A Complete Example of an Optimal Two-Bracket Income Tax

Jean-François Wen

Department of Economics

University of Calgary

March 6, 2014

Abstract

I provide a simple model that is solved analytically to yield tidy expressions for the Pareto efficient tax structures and the optimal two-backet marginal tax rates. It is for the special case of equally-sized groups of two skill types and no exogenous spending requirements of the government. The results and the exposition give a self-contained treatment of the central ideas of optimal income taxation.

JEL codes: A22, A23, H21

Keywords: economic education, optimal taxation

1 Introduction

The theory of optimal income taxation examines the tradeoff between equity and efficiency in designing the personal income tax schedule. It forms a cornerstone of modern public finance and garnered a Nobel Prize for James Mirrlees and William Vickery in 1994. The subject is very much alive today in both public discourse and academic research. For example, French president Francois Hollande kept to his campaign pledge by imposing a 75 percent tax on earnings over 1 million euros at the end of 2013, while a study by Diamond and Saez (2011) suggests that the top optimal income tax rate in the United States should be 73 percent, which is much higher than the current top rate.

The theory of optimal income taxation appears in undergraduate textbooks (e.g. Rosen et al. 2012) and in more advanced textbooks (e.g. Salanié 2011). However, the mathematical treatment of optimal income taxation relies on the theory of optimal control, which is often not taught in economics curriculla. A simple pedagogical example can therefore be useful for demonstrating some key ideas of the theory. In this spirit, Slemrod et al. (1994) use indifference curve diagrams to construct Pareto efficient tax structures and they provide a *numerical* simulation of the social welfare maximizing tax rates, when only two tax brackets are permitted and there are two (equally-sized) classes of workers ("high" and "low" skill); and there are no exogenous spending requirements.[1] This is an instructive example, but

[1]The authors also provide optimal tax simulations for the case of a large number of skill classes. The restriction to two tax brackets is advocated by Slemrod et al. as a means

the reliance on numerical solutions may be less satisfying or convincing for students than an algebraic result. Hence, in this article, I construct a similar two-bracket example but with a complete analytical solution. In deriving the social optimum two-bracket tax structure, I use and explain Slemrod et al.'s proposition regarding Pareto efficient tax structures. I derive the Pareto efficient tax schedules for the increasing marginal tax rate case and the decreasing marginal tax rate case, and then I show that the social welfare maximizing tax schedule has decreasing marginal tax rates, regardless of the relative weight attached to the low-skilled group. The expressions for the optimal tax rates are very simple.

My example model follows the typical optimal tax setup. A worker is endowed with a given skill level which is reflected in the wage she receives. The government would like to engage in redistributive taxation from the high-skilled to the low-skilled, but is assumed to be unable to condition tax liabilities directly on skill level. Hence, the second fundamental welfare theorem is off the table. Taxing a worker's observed income conflates the endowment of skill (wage) with the choice of hours of labor effort, thereby distorting the labor-leisure margin and inducing excess burden. In this context, the literature has produced several general insights on the structure of the optimal tax schedule. In particular, there is the famous "zero marginal tax rate at the top" result and the somewhat less known "zero marginal tax rate at the bottom" result.[2] My example focuses on these two insights.

of economizing on taxpayer compliance costs.

[2]These results necessarily hold only when certain boundedness conditions apply on the

2 Policy Insights

I briefly discuss the two main insights before turning to the illustrative model.

2.1 Zero at the Top

If no one in the population has an income (or more precisely a wage rate) higher than a certain level, then the marginal tax rate should be zero at the top of the income scale (Mirrlees 1971, Seade, 1977). Suppose, on the contrary, that the top income tax rate is positive. Lowering the rate a little bit increases the top earner's reward, which will induce him to work a little more. He will be better off (by revealed preference) and more tax revenue will be generated, which can be used to redistribute income to lower earners. This argument can be repeated until the top marginal tax rate is zero. A key assumption for the result is that the highest attainable skill level (wage) is known. When this is not the case, simulations by Mirrlees (1971) suggest that, while the optimal top tax rate may not be zero, the optimal tax structure is characterized by decreasing marginal tax rates near the top. A marginal tax rate schedule with this "degressive" feature was introduced in the Swiss cantons of Schaffhausen in 2004 and Obwalden in 2006, although degressive income taxation was rejected by the Swiss federal court as unconstitutional in 2007.

income distribution. This will be clarified below. The practical importance of these two results is disputed since they provide little guidance on the structure of optimal taxation outside the top and bottom extremes. However, students of public finance must at least understand the results.

2.2 Zero at the Bottom

If no one in the population earns zero income (i.e., no one is idle) in the optimal arrangement, then the marginal tax rate should be zero at the bottom of the income scale (Seade 1977). Suppose, on the contrary, that the bottom rate is positive. The revenue raised by this marginal tax rate cannot be used to finance redistribution downwards because there is no one further down the scale. Hence, the revenue loss has no equity implication, so that only efficiency matters for the marginal tax rate at the bottom. This implies the bottom tax rate should be zero. The key assumption here is that everyone earns positive income under the optimal tax scheme, which in turn requires the wage rate at the bottom of the income scale not to be too low. Otherwise, it may be optimal to tolerate idleness among the least productive workers in order to generate higher tax revenues from the more productive workers by setting a positive the marginal tax rate at the bottom.

2.2.1 Pareto Efficient Taxation

When there are only two marginal tax rates allowed in the system, then obviously "zero at the top" and "zero at the bottom" cannot both hold simultaneously, as there would be no tax revenue. As we shall see, Pareto efficient tax structures in a two-bracket tax system will feature one or the other of these two "end of the scale" results. A Pareto efficient tax structure means that there is no change in the tax system that preserves budgetary balance while making at least one person better off without harming others.

Any tax structure that is not Pareto efficient obviously cannot be social welfare maximizing. The specific social welfare function and the distribution of skill levels then determine which of the Pareto optimal tax structures is socially best.

3 A Simple Model

3.0.2 Preferences

Suppose the utility function is quasi-linear in consumption (c) and labor (L):[3]

$$u(c, L) = \ln c - \varepsilon L. \tag{1}$$

The budget constraint of an individual requires consumption expenditures to equal after-tax income:

$$c = m - T(m) \tag{2}$$

where $m = wL$ and $T(m)$ is the tax liability at the income level m and the price index for consumption goods has been normalized to equal one. A useful device for optimal income tax analysis is to rewrite the utility function in terms of consumption and before-tax income (instead of labor effort) by

[3]This is equivalent to the utility function $u(c, l) = \ln c - \varepsilon(1 - l)$, where l is leisure, after substituting for l using the time constraint $l + L = 1$. For the utility function (1), the elasticity of substitution between labor and consumption is 1. The uncompensated elasticity of labor supply is 0, while the compensated elasticity of labor supply is -1. The utility function (1) differs from the one used by Slemrod et al. (1994).

substituting m/w for L into (1) to obtain

$$u(c, m/w) = \ln c - \varepsilon m/w \tag{3}$$

where the wage rate w is treated as a parameter of the utility function and m is a choice variable. Note that utility is decreasing in before-tax income because a higher level of m requires more labor effort. On a diagram with c on the vertical axis and m on the horizontal axis the direction of increasing utility is toward the north-west, where consumption is highest and before-tax income—i.e., labor effort—is lowest. Each indifference curve slopes upward at an increasing rate. That is, the marginal rate of substitution between c and m is given by

$$MRS_{m,c} = \frac{dc}{dm} = \frac{\varepsilon/w}{1/c} = \frac{\varepsilon c}{w} > 0$$

and the total derivative of the $MRS_{m,c}$ with respect to m is

$$\frac{d\left(dc/dm\right)}{dm} = \frac{\varepsilon}{w}\frac{dc}{dm} = \left(\frac{\varepsilon}{w}\right)^2 c > 0.$$

3.0.3 Single-Crossing Property

An important property of the transformed utility function (3) is the so-called single-crossing property, which will enable the social planner to design the tax schedule in a manner that separates the different skill types in the equilibrium. The single-crossing property states that the indifference curves are less steep the higher the wage:

$$\frac{\partial}{\partial w}\left(\frac{dc}{dm}\right) = -\frac{\varepsilon c}{w^2} < 0. \tag{4}$$

Figure 1 compares indifference curves for a high- versus low-skilled worker that pass through a given point (c', m').

