TAMPERE ECONOMIC WORKING PAPERS
NET SERIES

ON OPTIMAL LIFETIME REDISTRIBUTION POLICY

Sanna Tenhunen
Matti Tuomala

Working Paper 50
January 2007
<http://tampub.uta.fi/econet/wp50-2007.pdf>

DEPARTMENT OF ECONOMICS AND ACCOUNTING
FI-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191
ISBN 978-951-44-6855-1

On optimal lifetime redistribution policy

by

Sanna Tenhunen

University of Tampere and FDPE

and

Matti Tuomala

University of Tampere

11.01.2007

Abstract: In this paper we examine various aspects of the optimal lifetime redistribution policy within a cohort. We characterise optimal redistribution policy when society consists of individuals who do not differ only in productivity, but also in time preference or myopia. We extend Diamond's (2003) analysis on nonlinear taxation of savings into the three and four type models. Our results provide a rationale for distortions (upward and downward) in savings behaviour in a simple two period model where high skilled and low skilled individuals have different non-observable time preferences beyond their earning capacity. If we interpret our model so that there is no private savings, but public provision of pension in period 2, then in different versions (and different parameterization) of three type model we find the U-shaped pattern of the replacement rates. Our numerical results suggest that the retirement consumption is less dispersed than the first period consumption in a paternalistic case, whereas in a welfarist case the ordering is reversed. Our numerical simulations also show that consumption when old should be less dispersed than consumption when young when some individuals are myopic. Moreover our numerical results suggest that when there are myopic individuals in the economy, a paternalistic government policy increases saving and makes saving larger than with paternalist government policy where there are no myopic individuals.

Keywords: Lifetime redistribution, retirement consumption, heterogeneous preferences, myopia.

1. Introduction

Publicly provided retirement programs, the largest single income source of the elderly, can be justified on several grounds. A standard justification for public pension systems is market failures generated by asymmetric information. Another possible justification for government intervention in retirement and saving decisions is redistributive grounds supplementing other redistributive instruments such as income taxation. A third standard justification for public pension systems or compulsory pension contributions relies on the assumption that some individual behave myopically, consuming “excessively” during their earnings years and then finding themselves with insufficient saving in retirement. The first two rationales for public pension program have received much more attention in public economics than the third one.

There is, however, some literature on pension policies that attempts to take into account the possible “undersaving” by households. Diamond (1977) discussed the case where individuals may undersave due to mistakes. Sheshinski (2003) proposes a general model with faulty individual decision making, where restricting individuals' choices leads to welfare improvements. Feldstein (1985) examined the case where individuals have higher discount rates than the government. Given this assumption he studied the optimal pay-as-you-go system in an OLG setting. Feldstein's (1985) model is extended by İmrohoroglu, İmrohoroglu and Joines (2003). They made numerical simulations in a pay-as-you-go model when individuals have hyperbolic discounting preferences. They conclude that pension system provides additional welfare for myopic individuals. There are also other models based on hyperbolic discounting. Diamond and Köszegi (2003) employ this kind of model to study the policy effects of endogenous retirement choices¹. In their analysis a public pension system is a commitment device.

Recent research in behavioural economics has demonstrated that individual's decision making often suffers from various biases. In these situations when there is the possible conflict between individual's preference for the long run and his or her short run behaviour the government may want to intervene. The normative analysis of such individual decision failures in the context of the design of pension system has not yet received much attention. One important exception is the recent book by Diamond (2003). Another study by Cremer et al (2006), closely related to our

¹ See also Cremer, Lozachmeur and Pestieau (2004).

paper, also examines non-linear social security scheme when the government has a paternalistic view and wants to help to overcome individuals' myopia problem. As in Diamond (2003) we also analyze lifetime redistribution of income across individuals within a cohort.² Thereby we avoid the dynamic complications that arise from intergenerational redistribution. More specifically, we consider a two-period variant of the Mirrlees income tax problem, where individuals work and then retire. As well known, the Mirrlees model as an annual tax system requires that the government can commit to ignore the information that has been revealed by individual's choices to it. Mirrlees (1971) suggests that one of the limitations of his analysis is that it does not address intertemporal problems. "in an optimum system, one would no doubt wish to relate tax payments to the whole life pattern of income...". Therefore reinterpreting the Mirrlees model as a model of lifetime redistribution we assume that the government can commit to a lifetime tax³. We will assume that everyone retires at the same age. We focus on different versions of such a model depending on assumptions on savings and labour supply. In other words are individuals time-consistent over their lifetime or not? The lifetime version of Mirrlees model can be interpreted either that the government controls the first and second period consumption and labour supply directly, subject to the self-selection constraints or that if we assume that there is no private savings we have a model of labour income taxation in period 1 and public provision of pension in period 2. The first interpretation means that we consider the many-good nonlinear tax model.

In the Mirrlees model the rich are different from the poor in only one way; they are endowed with the ability to command a higher market wage rate, which is assumed to reflect a higher productivity of their labour effort. In fact there is a variety of other reasons to why some people end up affluent and other do not. One might well argue that both high-income and low-income people do not owe their (un)success entirely to ability, but part of the income differentials are

² Vickrey (1939) is the famous early contribution to the normative theory of lifetime income taxation. He proposed a cumulative averaging system for personal income taxation. Vickrey argued that if the tax schedule is convex then individuals with fluctuating incomes would pay more taxes on average than individuals with steadier incomes. Vickrey based his argument on horizontal equity.

