Online Appendix: The Marginal Cost of Public Funds is One at the Optimal Tax System

Bas Jacobs*

December 18, 2017

A Marginal cost of public funds in terms of compensating variations

This Appendix shows that the traditional and Diamond-based marginal cost of funds can be written as a function of the compensating variation and the uncompensated and compensated revenue changes of a marginal tax change. The proof assumes that the consumption tax is normalized to to zero. Moreover, the government has no preference for income redistribution ($\xi_t^s = 0$).

Under the standard approach, the first-order condition of the marginal tax rate can be written as (using $\xi_t^s = 0$):

$$MCF_t^s = \frac{\eta}{\int_{\mathcal{N}} \lambda(n)\mathrm{d}F(n)} = \frac{1}{1 + \frac{t}{1-t}\bar{\varepsilon}_{lt}^u}. \tag{44}$$

By using the properties of the indirect utility function $v(t, \tau, g, G, n)$, the compensating variation $CV(n)$ for individual n due to a marginal tax increase $\mathrm{d}t$ as (at $\tau = 0$) can be written as:

$$CV(n) = -z(n)\mathrm{d}t. \tag{45}$$

The total compensating variation is given by:

$$CV \equiv \int_{\mathcal{N}} CV(n)\mathrm{d}F(n). \tag{46}$$

From the government budget constraint follows the uncompensated change in revenue (note that $\tau = 0$):

$$R \equiv \int_{\mathcal{N}} tz(n)\mathrm{d}F(n) - g - pG/N, \tag{47}$$

$$\mathrm{d}R = \left(1 + \frac{t}{1-t}\bar{\varepsilon}_{lt}^u\right)\int_{\mathcal{N}} z(n)\mathrm{d}F(n)\mathrm{d}t. \tag{48}$$

*Bas Jacobs, Erasmus University Rotterdam, Tinbergen Institute, and CESifo. Address: Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Phone: +31–10–4081452/1441. Fax: +31–10–4089161. E-mail: bjacobs@ese.eur.nl. Homepage: http://personal.eur.nl/bjacobs.

Hence, the first part of the proof is found:

$$MCF_t^s = \frac{\eta}{\int_{\mathcal{N}} \lambda(n)\mathrm{d}F(n)} = \frac{1}{1+\frac{t}{1-t}\bar{\varepsilon}_{lt}^u} = -\frac{CV}{\mathrm{d}R}. \tag{49}$$

The first-order condition for the optimal tax rate can also be written in terms of the Diamond-based measure for the marginal cost of public funds:

$$MCF_t = \frac{\eta}{\int_{\mathcal{N}} \alpha(n)\mathrm{d}F(n)} = \frac{1}{1+\frac{t}{1-t}\bar{\varepsilon}_{lt}^c}. \tag{50}$$

The compensating variation for individual n is the same as above:

$$CV(n) = -z(n)\mathrm{d}t. \tag{51}$$

From the government budget constraint follows that the compensated change in revenue is given by:

$$\mathrm{d}R^c = \left(1+\frac{t}{1-t}\bar{\varepsilon}_{lt}^c\right)\int_{\mathcal{N}} z(n)\mathrm{d}F(n)\mathrm{d}t. \tag{52}$$

Hence, the second part of the proof is found:

$$MCF_t = \frac{\eta}{\int_{\mathcal{N}} \alpha(n)\mathrm{d}F(n)} = \frac{1}{1+\frac{t}{1-t}\bar{\varepsilon}_{lt}^c} = -\frac{CV}{\mathrm{d}R^c}. \tag{53}$$

B Non-linear tax instruments

This Appendix derives optimal policies with non-linear taxes. We assume that income is taxed and the consumption tax is normalized to zero. The non-linear tax schedule is given by $T(z(n))$, where $T'(z(n)) \equiv \mathrm{d}T(z(n))/\mathrm{d}z(n)$ denotes the marginal income tax rate at income $z(n)$. The tax function is assumed to be continuous and differentiable. The government can only verify total income of an individual ($z(n) = nl(n)$), not its labor supply $l(n)$ or ability n, which rules out individualized lump-sum taxes. The individual budget constraint is modified to:

$$c(n) = z(n) - T(z(n)), \quad \forall n. \tag{54}$$

The first-order condition for labor supply of each individual under non-linear income taxation reads as

$$\frac{-u_l(c(n), l(n), G, n)}{u_c(c(n), l(n), G, n)} = (1 - T'(z(n)))n, \quad \forall n. \tag{55}$$

This first-order condition is the same as before, see equation (3), except that the non-linear marginal tax rate replaces the linear one.

The social welfare function remains the same as in equation (4). In the Mirrlees setup it is more convenient to work with the economy's resource constraint rather than the government budget constraint.[1] The economy's resource constraint is:

$$N\int_{\mathcal{N}}(nl(n)-c(n))\mathrm{d}F(n)=pG. \tag{56}$$

To determine the non-linear tax schedule $T(\cdot)$ and optimal public goods provision G, a standard mechanism-design approach is employed. Any second-best allocation must satisfy the incentive-compatibility constraints. Since n is not observable by the government, every bundle $\{c(n),z(n)\}$ for individual n must be such that each individual self-selects into this bundle and does not prefer another bundle $\{c(m),z(m)\}$ intended for individual $m\neq n$. By adopting the first-order approach, as in Mirrlees (1971), the incentive-compatibility constraints can be summarized by a differential equation on utility:

$$\frac{\mathrm{d}u(c(n),l(n),G,n)}{\mathrm{d}n}=u_n(c(n),l(n),G,n)-\frac{l(n)}{n}u_l(c(n),l(n),G,n),\quad \forall n. \tag{57}$$

The first-order approach yields an implementable allocation only if the second-order conditions for utility maximization are fulfilled at the optimum allocation. Lemma 2 states the conditions under which the first-order approach is both necessary and sufficient to describe the optimum.[2]

Lemma 2 *The optimal allocation derived under the first-order approach is implementable with the non-linear income tax function if i) utility* $U(c(n),z(n),G,n)\equiv u(c(n),z(n)/n,G,n)$ *satisfies the Spence-Mirrlees condition, i.e.,* $\mathrm{d}(-U_z/U_c)/\mathrm{d}n<0$*, and ii) gross incomes are non-decreasing with skill* n*, i.e.,* $\mathrm{d}z(n)/\mathrm{d}n\geq 0$.

