Online Appendix: Derivations of Optimal Tax Formulas

Derivation of the optimal marginal tax rate at income level z in the Mirrlees model (Figure A1)

Figure A1 depicts the optimal marginal tax rate derivation at income level z. Again, the horizontal axis in Figure A1 shows pre-tax income, while the vertical axis shows disposable income. Consider a situation in which the marginal tax rate is increased by $\Delta\tau$ in the band from z to $z + \Delta z$, but left unchanged anywhere else. The tax reform has three effects. First, the mechanical tax increase, leaving aside behavioral responses, will be the gap between the solid and dashed lines, shown by the vertical arrow equal to $\Delta\tau\Delta z$. The total mechanical tax increase is $\Delta M = \Delta\tau\Delta z[1 - H(z)]$ as $1 - H(z)$ is the fraction of individuals above z. Second, this tax increase creates a social welfare cost of $\Delta W = -\Delta\tau\Delta z[1 - H(z)]G(z)$ as $G(z)$ is defined as the average social marginal welfare weight for individuals with income above z. Third, there is a behavioral response to the tax change. Those in the income range from z to $z + \Delta z$ have a behavioral response to the higher marginal tax rates, shown by the horizontal line pointing left. Assuming away income effects, this is the only behavioral response; those with income levels above $z + \Delta z$ face no change in marginal tax rates and hence have no behavioral response. The $h(z)\Delta z$ taxpayers in the band reduce their income by $\delta z = -\Delta\tau ez/(1 - T'(z))$ where e is the elasticity of earnings z with respect to the net-of-tax rate $1 - T'$. This response leads to a tax loss equal to $\Delta B = -h(z)ez\,T'(z)/(1 - T'(z))\Delta z\Delta\tau$. At the optimum, the three effects cancel out so that $\Delta M + \Delta W + \Delta B = 0$. After introducing $\alpha(z) = zh(z)/(1 - H(z))$, this leads to the optimal tax formula presented in the main text:
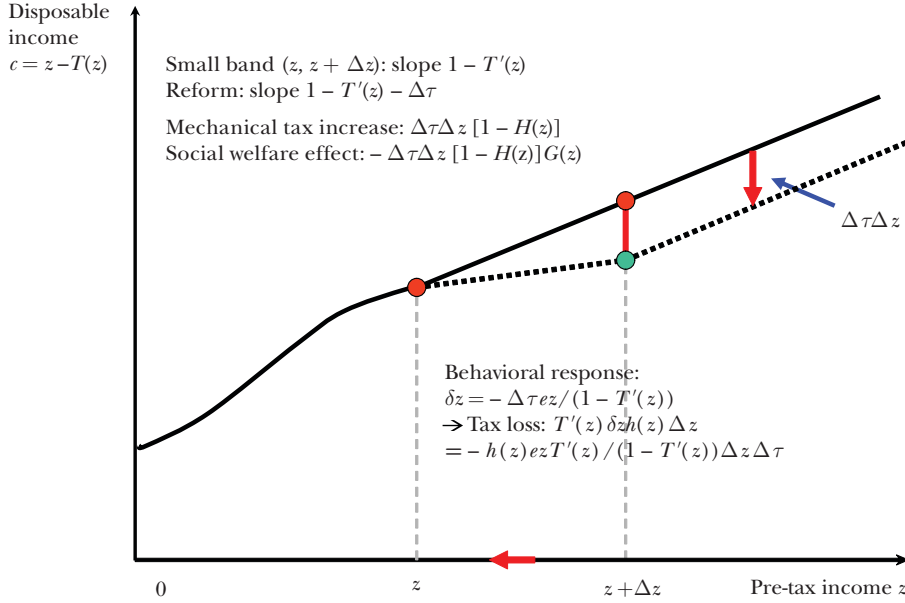
A1) $$T'(z)/(1 - T'(z)) = (1/e)(1 - G(z))(1 - H(z))/(zh(z)),$$

or

$$T'(z) = [1 - G(z)]/[1 - G(z) + \alpha(z)e].$$

Figure A1

Derivation of the Optimal Marginal Tax Rate at Income Level z in the Mirrlees Model