3.0.4 Two-Bracket Tax Schedule

Suppose the tax system is restricted to having two marginal tax rates: τ_1 on income below a threshold of $\overline{m}$ and τ_2 on incomes above the threshold. To allow for redistribution a demogrant (i.e. a lump-sum transfer) B is paid to everyone. Thus the two-bracket tax system is described by

$$T(m) = -B + \tau_1 m, \text{ if } m < \overline{m} \tag{5}$$

$$T(m) = -B + \tau_1 \overline{m} + \tau_2 (m - \overline{m}), \text{ if m} \geq \overline{m}. \tag{6}$$

Figure 2 illustrates the after-tax budget constraint (2) for an individual in the space of c and m. Its slope is $dc/dm = 1 - T'(m)$, where $T'(m)$ is the marginal tax rate, which equals τ_1 for $m < \overline{m}$ and equals τ_2 for $m \geq \overline{m}$. In the absence of taxation the budget constraint would be given by the 45 degree ray from the origin. For example, earning m_B would enable a consumption level c_B as indicated by the point B''. However, as a result of taxation, the income m_B buys the lesser amount $\hat{c}_B$ as shown by point B. Hence, the vertical difference

BB'' is the tax revenue collected from an individual earning m_B. Notice that this is equivalent to the horizontal distance to the 45 degree line indicated by BB'. Now consider a different income level, m_A, which in Figure 2 lies to the left of where the after-tax budget constraint crosses the 45 degree line. An individual earning m_A has a negative tax liability. That is, he receives a positive net transfer $(B - \tau_1 m_A > 0)$ from the government, equal to the horizontal distance AA'.

3.0.5 Social Welfare Maximization

For any given tax schedule an individual maximizes utility by finding the point in the space (c, m) that puts her on the highest feasible indifference curve. I shall assume that there are only two skill levels, high (H) and low (L), characterized by their wages $w_H > w_L$. Given a tax schedule, denote the corresponding utility maximizing income levels of the high- and low-skilled workers as m_H^* and m_L^* and the optimal consumption levels as c_H^* and c_L^*. If there are n_H high-skilled workers and n_L low skilled workers and the government has an exogenous revenue requirement R then the equation for budgetary balance is

$$n_H T(m_H^*) + n_L T(m_L^*) = R. \tag{7}$$

Assume that social welfare is given by the function

$$\omega = n_H u(c_H^*, m_H^*/w_H) + \delta n_L u(c_L^*, m_L^*/w^L) \tag{8}$$

where δ is a welfare weight on the utility of low-skilled individuals. The objective of optimal income taxation in this example is to choose the vector $\Theta = \{\tau_1, \tau_2, \overline{m}, B\}$ to maximize (8) subject to (7). I will impose further that $R = 0$ and $n^H = n^L$. The first assumption means that the only motive for taxation is redistribution. This assumption combined with the second one allows me to easily depict budgetary balance in a diagram.

4 Pareto Efficient Tax Structures

To characterize the optimal two-bracket income tax system, I begin by identifying the characteristics of *Pareto efficient* income tax structures. I shall only consider progressive tax structures with non-negative marginal tax rates ($\tau_1 \geq 0$, $\tau_2 \geq 0$) and a positive demogrant ($B > 0$).[4] *A progressive tax structure is one where the share of income paid in taxes is rising with income*: $\frac{d[T(m)/m]}{dm} > 0$. The two-bracket system is progressive if $(\tau_1 - \tau_2)\overline{m} < B$. A basic proposition of Slemrod, Yitzhaki, Mayshar, and Lundholm (1994), which tailors the analysis of Sadka (1976) and Stiglitz (1982) to a two-bracket tax system, is the following.

[4] It is possible for an optimal tax structure to consist of negative marginal tax rates and a negative demogrant. These are not seen in reality and I ignore the possibility.

Let m_L^* denote the before-tax income of the low-skilled at the low-skilled consumer's optimum point, and let u_H^* denote the indifference curve of the high-skilled corresponding to the high-skilled's consumer optimum point. Then the socially optimal two-bracket income tax system $\Theta = (\tau_1, \tau_2, \overline{m}, B)$ can be restricted to one of the following two cases:

(1) **Decreasing marginal tax rates**

(1) $\tau_1 > \tau_2 = 0$, with u_H^* touching both branches of the budget constraint.

(2) **Increasing marginal tax rates**

(2) $0 = \tau_1 < \tau_2$, with $m_L^* = \overline{m}$

Part (1) of the proposition is a case of a *progressive decreasing* marginal tax rate schedule ($\tau_1 > \tau_2$). Part (2) is a case of a *progressive increasing* marginal tax rate schedule ($\tau_1 < \tau_2$). Both cases are progressive if the low-skilled are net transfer recipients and the high-skilled are net taxpayers (assuming there is no government spending except for the demogrant). Each part of the proposition contains two statements: a statement about which tax rate is optimally set to zero; and a statement about tangency conditions between an indifference curve and the budget constraint. Figure 3 illustrates Part (1) of the proposition and figure 4 illustrates Part (2).

Each part of the proposition can be proven with diagrams similar to figures 3 and 4. In order to make this exposition self-contained, I provide the geometric proofs in the appendix, although they are available in Slemrod et al. (1994). I now use the proposition to derive analytical expressions for the Pareto efficient tax structures and the optimal two-bracket marginal tax rates.

5 An Algebraic Solution

I now characterize algebraically the two Pareto efficient tax structures for the model and then I will determine which one generates the highest social welfare.

5.1 Decreasing Marginal Tax Rates ($\tau_1 > \tau_2 = 0$)

From Part (1) of the proposition for Pareto efficient tax rates, we can impose $\tau_2 = 0$. Then the low-skilled's utility maximization problem is to find

$$u_L^* = \max_{m_L} \ln\left[(1-\tau_1)\, m_L + B\right] - \varepsilon \frac{m_L}{w_L} \tag{9}$$

giving the solution[5]

$$m_L^* = \frac{w_L}{\varepsilon} - \frac{B}{1-\tau_1}. \tag{10}$$

Substituting (10) back into the utility function (9) gives an expression for the indirect utility function. To simplify the notation, from herein I shall

[5] We can verify later that the optimal tax structure indeed implies that $m_L \geqslant 0$.

normalize the low-skilled wage to unity: $w_L = 1$. Hence

$$u_L^* = -\ln(1+\varepsilon) + \ln(1-\tau_1) + \varepsilon B/(1-\tau_1). \tag{11}$$

Recall that for the high-skilled we need to find two tangency points between an indifference curve and the budget constraint. One tangency point is on the segment with the marginal tax rate $\tau_1 > 0$; the other tangency point is along the segment with $\tau_2 = 0$.

5.1.1 First Segment Tangency

Analogous to the low-skilled's optimization problem (9) the high-skilled worker solves

$$u_H^1 = \max_{mH} \ln\left[(1-\tau_1)\, m_H + B\right] - \varepsilon \frac{m_H}{w_H}, \tag{12}$$

yielding

$$m_H^1 = \frac{w_H}{\varepsilon} - \frac{B}{1-\tau_1} \tag{13}$$

and

$$u_H^1 = -\ln(1+\varepsilon) + \ln(1-\tau_1) + \ln w_H + \varepsilon \frac{B}{(1-\tau_1) w_H}, \tag{14}$$

where I have replaced the superscript asterisk on u_H^* and m_H^* with "1" to indicate that these are the levels of utility and before-tax income corresponding to the first segment of the budget constraint.