³ In fact Roberts (1984, proposition 4) in an optimal dynamic taxation models without government commitment raises doubts about the government's ability to use information from early periods of life to accomplish redistribution in later periods of life with lower welfare cost. Gaube (2005) in turn interprets the practice of taxing current income in each period as a partial commitment device. Roberts (1984) and Gaube (2005) do not, however, consider taxes on savings as a possible tax instruments. See also Berliant and Ledyard (2005) for a more recent contribution.

due to luck, different time preferences and inherited resources.⁴ Hence unlike in the original Mirrlees model we assume that individuals do not differ only in ability, but also time preferences. It is quite plausible to assume that in reality for all sorts of reasons both ability and time preferences are not observable. First best taxation is not possible in this economy, because we cannot distinguish ex ante between the two types.

Economists such as Wicksell, Pigou, Ramsey and Allais⁵ have argued for the case that the society or social planner should be more patient than individuals. For example, Ramsey (1928) claimed that it was “ethically indefensible” to discount the future. Any such argument can be called paternalistic. The notion that individuals may not make the best choices for themselves raises difficult issues. Individual may be fully rational and they just happen to have a high preference for the present, which causes them to save little, because too little weight is given to future contingencies. Should the government maximise individual welfare, as the individual sees it. In other words should it be welfarist? Or, as suggested in recent behavioural public economics literature, should it be paternalistic or non-welfarist and discount the future at a lower rate than individuals. In fact individuals may want the government to intervene, to induce behaviour that is closer to what individuals wish they were doing. For example, in models with quasi-hyperbolic individuals it is typically desirable to impose a particular savings plan on individuals.

The object of this paper is to characterize optimal lifetime redistribution policy within cohort. We attempt to answer following questions both in welfarist and paternalistic (non-welfarist) cases. Should returns to saving be taxed or subsidised if individuals differ both in abilities and discount rates or myopia? How does the replacement rate depend on earnings? Is there less dispersion in retirement consumption than in the first period consumption? Is there less

⁴ There are also considerable variations in the rate of return people receive on their capital. It is quite possible that some of variations in the return on capital are the result of the application of skill and effort; but most is clearly the result of risky outcomes (luck).

⁵ ”In actual societies it seems to be common that social choices deviate from consumer preferences in the assessment of the relative importance of future needs with respect to present needs...Public saving and legal arrangement such as compulsory pension schemes allow this objective to be realised. ...It was in order to generalise optimum theory to such a collective attitude that M. Allais put forward the concept of ‘rendement social généralisé’. His idea is to define and investigate a notion of optimum in which individual preferences are retained for the choice between consumption relating to the same date, but not necessarily between those relating to different dates.” (Malinvaud, 1972, p. 244)

dispersion in retirement consumption when labour supply decisions are myopic? To gain better understanding about these questions we also solve some numerical examples

The structure of the paper is as follows. Section 2 considers a two period and two type version of the Mirrlees model with a positive correlation between skill and discount factor and with paternalistic (i.e. the government's and individual's discount rates differ) and welfarist governments. Section 3 in turn considers a four type model with different time preferences and computes some numerical examples. It also examines case with myopic individuals. We don't choose binding self selection constraints a priori. We simply determine them by solving this problem numerically. Based on lessons from numerical simulations of a four type model section 4 derives some analytical results in a three type model and also solves numerical examples. Section 5 concludes.

2. A benchmark model: two types with a positive correlation between skill and discount factor

Unlike in the original Mirrlees model we assume that individuals do not differ only in productivity, but also in discount rate or time preference⁶. In the simple two-type framework model, similar to the much used two type model first introduced by Stern (1982) and Stiglitz (1982). Each individual has a skill level reflecting his or her wage rate, denoted by n , and a discount factor, denoted by $\delta = \frac{1}{1 + \xi}$, where ξ is discount rate or time preference. We denote low skill types by superscript L and high skill types by superscript H. The assumption of positive correlation implies that $\delta^L \leq \delta^H$. The proportion of individuals of type i in the population is $N^i > 0$, with $\sum N^i = 1$. The social planner has a discount factor, δ^g , which can be assumed to be at least as high as the highest individual discount factor. In a welfarist case $\delta^g = \delta^i$, but in a non-welfarist or paternalist case the discount factor of the social planner may differ from the individuals' discount factors.

⁶ In Sandmo (1993) people are endowed with the same resources, but differ in preferences. Tarkiainen and Tuomala (1999, 2007) in turn numerically simulated optimal non-linear tax schedules for a continuum of taxpayers simultaneously distributed by skill and preferences for leisure and income.

As is well known, due to Atkinson-Stiglitz (1976), under a mild separability assumption, income taxation does not need be supplemented by other taxes. Saez (2002) argues that the Atkinson-Stiglitz result of commodity taxes holds when each individual have identical discount rates. He also argues that individuals with higher earning save relatively more, which suggests that high skill individuals are more likely to have higher discount factors. Hence discount factor is positively correlated with productivity level. For this reason we take as a starting point separable utility with different time preference.⁷ The life-time utility of an individual of type i is additive in the following way:

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

where c and x denote consumption when young and when retired, respectively, and y is labour supply when young. It is increasing in each argument and $u', v', \psi' > 0$ and $u'', v'' < 0$ and strictly concave. We also assume that all goods are normal.

