Optimal Dual Income Taxation

Patricia Apps
University of Sydney Law School and IZA

Ray Rees
University of Munich, University of Oslo, University of Warwick and CESifo

19 January 2014

Abstract

There is a strong case for arguing that, in contradiction to some traditional arguments in public economics, capital income should be taxed, though not necessarily under the same rate schedule as labour earnings. The Dual Income Taxation (DIT) system consists of two parallel piecewise linear tax schedules, in which that for capital income has a zero lower bracket rate and an upper bracket rate equal to the "standard rate" under the labour earnings system. Pioneered in the Nordic countries, it has also been introduced in a number of other OECD countries, including Germany. This paper for the first time analyses this system from the point of view of optimal tax theory, and sheds considerable doubt on the optimality of the particular structure of the DIT, though it confirms the general principle of taxing capital income, and provides a firm analytical foundation for further consideration of how this should be undertaken.

JEL numbers: D14, H21, H24.

1 Introduction

Traditionally, the taxation of capital income has often been discussed as part of the problem of choosing a tax base consisting of either consumption expenditure or some comprehensive measure of income. One way of calculating the consumption tax base is to exclude from the measure of a household’s taxable income the income it derives from saving. This approach has been prominent in the literature because of the apparently strong theoretical support for the proposition that the return to saving should indeed not be taxed.

In a standard model of the intertemporal labour supply/consumption/saving decision, with weak separability between consumption and labour supply/leisure in each period, identical preferences across households, and an optimal nonlinear tax on labour earnings, the Atkinson-Stiglitz Theorem implies that there is no case for taxing the income derived from saving. An extension of this theorem

\[^1\text{See Meade (1978) as a classic example}\]
by Konishi (1995), Laroque (2005) and Kaplow (2006) replaces the assumption of an optimal nonlinear tax system with that of the planner being able to choose any smooth function from a consumer’s gross to net income, and shows that an allocation with both direct and indirect taxation can always be Pareto-dominated by one with direct taxation alone. This in the intertemporal context again implies no taxation of the return to saving. Further contributions in an explicitly dynamic framework by, among others, Chamley (1986) and Judd (1985) also yield the result that the optimal rate of tax on capital income is zero. Thus taxation of capital income is claimed to be at best superfluous and at worst non-optimal.

There is however a more recent body of literature that contests this view. The assumptions of weak separability, identical preferences and optimal or unconstrained nonlinear taxation on which the first set of contributions is based are of course very strong, as are those underlying the explicitly dynamic models. A number of studies have shown that the result, which in policy-related discussions is often simply expressed without reservation or qualification as "capital income ought to be untaxed", is not robust to their relaxation. Indeed, Banks and Diamond (2008) argue that the models are based on "considerations of economic behaviour and the nature of economic environments that are too restrictive when viewed in the context of both theoretical findings in richer models and the available econometric evidence".

In the earlier tax base discussion, it was generally assumed that with a comprehensive income measure as the tax base all forms of income would be subjected to the same tax schedule, and the arguments about the relative merits of the two tax bases was carried on in these terms. However, if we accept that the arguments for excluding capital income from taxation are not robust, so that the question of how exactly to tax it arises, we can question whether it is optimal to tax all forms of income at the same rate or, more generally, we can seek to determine the optimal tax schedules for labour and capital income treated separately. The general argument for doing so is one of the second best: given taxes on labour income and therefore distortions in the labour market, it is not in general second best optimal to leave the capital market undistorted, given the interrelationship between household decisions on saving and labour supply. A given revenue requirement is met with a smaller overall welfare loss, and equity objectives are more efficiently met, when both sectors are taxed. This is reflected in the analysis presented in this paper.

Dual income taxation (DIT) is the system whereby labour earnings and capital income are taxed according to separate - but closely related - schedules, and, pioneered in the Nordic countries, has either been introduced or is under active consideration in a number of other OECD countries. The central idea is

2 For a concise but comprehensive survey of this as well as of the earlier literature see Boadway (2012), Ch. 3.

3 See for example Mankiw et al (2009).

that alongside a standard piecewise linear progressive labour income tax schedule there is a flat rate tax on capital income, where the rate is typically at or around the "standard rate" on labour earnings. Since there is also a tax-free capital income threshold before the flat rate is applied, the DIT effectively involves a two-bracket piecewise linear capital income tax with a zero marginal rate in the lower bracket. Sørensen (1994), (2005) gives a thorough discussion of the practical problems involved with this system and the theoretical issues surrounding its desirability in both closed and open economies. He argues, in our view convincingly, that there is a strong case for such a tax.

However, there appears not to have been a formal analysis of the optimal structure of the DIT in terms of the standard trade-off between equity and efficiency. In particular, there has not been an analysis of how the marginal rates in the two parallel piecewise linear systems should optimally be related, of how the bracket limits should be set, and of the conditions under which either or both of the marginal tax rates on capital income should be zero.\(^5\) In this paper, we take an economy in which labour earnings and capital income, or, more precisely, second period consumption,\(^6\) are subject to separate two-bracket piecewise linear tax systems,\(^7\) derive their optimal parameters, and then consider conditions under which setting a capital income tax with a zero lower rate and upper bracket rate at the lower earnings tax rate would be approximately optimal.

The approach we take is based not on optimal nonlinear taxation, but rather, reflecting the reality of these DIT systems as well as of virtually all other formal tax structures, on a situation in which households are pooled and offered the same tax schedule or budget constraint in the gross income/net income plane. This means that the form of the optimal tax results and the intuitions underlying their interpretation are much closer to the optimal linear taxation models of Sheshinski (1972), (1989)\(^8\) than to those of optimal nonlinear taxation, which are driven by the presence of binding self selection constraints in a separating equilibrium. We do however continue to assume that all consumers have identical preferences, since this finesse the problem of making interpersonal comparisons when preferences differ.\(^9\)

\(^5\)Neilson and Sørensen (1997) take the presence of the flat rate capital tax as given and show that it is then optimal to have a progressive income tax in an economy with endogenous investment in human capital, but do not consider the optimality of the capital tax structure itself.

\(^6\)To model capital income taxation as literally a tax on the return to saving would lead to differences in detail but in the simple model used here to no essential differences in the results.

\(^7\)Note that this restriction to piecewise linear tax systems takes our analysis outside the domain of the Atkinson/Stiglitz Theorem and allows us therefore to adopt utility functions that possess strong separability properties.


\(^9\)In contrast to Diamond and Spinnewijn (2011) and Saez (2002) who assume differences in time preferences that may be correlated with the taxpayer's productivity type. This assumption also takes their analysis outside the domain of the Atkinson/Stiglitz Theorem.
2 The Model

We assume a continuum of wage types defined by an interval $\Omega = [w_0, w_1] \subset \mathbb{R}_{++}$, distributed on $\Omega$ according to the cdf $\Phi(w)$, with $\Phi'(w) = \phi(w) > 0 \ \forall w \in \Omega$. We further assume for simplicity additive separability in within-period as well as across-period utility and write the utility function of wage type $w$ as

$$U(w) = u(x_0(w)) - \nu(L(w)) + \rho u(x_1(w))$$

where $x_t$ is consumption in period $t = 0, 1$, with $x_0$ as the untaxed numeraire, $L$ is labour supply, assumed to be zero in the second period, $u(\cdot)$ is strictly concave, $v(\cdot)$ strictly convex, and both are strictly increasing. It is however more convenient for the optimal tax analysis to work with earnings rather than labour supply, and so given gross labour income $y = wL(w)$ we define $v(y/w)$ and re-write the utility function as

$$U(w) = u(x_0(w)) - \psi(y(w), w) + \rho u(x_1(w))$$

Now $y$ and $x_1$ are to be taxed separately according to a piecewise linear structure where $t_1, t_2$ are the marginal tax rates for taxation of labour income and $\hat{y}$ is the bracket limit, and $\tau_1, \tau_2$ are the marginal tax rates on second period consumption and $\hat{x}$ is the bracket limit. More formally we can describe the tax system by the functions $T_x(x_1(w)), T_y(y(w))$, where:

$$T_x(x_1(w)) = \tau_1 x_1(w) \quad x_1(w) \leq \hat{x}$$
$$T_x(x_1(w)) = \tau_2[x_1(w) - \hat{x}] + \tau_1 \hat{x} \quad x_1(w) > \hat{x}$$
$$T_y(y(w)) = t_1 y(w) \quad y(w) \leq \hat{y}$$
$$T_y(y(w)) = t_2[y(w) - \hat{y}] + t_1 \hat{y} \quad y(w) > \hat{y}$$

and so the consumer’s wealth constraint takes the general form

$$x_0(w) + \delta x_1(w) \leq g + y(w) - T_y(y(w)) - \delta T_x(x_1(w))$$

where $g > 0$ is a lump sum payment per capita and $\delta = (1 + r)^{-1}$ is the market discount factor.