Proof. See Jacobs (2010). ■

Since the public good is identical for all agents, incentive compatibility constraints are not affected by introducing public goods, even if different individuals exhibit a different valuation for the public good, see also Kreiner and Verdelin (2012). Intuitively, one individual cannot mimic another individual's ability to benefit from the public good, since his utility function remains the same. In the remainder of this Appendix it's assumed that Lemma 2 holds.

To characterize the optimal second-best policy under non-linear income taxation, Proposition 8 provides the conditions for the optimal non-linear income tax, the marginal cost of public funds, and the optimal provision of public goods. The optimal non-linear tax and modified Samuelson rule are expressed in terms of sufficient statistics, i.e., the earnings distribution, the willingness to pay for public goods, behavioral elasticities, and social welfare weights. The earnings distribution equals the skill distribution, i.e., $\tilde{F}(z(n))\equiv F(n)$, in view of the one-to-one mapping between earnings and skills (Saez, 2001). The corresponding density of earnings equals $\tilde{f}(z(n))$.

[1] If the government maximizes social welfare subject to the resource constraint, and all households respect their budget constraints, then the government budget constraint is automatically satisfied by Walras' law.

[2] Lemma 2 implies that the second-order conditions are locally satisfied, which is a standard result in the Mirrlees model without public goods provision, see also Hellwig (2010).

Proposition 8 *The optimal non-linear income tax, the marginal cost of public funds, and optimal public goods provision are given by*

$$\frac{T'(z(n))}{1-T'(z(n))} = \frac{1}{-\varepsilon^c_{lT'}} \frac{\int_{z(n)}^{z(\overline{n})}\left(1-\frac{\alpha(m)}{\eta}\right)\tilde{f}(z(m))\mathrm{d}z(m)}{1-\tilde{F}(z(n))} \frac{1-\tilde{F}(z(n))}{z(n)\tilde{f}(z(n))}, \quad \forall n, \tag{58}$$

$$MCF \equiv \frac{\eta}{\int_{z(\underline{n})}^{z(\overline{n})}\alpha(m)\tilde{f}(z(m))\mathrm{d}z(m)} = 1, \tag{59}$$

$$(1-\xi_G)N\int_{\mathcal{N}}\frac{u_G(\cdot)}{u_c(\cdot)}\mathrm{d}F(n) = p\left(1-\int_{\mathcal{N}}\gamma_n\varepsilon^c_{lG}\mathrm{d}F(n)\right), \quad \gamma_n \equiv \frac{NT'(nl(n))nl(n)}{pG}. \tag{60}$$

Proof. See Appendix B. ■

The expression for the optimal non-linear income tax in equation (58) is identical to Saez (2001). This expression is well known in the literature and the reader is referred to Mirrlees (1971), Seade (1977), Tuomala (1984), Diamond (1998) and Saez (2001) for more elaborate discussions of the optimal non-linear tax.

At the optimum, the marginal cost of public funds in equation (59) is once again equal to one. The formulation of optimal tax policy in terms of sufficient statistics formally demonstrates that the marginal cost of public funds equals one also in the Mirrlees (1971) model. The marginal cost of public funds for non-distortionary taxes should be equal to the marginal cost of public funds for all distortionary taxes. Thus, tax distortions should be equal to distributional gains for marginal tax rates at each point in the income distribution. This can be seen by rewriting the expression for the optimal non-linear income tax in equation (58):

$$MEB(n)\cdot z(n)\tilde{f}(z(n)) = \int_{z(n)}^{\bar{z}}\left(1-\frac{\alpha(m)}{\eta}\right)\tilde{f}(z(m))\mathrm{d}z(m), \tag{61}$$

$$MEB(n) \equiv -\frac{T'(z(n))}{1-T'(z(n))}\varepsilon^c_{lT'}.$$

Clearly, the total marginal excess burden of a higher marginal tax rate at earnings level $z(n)$ equals the marginal distributional gain of a higher marginal tax rate at earnings level $z(n)$. The excess burden per unit of tax base $MEB(n)$ is multiplied by the tax base $\tilde{f}(z(n))z(n)$. The distributional gain of a higher marginal tax rate at income $z(n)$ equals the social value of extracting a marginal unit of revenue from all individuals above $z(n)$. The latter equals 1 – the unit of revenue – minus the Diamond (1975) social marginal value of income $\alpha(n)/\eta$ – the utility cost in money equivalents of paying a unit more income in tax.

If marginal tax rates are not optimized, equation (61) does not hold with equality. In that case, the marginal excess burden of a marginal tax rate is not equal to the marginal distributional benefit of the marginal tax rate. Moreover, if equation (61) does not hold with equality at all points in the income distribution, the marginal cost of public funds is no longer equal to one, i.e., $MCF \equiv \eta\left[\int_{z(\underline{n})}^{z(\overline{n})}\alpha(m)\tilde{f}(z(m))\mathrm{d}z(m)\right]^{-1} \neq 1$.