Disposable income
$c = z - T(z)$

Small band $(z, z + \Delta z)$: slope $1 - T'(z)$
Reform: slope $1 - T'(z) - \Delta\tau$

Mechanical tax increase: $\Delta\tau\Delta z \, [1 - H(z)]$
Social welfare effect: $- \Delta\tau\Delta z \, [1 - H(z)] G(z)$

$\Delta\tau\Delta z$

Behavioral response:
$\delta z = - \Delta\tau e z / (1 - T'(z))$
$\rightarrow$ Tax loss: $T'(z) \, \delta z h(z) \, \Delta z$
$= - h(z) e z T'(z) / (1 - T'(z)) \Delta z \Delta\tau$

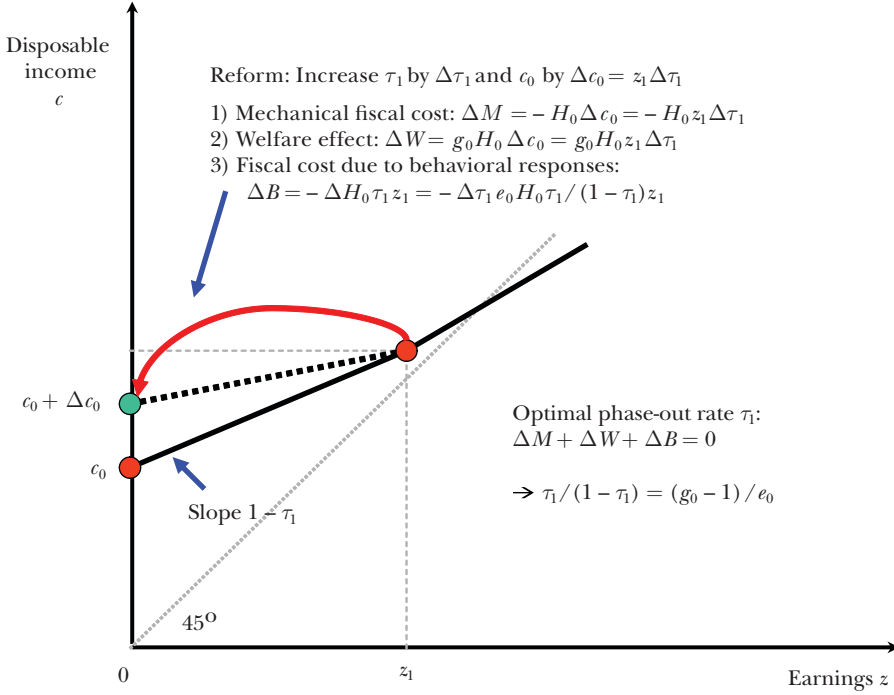0 z $z + \Delta z$ Pre-tax income z

Notes: The figure depicts the optimal marginal tax rate derivation at income level z by considering a small reform around the optimum, whereby the marginal tax rate in the small band $(z, z + \Delta z)$ is increased by $\Delta\tau$. This reform mechanically increases taxes by $\Delta\tau\Delta z$ for all taxpayers above the small band, leading to a mechanical tax increase $\Delta\tau\Delta z[1 - H(z)]$ and a social welfare cost of $-\Delta\tau\Delta z[1 - H(z)]G(z)$. Assuming away income effects, the only behavioral response is a substitution effect in the small band: the $h(z)\Delta z$ taxpayers in the band reduce their income by $\delta z = -\Delta\tau e z/(1 - T'(z))$ leading to a tax loss equal to $-h(z)e z T'(z)/(1 - T'(z))\Delta z\Delta\tau$. At the optimum, the three effects cancel out leading to the optimal tax formula $T'(z)/(1 - T'(z)) = (1/e)(1 - G(z))(1 - H(z))/(zh(z))$, or equivalently $T'(z) = [1 - G(z)]/[1 - G(z) + \alpha(z)e]$ after introducing $\alpha(z) = zh(z)/(1 - H(z))$.