The presentation of the results is greatly simplified if we restrict attention to tax structures for which $t_2 > t_1, \tau_2 > \tau_1$, so that the case of nonconvex budget sets for consumers is excluded.$^{10}$

For the set of tax systems we consider, the following assumption holds:

**Assumption 1**: The set of consumers/wage types can be partitioned into five non-empty subsets $\Omega_1 - \Omega_5$ according to the tax brackets they choose to be in under that tax system, and these subsets are defined as follows:

$$\Omega_1 = \{ w \in \Omega \mid x_1(w) < \hat{x}, \quad y(w) < \hat{y} \}$$

$^{10}$For analysis of the nonconvex budget set case see Apps, Long and Rees (2011) and the evidence given there that ruling this out is an empirically reasonable assumption.
\[ \Omega_2 = \{ w \in \Omega \mid x_1(w) = \hat{x}, \ y(w) < \hat{y} \} \] (9)

\[ \Omega_3 = \{ w \in \Omega \mid x_1(w) > \hat{x}, \ y(w) < \hat{y} \} \] (10)

\[ \Omega_4 = \{ w \in \Omega \mid x_1(w) > \hat{x}, \ y(w) = \hat{y} \} \] (11)

\[ \Omega_5 = \{ w \in \Omega \mid x_1(w) > \hat{x}, \ y(w) > \hat{y} \} \] (12)

Intuitively, as we move from \( w_0 \) through the set of wage types, at the lowest wage levels individuals are in the lower tax brackets for both labour and capital income, corresponding to subset \( \Omega_1 \). Then, some remain at the bracket limit on capital income, which represents a kink in their wealth constraint, while their labour income still increases with wage type. Individuals in \( \Omega_3 - \Omega_5 \) are in the higher capital income tax bracket but vary across the labour earnings brackets, with those in \( \Omega_3 \) still in the lower bracket, those in \( \Omega_4 \) at the kink of the gross/net labour income constraint and those in \( \Omega_5 \) in the higher brackets of both taxes. This pattern is facilitated by the fact that, given the assumption on the utility function, we have that the functions \( x^*(w) \) and \( y^*(w) \), understood as giving the optimal values chosen by consumers under the tax system as functions of their wage type, are everywhere continuous,\(^{11}\) while \( x^*(w) \) is strictly increasing with \( w \) over all subsets except \( \Omega_2 \), where it is constant at \( \hat{x} \), and \( y^*(w) \) is strictly increasing with \( w \) over all subsets except \( \Omega_4 \), where it is constant at \( \hat{y} \).

Of course other patterns are possible: for example there could be a subset for which both bracket limits bind, or the labour earnings bracket limit could occur before that of second period consumption. However, nothing essential would change in the results we derive for the case assumed. From the point of view of the appraisal of the DIT, it is useful to take the case where the lower labour income bracket, with what we interpret as the "standard" income tax rate, intersects with the upper capital income bracket. A key point is that in our model maximised utility is a strictly increasing and continuous function of the wage for all wage types under this kind of tax system.

For the optimal tax analysis, we require the indirect utility functions and their derivatives for households in each of the five subsets. These are as follows\(^{12}\)

\( \Omega_1 : \)

The indirect utility function\(^{13}\) is \( V(g, t_1, \tau_1; w) \) with derivatives

\[ V_g = \lambda; V_{t_1} = -\lambda y; V_{\tau_1} = -\lambda \delta x_1 \] (13)

\( \Omega_2 : \)

The indirect utility function is \( V(g, t_1, \tau_1, \hat{x}; w) \) with derivatives as for \( \Omega_1 \) except

\[ V_{\tau_1} = -\lambda \delta \hat{x}; V_{\hat{z}} = \alpha(w) = \rho u'(\hat{x}) - \lambda \delta (1 + \tau_1) \] (14)

\(^{11}\)It is the failure of this property to hold that causes the technical difficulties in the non-convex case.

\(^{12}\)The Appendix gives the details of the analysis underlying all the results presented in this paper.

\(^{13}\)Derivatives of tax parameters not included in the indirect utility function are understood to be zero.
The indirect utility function is \( V(g, t_1, \tau_1, \tau_2, \hat{x}; w) \) with derivatives as for \( \Omega_2 \), except
\[
    V_{\tau_2} = -\lambda \delta (x - \hat{x}); \quad V_{\hat{x}} = \lambda \delta (\tau_2 - \tau_1) \tag{15}
\]
\( \Omega_1 \) : The indirect utility function is \( V(g, t_1, \tau_1, \tau_2, \hat{x}, \hat{y}; w) \) with derivatives as for \( \Omega_3 \), except
\[
    V_{t_1} = -\lambda \hat{y}; \quad V_{\hat{y}} = \beta(w) = \lambda (1 - t_1) - \psi'(\hat{y}) \tag{16}
\]
\( \Omega_5 \) : The indirect utility function is \( V(g, t_1, t_2, \tau_1, \tau_2, \hat{x}, \hat{y}; w) \) with derivatives as for \( \Omega_4 \), except
\[
    V_{t_2} = -\lambda (y - \hat{y}); \quad V_{\hat{y}} = \lambda (t_2 - t_1) \tag{17}
\]
For the most part these derivations follow from straightforward applications of the Envelope Theorem. The derivatives with respect to the bracket limits, \( V_{\hat{x}} \) and \( V_{\hat{y}} \), are of particular interest. Wage types in these subsets are effectively constrained: they would like to earn (save) more if it would be taxed at the lower bracket rate, but not at the higher bracket tax rate, so that a marginal increase in the bracket limit would increase their utility, and \( \alpha(w), \beta(w) \) measure their net utility in these cases. Note finally that \( t_1, \hat{y} \) have income effects on the demands for \( x_1 \) and \( y \) for consumers in \( \Omega_5 \), and \( \tau_1, \hat{x} \) similarly have income effects on the demands for \( x_1 \) and \( y \) for consumers in subsets \( \Omega_3 - \Omega_5 \). Likewise \( \hat{x} \) has income effects on \( y \) in \( \Omega_2 \) and \( \hat{y} \) has income effects on \( x_1 \) in \( \Omega_4 \).