The government wishes to design a lifetime tax system that may redistribute income between individuals in the same cohort. There is asymmetric information in a sense that tax authority is informed neither about individual skill levels, labour supply nor discount rates. It can only observe before-tax income, ny . To introduce return to capital and the possible taxation thereof, it is useful to consider a two-period model. The economy lasts two periods. Individuals are free to divide their first period (when young) income between consumption, c and savings, s . Each unit of savings yields a consumer an additional $1+\theta$ units of consumption in the second period after tax income, x . As further simplifications we assume that there is a fixed rate of return to savings, which may be justified by assuming that we consider a small open economy facing a world capital market. Consumption in each period is given by $c^i = n^i y^i - T(n^i y^i) - s^i$ and $x^i = (1 + \theta)s^i$, $i = L, H$.

In this setting where taxes on both earnings and savings income are available we examine whether or not saving ought to be taxed. The separability assumption makes it possible to isolate the significance of variations in time preferences.

⁷ Alternatively the same outcome could be reached by assuming homothetic preferences and linear Engel curves.

2.1. The welfarist government

First we analyse a model in which the government does respect the individual sovereignty principle and evaluates individuals' well-being using their own discount rates. Now let's assume that the government controls c^i , x^i and y^i directly, subject to the self-selection constraint⁸. Alternatively if we assume that there is no private savings we have a model of labour income taxation in period 1 and public provision of pension in period 2. In the welfarist case, where $\delta^g = \delta^i$ $i = L, H$, the optimisation problem is maximisation of a following social welfare function

$$\sum N^i (u(c^i) + \delta^i v(x^i) + \psi(1 - y^i)) \quad (2)$$

subject to the revenue constraint

$$\sum N^i (n^i y^i - c^i - r x^i) = R \quad (3)$$

where $r = \frac{1}{1 + \theta}$, and self-selection constraint

$$u(c^H) + \delta^H v(x^H) + \psi(1 - y^H) \geq \hat{u}(c^L) + \delta^H \hat{v}(x^L) + \hat{\psi}\left(1 - \frac{n^L}{n^H} y^L\right)^9 \quad (4)$$

Note that in the constraint a mimicker values future consumption with his own discount factor instead of using the discount factor of the mimicked. Multipliers λ and μ are attached to the budget and self-selection constraints, respectively. The Lagrange function of the optimization problem is

$$\begin{aligned} L = & \sum_i N^i \left[u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \right] + \lambda \left[\sum N^i (n^i y^i - c^i - r x^i) - R \right] \\ & + \mu \left[u(c^H) + \delta^H v(x^H) + \psi(1 - y^H) - \hat{u}(c^L) - \delta^H \hat{v}(x^L) - \hat{\psi}\left(1 - \frac{n^L}{n^H} y^L\right) \right] \end{aligned} \quad (5)$$

⁸ The direction of the binding self-selection constraint is assumed to be, following the tradition in two-type model literature, from high-skilled individual towards low-skilled individual. This pattern is also confirmed by the numerical simulations we present in Appendix B1.

⁹ The terms with hat refer to mimicking behaviour. Note that regardless of different notation, the functional form is the same with or without a hat.

Optimality conditions are formulated in terms of marginal rates of substitution rather than explicit tax rates, but as has become conventional in the literature we may interpret the marginal rate of substitution between gross and net income as one minus the marginal income

tax $\frac{\psi'(\frac{ny}{n})}{nu_c(c)} = 1 - T'(ny)$, which would be equivalent to the characterisation of the labour

supply of an unspecified type of agent facing an income tax function $T(ny)$. Marginal labour income tax rates satisfy the usual properties; $T'(n^L y^L) > 0$ and $T'(n^H y^H) = 0$. Our main interest is in marginal taxation savings. For this purpose the first order conditions can be written

in the form $\left(\frac{u_c}{v_x}\right)^i = \frac{\delta^i}{r} [1 - d^i]$, where the left hand side is individual i 's marginal rate of

substitution between consumption in period one and consumption in period two and d^i is the distortion. Positive (negative) d^i implies that type i should have implicit tax (subsidy) on

savings. It is handy to define relative difference in discount factors as $\Delta^{ij} \equiv \frac{\delta^i - \delta^j}{\delta^j}$, for any

pair of discount factors $i, j = L, H, g$. The first order conditions (Appendix A) now imply that

$$\begin{aligned} d^L &= (\varphi^1 - 1)\Delta^{HL} \\ d^H &= 0 \end{aligned} \tag{6}$$

where $\varphi^1 = \frac{N^L}{N^L - \mu}$. Returns to savings of type i should not be taxed when d^i is zero. With

$d^H=0$, the optimal implicit marginal tax rate for high skill type is zero. *When we assume, empirically plausibly, $\delta^H > \delta^L$, we have $d^L > 0$ implying implicit taxation of savings for low skill type.* This is just the same result as in Diamond (2003).

Because of the two-dimensional heterogeneity, a tax on capital income is an effective way to relax an otherwise binding self-selection constraint. This is because even under separability mimicker and mimicking individual do not save the same amount. In other words distortions generate second-order efficiency costs but first-order redistributinal benefits.

The theory of optimum taxation is typically assumed that society favours redistribution from the high skilled to the low skilled individuals. In our model the high skilled types are also those

with high discount rate. Should this increase or decrease redistribution above the extent possible when we have only labour income tax available? This is an important normative question. And we think there is no easy answer. Another way to look at this question is to assume that the government has a paternalistic view.

