On optimal income taxation when inherited wealth differs

Terhi Ravaska*, Labour Institute for Economic Research, Helsinki, Finland

Matti Tuomala, Tampere University, Finland

This version 9.8.2019, Work in progress, Please do not circulate or cite!

Abstract

In this paper we study a multidimensional optimal taxation problem when individuals have differences in skills and in initial wealth (inheritance). In a two-period model with one cohort we derive the optimal distortions for the saving decision in two- to four-types economies. Numerical methods are used for solving the binding self-selection constraints and optimal net and gross incomes for each type. We also extend the model to include income shifting.

Our findings support the view that there should be non-linear capital income tax. In the simplest case of two-types, the saving decisions of the low-ability and low-wealth type is taxed at the margin. In the 3- to 4-type settings high initial wealth types are subsidized at the margin. The subsidy relax the self-selection constraint which prevent the high-wealth types mimicking to be low-wealth types. For the type of low-wealth and high-productivity the marginal distortion on the saving decision depends upon the degree of correlation between ability and initial wealth and the chosen social welfare function.

Keywords: Optimal taxation, lifetime redistribution, multidimensional tax problem, het-

*terhi.ravaska@labour.fi. We thank Albert Jan Hummel for the comments for the earlier version of this paper. We also thank the seminar participants in the IIPF conference 2016 in Lake Tahoe, at the Tampere Allecon meeting in 2016 and at the ECINEQ meeting in 2019.
1 Introduction

In many developed countries one highly significant phenomenon in recent years has been the ending of the downward trend in wealth concentration. Piketty and Zucman (2014) have estimated wealth-to-income ratios for eight advanced economies and their estimates reveal some striking trends. Wealth to income in these nations climbed from a range of 200 to 300 percent in 1970 to a range of 400 to 600 percent in 2010. The wealth differences for any given cohort will reflect income differences if individuals save for life-cycle smoothing purposes and everyone has the same preferences. However, this is not the only way in which people receive capital since some people inherit it. Hence, capital income inequality stems from differences in wealth due to past saving behaviour, inheritances received, and in rates of return that have varied dramatically over time and across assets.

Mirrlees (1971) states that "In an optimum system, one would no doubt wish to relate tax payments to the whole life pattern of income, and to initial wealth". In practice, taxation is not based on life-cycle but on annual incomes and initial wealth differences are only partly accounted in inheritance taxation. Especially the initial wealth differences have received only little attention as a source of heterogeneity in the optimal income taxation literature. Most of the optimal income taxation literature has focused solely on differences in productivities, and only recently has heterogeneity in other dimensions\(^1\) been incorporated into the models.

In the optimal inheritance taxation literature the center of interest is to determine how to tax the bequests left behind usually by parents to their children. This is a one-off occasion from the perspective of taxation. There is an ample set of models studying this issue, taking into account different assumptions about preferences for saving and bequest (see Kopczuk (2013) and Piketty and Saez (2013) and the references therein). In these studies the optimal inheritance tax rates vary considerably from zero to extreme high. However, in this paper we are not interested in taxation of this kind of one-off event but instead focus on the question how initial wealth affect the structure of income taxation. Especially we study whether savings should be taxed or not.

\(^1\)For example differences in preferences have been incorporated into optimal taxation models by Ravaska et al. (2018), Golosov et al. (2013), Diamond and Spinnewijn (2011) and Tenhunen and Tuomala (2010).
This approach is also motivated by the empirical fact that inheritance tax is one of the least popular forms of taxes and many countries have abolished it altogether with other net wealth taxes (Drometer and Frank, 2018). The same kind of economy is studied for example in Cremer et al. (2001) and Christiansen and Tuomala (2008). These papers study whether linear capital income tax should be added to the optimal tax mix. Our study differs from these two studies by studying non-linear capital income tax.

Non-linear capital income taxation is relatively understudied. Capital income taxation is often studied in the framework of commodity taxation and for many commodities the non-linear taxation is impractical for the tax arbitrage reasons. However, we argue that non-linear capital income taxation is possible especially in developed countries where also capital income is often reported by third-parties. Several countries have tax policies that effectively vary the capital income tax rate based on the total income (USA) or annual capital income (Finland). This makes our case for studying this tax instrument. For other studies on non-linear capital income taxation, see for example Golosov et al. (2013).

In one-dimensional world (assuming only differences in labour productivity) the Atkinson-Stiglitz result (Atkinson and Stiglitz, 1976) says that by assuming a mild separability between consumption and labour supply, the non-linear labour income tax does not need to be supplemented with other taxes, like capital income tax. There are very few papers on the role of capital income taxation in the Atkinson-Stiglitz type of economy with differences in unobserved inherited wealth. Such economies are studied in Boadway et al. (2000) and Cremer et al. (2003). Boadway et al. (2000) study overlapping generations model where savings tax is linear and labour income tax non-linear. They show that with accidental bequests, tax on the interest income can indirectly tax the unobserved inherited wealth and is thus desirable. Cremer et al. (2003) derive the same result with another bequest motive, ‘joy of giving’. They conclude that capital income tax can be a desirable additional instrument to bring more information about the unobserved inheritances. With positive correlation between inherited wealth and productivity, marginal tax rate on capital income is likely a positive one.