## 3 Optimal Dual Income Taxation

We are now in a position to solve the optimal tax problem. The planner wants to maximise the utilitarian social welfare function
\[
    \int_{\Omega} V(.; w) d\Phi \tag{18}
\]
subject to the government budget constraint. We derive this constraint by first defining the tax revenue raised from each subset of wage types:
\[
    R_1(w) = t_1 y(.; w) + \delta \tau_1 x_1 (.; w) \quad w \in \Omega_1 \tag{19}
\]
\[
    R_2(w) = t_1 y(.; w) + \delta \tau_1 \hat{x} \quad w \in \Omega_2 \tag{20}
\]
\[
    R_3(w) = t_1 y(.; w) + \delta [\tau_2 x_1 (.; w) - (\tau_2 - \tau_1) \hat{x}] \quad w \in \Omega_3 \tag{21}
\]
\[
    R_4(w) = t_2 \hat{y} + \delta [\tau_1 x_1 (.; w) - (\tau_2 - \tau_1) \hat{x}] \quad w \in \Omega_4 \tag{22}
\]
\[
    R_5(w) = [t_2 y(.; w) - (t_2 - t_1) \hat{y}] + \delta [\tau_2 x_1 (.; w) - (\tau_2 - \tau_1) \hat{x}] \quad w \in \Omega_5 \tag{23}
\]
and then writing the constraint as:
\[
    \sum_{i=1}^{5} \int_{\Omega_i} R_i(w) d\Phi \geq g + G \tag{24}
\]
where $G$ is a per capita revenue requirement. We then make:

**Assumption 2:** This problem has a unique interior global optimum in which all the subsets $\Omega_1 - \Omega_5$ are non-empty.

We characterise the solution in the following results:

**Result 1:** The condition corresponding to $g$ takes the form

$$\sum_{i=1}^{5} \int_{\Omega_i} (\sigma_i(w) - 1) d\Phi = 0$$

where $\sigma_i(w)$ is the net marginal social utility of income to individuals of type $w \in \Omega_i$, defined to include the effects on tax revenue resulting from the income effects of changes in the respective tax rates. This latter feature, as well as the fact that tax rates vary across subsets, means that the exact definitions of the $\sigma_i$ also vary across brackets. We denote by $\sigma(w)$ the value for any $w \in \Omega$, and so we can rewrite this condition as

$$\int_{\Omega} (\sigma(w) - 1) d\Phi = 0$$

The condition implies that the optimal $g$ equates the average value of the net marginal social utility of income across the population to its marginal social cost of 1, a result familiar from standard optimal linear tax theory. Since, other things equal, we expect $\sigma(w)$ to fall with the individual’s utility and therefore wage type, it is reasonable to assume that at relatively low wage levels $\sigma(w) - 1 > 0$ and at high wage levels $\sigma(w) - 1 < 0$.

Note also, as a corollary of Result 1, that

$$\sum_{i=1}^{k} \int_{\Omega_i} (\sigma_i(w) - 1) d\Phi = - \sum_{i=k+1}^{5} \int_{\Omega_i} (\sigma_i(w) - 1) d\Phi$$

for $k = 1, \ldots, 4$.

We denote the compensated derivatives of labour and capital income with respect to the relevant tax rates by $s_{ab}$, $a = x_1, y, b = t_j, \tau_j, j = 1, 2$. We then have:

**Result 2:**

The first order conditions characterising the four marginal tax rates $t_1-\tau_2$ respectively for given optimal bracket limits $\hat{x}, \hat{y}$ can be written as:

$$t_1 \sum_{i=1}^{3} \int_{\Omega_i} s_{yt_i} d\Phi + \tau_1 \delta \int_{\Omega_i} s_{x_1t_i} d\Phi + \tau_2 \delta \int_{\Omega_3} s_{x_1t_i} d\Phi = \sum_{i=1}^{3} \int_{\Omega_i} (\sigma_i(w) - 1)(y - \hat{y})$$

(28)

$$t_1 \int_{\Omega_1} s_{yt_1} d\Phi + \tau_1 \delta \int_{\Omega_1} s_{x_1t_1} d\Phi = \delta \int_{\Omega_1} (\sigma_1(w) - 1)(x_1 - \hat{x}) d\Phi$$

(29)

$$t_2 \int_{\Omega_5} s_{yt_2} d\Phi + \tau_2 \delta \int_{\Omega_5} s_{x_1t_2} d\Phi = \int_{\Omega_5} (\sigma_1(w) - 1)(y - \hat{y}) d\Phi$$

(30)

---

14 See the Appendix for the exact expressions.
\[
t_1 \int_{\Omega_3} s_{yr_3} d\Phi + t_2 \int_{\Omega_5} s_{y\tau_2} d\Phi + \tau_2 \delta \sum_{i=3}^{5} \int_{\Omega_i} s_{x_1\tau_3} d\Phi = \delta \sum_{i=3}^{5} \int_{\Omega_i} (\sigma_i(w)_1)(x_1 - \hat{x}) d\Phi
\]
\[31\]

**Result 3:**
Those characterising \(\hat{x}\) and \(\hat{y}\) respectively are

\[
\int_{\Omega_2} \left( \frac{\alpha(w)}{\mu} + \delta t_1 + t_1 \frac{\partial y}{\partial x} \right) d\Phi = -\delta(t_2 - t_1) \sum_{i=3}^{5} \int_{\Omega_i} (\sigma_i(w)_1) d\Phi
\]
\[32\]

\[
\int_{\Omega_4} \left( \frac{\beta(w)}{\mu} + t_1 + \delta t_2 \frac{\partial x_1}{\partial y} \right) d\Phi = -(t_2 - t_1) \int_{\Omega_6} (\sigma_i(w)_1) d\Phi
\]
\[33\]

To give some intuition for the first four conditions, begin by assuming that the cross derivatives \(s_{yr_1}, s_{x_1t_1} = 0, j = 1, 2\). We can then write the conditions, also using (26) to obtain the expressions in the numerators, as

**Result 4:**

\[
t_1 = \frac{\sum_{i=1}^{3} \int_{\Omega_i} (\sigma_i(w)_1) - 1)_y d\Phi + \hat{y} \sum_{i=4}^{5} \int_{\Omega_i} (\sigma_i(w)_1) - 1)_d\Phi}{\sum_{i=1}^{3} \int_{\Omega_i} s_{x_1t_1} d\Phi}
\]
\[34\]

\[
\tau_1 = \frac{\int_{\Omega_1} (\sigma_1(w)_1) - 1)_\delta x_1 d\Phi + \delta \hat{x} \sum_{i=2}^{5} \int_{\Omega_i} (\sigma_i(w)_1) - 1)_d\Phi}{\int_{\Omega_1} s_{x_1t_1} d\Phi}
\]
\[35\]

\[
t_2 = \frac{\int_{\Omega_2} (\sigma_1(w)_1) - 1)(y - \hat{y}) d\Phi}{\int_{\Omega_5} s_{y_\tau_2} d\Phi}
\]
\[36\]

\[
\tau_2 = \frac{\delta \sum_{i=3}^{5} \int_{\Omega_i} (\sigma_i(w)_1) - 1)(x_1 - \hat{x}) d\Phi}{\sum_{i=3}^{5} \int_{\Omega_i} s_{x_1\tau_3} d\Phi}
\]
\[37\]

Each of these defines the optimal tax rate as determined by a trade off between distributional equity, represented by the numerator term, and allocative efficiency, the denominator term. The latter is the sum over the relevant subset(s) of the compensated derivative of respectively earnings (labour supply) or first period consumption (saving) with respect to the corresponding tax rate, as a measure of marginal deadweight loss.

We give the interpretation of the numerator terms for the case of earnings taxation, those for capital income taxation follow similarly. In (33), the first term in the numerator is a measure of the closeness of the association between the marginal social utility of income \(\sigma_i(w)_1\) and labour earnings over the individuals in the lower tax bracket. Unlike the case of straightforward linear taxation, this is not strictly a covariance, since 1 is the population average of the \(\sigma_i(w)_1\).