To gain more intuition for the reasons why the modified Samuelson rule differs from the standard Samuelson rule, it can be expressed as well in terms of the model's primitives (see Appendix B):

$$N\int_{\mathcal{N}}(1+\Delta(n))\frac{u_G(\cdot)}{u_c(\cdot)}\mathrm{d}F(n) = p, \tag{62}$$

where

$$\Delta(n) \equiv \frac{T'(z(n))}{1-T'(z(n))}\frac{-\varepsilon^c_{lT'}}{\varepsilon^u_{zn}}\left(\frac{\partial \ln(u_G/u_c)}{\partial \ln l(n)} - \frac{\partial \ln(u_G/u_c)}{\partial \ln n}\right), \quad \forall n. \tag{63}$$

The modified Samuelson rules in equations (60) and (62) are the same, except that equation (60) is expressed in terms of sufficient statistics rather than the model's primitives, i.e. the skill distribution and the utility function (recall that $1 - T'(z(n)) = -u_l(\cdot)/u_c(\cdot)$). $\Delta(n)$ indicates the extent to which public goods are overprovided relative to the first-best Samuelson rule, cf. equation (26). $\Delta(n)$ can be interpreted as an implicit subsidy on public goods provision at skill level n. $\Delta(n)$ is different from zero if the marginal willingness to pay for the public good $\frac{u_G(\cdot)}{u_c(\cdot)}$ varies with labor supply or with ability.

Whether public goods are under- or overprovided compared to the first-best rule is determined by the presence of the distortionary income tax ($T'(z(n)) > 0$). The more distortionary is income taxation, as indicated by a larger tax rate $T'(z(n))$ or a larger compensated elasticity $\varepsilon^c_{lT'}$, the more public goods provision deviates from the first-best policy rule. In the absence of redistributional concerns, marginal tax rates are zero ($T'(z(n)) = 0$), and so is the implicit subsidy on public goods provision ($\Delta(n) = 0$). In that case, public goods provision satisfies the first-best Samuelson rule in equation (26).

The public good is more (less) complementary to work than private consumption, if the marginal willingness to pay for the public good rises (falls) with labor effort (i.e., $\frac{\partial \ln(u_G/u_c)}{\partial \ln l} > 0$ (< 0)).[3] The provision of public goods then increases. The intuition is the same as in Atkinson and Stiglitz (1976). The government should provide more public goods if they are more complementary to work than private goods are, and provide fewer public goods if they are more complementary to leisure than private goods are. In doing so, the government alleviates the distortions of the labor income tax on work effort. To put it differently, if $\frac{\partial \ln(u_G/u_c)}{\partial \ln l} > 0$ (< 0), a higher (lower) level of public goods provision relaxes the incentive constraints associated with the redistribution of income, as individuals with a high ability are less tempted to mimic individuals with a lower ability. The Atkinson and Stern (1974) term capturing the interaction of the public goods with labor supply is therefore also present under non-linear taxation. See also Christiansen (1981) and Boadway and Keen (1993). The term $\frac{\partial \ln(u_G/u_c)}{\partial \ln l}$ in equation (62) is associated with the compensated cross-elasticity ε^c_{lG} of labor supply with respect to the public good in the modified Samuelson rule in equation (60).

If the utility function differs across individuals, public goods provision should also be employed for redistribution. Public goods provision can extract additional information on the skill level n. Intuitively, the willingness to pay for public goods varies with ability even if all individuals would have the same labor earnings. Therefore, the willingness to pay for public goods reveals information about earnings ability that is independent from labor income. This makes over- or underprovision of public goods attractive for redistribution. Over- or under-provision of public goods – relative to the first-best Samuelson rule – results in more redistribution than the government can achieve with the non-linear income tax alone. Naturally, under- or overpro-

[3]The term $\frac{\partial \ln(u_G/u_c)}{\partial \ln l}$ can be rewritten as $\omega_l(\rho_{lG} - \rho_{lc})$, where $\omega_l \equiv \frac{-lu_l}{u} > 0$ is the utility share of labor, $\rho_{lG} \equiv \frac{u_{lG}u}{(-u_l)u_G}$ is the Hicksian partial elasticity of complementarity between public goods and labor, and $\rho_{lc} \equiv \frac{u_{lc}u}{(-u_l)u_c}$ is the Hicksian partial elasticity of complementarity between private consumption and labor. Thus, the willingness to pay for the public good increases with labor supply if $\rho_{lG} > \rho_{lc}$. That is, if the public good is a stronger Hicksian complement to labor than private consumption is.

vision causes inefficiencies, and these need to be traded off against the distributional gains. If the marginal willingness to pay for the public good rises (falls) with ability (i.e., $\frac{\partial \ln(u_G/u_c)}{\partial \ln n} > 0$ (< 0)), then the public good benefits the individuals with higher skill levels relatively more (less).[4] Consequently, the optimal level of public goods provision falls (increases) for redistributional reasons. The term $\frac{\partial \ln(u_G/u_c)}{\partial \ln n}$ in equation (62) is thus associated with the distributional characteristic ξ_G of the public good in the modified Samuelson rule under linear taxation in equation (60).

Alternatively, one can understand this result by looking at the incentive constraints. If $\frac{\partial \ln(u_G/u_c)}{\partial \ln n} > 0$ (< 0), then high-skill types are less (more) tempted to mimic low-skill types if public goods provision expands, since they have a stronger (weaker) preference for public goods. Consequently, smaller public goods provision relaxes (tightens) the incentive-compatibility constraints. Consequently, reducing public goods provision below the first-best rule reduces (increases) the distortions of redistributing income.