2.2. Government's and individuals' discount rates differ

Here we analyse a model in which the government does not respect the individual sovereignty principle and evaluates individuals' well-being using discount rates different from those of individuals. Now the paternalistic government with a discount factor δ^g maximises $\sum N^i (u(c^i) + \delta^g v(x^i) + \psi(1 - y^i))$.

The government evaluates outcomes with respect to its own objective function. In the first best situation where there are no redistributive goals, there would be a subsidy on savings, equal to Δ^{gi} , that exactly corrects the difference in discount factors. In the second best case with non-linear taxation the outcomes generated by individual preferences enter government's budget constraint and thus they also have to be taken into account. Another way to look at this is that the self-selection constraint that the government faces has the individual's preferences as an integral part.

The form giving the implicit marginal taxes for savings is now $\left(\frac{u_c}{v_x}\right)^i = \frac{\delta^g}{r} [1 - \alpha^i]$, where α^i gives the distortion. However, from the perspective of the individual the distortion is still given by d^i in $\left(\frac{u_c}{v_x}\right)^i = \frac{\delta^i}{r} [1 - d^i]$. There is a correspondence between these two distortions, given by $\alpha^i = d^i + \Delta^{ig}(d^i - 1)$, for all types i . It can be noticed, that as long as $d^i < 1$ (i.e. implicit distortion for savings from individuals' point of view is less than 100 per cent), $\alpha^i > d^i$. In other words, positive (negative) α^i implies that the savings decision of type i is under-subsidized (over-subsidized) relative to first best case from individuals point of view¹⁰

¹⁰ In the first best situation the distortion is $-\Delta^{gi}$, which corrects the effect of the difference in discount rates.

From the first order conditions of government's problem (Appendix A) we get the following expressions:

$$\begin{aligned}\alpha^L &= (\varphi^1 - 1)\Delta^{Hg} \\ \alpha^H &= (\varphi^2 - 1)\Delta^{Hg}\end{aligned}\tag{7}$$

where $\varphi^1 = \frac{N^L}{N^L - \mu}$ and $\varphi^2 = \frac{N^H}{N^H + \mu}$.

From the equation (7) we see that the returns to savings should be distorted: α^L is negative and α^H is positive, when type H individual's discount factor is smaller than that of the social planner. In a special case, where $\delta^H = \delta^g$, α^H reduces to zero. This implies the following proposition¹¹

Proposition 1: *As long as $\delta^H < \delta^g$, relative to the first best situation the savings of the high skill type should be implicitly under-subsidized at the margin while savings of the low skill type should be implicitly over-subsidized.*

Relaxing self-selection constraint in this paternalistic case means the upward distortion of the savings of type 1 and the downward distortion of the savings of type 2. Contrary to a welfarist case, savings of the low skill type are now subsidized. In other words, compared to the savings required to the first best optimum, it is socially beneficial to have the high skill type to save less and the low skill type to save more.

2.3 Numerical simulations

To investigate further the properties of optimal redistribution we now turn to numerical examples. We assume the following separable forms of utility functions;

$$U^i = \log c^i + \delta^i \log x^i + \log(1 - y^i) \text{ (CD) and } U^i = -\frac{1}{c^i} - \delta^i \frac{1}{x^i} - \frac{1}{1 - y^i} \text{ (CES).}$$

We choose the following parameterization:

¹¹ We are grateful to Thomas Gaube for suggesting this interpretation.

fractions; $N^i = 0.5$, $i=L,H$,
wages; $n^L = 2$, $n^H = 3$,
discount factors; $\delta^L = 0.6$, $\delta^H = 0.8$, $\delta^g = 1$, $r = 0.95$

No a priori assumptions of the binding self-selection constraints are made in numerical simulation. The results verify the general assumption that only the upwards self-selection constraint binds in the optimum¹².

Tables 1 a,b,c present the numerical results for a welfarist, paternalist and Rawlsian government (maximizing utility of type L)¹³. In Tables we present the values of the marginal rates of labour income, T' , and of savings, α , utilities and replacement rates, in terms of second period consumption relative to first period gross income (x/ny), at the optimum¹⁴.

One way to see how the overall dispersion of consumption is in different cases is to consider standard Gini-coefficients in both periods. It turns out that we can also make welfare rankings between the distributions of first and second period consumption based on the results by Atkinson (1970) and Shorrocks (1980). They determine welfare ranking with help of Lorenz dominance which requires less normative assumptions on the social welfare function. According to Atkinson's result welfare is greater at period j when mean consumption is greater and Lorenz curve L_j is inside the Lorenz curve of the other period. Shorrocks' result can be applied when Lorenz curves cross or Lorenz dominant distribution has lower mean. It uses generalized Lorenz curve of period j , GL_j , for comparison; higher GL in all points implies higher welfare.

¹² The slackness of the other self-selection constraints is also checked by calculating the difference in utilities when mimicking and when not.

¹³ This interpretation does not adequately reflect the original idea of the difference principle in Rawls (1971). It is a kind of welfarist variation of the original principle.

¹⁴ We provide further information on numerical simulations in appendix B.