We study optimal taxation in a static setting in a sense that we take the endowment or initial wealth as exogenous. This means that we study only one cohort who are entitled to a bequest from the previous cohort but leave no bequest. In our model, people differ in terms of productivity and in initial wealth which are both unobservable to the policy-maker\(^2\). Heterogeneity in

\(^2\)In reality some of the bequests are observable to the government but there are also transfers that are either
initial endowments put individuals in a different starting point in their life already before the productivity type is revealed. In a world without initial wealth differences, redistribution would occur only between productivity types but with initial endowments the direction of redistribution between types is ambiguous.

The focus is put on the optimal distortion for savings and hence for the question whether or not to tax capital income. To learn more about the direction of redistribution we solve numerical examples. We also comment on the optimal levels of the labour supply distortions in the numerical simulations. We contribute to the earlier literature by studying non-linear savings tax and extend the model to include also income shifting.

Our analytical results (assuming the direction of binding self-selection constraints) and numerical solution reveal that the saving decision is distorted at the margin if there are differences in initial wealth between individuals. The common pattern in most specifications is that there is a tax at the margin for the individuals with low initial wealth and a subsidy for the individuals with high initial wealth. In the numerical simulations we also consider the role of wealth inequality, correlation between ability and initial wealth and different social objectives. Wider wealth inequality requires less distortions for the saving decision. The correlation between the unobserved factors affect the optimal distortion in a non-monotonic way and the social objectives matter for which type is distorted at the margin.

The rest of the paper is organized as follows. In section 2 we introduce the benchmark model. In section 3 we extend the type-space to three and four types and discuss the numerical simulations. In section 4 we extend the model to include income shifting. Section 5 concludes.

2 Benchmark model

We will consider a discrete type version of the Mirrlees model in the spirit of Stern (1982) and Stiglitz (1982). We assume that individuals differ in their productivity and in their initial wealth level. First we consider a simple two-type economy where productivity and initial wealth are perfectly correlated. Table 1 shows evidence for Finland that wealth and capital income indeed are strongly correlated and the correlation has got stronger over time making this a reasonable assumption for the benchmark model. In later parts this assumption of perfect correlation is unobservable or unidentifiable.
Table 1: Correlation between income items and net wealth. Data sources: Wealth Study Statistic Finland. Source Riihelä et al. (2007), updated.

In our model the initial wealth is exogenous lump-sum, which is received in the first period before labour supply decision is made. Types are exogenous and the distribution is known by the tax planner. This exogeneity assumption is done to simplify the exposition and notation without loss of generality. There is asymmetric information in a sense that the tax authority can observe neither the endowments, productivities nor labour supply and this rules out first best taxation. Tax authorities can observe the savings from the first period to the second period and the earned income in the second period. These assumptions are made in order to study a fully non-linear tax system.

The government wants to create a lifetime tax system which redistributes income between the individuals in the same cohort and we assume it can commit to the chosen tax-and-transfer system. The government can now use both labour income tax and capital income tax. In this setting we study whether capital income tax is needed in the optimal tax mix.

2.1 Individuals

Each individual has a skill level, $n^i$, reflecting his wage level and an initial wealth level, $e^i$ reflecting the consumption potential in the first period. Low-skilled and/or low-endowed are denoted with superscript L and high-skilled and/or high-endowed are denoted with superscript H. The benchmark assumption of the positive correlation implies that we have two types of $(e^H, n^H)$ and $(e^L, n^L)$. The proportion of each type $i$ in the population is $N^i \geq 0$ and $\sum N^i = 1$.

We consider a two-period model. In order to say something about the Atkinson-Stiglitz
result, the life-time utility of an agent $i$ is separable and additive in the following way:

$$U^i = u(c^i) + \delta v(x^i) + \psi(1 - y^i), \quad (1)$$

where $c$ and $x$ denote consumption in the first and the second period, respectively, and $y$ is the labour supply during the second period. $\delta$ denotes the discount factor. It is assumed that $U$ is a strictly concave, continuously differentiable and utility is strictly increasing in $c$ and $x$, and strictly decreasing in $y$, and (partial derivatives) $u'$, $v'$, $\psi'$, $\psi'' > 0$ and $u''$, $v'' < 0$. We also assume that all goods are normal and technologies are linear, that is one extra unit of labour produces one unit of commodity good. Markets are assumed to be competitive.

In the first period individuals divide the initial wealth between consumption, $c$, and saving, $s$. Each unit of savings yields an additional $1 + \theta$ units of consumption in the second period. We simplify the analysis by considering a small open economy facing world capital markets which implies that the return to savings, $\theta$, is fixed. Consumption in the first period is $c^i = e^i - s^i$ and in the second period $x^i = (1 + \theta)s^i + B^i$, where after-tax labour income is denoted by $B^i = n^i y^i - T(n^i y^i)$. The inter-temporal budget constraint then is $c^i + r x^i = e^i + r B^i$, where $r = \frac{1}{1+\theta}$.

