i}{\partial b} - 1 \right) x_k^i. \end{aligned} \tag{53}$$

Here we can substitute for $\frac{1}{\lambda}$ from Eq. (52) in the first term at the lower row of Eq. (53). Following Deaton (1979, pp. 359–360), when preferences are separable and Engel curves are linear, demand is written as $x_j^i = \delta_j^i(q) + \theta_j(q)c^i$; hence, the derivative of demand with respect to disposable income c or transfer b is $\theta_j(q)$, i.e., independent of the person i . By writing out explicitly the solution that the derivative of demand w.r.t b is independent of i and write $\frac{\partial x_j^i}{\partial b} = \theta_j(q)$, we have:

$$\begin{aligned} \frac{1}{N} \sum_i \sum_j t_j \frac{\partial \tilde{x}_k^i}{\partial q_j} &= -\frac{1}{\lambda} \frac{1}{N} \sum_i D_c x_k^i - \frac{1}{\lambda} \frac{1}{N} \sum_i D_c \sum_j q_j \frac{\partial \tilde{x}_k^i}{\partial q_j} \\ &+ \frac{1 - \frac{1}{N} \sum_i \sum_j t_j \theta_j(q)}{\sum_i D_c \left(\sum_j q_j \theta_j(q) \right)} \sum_i D_c x_k^i \left(\sum_j q_j \theta_j(q) \right) \\ &+ \frac{1}{N} \sum_i x_k^i \left(\sum_j t_j \theta_j(q) - 1 \right), \end{aligned} \tag{54}$$

where in the second row we can cancel out the $\sum_j q_j \theta_j(q)$ terms and rewrite $\sum_i \sum_j t_j \theta_j(q) = N \sum_j t_j \theta_j(q)$ in the numerator because the term is independent over i :

$$\begin{aligned} \frac{1}{N} \sum_i \sum_j t_j \frac{\partial \tilde{x}_k^i}{\partial q_j} &= -\frac{1}{\lambda} \frac{1}{N} \sum_i D_c x_k^i - \frac{1}{\lambda} \frac{1}{N} \sum_i D_c \sum_j q_j \frac{\partial \tilde{x}_k^i}{\partial q_j} \\ &+ \frac{1 - \sum_j t_j \theta_j(q)}{\sum_i D_c} \sum_i D_c x_k^i + \frac{1}{N} \sum_i x_k^i \left(\sum_j t_j \theta_j(q) - 1 \right). \end{aligned} \tag{55}$$

Note next that due to homogeneity of degree 0 of compensated demand, $\sum_j q_j \frac{\partial \bar{x}_k^i}{\partial q_j} + w_i \frac{\partial \bar{x}_k^i}{\partial w_j} = 0$. This, together with the observation that if a uniform commodity tax t was a solution to the problem at hand, this would mean that the left-hand side of (53) could be written as $-\frac{t}{N} \sum_i w_i \frac{\partial \bar{x}_k^i}{\partial w}$. Because of separability, the substitution response is linked to the full income derivative, so that $\frac{\partial \bar{x}_k^i}{\partial w} = \phi^i \theta_j(q)$. Because of these arguments, (53) becomes

$$-\frac{t}{N} \theta_j(q) \sum_i w_i \phi^i = -\frac{1}{\lambda} \frac{1}{N} \sum_i D_c x_k^i - \frac{1}{\lambda} \frac{1}{N} \theta_j(q) \sum_i D_c w_i \phi^i + \frac{1-t}{\sum_i D_c} \sum_j \theta_j(q) \sum_i D_c x_k^i + \frac{1}{N} \sum_i x_k^i \left(t \sum_j \theta_j(q) - 1 \right). \tag{56}$$

Note that terms incorporating $\theta_j(q)$ cannot be canceled out from the equation so the result remains dependent on j . In addition, even if the terms were canceled, the term $\sum_i D_c \frac{x_k^i}{N}$ still depends on j . This shows that uniform commodity taxation is not optimal when the objective function of the government is to minimize poverty.