If the utility function is identical for all individuals n, the distributional term drops out (i.e., $\frac{\partial \ln(u_G/u_c)}{\partial \ln n} = 0$). Intuitively, if differences in earning ability are the only source of heterogeneity, and everyone has the same utility function, the willingness to pay for public good is the same if labor earnings are the same. Therefore, differences in the willingness to pay for public goods only originate from differences in labor earnings. In that case, under- or or overprovision of the public good – relative to the first-best Samuelson rule – does not result in more income redistribution than can be achieved with the non-linear income tax alone. However, under- or overprovision of public goods – relative to the first-best rule – causes inefficiencies. These distortions can be avoided by organizing all redistribution via the non-linear income tax and providing public goods according to the Samuelson rule. See also Christiansen (1981) and Boadway and Keen (1993) for the case of homogeneous preferences. Hence, when the utility function is the same for all individuals, public goods are not provided for redistributional reasons.

The finding that public goods are not used for redistribution if utilities are identical contrasts with the linear case. Under linear income taxation, optimal provision of the public good is *always* found to be dependent on the distributional impact of the public good, except for the (trivial) cases in which the marginal willingness to pay is equal or is linear in income for all individuals. With linear taxation, the government uses an informationally inferior instrument to redistribute income by ignoring the information on individual earnings. Hence, the government optimally complements the income tax by using indirect instruments for redistribution, such as public goods provision.

Christiansen (1981), Boadway and Keen (1993) and Kaplow (1996) discuss the case in which the utility function is identical for all individuals and is (weakly) separable between private/public goods and leisure. They find that the optimal provision of public goods then follows the first-best Samuelson rule. The next Proposition demonstrates this formally. Naturally, the allocations will differ between first- and second-best.

[4] The term $\frac{\partial \ln(u_G/u_c)}{\partial \ln n}$ can be rewritten as $\omega_n(\rho_{nG} - \rho_{nc})$, where $\omega_n \equiv \frac{n u_n}{u} > 0$ is the utility share of ability, $\rho_{nG} \equiv \frac{u_{nG} u}{u_n u_G}$ is the Hicksian partial elasticity of complementarity between public goods and ability, and $\rho_{nc} \equiv \frac{u_{nc} u}{u_n u_c}$ is the Hicksian partial elasticity of complementarity between private consumption and ability. The willingness to pay for the public good increases with ability if $\rho_{nG} > \rho_{nc}$. That is, if public goods are stronger Hicksian complements to ability than private goods are.

Proposition 9 *If utility is given by* $u(n) \equiv u(h(c(n), G), l(n))$, $\forall n$, *then the optimal provision of public goods follows the first-best Samuelson rule in second-best settings with optimal distortionary taxation:*

$$N \int_{\mathcal{N}} \frac{u_G(\cdot)}{u_c(\cdot)} \mathrm{d}F(n) = p. \tag{64}$$

Proof. If $u(n) \equiv u(h(c, G), l)$, it is immediately established that $\frac{\partial \ln(u_G/u_c)}{\partial \ln l} = \frac{\partial \ln(u_G/u_c)}{\partial \ln n} = 0$. Substitution in (62) yields (64). ■

Proposition 9 is the non-linear counterpart of Proposition 3. It demonstrates that, once second-best interactions of the public good with labor supply are absent, or when redistribution via public goods has no value added over direct income redistribution, the optimal rule for public good provision follows the first-best Samuelson rule. Hence, there should be no correction for the marginal cost of public funds as distortions of taxation cancel against the distributional gains of taxation.

C Proof Proposition 8

The proof of Proposition 8 follows in three steps. First, we derive the behavioral elasticities. Second, we derive the first-order conditions using a Hamiltonian approach. Third, we rewrite these first-order conditions using the elasticities derived in the first step of the proof.

C.1 Elasticities

First, the behavioral elasticities of model are derived. These elasticities will be used later in deriving the optimal tax expressions. As in Jacquet et al. (2013), define the following *shift function* – omitting the indices n:

$$\begin{aligned} L(l, G, n, \tau, \rho) \equiv\, & n(1 - T'(nl) - \tau) u_c(nl - T(nl) - \tau(nl - nl(n)) + \rho, l, G, n) \\ & + u_l(nl - T(nl) - \tau(nl - nl(n)) + \rho, l, G, n). \end{aligned} \tag{65}$$

$L(l, n, \tau, \rho, G)$ measures a *shift* in the first-order condition for labor supply when one of the variables l, G, n, τ, or ρ changes. τ captures an exogenous increase in the marginal tax rate (i.e., for any level of earnings). ρ is introduced to retrieve the income effect when the individual receives an exogenous amount of income ρ. The first-order condition for labor supply of the individual n is thus equivalent to $L(l, G, n, 0, 0) = 0$. The following partial derivatives of the shift function L are found, using the first-order condition $-u_l = n(1 - T')u_c$:

$$L_l(l, G, n, 0, 0) = u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\left(\frac{u_l}{u_c}\right) u_{cl} + n u_l \frac{T''}{(1 - T')}, \tag{66}$$

$$L_n(l, G, n, 0, 0) = \left(\frac{-u_l}{l} + n u_l \frac{T''}{(1 - T')} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - \left(\frac{u_l}{u_c}\right) u_{lc}\right) \frac{l}{n} + \left(\frac{u_{ln}}{u_l} - \frac{u_{cn}}{u_c}\right) u_l, \tag{67}$$

$$L_\tau(l, G, n, 0, 0) = -n u_c, \tag{68}$$

$$L_\rho(l, G, n, 0, 0) = \frac{u_{lc} u_c - u_l u_{cc}}{u_c}. \tag{69}$$

The envelope theorem gives the following formula for the partial derivatives: $\frac{\partial l}{\partial q} = -\frac{L_q}{L_l}$, $q = n, \tau, \rho$. The uncompensated *wage* elasticity of labor supply ε_{ln}^u is equal to:

$$\varepsilon_{ln}^u \equiv \frac{\partial l^u}{\partial n}\frac{n}{l} = -\frac{L_n}{L_l}\frac{n}{l} = -\frac{\left(\frac{-u_l}{l} + nu_l\frac{T''}{(1-T')} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - \left(\frac{u_l}{u_c}\right)u_{lc}\right) + \left(\frac{nu_{ln}}{u_l} - \frac{nu_{cn}}{u_c}\right)\frac{u_l}{l}}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\left(\frac{u_l}{u_c}\right)u_{cl} + nu_l\frac{T''}{(1-T')}}. \tag{70}$$