		U	T'	α	x/ny	Consumption dispersion
CD	type L	-1.4325	0.0356	0.0369	44.61	means: $\bar{c} = 0.9169$ $\bar{x} = 0.6756$
	type H	-1.2131	0	0	42.62	Lorenz dominance: $L_c > L_x$ Gini coefficients: $G_c = 0.050$ $G_x = 0.128$
CES	type L	-4.6711	0.0630	0.0672	51.20	means: $\bar{c} = 0.7283$ $\bar{x} = 0.6183$
	type H	-4.5626	0	0	44.42	Lorenz dominance: $L_c > L_x$ Gini coefficients: $G_c = 0.042$ $G_x = 0.086$

TABLE 1.a. Welfarist case: utility levels, marginal tax rates on labour income and saving, replacement rates and consumption dispersion

		U	T'	α	x/ny	Consumption dispersion
CD	type L	-1.5533	0.0683	-0.0440	59.33	means: $\bar{c} = 0.8192$ $\bar{x} = 0.8626$
	type H	-1.1766	0	0.0305	48.54	Lorenz dominance: $L_c < L_x$ Gini coefficients: $G_c = 0.090$ $G_x = 0.072$
CES	type L	-4.7919	0.1107	-0.0747	58.87	means: $\bar{c} = 0.6862$ $\bar{x} = 0.7065$
	type H	-4.5169	0	0.0428	47.62	Lorenz dominance: $L_c < L_x$ Gini coefficients: $G_c = 0.069$ $G_x = 0.055$

TABLE 1.b. Paternalistic case: utility levels, marginal tax rates on labour income and saving, replacement rates and consumption dispersion

		U	T'	α	x/ny	Consumption dispersion
CD	type L	-1.3606	0.2779	0.3830	42.06	means: $\bar{c} = 0.8809$ $\bar{x} = 0.5558$
	type H	-1.4044	0	0	38.52	Lorenz dominance: $L_c > L_x$ Gini coefficients: $G_c = 0.035$ $G_x = 0.213$
CES	type L	-4.5834	0.2943	0.4171	53.07	means: $\bar{c} = 0.7119$ $\bar{x} = 0.5488$
	type H	-4.7943	0	0	40.62	Lorenz dominance: $L_c > L_x$ Gini coefficients: $G_c = 0.028$ $G_x = 0.129$

TABLE 1.c. Rawlsian case¹⁵: utility level at the optimum, marginal tax rates on labour income and saving replacement rates and consumption dispersion

¹⁵ Optimisation with Cobb-Douglas utility function fails when size of the both groups is exactly 0.5. Thus we have solved problem in Cobb-Douglas case with closest possible combination, where the size of group 1 is 0.4986. In the case of CES utility function there occurs no problem due to size of the groups.

Our numerical simulations just confirm our analytical marginal tax rate results. The tables suggest that the retirement consumption, measured in terms of the Gini coefficient, is less dispersed than the first period consumption in a paternalistic case, whereas in welfarist and Rawlsian cases the ordering is reversed. As a corollary of Atkinson theorem we know with different mean incomes, the unambiguous welfare ranking survives only when it is the Lorenz-dominant distribution that has the higher means. It turns out to be so that in the welfarist and Rawlsian case the distribution of first period consumption is the Lorenz dominant distribution that has the higher mean. In the paternalist case it is the other way round. According to the numerical results the replacement rate (x/ny) decreases in earnings. Moreover, our numerical results suggest that a paternalistic government policy increases saving and makes saving larger than with a welfarist government policy (see Appendix B1, Tables B1.1 and B1.3)).

3. The 4-type case

In this section we generalise the previous model to a four type economy by giving up of the assumption that productivity and time preference are perfectly correlated. There are now four types of individuals that differ both in productivity and time preference. Our four types are presented in table 1:

	low skilled	high skilled
low delta	type 1	type 3
high delta	type 2	type 4

TABLE 2: Types of individuals

The maximisation problem is similar to the two type case: in the utilitarian or welfarist case the social planner maximises a sum of utilities, $\sum N^i [u(c^i) + \delta^i v(x^i) + \psi(1 - y^i)]$, subject to the revenue constraint $\sum N^i (n^i y^i - c^i - rx^i) = R$ $i=1,2,3,4$ and self-selection constraints $u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \geq \hat{u}^{ij}(c^j) + \delta^i \hat{v}^{ij}(x^j) + \hat{\psi}^{ij} \left(1 - \frac{n^j y^j}{n^i}\right)$ for $i,j=1,2,3,4$ and $i \neq j$.¹⁶

¹⁶ We don't choose the direction of redistribution a priori, i.e. bindingness of the self-selection constraints is not restricted a priori. We simply determine them by solving this problem numerically.

Our simulations were also carried out here for CD and CES functions. In the central variant specification we assume uniform distribution of types. Following assumptions of the parameters are made:

Fraction of individuals in each group	$N^i = 0.25$ for $i=1,2,3,4$
Discount factors	$\delta^L = 0.6, \delta^H = 0.8, r = 0.95$
Productivities (wages)	$n^L = 2, n^H = 3$

TABLE 3: Parameter values

In Table 4 the results for the case of welfarist or utilitarian government are presented¹⁷. Tables 5 and 6 present the results for a paternalist government with and without myopia. In Tables we have given the values of the marginal rates on labour income, (T'), and of savings, (α), utilities, (U) replacement rates (x/ny) of each type. We also present results on inequality in the first period consumption and retirement consumption. Moreover, we report in each case the pattern of binding self-selection constraints.