5.1.2 Second Segment Tangency

Along the second segment of the budget constraint $\tau_2 = 0$ and the optimal income m_H^2 exceeds the threshold $\overline{m}$. The utility maximization problem is

$$u_H^2 = \max \ln\left(m_H - \tau_1 \overline{m} + B\right) - \varepsilon \frac{m_H}{w_H}. \tag{15}$$

The solution is

$$m_H^2 = \frac{w_H}{\varepsilon} + \tau_1 \overline{m} - B, \tag{16}$$

where again the superscript "2" on m_H indicates optimality on the second segment of the budget constraint. After substituting (16) into (15) and simplifying the expression, the indirect utility function is

$$u_H^2 = -\ln(1+\varepsilon) + \ln w_H + \varepsilon \frac{B - \tau_1 \overline{m}}{w_H}. \tag{17}$$

5.1.3 Double Tangency Condition

The next step is to equate the utilities u_H^1 and u_H^2 since they are associated with the same indifference curve. After some rearrangement of terms, I obtain one of the key equations of the solution to the efficient decreasing marginal tax rates.

$$\ln(1 - \tau_1) = -\frac{\varepsilon}{w_H}\left(\frac{\tau_1 B}{1 - \tau_1} + \tau_1 \overline{m}\right). \tag{18}$$

5.1.4 Budget Balance

The second key equation for the solution is the requirement that government expenditures equal tax revenues, when the low-skill have income m_L^* given by (10) (with $w_L = 1$) and the high-skill have income m_H^2 given by (16). Thus, $T(m_L^*) + T(m_H^2) = 0$ can be written, after some rearrangement of terms, as

$$B = \tau_1(1/\varepsilon + \overline{m})\frac{(1-\tau_1)}{(2-\tau_1)}. \tag{19}$$

5.1.5 Pareto Efficient Decreasing Tax Rates

I can now use the equations (19) and (18) to derive expressions for B and $\overline{m}$ in terms of τ_1 and the parameters of the model (w_H and ε). After some tedious algebra, I obtain

$$\overline{m} = -\frac{\tau_1}{2\varepsilon} - \frac{w_H}{2\varepsilon}\frac{(2-\tau_1)}{\tau_1} \times \ln(1-\tau_1) \tag{20}$$

$$B = \frac{\tau_1(1-\tau_1)}{\varepsilon(2-\tau_1)} - \frac{(\tau_1)^2(1-\tau_1)}{2\varepsilon(2-\tau_1)} - \frac{(1-\tau_1)w_H}{2\varepsilon} \times \ln(1-\tau_1). \tag{21}$$

Finally, the solution for the Pareto efficient value of $\tau_1 > 0$ can be completed by choosing τ_1 to maximize the indirect utility function of a representative low-skilled individual u_L^* given by (11), after substituting the right-hand side of (21) for B. This is equivalent to

$$u_L^{**} = \max_{\tau_1}(2-w_H) \times \ln(1-\tau_1) + \tau_1 \tag{22}$$

yielding the elegant solution

$$\tau_1^* = w_H - 1 \tag{23}$$

which is positive as $w_H > 1$ by assumption. To ensure that $\tau_1^* < 1$, an upper bound $w_H < 2$ is required on the high-skilled wage. Hence, $1 = w_L < w_H < 2$ is imposed on the model. Notice that as w_H approaches 1 the optimal tax rate τ_1^* goes to zero. That is, as inequality vanishes, so does the need for redistribution.

The solutions for $\overline{m}$ and B follow from substituting τ_1^{**} into (20) and (21).

$$\overline{m}^* = -\frac{1}{2\varepsilon}\left(w_H \times \frac{(3-w_H)}{(w_H-1)} \times \ln(2-w_H) + (w_H - 1)\right) \tag{24}$$

$$B^* = \frac{(2-w_H)}{2\varepsilon}\left((w_H-1) - w_H \ln(2-w_H)\right). \tag{25}$$

5.1.6 Low-Skilled Labor Supply

Before proceding to the welfare calculation, we must verify whether the low-skilled workers engage in non-negative labor supply at the proposed solution. That is, we must check that $m_L^* \geq 0$ in (10) after setting $w_L = 1$, $\tau_1^* = w_H - 1$, and substituting for B using (25). Low-skilled workers supply non-negative labor until the high-skilled wage w_H reaches approximately 1.59.[6] When w_H exceeds this amount, $m_L^* = 0$, since negative labor supply is meaningless. This means that our solution is only applicable for high-skilled wages below

[6] The critical value for w_H is given by the equation $w_H(1 - \ln(2 - w_H)) = 3$.

the critical point. For higher values of w_H, the demogrant and optimal first-bracket tax rate τ_1^* must be recalculated based only on the income m_H^* of the high-skilled. I shall ignore this discontinuity in the solution by further restricting the high-skilled wage to be below the critical level, call it w_H^c, such that low-skilled workers supply positive labor. This implies a critical maximum value for $\tau_1^* = w_H^c - 1$.

5.1.7 Social Welfare with Decreasing Tax Rates

The optimized utility levels are obtained by substituting the solution for τ_1^* into (22) to obtain u_L^{**} and the solutions for τ_1^*, $\overline{m}$, and B into (17). This gives

$$u_L^{**} = -(1 + \ln \varepsilon) + \frac{(w_H - 1)}{2} + \frac{(2 - w_H)}{2} \times \ln(2 - w_H) \qquad (26)$$

$$u_H^{**} = u_L^{**} + \ln w_H. \qquad (27)$$

Social welfare under a Pareto efficient decreasing tax structure is then

$$\omega^{**} = u_H^{**} + \delta u_L^{**} = (1 + \delta) u_L^{**} + \ln w_H \qquad (28)$$

$$= (1 + \delta) \left[-(1 + \ln \varepsilon) + \frac{(w_H - 1)}{2} + \frac{(2 - w_H)}{2} \times \ln(2 - w_H) \right] + \ln w_H. \qquad (29)$$

5.2 Increasing Marginal Tax Rates ($0 = \tau_1 < \tau_2$)

From Part (2) of the proposition for Pareto efficient tax rates, we can impose $\tau_1 = 0$. The low-skilled's utility maximization problem is

$$u_L^* = \max_{m_L} \ln (m_L + B) - \varepsilon \frac{m_L}{w_L}, \tag{30}$$

which, after inserting $w_L = 1$, yields,

$$m_L^* = \frac{1}{\varepsilon} - B \tag{31}$$

and

$$u_L^* = -(1 + \ln \varepsilon) + \varepsilon B \tag{32}$$

The high-skilled worker pays a tax rate of τ_2 only on income exceeding $\overline{m}$. Her objective is

$$u_H^* = \max_{m_H} \ln ((1 - \tau_2) m_H + \tau_2 \overline{m} + B) - \varepsilon \frac{m_H}{w_H}, \tag{33}$$

resulting in

$$m_H^* = \frac{w_H}{\varepsilon} - \frac{\tau_2 \overline{m}}{1 - \tau_2} - \frac{B}{1 - \tau_2} \tag{34}$$

and

$$u_H^* = -(1 + \ln \varepsilon) + \ln w_H + \ln(1 - \tau_2) + \frac{\varepsilon \overline{m} \tau_2}{(1 - \tau_2) w_H} + \frac{\varepsilon B}{(1 - \tau_2) w_H}. \tag{35}$$

5.2.1 Budgetary Balance and Tangency Condition

Budgetary balance requires $T(m_L^*)+T(m_H^*)=0$ where m_L^* and m_H^* are given by (31) and (34). This implies

$$\tau_2 m_H^* - \tau_2 \overline{m} = 2B. \tag{36}$$

From Part (2) of the proposition, it is required that $m_L^* = \overline{m}$. Combining this requirement with (36), and using the expression for m_H^*, gives

$$\tau_2 \left(\frac{w_H}{\varepsilon} - \frac{\tau_2 \overline{m}}{1-\tau_2} - \frac{B}{1-\tau_2} \right) - \tau_2 \left(\frac{1}{\varepsilon} - B \right) = 2B, \tag{37}$$

which can be rearranged to obtain a second expression for $\overline{m}$:

$$\overline{m} = \frac{(w_H - 1)(1-\tau_2)}{\varepsilon \tau_2} - B \left(\frac{(\tau_2)^2 - 2\tau_2 + 2}{(\tau_2)^2} \right). \tag{38}$$

The expression (38) for $\overline{m}$ can be equated to m_L^*, given the tangency condition of Part (2), to deliver a solution for B in terms of τ_2:

$$B = \frac{(w_H - 1)\tau_2}{2\varepsilon} - \frac{(\tau_2)^2}{2\varepsilon(1-\tau_2)}. \tag{39}$$

Our final task for solving the optimal tax rate is to maximize the uility of the low-skilled, u_L^* given by equation (32), after substituting for B using (39).[7]

[7] Note that $m_L^* > 0$ when (39) is used to substitute for B in (31). Hence, the low-skilled supply positive labor effort in equilibrium.