The income elasticity of labor supply $\varepsilon_{l\rho}$ is defined as:

$$\varepsilon_{l\rho} \equiv (1-T')n\frac{\partial l}{\partial \rho} = -(1-T')n\frac{L_\rho}{L_l} = \frac{\frac{u_l}{u_c}\left(u_{lc} - \frac{u_l u_{cc}}{u_c}\right)}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\left(\frac{u_l}{u_c}\right)u_{cl} + nu_l\frac{T''}{(1-T')}}. \tag{71}$$

The compensated *wage* elasticity of labor supply ε_{ln}^c is defined residually by the Slutsky equation ($\varepsilon_{ln}^c \equiv \varepsilon_{ln}^u - \varepsilon_{l\rho}$):

$$\varepsilon_{ln}^c \equiv \frac{\partial l^c}{\partial n}\frac{n}{l} = \varepsilon_{ln}^u - \varepsilon_{l\rho} = \frac{\frac{u_l}{l} - nu_l\frac{T''}{(1-T')} - \left(\frac{nu_{ln}}{u_l} - \frac{nu_{cn}}{u_c}\right)\frac{u_l}{l}}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\left(\frac{u_l}{u_c}\right)u_{cl} + nu_l\frac{T''}{(1-T')}}. \tag{72}$$

The compensated *tax* elasticity of labor supply $\varepsilon_{lT'}^c$ is:

$$\varepsilon_{lT'}^c \equiv \frac{\partial l^c}{\partial \tau}\frac{(1-T')}{l} = -\frac{L_\tau}{L_l}\frac{(1-T')}{l} = \frac{-u_l/l}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\left(\frac{u_l}{u_c}\right)u_{cl} + nu_l\frac{T''}{(1-T')}}. \tag{73}$$

Note that the compensated *tax* elasticity of earnings supply $\varepsilon_{zT'}^c$ equals the compensated tax elasticity of labor supply (since the wage rate n is not affected by an increase in marginal taxes, only labor supply). The uncompensated *wage* elasticity of *earnings* supply ε_{zn}^u is equal to:

$$\varepsilon_{zn}^u \equiv \frac{\partial z}{\partial n}\frac{n}{z} = 1 + \varepsilon_{ln}^u = \frac{\left(1 + \left(\frac{lu_{ll}}{u_l} - \frac{lu_{cl}}{u_c}\right) - \left(\frac{nu_{ln}}{u_l} - \frac{nu_{cn}}{u_c}\right)\right)\frac{u_l}{l}}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\left(\frac{u_l}{u_c}\right)u_{cl} + nu_l\frac{T''}{(1-T')}}. \tag{74}$$

C.2 Hamiltonian

Consumption is defined as a function of the allocation: $c(n) \equiv c(l(n), u(n), G, n)$, which is obtained by inverting the utility function $u(c(n), l(n), G, n)$. The derivatives of $c(\cdot)$ are found by applying the implicit function theorem: $c_u = 1/u_c$, $c_l = -u_l/u_c$, $c_G = -u_G/u_c$, $c_n = -u_n/u_c$. For notational compactness the incentive compatibility constraint can be rewritten as:

$$\begin{aligned}\frac{\mathrm{d}u(n)}{\mathrm{d}n} &= u_n(c(l(n), u(n), G, n), l(n), G, n) - \frac{l(n)}{n}u_l(c(l(n), u(n), G, n), l(n), G, n) \\ &\equiv \varphi(l(n), u(n), G, n),\end{aligned} \tag{75}$$

where the function $\varphi(l(n), u(n), G, n)$ has partial derivatives – omitting the n-indices:

$$\varphi_l = -\frac{u_l}{n}\left(1 + \left(\frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c}\right) - \left(\frac{nu_{nl}}{u_l} - \frac{nu_{nc}}{u_c}\right)\right), \tag{76}$$

$$\varphi_u = \frac{1}{n}\left(\frac{nu_{nc}}{u_c} - \frac{lu_{lc}}{u_c}\right), \tag{77}$$

$$\varphi_G = u_{nG} - u_{nc}\frac{u_G}{u_c} - \frac{l}{n}\left(u_{lG} - u_{lc}\frac{u_G}{u_c}\right). \tag{78}$$