3.1 Welfarist case

		U	T'	α	x/ny	Consumption dispersion
CD	type 1	-1.4782	0.0440	0	44.28	means: $\bar{c} = 0.9102$ $\bar{x} = 0.6707$ Lorenz dominance: $GL_c > GL_x$ Gini coefficients: $G_c = 0.098$ $G_x = 0.102$
	type 2	-1.5931	0.0863	-0.0708	54.18	
	type 3	-1.1117	0	0.0320	36.14	
	type 4	-1.1884	0	0	43.20	
CES	type 1	-4.7327	0.0780	0	50.88	means: $\bar{c} = 0.7195$ $\bar{x} = 0.6181$ Lorenz dominance: $GL_c > GL_x$ Gini coefficients: $G_c = 0.073$ $G_x = 0.072$
	type 2	-5.1033	0.1317	-0.1138	56.40	
	type 3	-4.2166	0	0.0514	41.29	
	type 4	-4.5273	0	0	45.06	

TABLE 4: Utility levels at the optimum, marginal tax rates, replacement rates and Gini coefficients in a welfarist case. The binding self-selection constraints are $(3,1)$, $(3,2)$, $(4,3)$

¹⁷ Further information of numerical solutions is presented in Appendix B1.

In the optimum there are positive marginal income tax rates on both low-skilled types. The savings decision of types 1 and 4 is not distorted. Type 2 consumers, who in terms of utilities are worst-off types, have positive marginal income tax and a marginal subsidy on savings, whereas type 3, the best-off type in terms of utilities, has a marginal tax on savings.

Asymmetric information suggests that those individuals that are mimicked in the optimum should be taxed. In the two type case we had a positive marginal tax on both labour income and savings for the mimicked type. The pattern of the binding self-selection constraints explains the differences in optimal marginal tax rates for savings. As the only constraint that binds type 1 as the mimicked individual is (3,1), and types 1 and 3 have the same discount rates, taxation of savings does not make a difference between the types. On the other hand, they have different productivities, so income tax is an efficient tool to prevent mimicking. Equivalently, mimicking type 3 is made unattractive for type 4 by taxing savings, as taxing income would not have the desired effect between types with the same skill level.

3.2 Paternalistic case

Using a sum of utilities as social objective function may face some ethical objections. Namely if individuals are identical in preferences, equal marginal utilities of all coincides with equal total utilities. Sen (1973) pointed out that with diversity of human beings (eg. different time preferences in our case) the two can pull in opposite directions. For those who prefer to think of the justification for redistribution as being based on inequality of opportunity, differences in preferences may provide a suitable basis for distinguishing economic rewards but differences in skills in turn do not. This point of view raises questions on the nature of the parameter δ . Namely it may be argued that both skills and time preferences are “circumstances of birth”. Therefore, it is far from clear how suitable the utilitarian criterion is in this case. These considerations may be one reason to consider paternalistic cases with different utility functions.

In paternalistic case government maximises $\sum N^i [u(c^i) + \delta^g v(x^i) + \psi(1 - y^i)]$, the sum of utilities using its own discount factor, δ^g . In numerical solution we use $\delta^g = 1$ and the other parameter values given in table 3.

		U	T'	α	x/ny	Consumption dispersion
CD	type 1	-1.5901	0.0795	-0.1037	60.08	means: $\bar{c} = 0.8237$ $\bar{x} = 0.8614$ Lorenz dominance: $L_c < L_x$ Gini coefficients: $G_c = 0.105$ $G_x = 0.063$
	type 2	-1.6453	0.0795	-0.1037	60.08	
	type 3	-1.1422	0	0.1035	46.74	
	type 4	-1.1557	0	0.0323	48.64	
CES	type 1	-4.8363	0.1281	-0.1763	59.58	means: $\bar{c} = 0.6829$ $\bar{x} = 0.7061$ Lorenz dominance: $L_c < L_x$ Gini coefficients: $G_c = 0.073$ $G_x = 0.049$
	type 2	-5.1484	0.1281	-0.1763	59.58	
	type 3	-4.2316	0	0.1413	46.54	
	type 4	-4.4963	0	0.0449	47.99	

TABLE 5: Utility levels at the optimum, marginal tax rates, replacement rates and Gini coefficients in paternalistic case. The binding self-selection constraints are **(3,1), (3,2), (4,3)**

In paternalistic cases types 1 and 2 turn out to be pooled, i.e. they always choose a common bundle of labour supply and consumption. Similarly like in our welfarist case, in terms of utilities type 2 is the worst-off type and type 3 is the best-off type. With the pattern of binding self-selection constraints we could conclude that type 3 mimicking the common choice of types 1 and 2 is prevented by setting a positive marginal income tax on types 1 and 2. This is compensated with a subsidy on savings. The marginal tax on savings for types 3 and 4 is driven by the paternalistic preferences of the social planner.

3.3 Myopic individuals

Our analysis has so far considered only situations where individuals are rational. Next we analyse the model in which all individuals don't save voluntarily and in their labour supply decisions they don't take into account the effect of higher earnings on future pension benefits. They simply ignore the implications of their earnings when young on the retirement income. In other words they are myopic. Myopic behaviour may be quite common for a substantial proportion of individuals who hardly save and rely almost entirely on public pension benefits. Does this model generate less dispersed retirement consumption than the models with rational behaviour? Diamond (2003) shows that incentives of myopic individuals are not harmed by equalisation of pension benefits. Diamond (2003, chapter 4) and Diamond and Mirrlees (2000) consider a benchmark situation where individuals do not save at all. In their model workers are

otherwise identical, but their skills differ, and the government's objective is to design optimal redistributive policy for the working age and for the retired. If the social welfare function exhibits inequality aversion, the optimal retirement consumption is shown to be higher for those whose lifetime income has been smaller.