Individual’s problem is to maximize inter-temporal utility with budget constraint:

$$\max_{c^i, x^i, y^i} \quad u(c^i) + v(x^i) + \psi(1 - y^i)$$

s.t.

$$c^i + r x^i = e^i + r B^i. \quad (2)$$

Without distortive taxation the well-known Euler equation emerges from the first order conditions:

$$\frac{u'}{v'} = \frac{\delta}{r}. \quad (3)$$

Comparing this condition to the Euler equation from the government problem we can infer whether the individuals saving decision is distorted.
2.2 Government

There are asymmetric information between the government and individuals. Using the direct approach, government assigns bundles of gross and net incomes for each type in both periods. The self-selection or incentive compatibility constraints make sure that the individual weakly prefers the bundle aimed at him or her over all the other bundles. That is the self-selection constraints make sure that there is no mimicking of another type in the optimum.

In the benchmark case we assume that there is a perfect correlation between productivity and endowment received. Here the single-crossing property holds and the only binding self-selection constraint is from the direction of high-type towards low-type. In preceding sections, when the perfect correlation assumption is relaxed, numerical methods are used for determining the binding self-selection constraints. The incentive constraints which prevent the mimicking behaviour are written as:

\[ u(c^H) + \delta v(x^H) + \psi(1 - y^H) \geq \hat{u}(c^L) + \delta v(x^L) + \psi(1 - \frac{n^L}{n^H}y^L) \equiv U^{ij}, \]

where hat denotes mimicking behaviour and specifically \( \hat{u}(c^L) = u(e^H - s^L) \). That is, in the case of mimicking the mimicker would enjoy higher first period consumption than truly low-ability type.

In the case of utilitarian social objective function government’s problem is to

\[
\max_{c^i, x^i, y^i} G(U) = \sum_{i=1}^{N} N^i(u(c^i) + \delta v(x^i) + \psi(1 - y^i)) \\
\text{subject to} \\
\sum_{i=1}^{N} N^i(n^iy^i + (1 + \theta)(e^i - c^i) - x^i) \geq G \\nU^H \geq U^{HL},
\]

where the first constraint is the resource constraint and second the aforementioned incentive compatibility constraint. An alternative social objective function is maximin, where government maximizes only the welfare of the low-ability type (defined as the individuals with low inheritance and productivity).
2.3 Analytical results

We can now derive the first order conditions by forming a Lagrangian of the government’s problem. The multiplier for the resource constraint is \( \lambda \) and for the self-selection constraint \( \mu^{HL} \). The Lagrangian expression is:

\[
\mathcal{L} = \sum_{i=1}^{N} N^i [u(c^i) + \delta v(x^i) + \psi(1 - y^i)] \\
+ \lambda \left[ \sum_{i=1}^{N} N^i (n^i y^i + (1 + \theta)(e^i - c^i) - x^i) - G \right] \\
+ \mu^{HL} [u(c^H) + \delta v(x^H) + \psi(1 - y^H) - \hat{u}(c^L) - \delta v(x^L) - \psi(1 - \frac{n^L}{n^H} y^L)].
\]

The first order conditions with respect to \( c^i, x^i \) and \( y^i \) are given by

\[
\frac{\partial \mathcal{L}}{\partial c^i} = N^i u' - \lambda N^i (1 + \theta) - \mu^{HL} \hat{u}' = 0 \tag{7}
\]
\[
\frac{\partial \mathcal{L}}{\partial x^i} = N^i \delta v' - \lambda N^i - \mu^{HL} \delta v' = 0 \tag{8}
\]
\[
\frac{\partial \mathcal{L}}{\partial y^i} = -N^i \psi' + \lambda N^i n^i + \mu^{HL} \psi' \frac{n^i}{n^H} = 0 \tag{9}
\]
\[
\frac{\partial \mathcal{L}}{\partial c^H} = N^H u' - \lambda N^H (1 + \theta) - \mu^{HL} u' = 0 \tag{10}
\]
\[
\frac{\partial \mathcal{L}}{\partial x^H} = N^H \delta v' - \lambda N^H - \mu^{HL} \delta v' = 0 \tag{11}
\]
\[
\frac{\partial \mathcal{L}}{\partial y^H} = -N^H \psi' + \lambda N^H n^H - \mu^{HL} \psi' = 0. \tag{12}
\]