We expect \(\sigma_i(w)_1\) to be decreasing with wage type, in which case this term could be positive which, since the denominator is negative, would suggest a negative tax rate. However, the second term is certainly negative and reflects the fact that varying \(t_1\) yields an intramarginal income gain to the individuals in the
subset $\Omega_2$ and in the higher tax brackets $\Omega_3 - \Omega_5$ that is proportional to $\dot{y}$, and this represents a social cost since the average of $\sigma_i(w)$ across these subsets must be less than the population average. In the case of the higher bracket tax rate the second type of term is of course absent because there is no higher bracket.

The conditions on $t_1$ and $\tau_1$ suggest an interesting tradeoff between the equity considerations in redistributing income within the lower tax bracket, which will tend to imply a relatively low tax rate, and the distributional gain to taxing the upper bracket consumers in what is essentially a lump sum way that involves no deadweight loss, which argues for a higher tax rate. In fact we show\(^{15}\) that this numerator is necessarily negative overall, so that the tax rate is necessarily positive. The intuition is straightforward. Suppose the tax rate $t_1$ (or $\tau_1$) = 0. Then, consumers’ choices of earnings (saving) in the lower tax bracket are undistorted. A marginal increase in the tax rate then has a zero first order effect on welfare but increases tax revenue because of the lump sum non-distortionary tax on consumers in the higher bracket. This gives a positive net welfare effect.

The two conditions (31) and (32) characterise the choice of bracket limits $\hat{x}$ and $\hat{y}$ respectively. The left hand sides represent the marginal social benefit of a small relaxation of the corresponding bracket limit. This consists first of the gain in utility to the wage types (in subset $\Omega_2$ for $\hat{x}$ and $\Omega_4$ for $\hat{y}$) who are effectively constrained at that bracket limit and are able now to increase their saving or consumption and achieve a net utility increase. Secondly there is a change in tax revenue resulting from the increase in saving or labour supply. In the case of a small increase $d\hat{x}$ ($d\hat{y}$) there is an increase in tax revenue $\delta t_1 d\hat{x}$ ($t_1 d\hat{y}$). Moreover we can show that\(^{16}\)

\[
\frac{\partial y}{\partial \hat{x}} = -\delta(1 + \tau_1) \frac{\partial y}{\partial g} \frac{\partial x_1}{\partial \hat{y}} = (1 - t_1) \frac{\partial x_1}{\partial g}
\]

This gives a further increase in tax revenue, since $\left(\frac{\partial x_1}{\partial g}\right) > 0$, $\left(\frac{\partial y}{\partial g}\right) < 0$.

The right hand sides of these conditions represent the marginal social costs of relaxing the bracket limits, which take the form of the lump sum income gains to those individuals in the higher tax brackets, for whom the average difference from the population average marginal social utility of income $\int_{\Omega_3}(\sigma_i(w) - 1)d\Phi$ is negative.

The full set of conditions in (27)-(30) reflect the fact that the tax rate on labour earnings in a given tax bracket will affect at the margin the saving of individuals in that bracket, and the tax on capital income will likewise affect labour supply, while at the same time lower bracket tax rates have only income effects on saving and labour supply decisions in the higher brackets. They also serve to emphasise the fact that under dual taxation all tax rates, on earnings from capital as well as labour, should be jointly determined. In other words, one should not in general introduce a DIT into a tax system with a given set of

\(^{15}\)The proof is given in the Appendix. It is based on the proof of a similar proposition in Apps, Long and Rees (2011).

\(^{16}\)See the Appendix.
tax rates and leave those tax rates unchanged, since the existence of the capital income taxes will in general change the optimal earnings tax rates.

We now consider the question: what does the analysis imply for the structure of the two piecewise linear taxes and, in particular, under what conditions would we have \( \tau_1 = 0, \tau_2 = t_1 \) at the optimum, even approximately?

First consider the relationship among the tax rates for the simpler case in which the cross derivatives \( s_{y\tau_j}, s_{x_1 t_j} = 0, j = 1, 2 \), represented by the conditions (33)-(36). We are interested in the conditions under which \( \tau_1 = 0 \) and \( \tau_2 = t_1 \). As already pointed out, we can prove that the optimal \( \tau_1 > 0 \). Moreover, the optimal conditions, even for this simple case, suggest that \( \tau_2 \neq t_1 \) in general.

That is, we would require

\[
\frac{\sum_{i=1}^{3} \int_{\Omega_i} (\sigma_i(w) - 1)(y - \hat{y})d\Phi}{\sum_{i=1}^{3} \int_{\Omega_i} s_{yt_1} d\Phi} \approx \frac{\delta \sum_{i=3}^{5} \int_{\Omega_i} (\sigma_i(w) - 1)(x_1 - \hat{x})d\Phi}{\sum_{i=3}^{5} \int_{\Omega_i} s_{x_1 t_2} d\Phi}
\]

(39)

There is no reason in general for this to hold and we have not been able to construct special cases in which it does so. Removal of the assumption \( s_{y\tau_1}, s_{x_1 t_1} = 0 \) does not increase the likelihood that the optimal conditions support the standard features of the DIT, but of course the equivalent condition to (39) becomes much more complicated, as can readily be seen from the conditions in Result 2.

4 Conclusions

The aim of this paper has been to analyse the optimal piecewise linear taxation of incomes from the supply of labour and capital, regarded as separate but complementary tax bases, and to use the results as a framework in which to consider the design of the Dual Income Tax. Within this framework, which emphasises the trade-off between equity and efficiency involved in setting taxes to meet a given revenue requirement, no support can be found for two main features of the practical applications of the DIT, the zero rate of tax in the lower bracket, and the equality between the standard rate of tax on labour earnings and the positive "flat rate" tax on capital income.

The rationale for these features may however be based on considerations arising outside the framework of the present analysis. Many households, particularly those at lower income levels (wage types), save mainly in the form of occupational pensions and housing, and have very small holdings of other income-bearing assets. It therefore seems sensible to exempt these households from a tax on those assets. The present analysis suggests that the cost of doing this is to rule out the possibility of using a non-distortionary lump sum tax \((\tau_1)\) on the capital income of households with much higher asset holdings. On the other hand, the expression in Result 4 shows that this tax, though positive, could optimally be quite small, if the two components of the numerator in condition (34) are largely offsetting, in which case not a lot would be lost by setting it to zero.
This then raises the further question: why is the tax on capital income ($\tau_2$) effectively a flat rate tax? Why not two or more tax brackets for capital income to achieve a degree of progression within the set of households paying capital income tax? As we have shown elsewhere,\(^{17}\) this question can be analysed within the framework presented here, and the answer depends ultimately on the nature of the distributions of wage types and incomes. It is in fact likely that more than one bracket would be optimal, in which case the conditions presented in this paper, which show how the successive tax rates should be determined, become directly relevant.

References


Appendix

In this Appendix we present the details of the analysis underlying all the results presented in the paper. The notation throughout is that used in the text of the paper so definitions will not be repeated here.

**Household equilibria under the given tax system.**

All consumers have the utility function

\[ U(w) = u(x_0(w)) - \psi(y(w), w) + \rho u(x_1(w)) \]  \hspace{1cm} (40)

Recall that \( \psi(y(w), w) \equiv \psi(y(w)/w) \) with \( \psi' > 0, \psi'' > 0 \).