It is worth stressing that myopic behaviour is distinct from the behaviour associated with high discounting of the future. If individuals have high discount rates, they will save little for their retirement consumption, but this reflects optimising behaviour. By contrast, if individuals are myopic, and are subjected to forced saving, their welfare will increase. Although the behavioural foundations of myopia differ essentially from those of time consistent utility maximisation, the analysis developed above can be used with minor modifications. We simply interpret that the discount factor δ being either 0 for perfectly myopic low skill types and 1 for completely rational high skilled individuals.¹⁸

Myopic labour supply implies that retirement consumption does not enter the incentive compatibility constraint of a myopic mimicker. The social welfare function depends on *ex post* utilities given by $u(c^i) + \delta^g v(x^i) + \psi(1 - y^i)$. In this case individuals don't take into account the effect of higher earnings on future pension benefits in their labour supply decisions.

By assuming that all types are myopic, we get a two-type version of the continuum model analysed by Diamond (2003). In this case the basic justification for a pension system is to guarantee some level of resources in the retirement period. A more interesting case is to extend Diamonds' analysis into a more realistic case in which some people save and some don't. In other words we consider a case where myopia and ability are imperfectly correlated. We assume that $\delta^L = 0$, but unlike in Diamond (2003), we allow social planner and type 4 individuals to have different discount factors. Table 5 gives results when some individuals are myopic.

Myopic types perceive only the apparent utility $u(c^i) + \psi(1 - y^i)$. Thus there is now no first best case, where the government could by a subsidy on savings induce myopic types to save

¹⁸ Notice that myopic behaviour does not necessarily implicate that $\delta=0$. However, the extreme case with a zero discount factor can be interpreted as a consequence of myopia. In Diamond (2003) myopia is assumed to be with respect to labour supply decisions, but it is not modelled in detail. Our case is just one possibility of the effect of myopia.

voluntarily (first best distortion $-\frac{\delta^g - \delta^i}{\delta^i}$ approaches $-\infty$ when δ^i goes to zero). As a result, the interpretation basing on private distortion d cannot be used here. However, in the optimum, from government point of view, there is a subsidy also for myopic types.

Myopic types take only the apparent utility into account, also when mimicking. But rational type 4 perceives the changes in second period consumption that would occur if he mimicked myopic types, so the part reflecting the utility from second period remains in self-selection constraints.

		U	T'	α	x/ny	Consumption dispersion
CD	type 1	-1.4765	0.0915	-0.2377	64.46	means: $\bar{c} = 0.9195$ $\bar{x} = 0.8612$ Lorenz dominance: $L_c < L_x$ Gini coefficients: $G_c = 0.121$ $G_x = 0.035$
	type 2	-1.6344	0.0915	-0.2377	64.46	
	type 3	-1.0186	0	0.2565	41.89	
	type 4	-1.1790	0	0.0313	48.09	
CES	type 1	-3.9647	0.1416	-0.4007	62.86	means: $\bar{c} = 0.6794$ $\bar{x} = 0.7063$ Lorenz dominance: $L_c < L_x$ Gini coefficients: $G_c = 0.091$ $G_x = 0.027$
	type 2	-5.1397	0.1416	-0.4007	62.86	
	type 3	-3.3485	0	0.3385	42.90	
	type 4	-4.5326	0	0.0437	47.48	

TABLE 6: Utility levels at the optimum, marginal rates, replacement rates and Gini coefficients in paternalistic case with myopia. The binding self-selection constraints are **(3,1), (3,2), (4,1), (4,3)**

There are several points in our numerical results to emphasize. Tables 4, 5 and 6 imply that multidimensional problems are not likely to fulfil a simple pattern of binding self-selection constraints. However, interestingly in all cases considered in the text the upwards self-selection constraints turned out to be redundant. Marginal labour income tax rates satisfy the usual properties; zero marginal rates for both high-skilled types 3 and 4, $T'(n^3 y^3) = 0$ and $T'(n^4 y^4) = 0$. As we expected on the basis of previous numerical simulations in one-dimensional continuum model (see Tuomala, 1990) the marginal tax rates of labour income are higher with CES utility function than with CD utility function. The government with paternalist views and with myopic individual yields higher marginal labour income and savings tax rates

than in the case without myopia. Also in the latter case marginal savings tax and subsidy rates are smaller.

In the welfarist case the Lorenz curves cross, but for welfare ranking we can use Shorrocks' result for the generalized Lorenz curve. For paternalistic cases we can again apply Atkinson's result. The distribution of income when old is the dominant distribution that has the higher mean. The Gini coefficients for inequality in the first period consumption and retirement consumption in Tables 4-6 show that in the second period consumption is less dispersed than in the first period in paternalistic case. This seems plausible, as these governments try to correct the time preference for the second period consumption. In the welfarist case the values of Gini coefficients are practically speaking the same. It may be surprising that when there are myopic individuals in the economy, paternalist government policy increases saving and makes saving larger than paternalist government policy without myopic individuals (Table B2.3 and B2.5 in appendix B2).

In all cases considered, a type 3 – high skilled and low discount factor – attains the highest utility level. This may be interpreted so that there is in all cases a kind of bias towards present. In paternalistic cases our simulations show that in the optimum type 1 and type 2 are always pooled. This result can be used as a justification for a three type model we consider next¹⁹. Numerical results in the 4 type case also help us to choose particular pattern of self-selection constraints.