The Hamiltonian for maximizing social welfare is:

$$\mathcal{H} = u(n)f(n) + \eta\left[nl(n) - c(l(n), u(n), G, n) - pG/N\right] f(n) - \theta(n)\varphi(l(n), u(n), G, n), \tag{79}$$

where $\theta(n)$ is the co-state variable associated with state variable $u(n)$, which has to satisfy incentive compatibility constraint (75). Furthermore, η denotes the Lagrange multiplier on the economy's resource constraint in equation (56). $\theta(n)$ is multiplied with a minus sign to obtain a positive multiplier $\theta(n)$. The first-order conditions for an optimal allocation are – omitting the n-indices:

$$\frac{\partial \mathcal{H}}{\partial l} = \eta\left(n + \frac{u_l}{u_c}\right) f + \frac{\theta u_l}{n}\left(1 + \left(\frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c}\right) - \left(\frac{nu_{nl}}{u_l} - \frac{nu_{nc}}{u_c}\right)\right) = 0, \quad \forall n, \tag{80}$$

$$\frac{\partial \mathcal{H}}{\partial u} = \left(1 - \frac{\eta}{u_c}\right) f - \frac{\theta}{n}\left(\frac{nu_{nc}}{u_c} - \frac{lu_{lc}}{u_c}\right) = \frac{\mathrm{d}\theta}{\mathrm{d}n}, \quad \forall n, \tag{81}$$

$$\int_{\mathcal{N}} \frac{\partial \mathcal{H}}{\partial G}\mathrm{d}n = \int_{\mathcal{N}} \left[\eta\left(\frac{u_G}{u_c} - \frac{p}{N}\right) f - \frac{\theta u_G}{n}\left(\left(\frac{nu_{nG}}{u_G} - \frac{nu_{nc}}{u_c}\right) - \left(\frac{lu_{lG}}{u_G} - \frac{lu_{lc}}{u_c}\right)\right)\right] \mathrm{d}n = 0, \tag{82}$$

where the derivatives of the $c(\cdot)$ function are used in each line. The transversality conditions for this optimal control problem are given by $\lim_{n\to\overline{n}} \theta(n) = 0$, and $\lim_{n\to\underline{n}} \theta(n) = 0$.

C.3 Optimal taxation

C.3.1 Optimal non-linear taxes

The first-order conditions for l in equation (80) can be rewritten using $\frac{-u_l}{u_c} = (1 - T'(z))n$:

$$\frac{T'}{1 - T'} = \frac{u_c\theta/\eta}{nf(n)}\left[1 + \left(\frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c}\right) - \left(\frac{nu_{nl}}{u_l} - \frac{nu_{nc}}{u_c}\right)\right]. \tag{83}$$

Next, employ the definitions for the elasticities for $\varepsilon^c_{lT'}$ in equation (73) and ε^u_{zn} in equation (74) to find:

$$\left[1 + \left(\frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c}\right) - \left(\frac{nu_{nl}}{u_l} - \frac{nu_{nc}}{u_c}\right)\right] = \frac{\varepsilon^u_{zn}}{-\varepsilon^c_{lT'}}. \tag{84}$$

Hence, the first-order condition for optimal taxes can be written as:

$$\frac{T'(z(n))}{1 - T'(z(n))} = \frac{u_c(\cdot)\theta/\eta}{nf(n)}\frac{\varepsilon^u_{zn}}{-\varepsilon^c_{lT'}}. \tag{85}$$

The first-order condition for u in equation (81) can be rewritten as:

$$\left(\frac{u_c}{\eta}-1\right)f(n)-\frac{\theta}{\eta}u_{cn}+\frac{\theta}{\eta}\frac{lu_{lc}}{n}=\frac{u_c}{\eta}\frac{\mathrm{d}\theta}{\mathrm{d}n}. \tag{86}$$

Introduce the composite multiplier $\Theta(n)$:

$$\Theta(n)\equiv\frac{\theta(n)u_c(c(n),l(n),G,n)}{\eta}=\frac{\theta(n)u_c(c(l(n),u(n),G,n),l(n),G,n)}{\eta}, \tag{87}$$

which has total derivative (note that G does not vary with n):

$$\frac{\mathrm{d}\Theta}{\mathrm{d}n}=\frac{\mathrm{d}\theta}{\mathrm{d}n}\frac{u_c}{\eta}+\frac{\theta u_{cc}}{\eta}\left(\frac{\partial c}{\partial l}\frac{\mathrm{d}l}{\mathrm{d}n}+\frac{\partial c}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}n}+\frac{\partial c}{\partial n}\right)+\frac{\theta}{\eta}u_{cl}\frac{\mathrm{d}l}{\mathrm{d}n}+\frac{\theta}{\eta}u_{cn}. \tag{88}$$

Simplify the last expression using the derivatives of the c-function and substituting equation (75) for $\frac{\mathrm{d}u}{\mathrm{d}n}$:

$$\frac{\mathrm{d}\Theta}{\mathrm{d}n}=\frac{\mathrm{d}\theta}{\mathrm{d}n}\frac{u_c}{\eta}-\frac{\theta}{\eta}\left(\frac{l}{n}\frac{u_l u_{cc}}{u_c}-u_{cn}\right)-\frac{\theta}{\eta}\left(\frac{u_l u_{cc}}{u_c}-u_{cl}\right)\frac{\mathrm{d}l}{\mathrm{d}n}. \tag{89}$$

Use the elasticity $\varepsilon_{l\rho}$ in equation (71) and the elasticity $\varepsilon^c_{lT'}$ in equation (73) to find an expression for $\frac{u_{cc}u_l}{u_c}-u_{cl}$:

$$\frac{u_l u_{cc}}{u_c}-u_{lc}=\frac{\varepsilon_{l\rho}}{\varepsilon^c_{lT'}}\frac{u_c}{l}. \tag{90}$$

Substitute equation (90) into equation (89), and use ε^u_{ln} from equation (70) to derive:

$$\frac{\mathrm{d}\Theta}{\mathrm{d}n}=\frac{\mathrm{d}\theta}{\mathrm{d}n}\frac{u_c}{\eta}-\frac{\theta}{\eta}\frac{l}{n}\frac{u_{cc}u_l}{u_c}+\frac{\theta}{\eta}u_{cn}-\frac{\theta}{\eta}\frac{\varepsilon_{l\rho}}{\varepsilon^c_{lT'}}\frac{u_c}{n}\varepsilon^u_{ln}. \tag{91}$$