Optimal income taxation with an informal sector

Welfarism

The welfarist Lagrangian, in the presence of informality, is $L = \sum W (V^i(a, b, e)) + \lambda((1 - a) \sum (z^i - e^i) - Nb - R)$. We can denote the effective tax base as $z^e = z - e$. The derivative of this tax base with respect to tax rate a is denoted $z_a^e = z_a - \frac{\partial e}{\partial a}$, where we assume $\frac{\partial e}{\partial a} < 0$ (whereas $\frac{\partial e}{\partial b} = 0$). The first-order conditions with respect to a and b are:

$$\sum W_V V_a^e = \lambda \left(\sum z^e - (1 - a) \sum z_a^e \right) \\ \sum W_V V_b = \lambda \left(N - (1 - a) \sum z_b \right),$$

where V_a^e is a shorthand for the derivative of the indirect utility function that takes individual evasion behavior into account. Should there be no evasion, the individual would maximize her utility over income $az + b$ and $V_a = \lambda z$. Under evasion, consumption is $a(z - e) + e + b$ and, by the envelope theorem, $V_a^e = \lambda(z - e) = \lambda z^e$. Roy’s theorem adapts in this case to: $V_a^e = V_b z^e$, and welfare-weighted average income can be denoted as $z^e(\beta) = \frac{\sum \beta^i z^{e,i}}{\sum \beta}$. The ratio of the first-order conditions is:

$$z^e(\beta) = \frac{\bar{z}^e - (1 - a)\bar{z}_a^e}{1 - (1 - a)\bar{z}_b},$$

and we can derive the optimal tax rate by following the same steps as in the model without evasion, by considering the evasion-modified tax base z^e instead of z :

$$\frac{\tau^*}{1 - \tau^*} = \frac{1}{\varepsilon^e} \left(1 - \frac{z^e(\beta)}{\bar{z}^e} \right).$$

The intuition behind the derivation and the tax rule is the same as before, but we must consider the relevant tax base in the context of evasion. Both the elasticity of labor income with respect to the tax rate and the relevant welfare concepts change when part of the income base evades taxation.

Poverty minimization

The derivation of Eq. (17) follows the same steps as presented above and in the poverty-minimization model without evasion. The first-order conditions with respect to a and b are:

$$\begin{aligned} \sum D_c \left(\frac{\partial e^i}{\partial a} + z^e + az_a^e \right) &= \lambda \left(\sum z^e - (1 - a) \sum z_a^e \right) \\ \sum D_c (1 + az_b^i) &= \lambda \left(N - (1 - a) \sum z_b^i \right). \end{aligned}$$

From the ratio of the two conditions we get the measure of relative deprivation impact under tax evasion, \tilde{D}^e :

$$\tilde{D}^e \equiv \frac{\sum D_c (z^e + az_a^e + \frac{\partial e}{\partial a})}{D_c (1 + az_b^i)} = \frac{\bar{z}^e - (1 - a)\bar{z}_a^e}{1 - (1 - a)\bar{z}_b}$$

which gives us Eq. (17) in the text. \tilde{D}^e measures the relative efficiency of taxes and transfers. The latter impact (the denominator) is the same as before, but the impact of taxation (numerator) is different in the presence of tax evasion.

Appendix 2: Measuring poverty

One of the most popular poverty measures is the P_α category developed by Foster et al. (1984). It is usually written in the form of $P_\alpha = \int_0^z \left(\frac{z-y}{z} \right)^\alpha f(y) d(y)$ where z is the poverty line and y is income. Defining the poverty index in terms of disposable income (as in Kanbur and Keen 1989 for example), the measure becomes: $P_\alpha = \int_0^{\bar{c}} \left(\frac{\bar{c}-c^i}{\bar{c}} \right)^\alpha dv(i)$. c^i is disposable income, which is defined in the Piketty–Saez model as $c^i = (1 - \tau)z^i + b = (1 - \tau)z^i + \tau Z(1 - \tau) - R$. We can use this specification of the functional form of deprivation to define the derivative $D_c = -\frac{\alpha}{\bar{c}} \left(\frac{\bar{c}-c^i}{\bar{c}} \right)^{\alpha-1}$ (note that $D_c < 0$ as long as $c^i < \bar{c}$). We can follow the same steps to arrive at the optimal tax rate $\tau^* = \frac{1-\beta-\beta^e}{1-\beta-\beta^e+\varepsilon}$ where now