In the case of maximin social objective function the last three equations become:

\[
\frac{\partial \mathcal{L}}{\partial c^H} = -\lambda N^H (1 + \theta) - \mu^{HL} u' = 0 \tag{13}
\]
\[
\frac{\partial \mathcal{L}}{\partial x^H} = -\lambda N^H - \mu^{HL} \delta v' = 0 \tag{14}
\]
\[
\frac{\partial \mathcal{L}}{\partial y^H} = \lambda N^H n^H - \mu^{HL} \psi' = 0. \tag{15}
\]
It should be noted that the consumption of the mimicker is $\Delta e + c_L$, where $\Delta e = e^H - e^L$. Define then the inter-temporal marginal substitution as $u'_i / v'_i$ and rearrange to get for the high-ability type

$$\frac{u'}{v'} = \frac{\delta}{r}, \quad (16)$$

that is, the saving decision is not distorted at the margin even when there are differences in initial wealth. However, for the low-ability type, the saving decision is not distorted only in the case where there is no differences in initial endowments (and so $u'_L = \hat{u}'$) or when incentive-constraint $\mu^{HL}$ is not binding. We can see this by rearranging the first order conditions as before:

$$\frac{N^L u' - \mu^{HL} \hat{u}'}{v'(N^L - \mu^{HL})} = \frac{\delta}{r}. \quad (17)$$

Rearranging equation 17 to $\frac{u'}{v'} = \frac{\delta}{r} (1 - d_L)$ and solving for $d_L$, we get the marginal distortion on the low-types saving decision (assuming the only binding constraint is from high-type towards the low-type):

$$d_L = \frac{\mu^{HL}}{N^L} (1 - \frac{r}{\delta} \frac{\hat{u}'}{v'}). \quad (18)$$

While equation 18 does not instantly reveal the sign of the distortion, we know that mimicker’s marginal utility of consuming during the first period is smaller than for the true low ability types. In order to prevent the mimicker for consuming too much during the first period, we need to distort the low-ability types saving decision which would make mimicking behaviour less beneficial. That means that the distortion is positive in the current case.

Solving for the optimal marginal labour tax\(^3\) reveals the typical result that the high-ability type’s labour decision is not distorted at the margin. As the labour supply decision is independent of the endowment we know that the common case of positive marginal labour income tax takes place for the low-ability type. This distortion equals

$$T'_L = \frac{\mu^{HL} (1 - \frac{n^L}{\hat{n}^L})}{N^1 - \mu^{HL}}. \quad (19)$$

We can state the following proposition:

\(^3\)The labour supply distortion is solved with the first order conditions as: $\frac{\psi'}{\delta (\mu)} = 1 - T'$.  

9
Proposition 1

In the two-type economy

(i) Both in utilitarian and maximin cases if $e^L = e^H$ and if preferences of individuals are additively separable, then there is no taxation of capital income at the optimum. Low-ability type faces a positive marginal labour income tax.

(ii) If $e^L < e^H$ and if preferences of individuals are additively separable and the single-crossing condition holds, then there is a case for taxing capital income at the optimum. The low-ability type faces a positive capital income tax at the margin for both utilitarian or maximin social objectives. Low-ability type faces also a positive marginal labour income tax.

3 Extension to type-space

Relaxing the assumption of perfect correlation between productivity and initial wealth leads to four-type economy as in the table 2. With zero correlation, there are equal amount of individuals in each type and for imperfect correlation cases the sizes of the types vary. Including more types leads to larger set of potentially binding self-selection constraints. In the next subsection we study 3-type case by assuming some of the binding constraint while the subsequent subsection utilizes numerical methods for the 4-type model. The government’s objective is either utilitarian or maximin with respect of productivity.

<table>
<thead>
<tr>
<th></th>
<th>$e^L$</th>
<th>$e^H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n^L$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$n^H$</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2: Definition of types by initial wealth and productivity
The Lagrange function in general form is:

\[
L = \sum_{i=1}^{N} N^i (u(c^i) + \delta v(x^i) + \psi(1 - y^i)) + \lambda \left[ \sum_{i=1}^{N} N^i (n^i y^i + (1 + \theta)(c^i - c^i) - x^i) - G \right] \\
+ \sum_{i,j=1, i \neq j}^{N} \mu^{ij} [u(c^i) + \delta v(x^i) + \psi(1 - y^i) - \hat{u}(c^j) - \delta v(x^j) - \psi(1 - \frac{n^i}{n^j} y^j)] \\
+ \sum_{i,j=1, j \neq i}^{N} \mu^{ji} [u(c^j) + \delta v(x^j) + \psi(1 - y^j) - \hat{u}(c^i) - \delta v(x^i) - \psi(1 - \frac{n^j}{n^i} y^i)]
\]

(20)

The first order conditions with respect to \(c^i\), \(x^i\) and \(y^i\) are

\[
\frac{\partial L}{\partial c_i} = N^i u' - \lambda (1 + \theta) N^i + \sum_{i,j=1, i \neq j}^{N} \mu^{ij} u' - \sum_{i,j=1, i \neq j}^{N} \mu^{ji} \hat{u}' = 0 \\
\frac{\partial L}{\partial x_i} = N^i \delta v' - \lambda N^i + \sum_{i,j=1, i \neq j}^{N} \mu^{ij} \delta v' - \sum_{i,j=1, i \neq j}^{N} \mu^{ji} \delta v' = 0 \\
\frac{\partial L}{\partial y_i} = -N^i \psi' + \lambda N^i n^i - \sum_{i,j=1, i \neq j}^{N} \mu^{ij} \psi' + \sum_{i,j=1, i \neq j}^{N} \mu^{ji} \psi' = 0.
\]