The characterisation of a household’s equilibrium depends on the subset to which it belongs, as defined in the text. We consider them in turn:

1. \( \Omega_1 = \{ w \in \Omega \mid x_1(w) < \hat{x}, y(w) < \hat{y} \} \)

The wealth constraint for a household in this subset is:

\[ x_0 + \delta(1 + \tau_1)x_1 \leq g + (1 - t_1)y \]  \hspace{1cm} (41)

and so the Lagrange function is

\[ L_1 = U(w) + \lambda(w)[g + (1 - t_1)y - x_0 - \delta(1 + \tau_1)x_1] \]  \hspace{1cm} (42)

The foc for the consumer’s optimisation problem are

\[ \frac{\partial L_1}{\partial x_0} = u'(x_0^*(w)) - \lambda^*(w) = 0 \]  \hspace{1cm} (43)

\[ \frac{\partial L_1}{\partial y} = -\psi_y(y^*(w), w) + \lambda^*(w)(1 - t_1) = 0 \]  \hspace{1cm} (44)

\[ \frac{\partial L_1}{\partial x_1} = \rho u'(x_1^*(w)) - \lambda^*(w)\delta(1 + \tau_1) = 0 \]  \hspace{1cm} (45)

\[ \frac{\partial L_1}{\partial \lambda} = -[x_0^* + \delta(1 + \tau_1)x_1^*] + g + (1 - t_1)y^* = 0 \]  \hspace{1cm} (46)

2. \( \Omega_2 = \{ w \in \Omega \mid x_1(w) = \hat{x}, y(w) < \hat{y} \} \)

The wealth constraint for a household in this subset is:

\[ x_0 + \delta(1 + \tau_1)x_1 \leq g + (1 - t_1)y \]  \hspace{1cm} (47)

and we also impose the constraint \( \hat{x} \geq x_1 \) and we assume it is binding at the optimum. Thus the Lagrange function is

\[ L_2 = U(w) + \lambda(w)[g + (1 - t_1)y - x_0 - \delta(1 + \tau_1)x_1] + \alpha(w)[\hat{x} - x_1] \]  \hspace{1cm} (48)
and the foc are as for $\Omega_1$ except that

$$\frac{\partial L_2}{\partial x_1} = \rho u'(x_1^*(w)) - \lambda^*(w)\delta(1 + \tau_1) - \alpha^*(w) = 0$$ (49)

so with a binding constraint we have $\alpha^*(w) > 0$.

Note that setting $\dot{x} = x_1^*$ the comparative statics of this problem gives

$$\frac{\partial y^*}{\partial \dot{x}} = -\delta(1 + \tau_1)\frac{\partial y^*}{\partial y}$$ (50)

$$\Omega_3 = \{ w \in \Omega \mid x_1(w) > \dot{x}, y(w) < \dot{y} \}$$

The wealth constraint in this subset is:

$$x_0 + \delta(1 + \tau_2)x_1 \leq \dot{g} + (1 - \tau_1)y$$ (51)

with $\dot{g} \equiv g + \delta(\tau_2 - \tau_1)\dot{x}$ and so the Lagrange function is

$$L_3 = U(w) + \lambda(w)[\dot{g} + (1 - \tau_1)y - x_0 - \delta(1 + \tau_2)x_1]$$ (52)

with foc as for $\Omega_1$ except that

$$\frac{\partial L_3}{\partial x_1} = \rho u'(x_1^*(w)) - \lambda^*(w)\delta(1 + \tau_2) = 0$$ (53)

and

$$\frac{\partial L_3}{\partial \lambda} = -[x_0 + \delta(1 + \tau_2)x_1] + \dot{g} + (1 - \tau_1)y$$ (54)

$$\Omega_4 = \{ w \in \Omega \mid x_1(w) > \dot{x}, y(w) = \dot{y} \}$$

The wealth constraint in this subset is:

$$x_0 + \delta(1 + \tau_2)x_1 \leq \dot{g} + (1 - \tau_1)y$$ (55)

and we also impose the constraint $y \leq \dot{y}$ so that the Lagrange function is

$$L_4 = U(w) + \lambda(w)[\dot{g} + (1 - \tau_1)y - x_0 - \delta(1 + \tau_2)x_1] + \beta(w)[\dot{y} - y]$$ (56)

The foc are as for $\Omega_3$ except for

$$\frac{\partial L_4}{\partial y} = -\psi_y(y^*(w), w) + \lambda^*(w)(1 - \tau_1) - \beta^*(w) = 0$$ (57)

with $\beta^*(w) > 0$ and thus $\dot{y} = y^*$.

$$\Omega_5 = \{ w \in \Omega \mid x_1(w) > \dot{x}, y(w) > \dot{y} \}$$

The wealth constraint in this subset is:

$$x_0 + \delta(1 + \tau_2)x_1 \leq \dot{g} + (1 - \tau_2)y$$ (58)

where $\dot{g} \equiv \dot{g} + (t_2 - \tau_1)\dot{y}$ and so the Lagrange function is

$$L_5 = U(w) + \lambda(w)[\dot{g} + (1 - \tau_2)y - x_0 - \delta(1 + \tau_2)x_1]$$ (59)
The foc are

\[
\frac{\partial L_5}{\partial x_0} = u'(x_0^*(w)) - \lambda^*(w) = 0
\]  
(60)

\[
\frac{\partial L_5}{\partial y} = -\psi'(y^*(w), w) + \lambda^*(w)(1 - t_2) = 0
\]  
(61)

\[
\frac{\partial L_5}{\partial x_1} = \rho u'(x_1^*(w)) - \lambda^*(w)\delta(1 + \tau_2) = 0
\]  
(62)

\[
\frac{\partial L_5}{\partial \lambda} = -[x_0^* + \delta(1 + \tau_2)x_1^*] + \hat{y} + (1 - t_2)y^* = 0
\]  
(63)

The results on the derivatives of the indirect utility functions presented in the text are derived by applying the Envelope Theorem to the Lagrange functions \( L_i \) to \( L_5 \). Note also that we have

\[
\frac{\partial L_i}{\partial w} = -\frac{\partial \psi(y^*(w), w)}{\partial w} = \frac{v'(y^*(w)/w)y^*(w)}{w^2} > 0 \quad i = 1, ..., 5
\]  
(64)

which establishes that the indirect utility function \( V(.) ; w \) is strictly increasing in \( w \) across the entire domain \( \Omega \).

**Optimal Tax Problem**

We are now in a position to solve the optimal tax problem. The planner wants to maximise the utilitarian social welfare function

\[
\int_{\Omega} V(.) ; w d\Phi
\]  
(65)

subject to the government budget constraint. We derive this constraint by first defining the tax revenue raised from each respective subset of wage types:

\[
R_1(w) = t_1y(w) + \delta \tau_1x_1(w) \quad w \in \Omega_1
\]  
(66)

\[
R_2(w) = t_1y(w) + \delta \tau_1\hat{x} \quad w \in \Omega_2
\]  
(67)

\[
R_3(w) = t_1y(w) + \delta[\tau_2x_1(w) - (\tau_2 - \tau_1)\hat{x}] \quad w \in \Omega_3
\]  
(68)

\[
R_4(w) = t_1\hat{y} + \delta[\tau_2x_1(w) - (\tau_2 - \tau_1)\hat{x}] \quad w \in \Omega_4
\]  
(69)

\[
R_5(w) = [t_2y(w) - (t_2 - t_1)\hat{y}] + \delta[\tau_2x_1(w) - (\tau_2 - \tau_1)\hat{x}] \quad w \in \Omega_5
\]  
(70)

and then writing the constraint as:

\[
\sum_{i=1}^{5} \int_{\Omega_i} R_i(w) d\Phi \geq g + G
\]  
(71)

where \( G \) is a per capita revenue requirement.