Derivation of the optimal marginal tax rate at the bottom in the Mirrlees model (Figure A2)

For expositional simplicity, let us consider a discrete version of the Mirrlees (1971) model developed in Piketty (1997) and Saez (2002).

As illustrated on Figure A2, suppose that low-ability individuals can choose either to work and earn z_1 or not work and earn zero. The government offers a transfer c_0 to those not working phased out at rate τ_1 so that those working receive on net $c_1 = (1 - \tau_1)z_1 + c_0$. In words, nonworkers would keep a fraction $1 - \tau_1$ of their earnings should they work and earn z_1. Therefore, increasing τ_1 discourages some low-income workers from working. Let us denote by H_0 the fraction of nonworkers in the economy and by $e_0 = -(1 - \tau_1)/H_0 \, \Delta H_0/\Delta(1 - \tau_1)$ the elasticity

Optimal Bottom Marginal Tax Rate with Only Intensive Labor Supply Responses

Reform: Increase τ_1 by $\Delta\tau_1$ and c_0 by $\Delta c_0 = z_1 \Delta\tau_1$

1) Mechanical fiscal cost: $\Delta M = -H_0 \Delta c_0 = -H_0 z_1 \Delta\tau_1$
2) Welfare effect: $\Delta W = g_0 H_0 \Delta c_0 = g_0 H_0 z_1 \Delta\tau_1$
3) Fiscal cost due to behavioral responses:
   $\Delta B = -\Delta H_0 \tau_1 z_1 = -\Delta\tau_1 e_0 H_0 \tau_1 / (1-\tau_1) z_1$

Optimal phase-out rate τ_1:
$\Delta M + \Delta W + \Delta B = 0$

$\rightarrow \tau_1/(1-\tau_1) = (g_0 - 1)/e_0$

Slope $1 - \tau_1$

45°

Notes: The figure depicts the derivation of the optimal marginal tax rate at the bottom in the discrete Mirrlees (1971) model with labor supply responses along the intensive margin only. Let H_0 be the fraction of the population not working. This is a function of $1-\tau_1$, the net-of-tax rate at the bottom, with elasticity e_0. We consider a small reform around the optimum where the government increases the maximum transfer by Δc_0 by increasing the phase-out rate by $\Delta\tau_1$ leaving the tax schedule unchanged for those with income above z_1, this creates three effects which cancel out at the optimum. At the optimum, we have $\tau_1/(1-\tau_1) = (g_0-1)/e_0$ or $\tau_1 = (g_0-1)/(g_0-1+e_0)$. Under standard redistributive preferences, g_0 is large implying that τ_1 is large.

of nonworkers H_0 with respect to the net-of-tax rate $1-\tau_1$, where the minus sign is used so that $e_0 > 0$.

Suppose now that the government increases both the maximum transfer by Δc_0 and the phase-out rate by $\Delta\tau_1$ leaving the tax schedule unchanged for those with income equal to or above z_1 so that $\Delta c_0 = z_1 \Delta\tau_1$ as depicted on Figure A2. The fiscal cost is $-H_0 \Delta c_0$, but the welfare benefit is $H_0 g_0 \Delta c_0$ where g_0 is the social welfare weight on nonworkers. If the government values redistribution, then $g_0 > 1$ and g_0 is potentially large as nonworkers are the most disadvantaged. Because behavioral responses take place along the intensive margin only in the Mirrlees model, with no income change above z_1, the labor supply of those above z_1 is not affected by the reform. By definition of e_0, a number $\Delta H_0 = \Delta\tau_1 e_0 H_0/(1-\tau_1)$ of low-income

workers stop working, creating a revenue loss of $\tau_1 z_1 \Delta H_0 = \Delta c_0 H_0 e_0 \tau_1 /(1 - \tau_1)$. At the optimum, the three effects sum to zero leading to the optimal bottom rate formula:

A2)
$$\tau_1 /(1 - \tau_1) = (g_0 - 1)/e_0$$

or

$$\tau_1 = (g_0 - 1)/(g_0 - 1 + e_0).$$

Because g_0 is large, τ_1 will also be large. For example, if $g_0 = 3$ and $e_0 = 0.5$ (an elasticity in the mid range of empirical estimates), then $\tau_1 = 2/2.5 = 80$ percent—a very high phase-out rate. Formula (A2) is the optimal marginal tax rate at zero earnings in the standard Mirrlees (1971) model when there is a fraction of individuals who do not work, which is the most realistic case (this result does not seem to have been noticed in the literature). As is well known since Seade (1977), the optimal bottom tax rate is zero when everybody works and bottom earnings are strictly positive, but this case is not practically relevant.

Derivation of the optimal bottom marginal tax rate with extensive labor supply responses (Figure A3)

Consider now a model where behavioral responses of low- and mid-income earners take place through the extensive elasticity only—i.e., whether or not to work—and that earnings when working do not respond to marginal tax rates. As depicted on Figure A3, suppose the government starts from a transfer scheme with a positive phase-out rate τ_1 and introduces an additional small in-work benefit Δc_1 that increases net transfers to low-income workers earning z_1. Let h_1 be the fraction of low-income workers with earnings z_1. Let us denote by e_1 the elasticity of h_1 with respect to the participation net-of-tax rate $1 - \tau_1$, so that $e_1 = (1 - \tau_1)/h_1 \Delta h_1/\Delta(1 - \tau_1)$. The reform has again three effects. First, the reform has a mechanical fiscal cost $\Delta M = -h_1 \Delta c_1$ for the government. Second, it generates a social welfare gain, $\Delta W = g_1 h_1 \Delta c_1$ where g_1 is the marginal social welfare weight on low-income workers with earnings z_1. Third, there is a tax revenue gain due to behavioral responses $\Delta B = \tau_1 z_1 \Delta h_1 = e_1 \tau_1/(1 - \tau_1)h_1 \Delta c_1$. If $g_1 > 1$, then $\Delta W + \Delta M > 0$. In that case, if $\tau_1 > 0$, then $\Delta B > 0$ implying that $\tau_1 > 0$ cannot be optimal. The optimal τ_1 is such that $\Delta M + \Delta W + \Delta B = 0$ implying that

A3)
$$\tau_1 /(1 - \tau_1) = (1 - g_1)/e_1$$

or

$$\tau_1 = (1 - g_1)/(1 - g_1 + e_1).$$

Figure A3

Optimal Bottom Marginal Tax Rate with Extensive Labor Supply Responses
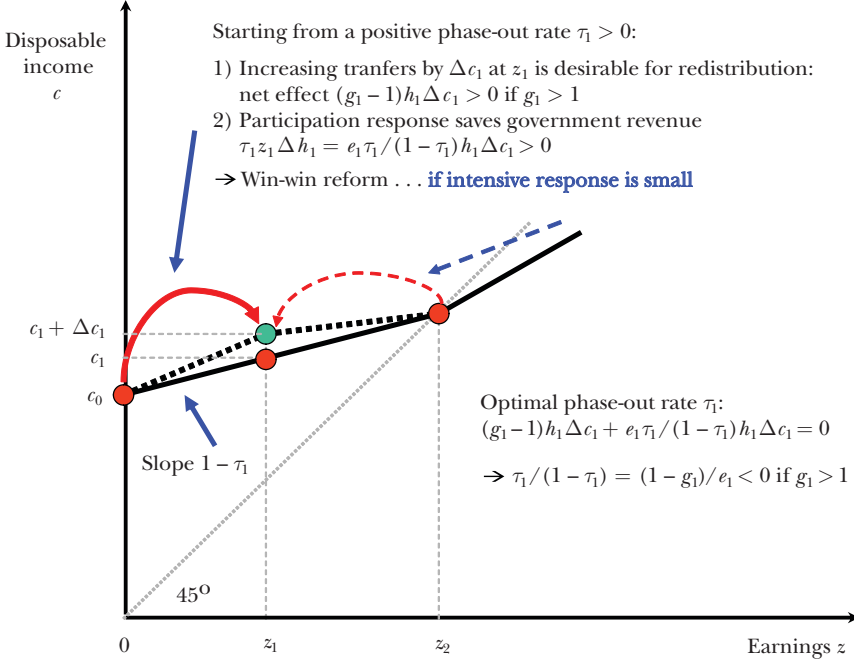
Notes: The figure depicts the derivation of the optimal marginal tax rate at the bottom in the discrete model with labor supply responses along the extensive margin only. Starting with a positive phase-out rate $\tau_1 > 0$, the government introduces a small in-work benefit Δc_1. Let h_1 be the fraction of low-income workers with earnings z_1, and let e_1 be the elasticity of h_1 with respect to the participation net-of-tax rate $1 - \tau_1$. The reform has three standard effects: mechanical fiscal cost $\Delta M = -h_1 \Delta c_1$, social welfare gain, $\Delta W = g_1 h_1 \Delta c_1$, and tax revenue gain due to behavioral responses $\Delta B = \tau_1 z_1 \Delta h_1 = e_1 (\tau_1 / (1 - \tau_1))$ $\times h_1 \Delta c_1$. If $g_1 > 1$, then $\Delta W + \Delta M > 0$. If $\tau_1 > 0$, then $\Delta B > 0$ implying that $\tau_1 > 0$ cannot be optimal. The optimal τ_1 is such that $\Delta M + \Delta W + \Delta B = 0$ implying that $\tau_1 / (1 - \tau_1) = (1 - g_1)/e_1$.