(21, 22, 23)

For the maximin social objectives the first term drops out for types 3 and 4. In this setting the distortions depend on which self-selection constraints are binding. Note particularly that the single-crossing property does not necessarily hold meaning that the self-selection constraint can bind towards both directions simultaneously. Whenever there is at least one binding self-selection constraint between types with different initial wealth, there is a distortion for the saving decision compared to the first best. Formally we see this from the condition

\[
\frac{(N^i + \sum_{i,j=1, i \neq j}^{N} \mu^{ij}) u' - \sum_{i,j=1, i \neq j}^{N} \mu^{ij} \hat{u}'}{(N^i + \sum_{i,j=1, i \neq j}^{N} \mu^{ij} - \sum_{i,j=1, i \neq j}^{N} \mu^{ji}) \psi'} = \frac{\delta}{r}
\]

(24)

for all \(i\) in the case of utilitarian social objectives and for the types 1 and 2 in the case of maximin.
social objectives. For types 3 and 4 under maximin the condition is:

\[
\frac{\sum_{i,j=1}^{N} \mu_{ij} u' - \sum_{i,j=1}^{N} \mu_{ji} u'}{\left( \sum_{i,j=1}^{N} \mu_{ij} - \sum_{i,j=1}^{N} \mu_{ji} \right) v'} = \frac{\delta}{r}.
\]  

(25)

3.1 3-type model

In the four-type model we cannot state which of the self-selection constraints are binding. It is especially difficult to prove whether the constraint between types 2 and 3 is binding upwards or downwards, or both ways. To get further insight from the analytical results we study a 3-type case where economy consists low-ability and low-wealth individuals (type 1) and high-ability individuals with varying initial wealth levels (types 3 and 4). We assume that the binding self-selection constraint are from types with higher resources towards types with lower resources meaning \( \mu_{41}, \mu_{43}, \mu_{31} > 0 \).

Now we can solve the distortions for savings:

\[
d_1 = \frac{\mu_{41} + \mu_{31} \left( 1 - \frac{r \hat{u}'}{\delta v'} \right)}{N^{1}} > 0
\]

(26)

\[
d_3 = 0
\]

(27)

\[
d_4 = 0.
\]

(28)

Distortions are the same under maximin social objectives\(^4\). That is with wealth heterogeneity the marginal tax on capital income is positive for the low-ability type in order to relax the otherwise binding self-selection constraints.

The optimal marginal labour income tax is non-zero for low-ability type and 0 otherwise, indicating that there is no distortion at the top. Specifically the marginal labour income tax for type 1 is:

\[
\hat{T}_1 = \frac{\left( \mu_{41} + \mu_{31} \right) \left( 1 - \frac{n_L}{n} \right)}{N^{1} - \left( \mu_{41} + \mu_{31} \right) \frac{n_L}{n}}
\]

(29)

We can state the following proposition:

\(^4\)Note that the levels of marginal distortions vary between the social objectives because the optimal bundles of consumption and leisure differ.
Proposition 2

In the three-type economy, where low-ability types have low initial wealth and the wealth level varies within the high-ability type, only the saving decision of the low-ability type is distorted at the margin both in the utilitarian and the maximin cases. The result hinges on the assumption that self-selection constraints are binding along decreasing resources (combining productivity and initial wealth). The ”no distortion at the top” result holds.

3.2 Numerical illustration

For the full four-type model we solve some numerical examples. These numerical examples show the qualitative properties of the optimal tax structure to better understand the analytical results. We choose a Cobb-Douglas utility function and parameter values following Cremer et al. (2001). The utility function is separable and represented in the following way: \( U^i = \log c^i + \delta \log x^i + \log (1 - y^i) \). The parameters for the baseline case are as in the table 3.

<table>
<thead>
<tr>
<th>Fraction of individuals in each group</th>
<th>( N^i = 0.25 \text{ for } i = 1, 2, 3, 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial wealth</td>
<td>( e^L = 2, e^H = 10 )</td>
</tr>
<tr>
<td>Discount rate</td>
<td>( \delta = 0.9 \text{ and } r = 0.95 )</td>
</tr>
<tr>
<td>Productivities</td>
<td>( n^L = 6, n^H = 9 )</td>
</tr>
</tbody>
</table>

Table 3: Parameters for the numerical simulations, baseline

Table 4 shows the optimization result in a baseline case. The upper panel is for the utilitarian social objective function and lower panel for the maximin case. For the benchmark we report the distribution of savings, labour income, lifetime consumption and the distortions for saving and labour income. The seventh column also presents the life-time utility. The binding self-selection constraints are also listed. In the last column we also show a simple inequality measures which are calculated by taking the absolute difference between the extreme types (types 1 and 4).