The maximization problem is

$$u_L^{**} = \max_{\tau_2} \tau_2(w_H - 1) - \frac{(\tau_2)^2}{1-\tau_2}. \tag{40}$$

The solution to the first-order condition is

$$\tau_2 = 1 \pm \sqrt{1/w_H}. \tag{41}$$

Only the negative root is admissable as otherwise $\tau_2 > 1$. Hence, the optimal value of τ_2 is

$$\tau_2^* = 1 - (w_H)^{-1/2}. \tag{42}$$

Observe again that if inequality vanishes ($w_H \to 1$) then τ_2^* approaches zero, since the motive for redistribution disappears.

By substituting τ_2^* into (39) and manipulating the terms, I obtain the solution for B^*:

$$B^* = \frac{\left(w_H^{1/2} - 1\right)^2}{2\varepsilon}. \tag{43}$$

In turn, B^* is substituted into (31), to find the optimal threshold, using the fact that $\overline{m} = m_L^*$:

$$\overline{m} = \frac{2 - \left((w_H)^{1/2} - 1\right)^2}{2\varepsilon}. \tag{44}$$

Finally, the optimized values of utility and social welfare are as follows:

$$u_L^{**} = -(1+\ln\varepsilon) + \frac{\left((w_H)^{1/2}-1\right)^2}{2} \tag{45}$$

$$u_H^{**} = -(1+\ln\varepsilon) + (1/2)\ln w_H + \frac{(w_H-1)}{2w_H} \tag{46}$$

$$\begin{aligned}\omega^{**} &= -(1+\delta)(1+\ln\varepsilon) + (1/2)\ln w_H + \frac{(w_H-1)}{2w_H} \\ &\quad +\frac{\delta}{2}\left((w_H)^{1/2}-1\right)^2.\end{aligned} \tag{47}$$

5.3 Welfare Optimum: Decreasing or Increasing Marginal Tax Rates?

The Pareto efficient tax structures have been solved for, but which one yields the highest level of social welfare? The answer requires a comparison of the social welfare values for the decreasing marginal tax rate (DMRT) case, given by (29), and the increasing marginal tax rate case (IMRT), given by (47). Thus,

$$\begin{aligned}\Delta &= \omega^{**}(\text{DMRT})\text{-}\omega^{**}(\text{IMRT}) && (48)\\ &= (1+\delta)\left[\frac{(w_H-1)}{2} + \frac{(2-w_H)}{2}\times\ln(2-w_H)\right] \\ &\quad +(1/2)\ln w_H - \frac{(w_H-1)}{2w_H} - \frac{\delta}{2}\left((w_H)^{1/2}-1\right)^2 \\ &= (1/2)\left[(w_H-1)^2/w_H + (2-w_H)\ln(2-w_H) + \ln w_H\right] && (49)\\ &\quad +(\delta/2)\left[(w_H-1) - \left((w_H)^{1/2}-1\right)^2 + (2-w_H)\ln(2-w_H)\right].\end{aligned}$$

An inspection of (49) reveals that every term in it is positive. Hence, the optimal decreasing marginal tax rate structure generates the highest social welfare, regardless of the size of the welfare weight δ. The "zero marginal tax rate at the top" case prevails in this example, which corroborates the observed tendency often reported by researchers using numerical simulations, that the welfare optimum features declining marginal tax rates near the top of the income distribution.

6 Conclusion

In this article, I have provided a simple model and demonstrated step-by-step how to solve it analytically to yield Pareto efficient tax structures and the social welfare maximizing tax policy. It is for the special case of two marginal tax rates, equally-sized groups of two skill types, and no exogenous spending requirements of the government. The results and the exposition give a self-contained treatment of the central ideas of optimal income taxation. The presentation builds on the analysis of Slemrod et al. (1994). In contrast to that paper, however, I present closed form solutions for the optimal tax rates. Closed form solutions to optimal income tax problems are rare in the literature, which usually deals with more complicated versions of the model. The optimal tax structure that arises from the analysis confirms the "zero at the bottom" result that is commonly found in optimal tax models. The current media focus on income inequality and taxation shows that the optimal income tax problem is very much alive in public debate.

References

[1] Diamond, P. and Emmanuel S. 2011. The case for a progressive tax: from basic research to policy recommendations. *Journal of Economic Perspectives* 25(4): 165-90.

[2] Mirrlees, J. A. 1971. An exploration in the theory of optimal income taxation. *Review of Economic Studies* 38(2): 175–208.

[3] Rosen, H.S., Wen, J.-F., Snoddon, T. 2012. *Public finance in Canada*. Toronto: McGraw-Hill Ryerson.

[4] Sadka, E. 1976. On income distribution, incentive effects and optimal income taxation. *Review of Economic Studies* 43(2): 261–67.

[5] Salanié, B. 2011, *The economics of taxation*. Cambridge, MA: MIT Press

[6] Seade, J. K. 1977. On the shape of optimal tax schedules. *Journal of Public Economics* 7(1): 203–36.

[7] Slemrod, J., Yitzhaki, S., Mayshar, J. and Lundholm, M. 1994. The optimal two-bracket linear income tax. *Journal of Public Economics* 53(2): 269–290.

[8] Stern, N.H. 1976. On the specification of models of optimal income taxation. *Journal of Public Economics* 6: 123–162.

[9] Stiglitz, J. 1982. Self-selection pareto efficient taxation. *Journal of Public Economics* 17: 213–40.

7 Appendix

7.1 Pareto Efficient Decreasing Marginal Tax Rates

7.1.1 Zero at the Top: Why $\tau_2 = 0$?

I will first show that if $\tau_1 > \tau_2$, then τ_2 must equal zero, as otherwise there is a Pareto-improving change to the income tax system. In figure 5 the initial budget constraint is given by OD'BE and has a kink at point D' (at income $\overline{m}'$), such that $\tau_1 \geq \tau_2 > 0$. Thus both marginal tax rates are positive along the segments forming OD'BE, contrary to the requirement of Part (1) for a Pareto efficient structure. The low-skilled locate at point A and the high-skilled locate at B, that is, where the respective indifference curves are tangent to the budget constraint. With our assumption of equal numbers high- and low-skilled workers, the net transfer to each low-skilled worker, shown as AA', must equal the net tax payment of each high-skilled worker, shown as BB'. Note that the single-crossing property is what ensures that a two-bracket schedule can be designed such that only the high-skilled locate on the second segment of the budget constraint, earning more before-tax income than do the low-skilled.