$$\beta^i = \frac{D_c}{\int_0^{\bar{c}} D_c \, dv(i)} = \frac{-\frac{\alpha}{\bar{c}} \left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1}}{\int_0^{\bar{c}} -\frac{\alpha}{\bar{c}} \left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1} \, dv(i)} = \frac{\left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1}}{\int_0^{\bar{c}} \left(\frac{\bar{c}-c^i}{\bar{c}}\right)^{\alpha-1} \, dv(i)}$$

and consequently $\bar{\beta} = \frac{\int_0^{\bar{c}} \beta^i z^i \, dv(i)}{Z}$ and $\bar{\beta}^\varepsilon = \frac{\int_0^{\bar{c}} \beta^i z^i \varepsilon^i \, dv(i)}{Z}$, as before. Everything else stays exactly the same as in the calculations of Appendix 1. Also in the case of Tuomala's and Dixit and Sandmo's models, the results stay the same, and we can plug in the explicit definition for D_c , the derivative of the poverty measure with respect to disposable income, into the results.

Poverty measurement in the context of public good provision

Employing the FGT poverty measure in the context of public good provision for poverty reduction is more complicated than in the case of just disposable income. In Sect. 3.2 the government's objective function was defined as $\min P = \int D(x^i, G, \bar{x}, \bar{G}) \, dv(i)$, that is, deprivation was measured both as deprivation in private consumption (i.e., disposable income) as well as with respect to the public good. But the FGT index is a uni-dimensional measure, measuring deprivation with respect to one dimension only (e.g., disposable income). If one wants to consider publicly offered goods such as education as separate from private consumption, a multidimensional FGT measure is needed. Multidimensionality, however, entails a difficult question of determining when a person should be determined as deprived.

There are several approaches to multidimensionality of FGT-type poverty measures.¹⁹ For example, [Besley and Kanbur \(1988\)](#), who consider the poverty impacts of food subsidies, employ the uni-dimensional FGT measure but define deprivation in terms of equivalent income: $P_\alpha = \int_0^z \left(\frac{z_E - y_E}{z_E}\right)^\alpha f(y) \, d(y)$, where y_E is equivalent income, defined implicitly from $V(p, y_E) = V(q, y)$, and z_E is the poverty line corresponding to equivalent income. But given our aim of defining optimal policy in terms of poverty reduction, irrespective of individual welfare, the use of equivalent income is problematic as it forces the solution to be such that, by definition, individuals are kept as well off as before. [Pirttilä and Tuomala \(2004\)](#) employ shadow prices in a poverty-minimizing context to allow for several goods in the poverty measure. For them, deprivation is measured as $D(z, y(q, w))$ where $z^h = s_x x^* - s_L^h L^*$ and $y^h(q, w^h) = s_x x(q, w^h) - s_L^h L(q, w^h)$. This approach requires determining shadow prices s_x, s_L for consumption and leisure in order to construct a reference bundle respective to which deprivation can be measured, but there is no clear guideline to the choice of the shadow prices.

The approach in [Bourguignon and Chakravarty \(2003\)](#) is more suitable for our purposes. They provide a multidimensional extension of the FGT measure, according to which a person is poor if she is deprived in at least one dimension. A simple example of such an extension of the FGT is

¹⁹ See [Foster et al. \(2010\)](#), pp. 504–5 for a brief overview of multidimensional FGT extensions that allow the inclusion of dimensions such as health, education, and nutrition in addition to other consumption.

$$P_\theta = \frac{1}{n} \sum_{j=1}^m \sum_{i \in S_j} a_j \left(\frac{z_j - x_{ij}}{z_j} \right)^{\theta_j},$$