For the baseline model in the utilitarian case the only slack self-selection constraints are \( \mu^{12}, \mu^{13} \) and \( \mu^{23} \) and for the maximin case \( \mu^{12}, \mu^{13}, \mu^{23}, \mu^{34} \) and \( \mu^{13} \). This demonstrates that the intuition from the one-dimensional case does not hold in a multidimensional screening problem.
For example constraint $\mu^{34}$ binds even though type 4 has a higher disposable income potential in the first best case without distortive taxation. Also it is interesting that $\mu^{41}$ binds as it links the extreme types and indicates that type 1 is taxed at the margin partly to prevent the high-ability high-wealth type from pretending to be both low-ability and low-wealth. Difference in the after-tax incomes between these two types can be reduced only until the self-selection constraint becomes binding.

Turning to the optimal distortions, the simulation exercise suggests that the saving decision is distorted with a positive tax for the low-wealth types at the margin. Types with high-wealth types are instead subsidized. The tax on type 1 relaxes all the self-selection constraints towards the type 1. The subsidy for high wealth types helps to mitigate the binding self-selection constraint as now the potential mimickers do not want to meet the high-wealth types’ savings level.

The marginal distortion for the labour supply are somewhat surprising, however, not uncommon in the multidimensional tax problems (Cremer et al., 2001). We observe that the negative marginal tax rates occur to the high-ability and high-wealth type. Comparing the social objectives tells that the saving distortion for type 1 and 4 are substantially affected by changing the redistributive preferences to maximin.

<table>
<thead>
<tr>
<th>Utilitarian</th>
<th>s</th>
<th>ny</th>
<th>$c + x$</th>
<th>d</th>
<th>$T'$</th>
<th>lifetime $U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>0.55</td>
<td>2.06</td>
<td>2.24</td>
<td>29.0</td>
<td>35.3</td>
<td>-0.25</td>
</tr>
<tr>
<td>Type 2</td>
<td>1.90</td>
<td>1.07</td>
<td>9.48</td>
<td>-17.2</td>
<td>32.3</td>
<td>2.18 difference in lifetime consumption: 8.72</td>
</tr>
<tr>
<td>Type 3</td>
<td>0.28</td>
<td>6.0</td>
<td>3.96</td>
<td>0.5</td>
<td>0</td>
<td>0.17 difference in lifetime utility: 2.65</td>
</tr>
<tr>
<td>Type 4</td>
<td>1.73</td>
<td>1.05</td>
<td>9.96</td>
<td>-24.0</td>
<td>-34.3</td>
<td>2.40</td>
</tr>
<tr>
<td>Binding SS</td>
<td>14, 21, 24, 31, 32, 41, 42, 43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maximin</th>
<th>s</th>
<th>ny</th>
<th>$c + x$</th>
<th>d</th>
<th>$T'$</th>
<th>lifetime $U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>0.7</td>
<td>2.22</td>
<td>1.86</td>
<td>45.3</td>
<td>47.4</td>
<td>-0.71</td>
</tr>
<tr>
<td>Type 2</td>
<td>1.66</td>
<td>1.74</td>
<td>9.28</td>
<td>-22.5</td>
<td>57.2</td>
<td>1.71 difference in lifetime consumption: 8.7</td>
</tr>
<tr>
<td>Type 3</td>
<td>0.53</td>
<td>5.94</td>
<td>3.19</td>
<td>0</td>
<td>0</td>
<td>-0.21 difference in lifetime utility: 2.67</td>
</tr>
<tr>
<td>Type 4</td>
<td>1.52</td>
<td>5.04</td>
<td>10.56</td>
<td>-12.8</td>
<td>-12.5</td>
<td>1.96</td>
</tr>
<tr>
<td>Binding SS</td>
<td>14, 21, 24, 31, 32, 41, 42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4**: Baseline model simulation

Table presents in the baseline model the optimal savings $s$, labour income $ny$, lifetime consumption $c + x$, saving distortion $d$, marginal labour income tax $T'$ and lifetime utility $U$. Differences in lifetime consumption and utility are calculated as absolute difference between the extreme types, i.e. 1 and 4. The binding self-selection constraints listed at the bottom.

The impact of wealth inequality
Next we will explore the role of wealth inequality. The overall wealth in the economy is kept constant but the distribution is changed. Other parameters are as in baseline case. The results on the variables we are the most interested in are shown in table 5.

It appears that the saving distortion increases for type 1 and 3 as the wealth inequality is reducing. At the same time the subsidy provided for the high-wealth types increases. In the utilitarian case the marginal labour income tax for the type 1 is minimally affected with decreasing wealth inequality. For high-initial-wealth types the optimal labour supply distortions are non-monotone.

Generally, as the wealth differences are narrower, the marginal distortions get larger. In the case that there is no initial wealth differences, we are back to (Mirrlees, 1971) result where there is no distortion for the intertemporal decision and the labour supply of the low ability type is distorted.