The derivatives of the tax revenue functions \( R_i(w) \) with respect to the tax instruments depend on the subset \( i \) in general and are given by:

\( \Omega_1 \):

\[
\frac{\partial R_1}{\partial y} = t_1 \frac{\partial y}{\partial y} + \delta \tau_1 \frac{\partial x_1}{\partial y}
\]  
(72)
\[
\frac{\partial R_1}{\partial t_1} = y(w) + t_1 \frac{\partial y}{\partial t_1} + \delta x_1 \frac{\partial x_1}{\partial t_1} \tag{73}
\]
\[
\frac{\partial R_1}{\partial \tau_1} = t_1 \frac{\partial y}{\partial \tau_1} + \delta(x_1(w) + \tau_1 \frac{\partial x_1}{\partial \tau_1}) \tag{74}
\]

\[\Omega_2:\]
\[
\frac{\partial R_2}{\partial y} = t_1 \frac{\partial y}{\partial y} \tag{75}
\]
\[
\frac{\partial R_2}{\partial t_1} = y(w) + t_1 \frac{\partial y}{\partial t_1} \tag{76}
\]
\[
\frac{\partial R_2}{\partial \tau_1} = t_1 \frac{\partial y}{\partial \tau_1} + \delta x \tag{77}
\]
\[
\frac{\partial R_2}{\partial \delta x} = t_1 \frac{\partial y}{\partial \delta x} + \delta \tau_1 \tag{78}
\]

\[\Omega_3:\]
\[
\frac{\partial R_3}{\partial y} = t_1 \frac{\partial y}{\partial y} + \delta \tau_2 \frac{\partial x_1}{\partial y} \tag{79}
\]
\[
\frac{\partial R_3}{\partial t_1} = y(w) + t_1 \frac{\partial y}{\partial t_1} + \delta \tau_2 \frac{\partial x_1}{\partial t_1} \tag{80}
\]
\[
\frac{\partial R_3}{\partial \tau_1} = t_1 \frac{\partial y}{\partial \tau_1} + \delta(\tau_2 \frac{\partial x_1}{\partial \tau_1} + \delta x) \tag{81}
\]
\[
\frac{\partial R_3}{\partial \tau_2} = t_1 \frac{\partial y}{\partial \tau_2} + \delta(x_1(w) + \tau_1 \frac{\partial x_1}{\partial \tau_2}) \tag{82}
\]
\[
\frac{\partial R_3}{\partial \delta x} = t_1 \frac{\partial y}{\partial \delta x} + \delta(\tau_2 \frac{\partial x_1}{\partial \delta x} - (\tau_2 - \tau_1)) \tag{83}
\]

\[\Omega_4:\]
\[
\frac{\partial R_4}{\partial y} = \delta \tau_2 \frac{\partial x_1}{\partial y} \tag{84}
\]
\[
\frac{\partial R_4}{\partial t_1} = \dot{y} + \delta \tau_2 \frac{\partial x_1}{\partial t_1} \tag{85}
\]
\[
\frac{\partial R_4}{\partial \tau_1} = \delta(\tau_2 \frac{\partial x_1}{\partial \tau_1} + \dot{x}) \tag{86}
\]
\[
\frac{\partial R_4}{\partial \tau_2} = \delta(x_1(w) + \tau_2 \frac{\partial x_1}{\partial \tau_2} - \dot{x}) \tag{87}
\]
\[
\frac{\partial R_4}{\partial \delta x} = \delta(\tau_2 \frac{\partial x_1}{\partial \delta x} - (\tau_2 - \tau_1)) \tag{88}
\]
\[
\frac{\partial R_4}{\partial \dot{y}} = \delta \tau_2 \frac{\partial x_1}{\partial \dot{y}} \tag{89}
\]

16
The Lagrange function for the optimal tax problem is 
\[ L = R V(g; t, w) + \mu \left[ \sum_{i=1}^{5} R_i(w) d\Phi - g - G \right]. \]

The following are the derivatives of the indirect utility function (also given in the text):

\( \Omega_1 \):

The indirect utility function is \( V(g, t_1, \tau_1; w) \) with derivatives
\[ V_g = \lambda; V_{t_1} = -\lambda y; V_{\tau_1} = -\lambda \delta x_1 \] (97)

\( \Omega_2 \):

The indirect utility function is \( V(g, t_1, \tau_1, \hat{x}; w) \) with derivatives as for \( \Omega_1 \) except
\[ V_{\hat{x}_1} = -\lambda \delta \hat{x}; V_{\hat{x}} = \alpha(w) = \rho u'(\hat{x}) - \lambda(\hat{x}) \delta(1 + \tau_1) \] (98)

\( \Omega_3 \):

The indirect utility function is \( V(g, t_1, \tau_1, \tau_2, \hat{x}; w) \) with derivatives as for \( \Omega_2 \) except
\[ V_{\tau_2} = -\lambda \delta(x - \hat{x}); V_{\hat{x}} = \lambda \delta(\tau_2 - \tau_1) \] (99)

\( \Omega_4 \) : The indirect utility function is \( V(g, t_1, \tau_1, \tau_2, \hat{x}, \hat{y}; w) \) with derivatives as for \( \Omega_3 \), except
\[ V_{t_1} = -\lambda \hat{y}; V_{\hat{y}} = \beta(w) = \lambda(1 - t_1) - \psi'(\hat{y}) \] (100)

\( \Omega_5 \) : The indirect utility function is \( V(g, t_1, t_2, \tau_1, \tau_2, \hat{x}, \hat{y}; w) \) with derivatives as for \( \Omega_4 \), except
\[ V_{t_2} = -\lambda(y - \hat{y}); V_{\hat{y}} = \lambda(t_2 - t_1) \] (101)

\( ^{18} \) Derivatives of tax parameters not included in the indirect utility function are understood to be zero.
We use these in deriving the first order conditions, which are

\[
\frac{\partial L}{\partial t_1} = \sum_{i=1}^{3} \int_{\Omega_i} \left[ (-\lambda(w) y) + \mu \frac{\partial R_i}{\partial t_1} \right] d\Phi = 0
\]

\[
\frac{\partial L}{\partial t_2} = \sum_{i=1}^{3} \int_{\Omega_i} \left[ (-\lambda(w) \delta x) + \mu \frac{\partial R_i}{\partial t_2} \right] d\Phi = 0
\]

\[
\frac{\partial L}{\partial \tau_1} = \sum_{i=1}^{3} \int_{\Omega_i} \left[ (-\lambda(w) \delta x) + \mu \frac{\partial R_i}{\partial \tau_1} \right] d\Phi = 0
\]

\[
\frac{\partial L}{\partial \tau_2} = \sum_{i=1}^{3} \int_{\Omega_i} \left[ (-\lambda(w) \delta x) + \mu \frac{\partial R_i}{\partial \tau_2} \right] d\Phi = 0
\]

For further interpretation we need the Slutsky equations, which are:

For \( \Omega_1 \),

\[
\frac{\partial y}{\partial t_1} = s_{yt_1} - y \frac{\partial y}{\partial g}; \quad \frac{\partial x_1}{\partial t_1} = s_{x_1 t_1} - y \frac{\partial x_1}{\partial g}
\]

\[
\frac{\partial y}{\partial \tau_1} = s_{yt_1} - \delta x_1 \frac{\partial y}{\partial g}; \quad \frac{\partial x_1}{\partial \tau_1} = s_{x_1 \tau_1} - \delta x_1 \frac{\partial x_1}{\partial g}
\]

For \( \Omega_2 \),

\[
\frac{\partial y}{\partial t_1} = s_{yt_1} - y \frac{\partial y}{\partial g}
\]

For \( \Omega_3 \),

\[
\frac{\partial y}{\partial t_1} = s_{yt_1} - y \frac{\partial y}{\partial g}; \quad \frac{\partial x_1}{\partial t_1} = s_{x_1 t_1} - y \frac{\partial x_1}{\partial g}
\]

\[
\frac{\partial y}{\partial \tau_2} = s_{yt_2} - \delta (x_1 - \hat{x}) \frac{\partial y}{\partial g}; \quad \frac{\partial x_1}{\partial \tau_2} = s_{x_1 \tau_2} - \delta (x_1 - \hat{x}) \frac{\partial x_1}{\partial g}
\]