Now increase the cutoff for the first tax bracket $\overline{m}'$ to $\overline{m}$ and reduce τ_2 to zero. The budget constraint becomes ODBE', with the segment DBE' cutting through the previous tangency point B. The lower ability individuals remain at the same equilibrium point as before (point A) but the higher ability individuals are better off somewhere along the segment DBE' than they are at their previous optimum at B. In the illustration the high-skilled now choose point F on the segment DBE' where they earn more income and they attain higher utility. This change in τ_2 has no tax revenue implications: the horizontal distance between the 45 degree line and point F is, by construction, the same as the horizontal distance BB'. Thus it could not have been optimal to have a positive τ_2 whenever it is optimal to have $\tau_2 \leq \tau_1$. Notice that, even though $\tau_2 = 0$, the tax schedule embodied in the budget constraint ODBE' remains redistributive from the high-skilled toward the low-skilled.

In the construction of figure 5, I noted that the low-skilled are *no worse off* by reducing τ_2 to zero while increasing the cutoff $\overline{m}$, but the high-skilled are better off. However, the tax system can also be changed in such a way that *both* skill types are strictly better off when τ_2 is reduced to zero. Consider figure 6, where points B and F correspond to the points shown previously in figure 5. Suppose the cutoff is increased all the way to $\overline{m}''$ while setting $\tau_2 = 0$. The high-skilled locate at point G on the segment D"GE", where they are clearly better off than at point B. However, *at G there is more tax revenue than at B*, since G lies further to the right of the 45 degree line than point B. The additional tax revenue can be distributed to both skill types in the form of a larger demogrant, B. Geometrically, a larger B means the entire budget constraint is shifted horizontally to the left. The low- and high-skilled will adjust their labour supply along the new budget constraint (i.e., their optimal incomes, m_L^* and m_H^*). The process stops when budgetary balance is restored. In the final equilibrium, everyone is on a higher indifference curve than they began at (i.e. when $\tau_2 > 0$). The important point, however, is that Pareto improvements are possible until τ_2 is reduced to zero.

7.1.2 Tangency Condition

We can now focus on the tangency condition in Part (1) to fully characterize the optimal two-bracket income tax system with decreasing marginal tax rates. Figure 7 shows a tax schedule ODD'E, where $\tau_2 = 0$ but the indifference curve u_H^* does not touch both branches of the budget constraint, as required by the tangency condition of Part (1). The low-ability types locate at point A and the high-ability types locate at point C. It is then possible to increase the cutoff from $\overline{m}'$to $\overline{m}$ and at the same time to lower τ_1 such that: (i) the new segment O'AD' cuts through point A (i.e. where the low-skilled's indifference curve u_L^* is tangent to the line OAD) and (ii) the indifference curve u_H^* is now tangent to the lower branch O'AD' at point B and tangent to the upper branch D'CE at point C. Assuming the high-ability individuals continue to choose point C, they are unperturbed by the tax reform; but the low-ability individuals are strictly better off at some points on the branch O'AD' than they are at A. In particular, they face a lower marginal tax rate and they work more than before, for example at point F, where indifference curve $u_L^{*\prime}$ is tangent to the line O'AD'. The amount of net transfers they receive declines, because point F must lie to the right of point A, and, since the net taxes paid by the high-skilled has not changed, the government has extra money on its hands, which it can disburse as a higher demogrant. The higher demogrant in turn would shift the entire budget line O'FD'CE to the left in parallel way, so that all individuals will be strictly better off compared to the original tax schedule.

Thus any degressive optimal two-bracket tax structure with two ability types requires u_H^* to touch both branches of the budget constraint, as required by Part (1). It is not optimal to make the line segment O'AD' any steeper than what is depicted in the figure, that is, to reduce τ_1 any further, as this would induce the high-skilled workers to abandon point C on the second budget segment, in favour of a point on the first budget segment, where their tax payments would be lower and the government could no longer afford its existing level of redistribution. Notice, too, the role of the single-crossing property in this construction. It is the relative shallowness of the high-skilled's indifference curves that leads to a tangency on the segment O'FD'at point B while the low-skilled achieve a tangency along the same segment at point F. Consequently, in the equilibrium, the high-skilled earn more income than the low-skilled; and redistribution from those endowed with w_H toward those with w_L can occur through the income tax system even without the government being able to identify beforehand the skill levels of the different individuals.

7.2 Pareto Efficient Increasing Marginal Tax Rates

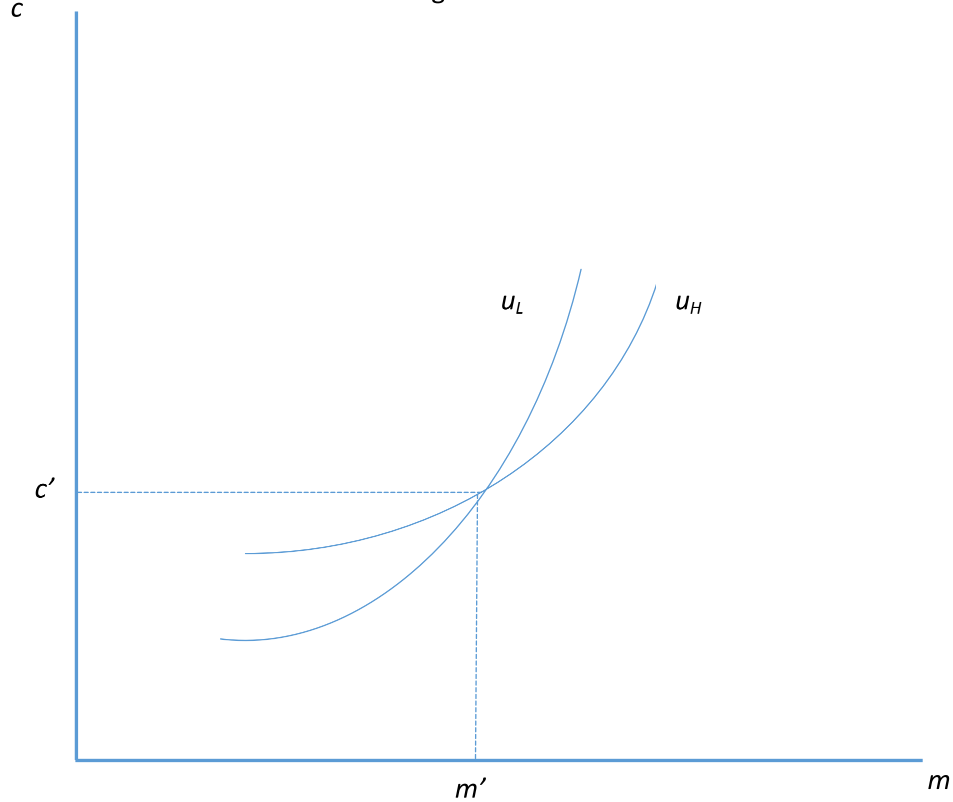
7.2.1 Zero at the Bottom: Why $\tau_1 = 0$?

Now let us see why the proposition requires $\tau_1 = 0$ when the marginal tax schedule is increasing. Figure 8 shows a situation where $\tau_2 \geq \tau_1 > 0$. The initial tax schedule is depicted as O'AD'E (with a kink at D'). In this case, one can lower the income threshold from $\overline{m}'$ to $\overline{m}$ and reduce τ_1 to zero, such that the budget constraint becomes OADE, with the segment OAD being parallel to the 45 degree line, reflecting $\tau_1 = 0$. This tax change will leave the high-skill individuals unperturbed at B while raising the utility of the low-skill individuals, without changing their tax benefits. The low-skilled now locate at point F on the line OAD, where their utility is higher than at point A. Thus Figure 8 establishes that an increasing two-bracket marginal tax rate schedule cannot be optimal unless the bottom rate is zero.