<table>
<thead>
<tr>
<th></th>
<th>$e^L$</th>
<th>$e^H$</th>
<th>$d^1$</th>
<th>$d^2$</th>
<th>$d^3$</th>
<th>$d^4$</th>
<th>$T^1$</th>
<th>$T^2$</th>
<th>$T^3$</th>
<th>$T^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10</td>
<td>29,0</td>
<td>-17,2</td>
<td>0,5</td>
<td>-24,0</td>
<td>35,3</td>
<td>32,3</td>
<td>0</td>
<td>-34,3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>41,9</td>
<td>-27,3</td>
<td>2,5</td>
<td>-28,3</td>
<td>45,1</td>
<td>33,2</td>
<td>-0,02</td>
<td>-46,2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>59,8</td>
<td>-73,6</td>
<td>16,6</td>
<td>-51,0</td>
<td>52,4</td>
<td>18,0</td>
<td>-24,4</td>
<td>-34,4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23,15</td>
<td>23,15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$e^L$</th>
<th>$e^H$</th>
<th>$d^1$</th>
<th>$d^2$</th>
<th>$d^3$</th>
<th>$d^4$</th>
<th>$T^1$</th>
<th>$T^2$</th>
<th>$T^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10</td>
<td>45,3</td>
<td>-22,5</td>
<td>0</td>
<td>-12,8</td>
<td>47,3</td>
<td>57,3</td>
<td>0</td>
<td>-12,5</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>65,2</td>
<td>-45,2</td>
<td>0</td>
<td>-25,3</td>
<td>70,4</td>
<td>45,2</td>
<td>0</td>
<td>-9,7</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>68,8</td>
<td>-67,3</td>
<td>14,3</td>
<td>-48,9</td>
<td>63,3</td>
<td>37,3</td>
<td>-7,9</td>
<td>-9,8</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>43,3</td>
<td>43,3</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 5: Impact of wealth inequality**

*Correlation between ability and wealth*

In table 6 we explore how the correlation between ability and wealth affects the optimal distortions keeping the wealth inequality and other parameters in the baseline specification. The above analysis have assumed zero correlation. Correlation 1 corresponds to the two-type case discussed in section 2 and verifies that the only binding constraint is towards the low-ability and low-wealth type. In the two-type case the government can recognize the true type more easily than in a case where the is less than perfect correlation\(^5\). For this reason the distortion for the

\(^5\)The perfect negative correlation is not shown because with this specification neither of the self-selection constraint is binding and so one cannot solve for the optimal distortion.
type 1 is smaller than with the zero or positive correlation cases. Compared to the linear tax case studied in Cremer et al. (2001), there are remarkable differences in the findings. While with linear savings tax and unobservable savings increasing the correlation indicates a higher tax, with the non-linear case there is no clear pattern but only that the distortions are non-monotone.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>$d^1$</th>
<th>$d^2$</th>
<th>$d^3$</th>
<th>$d^4$</th>
<th>$T^1$</th>
<th>$T^2$</th>
<th>$T^3$</th>
<th>$T^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>positive</td>
<td>37.7</td>
<td>-20.2</td>
<td>0</td>
<td>-22.3</td>
<td>52.9</td>
<td>23.7</td>
<td>0</td>
<td>-32.4</td>
</tr>
<tr>
<td>0</td>
<td>29.0</td>
<td>-17.2</td>
<td>0.5</td>
<td>-24.0</td>
<td>38.3</td>
<td>32.3</td>
<td>0</td>
<td>-35.0</td>
</tr>
<tr>
<td>negative</td>
<td>12.0</td>
<td>-11.0</td>
<td>2.1</td>
<td>-29.4</td>
<td>10.5</td>
<td>39.0</td>
<td>-1.45</td>
<td>-36.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation</th>
<th>$d^1$</th>
<th>$d^2$</th>
<th>$d^3$</th>
<th>$d^4$</th>
<th>$T^1$</th>
<th>$T^2$</th>
<th>$T^3$</th>
<th>$T^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>52.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>35.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>positive</td>
<td>63.1</td>
<td>-25.3</td>
<td>6.8</td>
<td>-9.8</td>
<td>62.2</td>
<td>54.7</td>
<td>-3.7</td>
<td>-6.7</td>
</tr>
<tr>
<td>0</td>
<td>45.3</td>
<td>-22.5</td>
<td>0</td>
<td>-12.8</td>
<td>51.5</td>
<td>57.2</td>
<td>0</td>
<td>-7.1</td>
</tr>
<tr>
<td>negative</td>
<td>0</td>
<td>-14.3</td>
<td>0</td>
<td>-18.5</td>
<td>38.4</td>
<td>55.2</td>
<td>0</td>
<td>-7.2</td>
</tr>
</tbody>
</table>

**Table 6:** Impact of correlation between ability and wealth

4 Including income shifting

Lastly, we extend our model to include income shifting. The same question with similar model has been studied in Christiansen and Tuomala (2008) but with a linear capital income tax. They study income shifting in a two-type and two period model when there is a perfect correlation between productivity and initial wealth. They solve the optimal tax rate analytically. For now, we will also consider a two-type model with perfect correlation and leave the numerical results and other cases of correlations for future. We are interested in whether the results from Christiansen and Tuomala (2008) extend to an economy where government can also tax savings with a non-linear tax schedule.