where θ_j and a_j are weights given to dimension j , and S_j is the group of people who are poor in dimension j . Alkire and Foster (2011) for their part provide a similar measure which uses a weighted count of dimensions in which the person is deprived to determine whether she is poor. An aspect of this is also whether the goods under consideration are complements or substitutes. Following the Bourguignon–Chakravarty approach and defining $x_{i1} = x_i$ as private consumption, $z_1 = \bar{x}$, $x_{i2} = G$ as the amount of public good, and $z_2 = \bar{G}$ would give us $P_\theta = \frac{1}{n} \sum_{i \in S_j} \left(a_1 \left(\frac{\bar{x} - x^i}{\bar{x}} \right)^{\theta_1} + a_2 \left(\frac{\bar{G} - G}{\bar{G}} \right)^{\theta_2} \right)$. Using this measure, $D_x = -\frac{\theta_1 a_1}{\bar{x}} \left(\frac{\bar{x} - x^i}{\bar{x}} \right)^{\theta_1 - 1}$ and $D_G = -\frac{\theta_2 a_2}{\bar{G}} \left(\frac{\bar{G} - G}{\bar{G}} \right)^{\theta_2 - 1}$. These can then be inserted to the public provision rules. For example, (9) becomes

$$\frac{\int \left(\frac{\theta_2 a_2}{\bar{G}} \left(\frac{\bar{G} - G}{\bar{G}} \right)^{\theta_2 - 1} + \frac{\theta_1 a_1}{\bar{x}} \left(\frac{\bar{x} - x^i}{\bar{x}} \right)^{\theta_1 - 1} (1 - \tau) \frac{\partial z^i}{\partial G} \right) dv(i)}{\int \frac{\theta_1 a_1}{\bar{x}} \left(\frac{\bar{x} - x^i}{\bar{x}} \right)^{\theta_1 - 1} dv(i)} = p - \tau \frac{dZ}{dG}$$

and (40) becomes

$$\frac{\sum \frac{\theta_2 a_2}{\bar{G}} \left(\frac{\bar{G} - G}{\bar{G}} \right)^{\theta_2 - 1} + \sum \frac{\theta_1 a_1}{\bar{x}} \left(\frac{\bar{x} - x^i}{\bar{x}} \right)^{\theta_1 - 1} a z_G^i}{\sum \frac{\theta_1 a_1}{\bar{x}} \left(\frac{\bar{x} - x^i}{\bar{x}} \right)^{\theta_1 - 1} (1 + a z_b^i)} = \frac{p - (1 - a) \sum z_G^i}{1 - (1 - a) \sum z_b^i},$$

from where it can be seen that the relative efficiency of the public good versus cash transfers on reducing poverty can be directly traced back to the magnitudes of θ_1 and θ_2 .

Poverty measurement in the context of commodity taxation

In the case of commodity taxes, we run into the same issues regarding deprivation measurement as with public goods. However, in Sect. 4.2 deprivation was measured only in terms of disposable income, c . We thus escape the multidimensionality issue and employing the FGT poverty measure is thus as simple as in the linear income tax case: we simply need to define $D = P^\alpha$ and thus $D_c = -\frac{\alpha}{c} \left(\frac{\bar{c} - c^i}{\bar{c}} \right)^{\alpha - 1}$ in Eq. (15). Potentially the government might also consider weighting different goods according to their importance to measured poverty.

References

Aitchison, J., & Brown, J. A. C. (1957). *The lognormal distribution*. Cambridge: Cambridge University Press.

- Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95, 476–487.
- Barrientos, A. (2013). *Social assistance in developing countries*. Cambridge: Cambridge University Press.
- Besley, T., & Kanbur, R. (1988). Food subsidies and poverty alleviation. *The Economic Journal*, 98, 701–719.
- Besley, T., & Persson, T. (2013). Taxation and development. In A. J. Auerbach, R. Chetty, M. Feldstein, & E. Saez (Eds.), *Handbook of public economics* (Vol. 5, pp. 51–110). Amsterdam: Elsevier.
- Bird R.M., & Gendron, P. (2007). *The VAT in developing and transitional countries*. Cambridge: Cambridge University Press.
- Boadway, R., Marchand, M., & Pestieau, P. (1994). Towards a theory of the direct–indirect tax mix. *Journal of Public Economics*, 55, 71–88.
- Bourguignon, F., & Chakravarty, S. (2003). The measurement of multidimensional poverty. *Journal of Economic Inequality*, 1, 25–49.
- Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105, 1–33.
- Deaton, A. (1979). Optimally uniform commodity taxes. *Economics Letters*, 2, 357–361.
- Diamond, P. (1975). A many-person Ramsey tax rule. *Journal of Public Economics*, 4, 335–342.
- Dixit, A., & Sandmo, A. (1977). Some simplified formulae for optimal income taxation. *Scandinavian Journal of Economics*, 79, 417–423.
- Farhi, E., & Gabaix, X. (2015). Optimal taxation with behavioral agents. Working Paper 21524, National Bureau of Economic Research.
- Fleurbaey, M., & Maniquet, F. (2007). Help the low skilled or let the hardworking thrive? A study of fairness in optimal income taxation. *Journal of Public Economic Theory*, 9, 467–500.
- Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52, 761–766.
- Foster, J., Greer, J., & Thorbecke, E. (2010). The Foster-Greer-Thorbecke (FGT) poverty measures: 25 years later. *Journal of Economic Inequality*, 8, 491–524.
- Gerritsen, A. (2016). Optimal taxation when people do not maximize well-being. *Journal of Public Economics*, 144, 122–139.
- Gordon, R., & Li, W. (2009). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics*, 93, 855–866.
- IFS & Mirrlees, J. (2011). *Tax by design*. Oxford: Oxford University Press.
- Kanbur, R. (2015). *Informality: Causes, consequences and policy responses*. Discussion Paper 10509, Centre for Economic Policy Research.
- Kanbur, R., & Keen, M. (1989). Poverty, incentives and linear income taxation. In A. Dilnot & I. Walker (Eds.), *The economics of social security* (pp. 99–115). Oxford: Clarendon Press.
- Kanbur, R., & Keen, M. (2014). Thresholds, informality, and partitions of compliance. *International Tax and Public Finance*, 21, 536–559.
- Kanbur, R., Keen, M., & Tuomala, M. (1994). Optimal non-linear income taxation for the alleviation of income-poverty. *European Economic Review*, 98, 701–719.
- Kanbur, R., Pirttilä, J., & Tuomala, M. (2006). Non-welfarist optimal taxation and behavioural public economics. *Journal of Economic Surveys*, 20, 849–868.
- Keen, M. (2009). What do (and don't) we know about the value added tax? A review of Richard M. Bird and Pierre-Pascal Gendron's *The VAT in developing and transitional countries*. *Journal of Economic Literature*, 47, 159–170.
- Keen, M. (2012). Taxation and development: Again. Working Paper 12/220, International Monetary Fund.
- Keen, M. (2014). Targeting, cascading and indirect tax design. *Indian Growth and Development Review*, 7, 181–201.
- Keen, M., Kim, Y., & Varsano, R. (2008). The “flat tax(es)”: Principles and experience. *International Tax and Public Finance*, 15, 712–751.
- Lee, D., & Saez, E. (2012). Optimal minimum wage policy in competitive labor markets. *Journal of Public Economics*, 96, 739–749.
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38, 175–208.
- Peichl, A. (2014). Flat-rate tax systems and their effect on labor markets. *IZA World of Labor*, 61.
- Piketty, T., & Saez, E. (2013). Optimal labor income taxation. In A. J. Auerbach, R. Chetty, M. Feldstein, & E. Saez (Eds.), *Handbook of Public Economics* (Vol. 5, pp. 391–474). Amsterdam: Elsevier.

- Pirttilä, J., & Tuomala, M. (2004). Poverty alleviation and tax policy. *European Economic Review*, 48, 1075–1090.
- Roemer, J. E., Aaberge, R., Colombino, U., Fritzell, J., Jenkins, S. P., Lefranc, A., et al. (2003). To what extent do fiscal regimes equalize opportunities for income acquisition among citizens? *Journal of Public Economics*, 87, 539–565.
- Saez, E., & Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, 106, 24–45.
- Sen, A. (1985). *Commodities and capabilities*. Amsterdam: North Holland.
- Slemrod, J. (1990). Optimal taxation and optimal tax systems. *Journal of Economic Perspectives*, 4, 157–178.
- Slemrod, J., & Gillitzer, C. (2014). Insights from a tax-systems perspective. *CESifo Economic Studies*, 60, 1–31.
- Stern, N. (1976). On the specification of models of optimum income taxation. *Journal of Public Economics*, 6, 123–62.
- Tuomala, M. (1985). Simplified formulae for optimal linear income taxation. *Scandinavian Journal of Economics*, 87, 668–672.
- Tuomala, M. (1990). *Optimal income tax and redistribution*. Oxford: Oxford University Press.