For \( \Omega_4 \),

\[
\frac{\partial x_1}{\partial \tau_2} = s_{x_1 \tau_2} - \delta (x_1 - \hat{x}) \frac{\partial x_1}{\partial g}
\]

For \( \Omega_5 \),

\[
\frac{\partial y}{\partial \tau_2} = s_{yt_2} - (y - \hat{y}) \frac{\partial y}{\partial g}; \quad \frac{\partial x_1}{\partial \tau_2} = s_{x_1 t_2} - (y - \hat{y}) \frac{\partial x_1}{\partial g}
\]
\[
\frac{\partial y}{\partial \tau_2} = s_y \tau_2 - \delta (x_1 - \hat{x}) \frac{\partial y}{\partial \tau_2} ;
\frac{\partial x_1}{\partial \tau_2} = s_x \tau_2 - \delta (x_1 - \hat{x}) \frac{\partial x_1}{\partial \tau_2}
\]

In addition, we have the following income effects:

For \( \Omega_2 \):

\[
\frac{\partial y}{\partial \tau_1} = -\delta \hat{x} \frac{\partial y}{\partial g} \frac{\partial x_1}{\partial \tau_1} = -\delta (1 + \tau_1) \frac{\partial y}{\partial g}
\]

For \( \Omega_3 \):

\[
\frac{\partial y}{\partial \tau_1} = -\delta \hat{x} \frac{\partial y}{\partial g} \frac{\partial x_1}{\partial \tau_1} = -\delta \hat{x} \frac{\partial x_1}{\partial \tau_1} ;
\frac{\partial y}{\partial \hat{x}} = \delta (\tau_2 - \tau_1) \frac{\partial y}{\partial \hat{x}} ;
\frac{\partial x_1}{\partial \tau_1} = \delta (\tau_2 - \tau_1) \frac{\partial x_1}{\partial \tau_1}
\]

For \( \Omega_4 \),

\[
\frac{\partial x_1}{\partial \hat{t}_1} = -\hat{y} \frac{\partial x_1}{\partial \hat{g}} ;
\frac{\partial x_1}{\partial \hat{g}} = (1 - t_1) \frac{\partial x_1}{\partial \hat{g}} ;
\frac{\partial x_1}{\partial \hat{x}} = \delta (\tau_2 - \tau_1) \frac{\partial x_1}{\partial \hat{x}}
\]

For \( \Omega_5 \),

\[
\frac{\partial y}{\partial \hat{t}_1} = -\hat{y} \frac{\partial y}{\partial \hat{g}} \frac{\partial x_1}{\partial \hat{t}_1} = -\hat{y} \frac{\partial x_1}{\partial \hat{g}} ;
\frac{\partial y}{\partial \hat{g}} = (t_2 - t_1) \frac{\partial y}{\partial \hat{g}} ;
\frac{\partial x_1}{\partial \hat{g}} = (t_2 - t_1) \frac{\partial x_1}{\partial \hat{g}}
\]

\[
\frac{\partial y}{\partial \hat{x}} = -\delta \hat{x} \frac{\partial y}{\partial \hat{g}} \frac{\partial x_1}{\partial \hat{x}} = -\delta \hat{x} \frac{\partial x_1}{\partial \hat{x}} ;
\frac{\partial y}{\partial \hat{x}} = \delta (\tau_2 - \tau_1) \frac{\partial y}{\partial \hat{x}} ;
\frac{\partial x_1}{\partial \hat{x}} = \delta (\tau_2 - \tau_1) \frac{\partial x_1}{\partial \hat{x}}
\]

Inserting the Slutsky equations and income effects into the derivatives of the tax revenue functions \( R_i(w) \) gives the optimal tax conditions.

\[
\frac{\partial L}{\partial g} = \int_{\Omega_1} \{ \lambda + \mu t_1 \frac{\partial y}{\partial g} + \delta \tau_1 \frac{\partial x_1}{\partial \tau_1} \} d\Phi + \int_{\Omega_2} \{ \lambda + \mu t_1 \frac{\partial y}{\partial g} \} d\Phi
\]

\[
+ \int_{\Omega_3} \{ \lambda + \mu t_1 \frac{\partial y}{\partial g} + \delta \tau_2 \frac{\partial x_1}{\partial \tau_2} \} d\Phi
\]

\[
+ \int_{\Omega_4} \{ \lambda + \mu \delta \tau_2 \frac{\partial x_1}{\partial \tau_2} \} d\Phi
\]

\[
+ \int_{\Omega_5} \{ \lambda + \mu t_2 \frac{\partial y}{\partial g} + \delta \tau_2 \frac{\partial x_1}{\partial \tau_2} \} d\Phi - \mu
\]

\[
= 0
\]
\[
\frac{\partial L}{\partial t_1} = \int_{\Omega_1} \{ -\lambda y + \mu[y + t_1(s_{y_{t_1}} - y \frac{\partial y}{\partial g}) \\
+ \delta \tau_1(s_{x_{t_1}} - y \frac{\partial x_1}{\partial g}) \} d\Phi + \\
\int_{\Omega_2} \{ -\lambda y + \mu[y + t_1(s_{y_{t_1}} - y \frac{\partial y}{\partial g})] \} d\Phi + \\
\int_{\Omega_3} \{ -\lambda y + \mu[y + t_1(s_{y_{t_1}} - y \frac{\partial y}{\partial g})] \} d\Phi + \\
+ \delta \tau_2(s_{x_{t_1}} - y \frac{\partial x_1}{\partial g}) \} d\Phi + \\
\int_{\Omega_4} \{ -\lambda \dot{y} + \mu[\dot{y} - \delta \tau_2 \frac{\partial x_1}{\partial g}] \} d\Phi + \\
\int_{\Omega_5} \{ -\lambda \dot{y} + \mu[\dot{y} - \delta \tau_2 \frac{\partial x_1}{\partial g}] \} d\Phi \\
= 0
\]

\[
\frac{\partial L}{\partial \tau_1} = \int_{\Omega_1} \{ -\lambda \delta x_1 + \mu[t_1(s_{y_{t_1}} - \delta x_1 \frac{\partial y}{\partial g}) + \delta(x_1 + \tau_1(s_{x_{t_1}} - x_1 \frac{\partial x_1}{\partial g})) \} d\Phi + \\
\int_{\Omega_2} \{ -\lambda \delta \dot{x} = \mu[-t_1 \delta \frac{\partial y}{\partial g} + \delta \dot{x}] \} d\Phi + \\
\int_{\Omega_3} \{ -\lambda \delta \dot{x} = \mu([-t_1 \delta \frac{\partial y}{\partial g}) \} d\Phi + \\
\int_{\Omega_4} \{ -\lambda \delta \dot{x} = \mu[\delta(-\tau_2 \frac{\partial x_1}{\partial g} + \dot{x})] \} d\Phi + \\
\int_{\Omega_5} \{ -\lambda \delta \dot{x} + \mu[-t_2 \delta \frac{\partial y}{\partial g} + \delta(-\tau_2 \frac{\partial x_1}{\partial g} + \dot{x})] \} d\Phi
\]

\[
\frac{\partial L}{\partial \tau_2} = \int_{\Omega_5} \{ -\lambda(y - \dot{y}) + \mu[y - \dot{y} + t_2(s_{y_{t_2}} - (y - \dot{y}) \frac{\partial y}{\partial g}) + \delta \tau_2(s_{x_{t_2}} - (y - \dot{y}) \frac{\partial x_1}{\partial g})) \} d\Phi \\
= 0
\]