Earlier we assumed that government can observe savings and income from labour. However, now we shift our focus on more realistic environment where savings and labour income are reported to the government. So, from the point of view of the government, the savings and income are only partially observed. It is costly for government to monitor if incomes are reported truthfully. This creates an incentive to shift part of the labour income to capital income if there are differences in the tax rates.
To model income shifting, we assume that individuals shift an amount of $\Delta$ of labour income to capital income tax base at a cost of $k(\Delta)$. We restrict only to the case where $k(\Delta) > 0$ and so rule out the income shifting from capital income to labour income. The cost is increasing and convex in $\Delta$, i.e $k'(\Delta) > 0$, $k''(\Delta) > 0$. Individual report labour income $z^R$ while the true labour income equals to $z = z^R + \Delta$. Individual’s actual labour supply is $y = \frac{z^R + \Delta}{n}$. Individuals shift capital income as long as the marginal saving in tax is larger than the cost of transforming one unit of labour income to capital.

Each type chooses how much to report as savings, how much to save from the first period and the labour supply to maximize his utility. To characterize Pareto efficient second best taxes, we set the government problem as to maximize

$$\max_{c^i, x^i, y^i} G(U) = \sum_{i=1}^{N} N^i (u(c^i) + \delta v(x^i) + \psi(1 - y^i))$$

subject to

$$\sum_{i=1}^{N} N^i (n^i y^i + (1 + \theta)(e^i - c^i) - x^i) \geq G$$

$$U^H \geq U^{HL},$$

where self-selection constraint are (assuming only $\mu^{HL}$ binds)

$$u(c^H) + \delta v(x^H) + \psi(1 - y^H) \geq \hat{u} + \delta \hat{v} + \psi(1 - \frac{n_L}{n^H} y^L) \equiv U^{ij}.$$  

As before $\hat{u}$ denotes the mimickers first period utility at $u(c^L + e^H - e^L)$. The new type of mimicking in this context is that the high-ability type may want to mimic the low-ability type and in reality have larger second period consumption due to the income shifting. In this case the second period utility equals to $\hat{v}(x^L + (\Delta^H - k(\Delta^H) - (\Delta^L - k(\Delta^L))$. Both types of mimicking behaviour needs to be accounted.

Solving the first order conditions and rearranging the terms as before for the saving distortion we notice that the high-ability and high-wealth type faces zero distortion as before but for the low-ability and low-wealth type the distortion is:
\begin{equation}
\frac{N^L u' - \mu^{HL} \tilde{w}'}{N^L v' - \mu^{HL} \tilde{v}'} = \frac{\delta}{r}.
\end{equation}

If there is no wealth differences, the nominator becomes \((N^L - \mu^{HL})u'\). The denominator takes into account the income shifting. Because \(\tilde{v}' < v'\) (assuming that high-ability individual shifts more income), the denominator is larger and the marginal rate of substitution between the two periods is smaller than without income shifting. This indicates that there is a positive marginal tax for the low-ability -type. If there are differences in the initial wealth, the sign of the distortion is ambiguous since nominator and denominator is working in opposite directions. This requires numerical simulations and are left for the future versions of the model.

\section{Conclusions}

Rising inequality is often (too often) discussed in differences in labour income. The standard optimal income tax analysis is also based on differences in earnings capacity. The wage distribution is certainly important but people differ also in wealth they have. Increasing wealth inequality has motivated us to study the non-linear taxation of labour and capital income in a two period model where agents differ in their earnings capacity and in their initial wealth levels.

Given the Atkinson-Stiglitz result capital income tax might be unnecessary in a setting where non-linear labour income tax is available. We have shown that there may, however, be a role for taxing savings. Non-linear marginal savings tax schedule alleviates the informational constraints. Our finding indicates that low-wealth types’ savings are taxed at margin while the high-wealth type’s saving decision is distorted upwards. This reveals that there is a case for non-linear capital income tax in the optimal tax instrument mix.

Because multidimensional problems can be difficult to grasp without numerical simulation, in particular because the analytical results turns out to be ambiguous, several model economies are set up for the numerical examples in this paper. We have studied the effect of wealth inequality on the optimal savings and labour supply distortion. We have also characterized the effect of correlation between ability and wealth. Lastly, we have studied the role of income shifting in an analytical framework.

We motivated the study with the observation that wealth and inheritance taxes have been
going out of favour in many developed country. While we have shown that the capital income tax belongs to the optimal tax mix, the extent of redistribution changes little between the maximin and the utilitarian social objectives. This might suggest that initial wealth might be easier to target with wealth taxation.

References and Notes