Substitute the first-order condition for u in equation (86) into equation (91) to find:

$$\frac{\mathrm{d}\Theta}{\mathrm{d}n}+\frac{\theta}{\eta}\frac{\varepsilon_{l\rho}}{\varepsilon^c_{lT'}}\frac{u_c}{n}\varepsilon^u_{ln}=\left(\frac{u_c}{\eta}-1\right)f(n)-\frac{\theta}{\eta}\left(\frac{u_{cc}u_l}{u_c}-u_{lc}\right)\frac{l}{n}. \tag{92}$$

Substitute equation (90) in equation (92) and note that $1+\varepsilon^u_{ln}=\varepsilon^u_{zn}$ from equations (70) and (74) to find:

$$\frac{\mathrm{d}\Theta}{\mathrm{d}n}+\frac{u_c}{u_l}\frac{\theta}{\eta}\frac{\varepsilon^u_{zn}}{\varepsilon^c_{lT'}}\frac{u_l}{n}\varepsilon_{l\rho}=\left(\frac{u_c}{\eta}-1\right)f(n). \tag{93}$$

Rewrite first-order condition for z_n in equation (85):

$$T'nf(n)=\frac{\theta}{\eta}\frac{u_l}{n}\frac{\varepsilon^u_{zn}}{\varepsilon^c_{lT'}}. \tag{94}$$

Substitute equation (94) in (93) to find:

$$\frac{\mathrm{d}\Theta}{\mathrm{d}n}=\left(\frac{u_c}{\eta}+\frac{T'}{1-T'}\varepsilon_{l\rho}-1\right)f(n). \tag{95}$$

Equation (95) can be integrated – using a transversality condition – to obtain:

$$\Theta(n) = \frac{\theta(n)u_c(c(n), l(n), G, n)}{\eta} = \int_n^{\overline{n}} \left(1 - \frac{u_c(m)}{\eta} - \frac{T'(z(m))}{1-T'(z(m))}\varepsilon_{l\rho}\right) f(m)\mathrm{d}m. \tag{96}$$

The ABC-formula for the optimal non-linear tax results upon substituting equation (96) in equation (85):

$$\frac{T'(z(n))}{1-T'(z(n))} = \frac{\varepsilon^u_{zn}}{-\varepsilon^c_{lT'}} \frac{\int_n^{\overline{n}} \left(1 - \frac{u_c(m)}{\eta} - \frac{T'(z(m))}{1-T'(z(m))}\varepsilon_{l\rho}\right) f(m)\mathrm{d}m}{1-F(n)} \frac{1-F(n)}{nf(n)}. \tag{97}$$

Finally, the optimal tax formula can be written in terms of the earnings density as in Saez (2001) and Jacquet et al. (2013). Define the earnings distribution as $\tilde{F}(z(n)) \equiv F(n)$, with corresponding density function $\tilde{f}(z(n))$. Then, it can be derived that $\varepsilon^u_{zn}\tilde{f}(z(n))z(n) = f(n)n$. By noting that there is a perfect mapping between earnings $z(n)$ and ability n the expression for the optimal income tax is written as:

$$\frac{T'(z(n))}{1-T'(z(n))} = \frac{1}{-\varepsilon^c_{lT'}} \frac{\int_{z(n)}^{z(\overline{n})} \left(1 - g(m) - \frac{T'(z(m))}{1-T'(z(m))}\varepsilon_{l\rho}\right) \tilde{f}(z(m))\mathrm{d}z(m)}{1-\tilde{F}(z(n))} \frac{1-\tilde{F}(z(n))}{z(n)\tilde{f}(z(n))}. \tag{98}$$

Using the definition of $\alpha(n) \equiv \lambda(n) + \eta\frac{T'(z(n))}{1-T'(z(n))}\varepsilon_{l\rho} = \lambda(n) + \eta n T'(z(n))\frac{\partial l(n)}{\partial(-T(0))}$, where the second step follows from $\varepsilon_{l\rho} = \varepsilon_{l(-T(0))}$, this can be rewritten as the expression in the main text.

C.3.2 Marginal cost of public funds

Use $\alpha(n) \equiv \lambda(n) + \eta\frac{T'(z(n))}{1-T'(z(n))}\varepsilon_{l\rho}$ and the transversality condition $\theta(\underline{n}) = 0$ in equation (96) to derive:

$$0 = \int_{\underline{n}}^{\overline{n}} (1 - \alpha(m)) f(m)\mathrm{d}m. \tag{99}$$

Hence, switching the bound of the integral, the marginal cost of public funds equals unity in the policy optimum:

$$MCF \equiv \frac{\eta}{\int_{z(\underline{n})}^{z(\overline{n})} \alpha(m)\tilde{f}(z(m))\mathrm{d}m} = 1. \tag{100}$$