\[
\frac{\partial L}{\partial \tau_2} = \int_{\Omega_3} \{ (-\lambda \delta(x_1 - \dot{x})) + \mu[t_2(s_{y_{t_2}} - \delta(x_1 - \dot{x}) \frac{\partial y}{\partial g}) \\
+ \delta(x_1 - \dot{x}) + \tau_2(s_{x_{t_2}} - \delta(x_1 - \dot{x}) \frac{\partial x_1}{\partial g})] \} d\Phi + \\
\int_{\Omega_4} \{ (-\lambda \delta(x_1 - \dot{x})) + \mu[\delta[(x_1 - \dot{x}) + \tau_2(s_{x_{t_2}} - (x_1 - \dot{x}) \frac{\partial x_1}{\partial g})] \} d\Phi + \\
\int_{\Omega_5} \{ -\lambda \delta(x_1 - \dot{x}) + \mu[t_2(s_{y_{t_2}} - \delta(x_1 - \dot{x}) \frac{\partial y}{\partial g}) \\
+ \delta(x_1 - \dot{x}) + \tau_2(s_{x_{t_2}} - (x_1 - \dot{x}) \frac{\partial x_1}{\partial g}) \} d\Phi \\
= 0
\]
\[
\frac{\partial L}{\partial x} = \int_{\Omega_2} \{\alpha(w) + \mu[\delta(t_1 - t_1(1 + \tau_1) \frac{\partial y}{\partial g})] \} d\Phi + \\
\int_{\Omega_3} \{\lambda \delta(t_2 - t_1) + \mu[t_1 \delta(t_2 - t_1) \frac{\partial y}{\partial g}] + \delta(t_2 \delta(t_2 - t_1) \frac{\partial x_1}{\partial g} - (t_2 - t_1)) \} d\Phi + \\
\int_{\Omega_4} \{\lambda \delta(t_2 - t_1) + \mu t_2 \delta(t_2 - t_1) \frac{\partial y}{\partial g}
+ \delta(t_2 \delta(t_2 - t_1) \frac{\partial x_1}{\partial g} - (t_2 - t_1)) \} d\Phi + \\
\int_{\Omega_5} \{\lambda \delta(t_2 - t_1) + \mu t_2 \delta(t_2 - t_1) \frac{\partial y}{\partial g}
+ \delta(t_2 \delta(t_2 - t_1) \frac{\partial x_1}{\partial g} - (t_2 - t_1)) \} d\Phi
\]

= 0

\[
\frac{\partial L}{\partial y} = \int_{\Omega_4} (\beta(w) + \mu \delta(t_2(1 - t_1) \frac{\partial x_1}{\partial g}) d\Phi + \\
\int_{\Omega_5} [\lambda(t_2 - t_1) + \mu(t_2(t_2 - t_1) \frac{\partial y}{\partial g} - (t_2 - t_1) + \delta(t_2(t_2 - t_1) \frac{\partial x_1}{\partial g})] d\Phi
\]

= 0

Then, by appropriate definition of the marginal social utility of income to consumers in each subset we simplify the expressions to those given in the text of the paper. These marginal social utilities of income are defined as:

For \( \Omega_1 \),
\[
\sigma_1(w) = \frac{\lambda(w)}{\mu} + t_1 \frac{\partial y}{\partial g} + \delta t_1 \frac{\partial x_1}{\partial g}
\]

For \( \Omega_2 \),
\[
\sigma_2(w) = \frac{\lambda(w)}{\mu} + t_1 \frac{\partial y}{\partial g}
\]

For \( \Omega_3 \),
\[
\sigma_3(w) = \frac{\lambda(w)}{\mu} + t_1 \frac{\partial y}{\partial g} + \delta t_2 \frac{\partial x_1}{\partial g}
\]

For \( \Omega_4 \),
\[
\sigma_4(w) = \frac{\lambda(w)}{\mu} + \delta t_2 \frac{\partial x_1}{\partial g}
\]

For \( \Omega_5 \),
\[
\sigma_5(w) = \frac{\lambda(w)}{\mu} + t_2 \frac{\partial y}{\partial g} + \delta t_2 \frac{\partial x_1}{\partial g}
\]

Proof that \( \tau_1 > 0 \)

Proof of the proposition that the lower bracket tax rate on capital income should be positive, under the assumption that \( s_{y_j}, s_{x_j}, t_j = 0, j = 1, 2 \).
Before proving the proposition we present a lemma that is simple enough not to need a proof here.

**Lemma:**
(a) Given \( \sigma'(w) < 0 \), the first order condition (26) implies \( \exists \hat{w} \in (w_0, w_1) \) such that \( \sigma(\hat{w}) - 1 = 0 \), and that \( (\sigma(w) - 1) \geq 0 \) according as \( w \leq \hat{w} \), \( \forall \omega \in [w_0, w_1] \)

(b) Given also that \( \phi(w) > 0 \), we have

\[
\int_{w_0}^{\omega} (\sigma(w) - 1)\phi(w)dw > 0 \quad \forall \omega \in [w_0, w_1]
\] (102)

and

\[
\int_{\omega}^{w_1} (\sigma(w) - 1)\phi(w)dw < 0 \quad \forall \omega \in (w_0, w_1]
\] (103)

**Proposition 1:** On the assumptions of the model, and given \( \sigma'(w) < 0 \), we have that \( \tau_1^* > 0 \).

**Proof:** we have to show that

\[
\delta \int_{\Omega_1} (\sigma_1(w) - 1)(x_1 - \hat{x})d\Phi < 0
\] (104)

or equivalently

\[
\delta \int_{\Omega_1} (\lambda(w) - \mu)(x_1 - \hat{x})d\Phi < 0
\] (105)

Recall that \( \Omega_1 = [w_0, \hat{w}] \). Then we have to consider two cases:

(a) \( \hat{w} \leq \hat{w} \):
If \( \hat{w} < \hat{w} \), we have \( \lambda(w) - \mu > 0 \) and, since \( x_1 - \hat{x} < 0 \), we have the result immediately.

Note in this case that we have

\[
\delta \int_{\Omega_1} (\lambda(w) - \mu)x_1d\Phi > 0
\] (106)

(b) \( \hat{w} > \hat{w} \):
Since \( \Phi'(w) \equiv \phi(w) > 0 \), \( \lambda(w) - \mu \phi(w) > 0 \) and \( \hat{x} - x_1(w) > \hat{x} - x_1(\hat{w}) \) over \( [w_0, \hat{w}] \), we have

\[
\int_{w_0}^{\hat{w}} \{ \lambda(w) - \mu \} [\hat{x} - x_1(w)]\phi(w)dw > [\hat{x} - x_1(\hat{w})] \int_{w_0}^{\hat{w}} \{ \lambda(w) - \mu \} \phi(w)dw
\] (107)

In addition, since \( \{ \lambda(w) - \mu \} \phi(w)dw < 0 \) and \( 0 < \hat{x} - x_1(w) < \hat{x} - x_1(\hat{w}) \) over \( (\hat{w}, \hat{w}) \), we have

\[
\int_{\hat{w}}^{\hat{w}} \{ \lambda(w) - \mu \} [\hat{x} - x_1(w)]\phi(w)dw > [\hat{x} - x_1(\hat{w})] \int_{\hat{w}}^{\hat{w}} \{ \lambda(w) - \mu \} \phi(w)dw
\] (108)

Adding these two inequalities gives

\[
\int_{\Omega_1} \{ \lambda(w) - \mu \} [\hat{x} - x_1(w)]\phi(w)dw > [\hat{x} - x_1(\hat{w})] \int_{\Omega_1} \{ \lambda(w) - \mu \} \phi(w)dw > 0
\] (109)
where the last inequality follows from applying the lemma with $\hat{w} \equiv \omega$. This then gives

$$\delta \int_{\Omega_1} \frac{(\lambda(w) - 1)(x_1(w) - \hat{x})}{\mu} d\Phi < 0$$

(110)

as required.