Derivation of the optimal asymptotic marginal tax rate with a random finite population

Actual distributions are both bounded and with a finite population that becomes progressively sparser in the upper tail. Moreover, the government does not know the exact realization of the earnings distribution when setting tax policy. Hence, the (known) bounded and finite model of optimal taxation does not seem practically useful for thinking about tax rates on top earners.

A more realistic scenario is that the government knows the distribution of realized earnings (conditional on a given tax policy). A simple way to model this is to assume that individual skills n are drawn from a known Pareto distribution that is unbounded and with density $f(n)$. Any finite draw will generate a distribution that

is both bounded and finite. We retain the key assumption that individuals know their skill n before making their labor supply decision.

As the government does not know the exact draw ex ante, a natural objective of the government is to maximize expected social welfare $SWF = \int G(u(n))f(n)\,dn$. The budget constraint for any particular draw is not met with an ex ante tax function. However, if the population is large enough, the actual budget is close to the expected budget as long as the share of income accruing to the very top earners is not too large. Hence, it is natural to assume that the budget constraint needs to hold in expectation $\int T(z)f(n)\,dn \geq 0$. Small fluctuations in debt from repeated realizations over time would justify use of a similar approximation. This replacement of the actual budget constraint by the expected budget constraint is the key point that generalizes the Mirrlees (1971) model to finite populations. Therefore, we are exactly back to the Mirrlees (1971) model, and hence the optimal tax system is given by the standard formulas.

This in particular implies that the optimal top tax rate $\tau^* = 1/(1 + ae)$ continues to apply with a the Pareto parameter of the expected earnings distribution. More concretely and coming back to our derivation presented in the text, recall that the optimal constant tax rate above z^* is given by formula $\tau^* = 1/(1 + ae)$ with $a = z_m/z^*/(z_m/z^* - 1)$ and z_m is the average income above z^*. Obviously, z_m (and hence a) are not defined if z^* is above the actual realized top. However, if the actual finite draw is unknown to the government when τ^* is set, the government should naturally replace z_m/z^* by Ez_m/z^*, i.e., the expected average income above z^* divided by z^*. Given the very close fit of the Pareto distribution up to the very top of the distribution (something that can actually be verified with actual rich lists that have been compiled by the press in a number of cases), the natural assumption is that Ez_m/z^* never converges to one.

Appendix Reference

Piketty, Thomas. 1997. "La Redistribution Fiscale face au Chomage." *Revue Francaise d'Economie*, 12(1): 157–201.