7.2.2 Tangency Condition

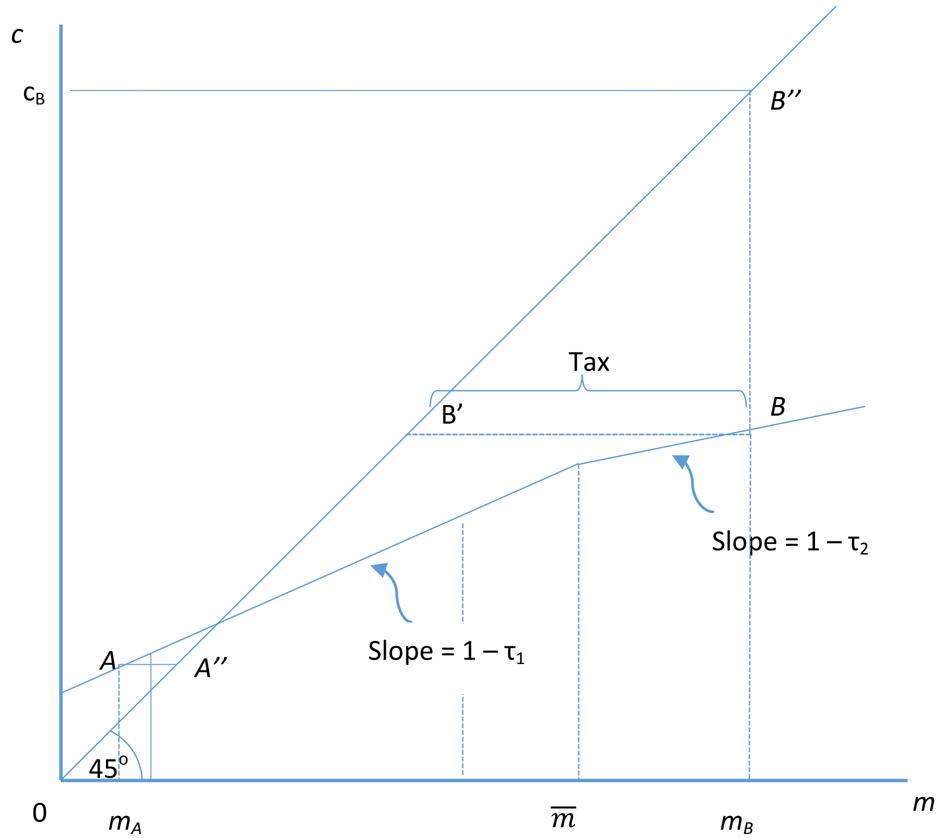
Pareto efficiency also requires that the low-skilled indifference curve be tangent to the budget constraint precisely at the cutoff income level $\overline{m}$. In figure 9 the schedule ODBE features $\tau_1 = 0$ but $m_L^* \neq \overline{m}$ thus violating the tangency condition of Part (2). As shown in figure 9, it is then possible to reduce $\overline{m}'$ to $\overline{m}$ such that the tax schedule now has a kink at point A. Low-ability individuals continue to locate at point A, where their indifference curve u_L^* is the highest achievable along OABE. Thus they get the same net transfer as before. At the same time, high-ability individuals become strictly better off by choosing a point such as B' along the budget constraint segment ABB'E' compared to their utility at point B. In particular, they face a lower marginal tax rate, they work more, pay more in net taxes, and they consume more. The additional tax payments can be used to raise the demogrant, which would shift OABB'E' to the left in a parallel fashion and make everyone better off. Thus $m_L^* = \overline{m}$ is a necessary condition for Pareto efficient tax structures, as required by Part (2).

Figure 1



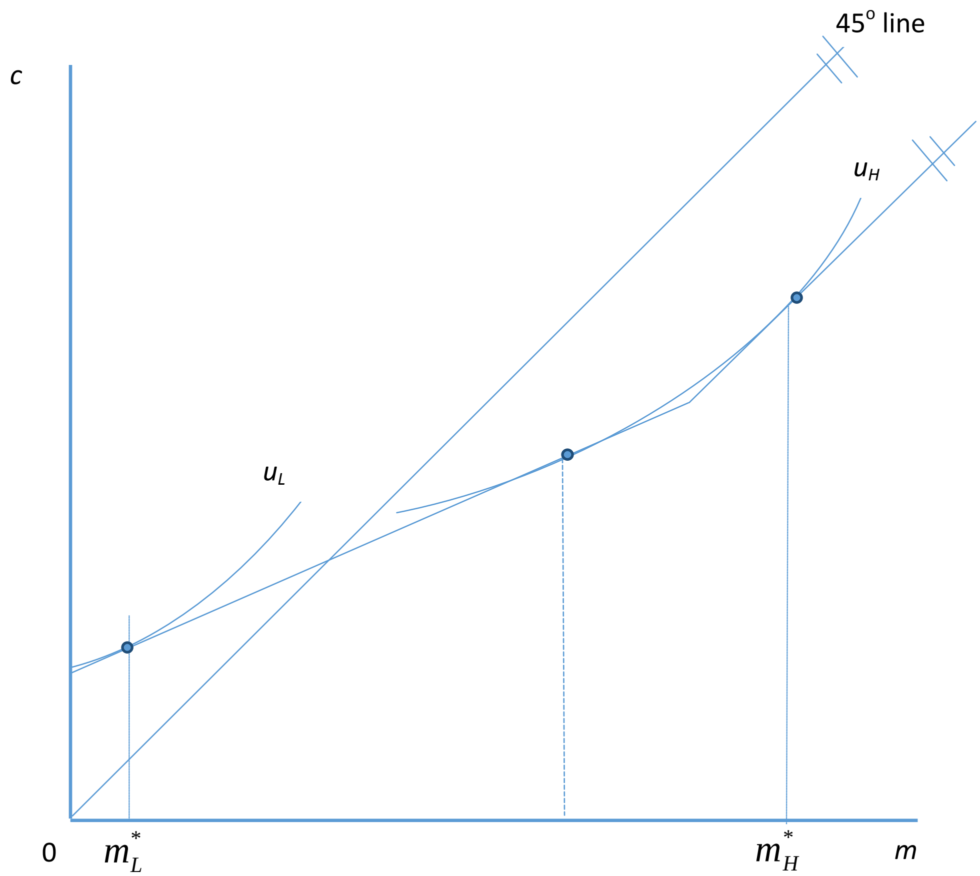
The indifference curves exhibit the single-crossing property.

Figure 2



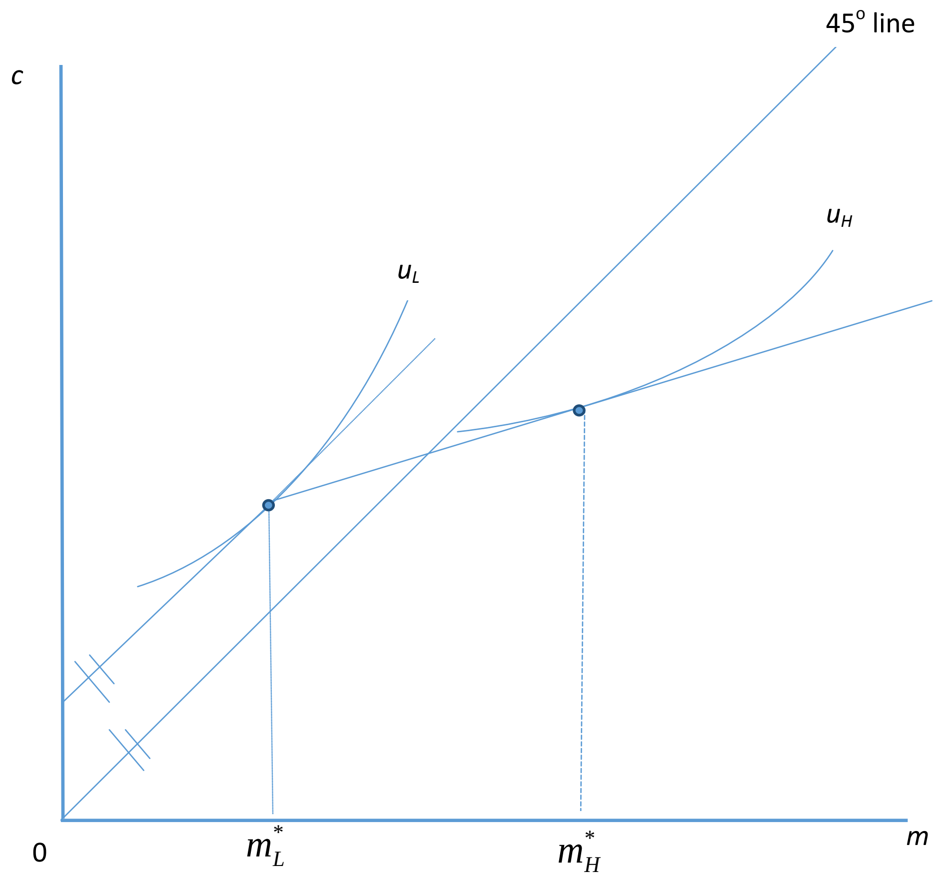
An illustration of the budget constraint.