C.3.3 Optimal public goods provision

In order to express the optimal formula for public goods provision in terms of sufficient statistics, the optimal level of public goods is evaluated along an optimized non-linear tax system $T^*(nl(n))$. The Lagrangian for maximizing social welfare with respect to the public good can then be formulated as:

$$\mathcal{L} \equiv N \int_{\mathcal{N}} v(n)\mathrm{d}F(n) + \eta \left(N \int_{\mathcal{N}} T^*(nl(n))\mathrm{d}F(n) - pG\right), \tag{101}$$

where $v(n)$ is indirect utility of the individual. The first-order condition for G is given by:

$$\frac{\partial \mathcal{L}}{\partial G} = N \int_{\mathcal{N}} \left(u_G + \eta n T^{*\prime}(nl(n)) \frac{\partial l(n)}{\partial G} \right) \mathrm{d}F(n) - p\eta = 0, \tag{102}$$

where the derivatives of indirect utility have been substituted. Substituting the Slutsky equation $\frac{\partial l^u(n)}{\partial G} = \frac{\partial l^c(n)}{\partial G} + \frac{u_G}{\lambda(n)} \frac{\partial l(n)}{\partial \rho}$ yields:

$$N \int_{\mathcal{N}} \frac{\alpha(n)}{\eta} \frac{u_G}{u_c} \mathrm{d}F(n) + N \int_{\mathcal{N}} \left(n T^{*\prime}(nl(n)) \frac{\partial l^c(n)}{\partial G} \right) \mathrm{d}F(n) = p, \tag{103}$$

Use the Feldstein characteristic for the public good ξ_G to derive:

$$(1 - \xi_G) \int_{\mathcal{N}} \frac{u_G(\cdot)}{u_c(\cdot)} \mathrm{d}F(n) = \int_{\mathcal{N}} \frac{\alpha(n)}{\eta} \frac{u_G(\cdot)}{u_c(\cdot)} \mathrm{d}F(n). \tag{104}$$

Hence, optimal provision of the public good follows from:

$$(1 - \xi_G) N \int_{\mathcal{N}} \frac{u_G(\cdot)}{u_c(\cdot)} \mathrm{d}F(n) = p \left(1 - \int_{\mathcal{N}} \gamma_n \varepsilon^c_{lG} \mathrm{d}F(n) \right), \tag{105}$$

where $\gamma_n \equiv \frac{NT^{*\prime}(nl(n))nl(n)}{pG}$.

Alternatively, one can derive the optimal rule for public goods provision in terms of the model's primitives, as in Jacobs (2010). Derive that $\frac{nu_{nG}}{u_G} - \frac{nu_{nc}}{u_c} = \frac{\partial \ln(u_G/u_c)}{\partial \ln n}$ and $\frac{lu_{lG}}{u_G} - \frac{lu_{lc}}{u_c} = \frac{\partial \ln(u_G/u_c)}{\partial \ln l}$. Substitute this in equation (82), and rearrange to obtain:

$$N \int_{\mathcal{N}} \frac{u_G(\cdot)}{u_c(\cdot)} \mathrm{d}F(n) = p + N \int_{\mathcal{N}} \frac{\theta(n) u_G(\cdot)/\eta}{nf(n)} \left(\frac{\partial \ln(u_G/u_c)}{\partial \ln n} - \frac{\partial \ln(u_G/u_c)}{\partial \ln l(n)} \right) \mathrm{d}F(n). \tag{106}$$

Substitute equation (85) for the optimal income tax in equation (106) to find:

$$N \int_{\mathcal{N}} \frac{u_G(\cdot)}{u_c(\cdot)} (1 + \Delta(n)) \mathrm{d}F(n) = p, \tag{107}$$

where

$$\Delta(n) \equiv \frac{T'(z(n))}{1 - T'(z(n))} \frac{-\varepsilon^c_{lT'}}{\varepsilon^u_{zn}} \left(\frac{\partial \ln(u_G/u_c)}{\partial \ln l(n)} - \frac{\partial \ln(u_G/u_c)}{\partial \ln n} \right). \tag{108}$$

References

Atkinson, Anthony B., and Joseph E. Stiglitz (1976) ‘The design of tax structure: Direct versus indirect taxation.’ *Journal of Public Economics* 6(1-2), 55–75

Atkinson, Anthony B., and Nicholas H. Stern (1974) ‘Pigou, taxation and public goods.’ *Review of Economic Studies* 41(1), 119–128

Boadway, Robin, and Michael Keen (1993) ‘Public goods, self-selection and optimal income taxation.’ *International Economic Review* 34(3), 463–478

Christiansen, Vidar (1981) ‘Evaluation of public projects under optimal taxation.’ *Review of Economic Studies* 48(3), 447–457

Diamond, Peter A. (1975) ‘A many-person ramsey tax rule.’ *Journal of Public Economics* 4(4), 335–342

Diamond, Peter A. (1998) ‘Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates.’ *American Economic Review* 88(1), 83–95

Hellwig, Martin (2010) ‘Incentive problems with unidimensional hidden characteristics: A unified approach.’ *Econometrica* 78(4), 1201–1237

Jacobs, Bas (2010) ‘The marginal cost of public funds is one.’ CESifo Working Paper No. 3250 Munich: CESifo

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden (2013) ‘The optimal marginal tax rates with both intensive and extensive responses.’ *Journal of Economic Theory* 148(5), 1770–1805

Kaplow, Louis (1996) ‘The optimal supply of public goods and the distortionary cost of taxation.’ *National Tax Journal* 49(4), 513–533

Kreiner, Klaus Thustrup, and Nicolai Verdelin (2012) ‘Optimal provision of public goods: A synthesis.’ *Scandinavian Journal of Economics* 114(2), 384–408

Mirrlees, James A. (1971) ‘An exploration in the theory of optimum income taxation.’ *Review of Economic Studies* 38(2), 175–208

Saez, Emmanuel (2001) ‘Using elasticities to derive optimal income tax rates.’ *Review of Economic Studies* 68(1), 205–229

Seade, Jesus K. (1977) ‘On the shape of optimal tax schedules.’ *Journal of Public Economics* 7(2), 203–235

Tuomala, Matti (1984) ‘On the optimal income taxation: Some further numerical results.’ *Journal of Public Economics* 23(3), 351–366