Figure 3



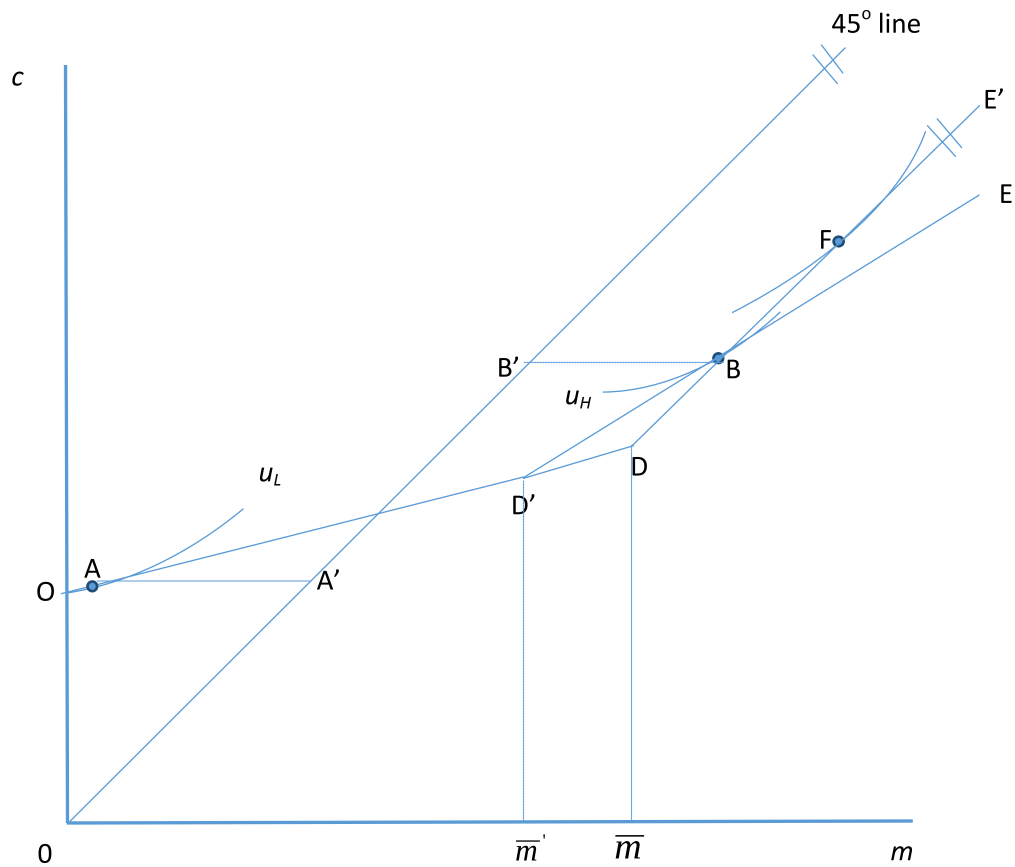
An illustration of a Case (1) optimal tax: $\tau_1 > 0$, $\tau_2 = 0$.

Figure 4



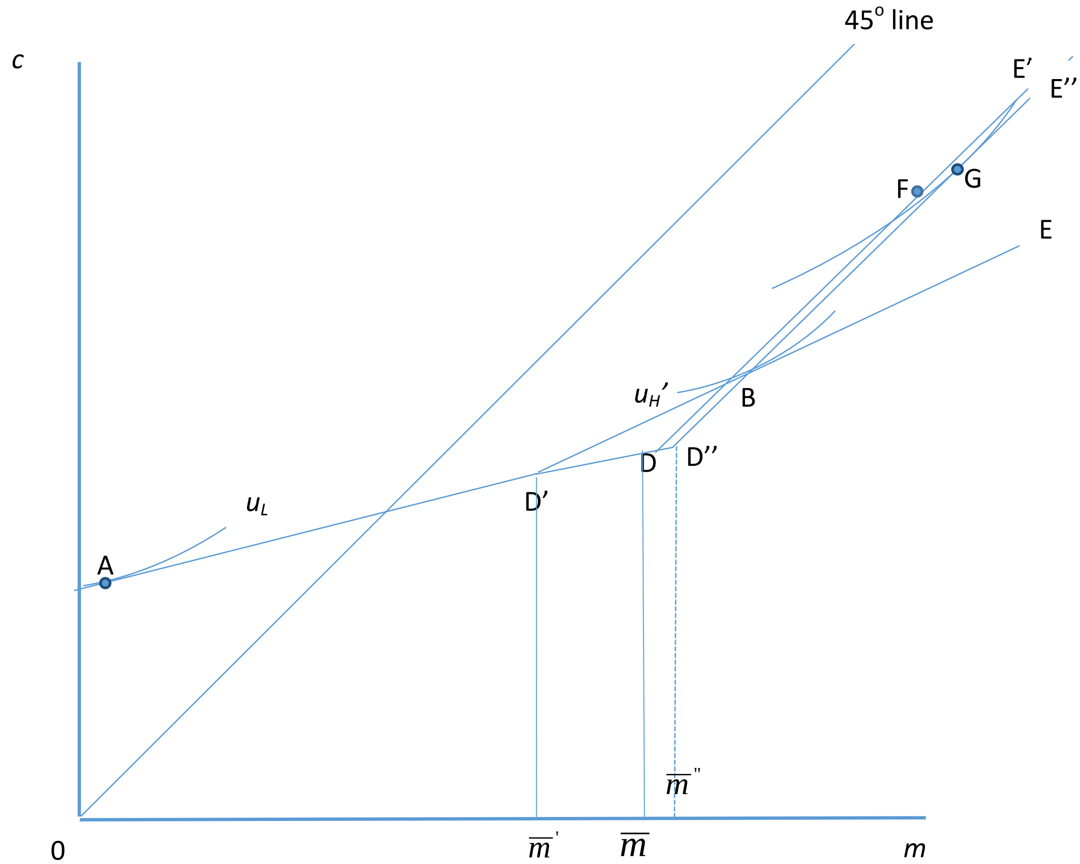
An illustration of a Case (2) optimal tax: $\tau_1 = 0$, $\tau_2 > 0$.

Figure 5



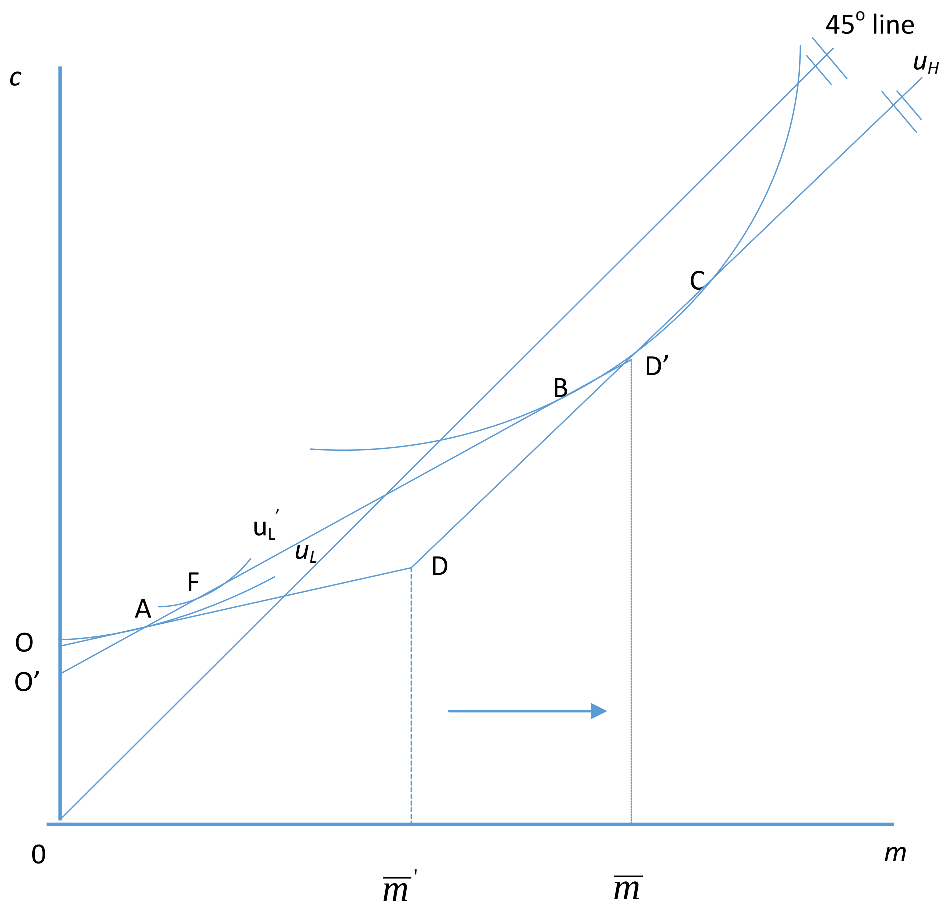
A progressive decreasing tax schedule: if $\tau_1^* > 0$, then $\tau_2^* = 0$ is optimal.

Figure 6



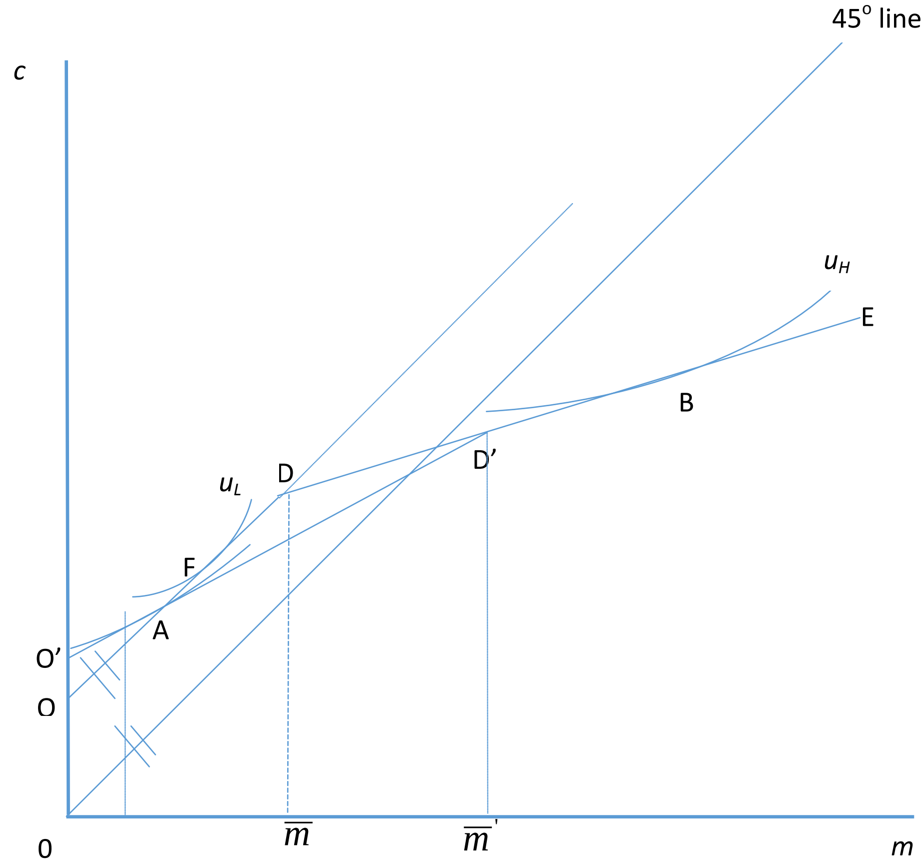
Points B and F are the same as in figure 5. The cut-off has been increased to $\overline{m}''$. The high-skilled now choose point G. They are better off than at point B and they pay more taxes.

Figure 7



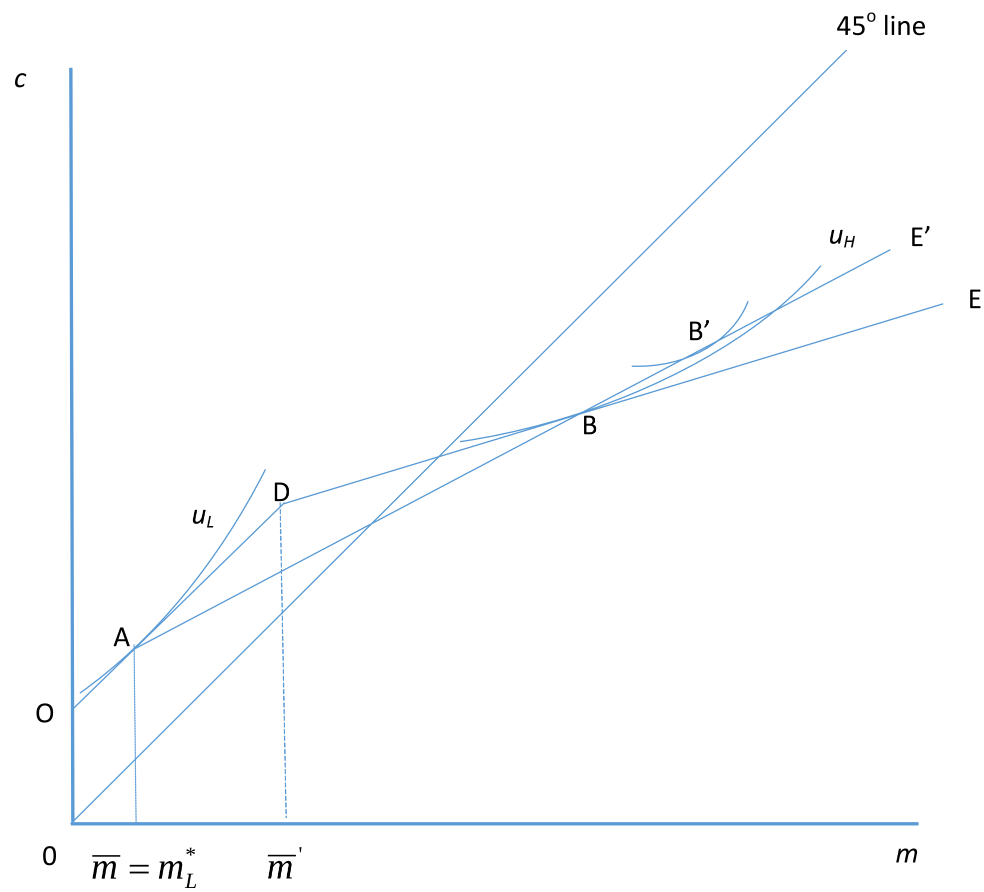
A progressive decreasing tax schedule. The indifference curve of the high-skilled touches both branches of the budget constraint and $\tau_2 = 0$.

Figure 8



A progressive increasing tax schedule: if $\tau_1 > 0$, then $\tau_2 = 0$ is optimal.

Figure 9



A progressive increasing tax schedule: u_L^* is tangent to the budget constraint where $m_L^* = \overline{m}$, and $\tau_1 = 0$.