4. The 3 type case

Assume now that the low skill types, indexed now as type 1, have low discount factor δ^L . A part of the high skill types, group of type 3, has the same lower discount factor as all of the low skill types. The rest of high skill types, group 4, have the higher discount factor δ^H . Hence a type 1 = (n^1, δ^L) , a type 3 = (n^3, δ^L) , and a type 4 = (n^4, δ^H) , where n stands for the wage rate, such that $n^1 = n^L$ and $n^3 = n^4 = n^H$, $n^L < n^H$. Assume also that utility is given by (1), the same additively separable form as in two-type case.

¹⁹ Blomquist and Christiansen (2004) and Cuff (2000) apply a three-type version of a Mirrlees (1971) optimal income tax model with heterogeneity in preferences between leisure and consumption.

In the 3-type case there are also several possibilities for mimicking. The bindingness of the self-selection constraints depends now on how labour supply and the distributional preferences of the government hinge on the time preferences and skill level. Without any assumptions of the mimicking behaviour there are six possible self-selection constraints:

$u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \geq \hat{u}^{ij}(c^j) + \delta^i \hat{v}^{ij}(x^j) + \hat{\psi}^{ij} \left(1 - \frac{n^j y^j}{n^i}\right)$ for $i, j = 1, 3, 4$ and $i \neq j$. Thus a large number of different cases may emerge.

4.1 The welfarist social planner

The welfarist government maximises a sum of utilities, $\sum N^i [u(c^i) + \delta^i v(x^i) + \psi(1 - y^i)]$, subject to the revenue constraint and self-selection constraints $u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \geq \hat{u}^{ij}(c^j) + \delta^i \hat{v}^{ij}(x^j) + \hat{\psi}^{ij} \left(1 - \frac{n^j y^j}{n^i}\right)$ for $i, j = 1, 3, 4$ and $i \neq j$.

Analytically it is not possible to determine which of the self-selection constraints are binding. Based on our simulations in 4-type case we found that none of the upwards self-selection constraints are binding. Thus we concentrate on the downwards constraints.²⁰ The possibly binding constraints are

$$u(c^4) + \delta^H v(x^4) + \psi(1 - y^4) \geq \hat{u}^{43}(c^3) + \delta^H \hat{v}^{43}(x^3) + \hat{\psi}^{43}(1 - y^2) \quad (8a)$$

$$u(c^3) + \delta^L v(x^3) + \psi(1 - y^1) \geq \hat{u}^{31}(c^1) + \delta^L \hat{v}^{31}(x^1) + \hat{\psi}^{31} \left(1 - \frac{n^1 y^1}{n^3}\right) \quad (8b)$$

$$u(c^4) + \delta^H v(x^4) + \psi(1 - y^3) \geq \hat{u}^{41}(c^1) + \delta^H \hat{v}^{41}(x^1) + \hat{\psi}^{43} \left(1 - \frac{n^1 y^1}{n^4}\right) \quad (8c)$$

However, it turns out that only (8a) and (8b) are binding in the optimum implying that the economy ends up to separating equilibrium. The other possibility is bunching optimum, where at least two of the types are indistinguishable. Bunching at the bottom is not an interesting case here, as it leads back to traditional two-type case with only difference in productivity analysed in earlier literature. Bunching at the top, where high skilled types 3 and 4 are indistinguishable is also a possible equilibrium.

²⁰ Numerical simulation of the 3-type case verifies that this assumption indeed holds.

Using numerical result of the binding self-selection constraints we can also solve the optimization problem analytically. The welfarist social planner, who maximises the sum of individuals' utilities, the Lagrange function of the optimization problem is

$$\begin{aligned}
L = & \sum N^i \left[u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \right] + \lambda \left[\sum N^i (n^i y^i - c^i - rx^i) - R \right] \\
& + \mu^{43} \left[u(c^4) + \delta^H v(x^4) + \psi(1 - y^4) - \hat{u}^{43}(c^3) - \delta^H \hat{v}^{43}(x^3) - \hat{\psi}^{43}(1 - y^3) \right] \\
& + \mu^{31} \left[u(c^3) + \delta^L v(x^3) + \psi(1 - y^3) - \hat{u}^{31}(c^1) - \delta^L \hat{v}^{31}(x^1) - \hat{\psi}^{31} \left(1 - \frac{n^1 y^1}{n^3} \right) \right]
\end{aligned} \tag{9}$$

Multipliers λ and μ are attached to the budget and binding self-selection constraints.

Following two-type analysis, we can rewrite the first order conditions as

$$\left(\frac{u_c}{v_x} \right)^i = \frac{\delta^i}{r} [1 - d^i], i = 1, 3, 4, \text{ where discount factors are } \delta^1 = \delta^3 = \delta^L \text{ and } \delta^4 = \delta^H, \text{ and the}$$

distortions are given by α^i :

$$\begin{aligned}
d^1 &= 0 \\
d^3 &= \frac{\mu^{43}}{N^3 - \mu^{43} + \mu^{31}} \Delta^{HL} \\
d^4 &= 0
\end{aligned} \tag{10}$$

where $\Delta^{HL} = \frac{\delta^H - \delta^L}{\delta^L} > 0$. Equation (10) implies the following proposition

Proposition 2. *Savings decisions of type 1 and type 4 are not distorted, and hence not taxed at the margin. Type 3 faces a positive marginal tax rate on savings.*

Proposition 2 implies that even when the government respects consumers' time preferences, there is a distortion for type 3.

Following earlier parts, also here we solve the problem numerically to be able to say something more of the taxes and consumption in the optimum. The parameters used are given in Table 7;

they are otherwise the same as before, except that the size of the group of type 1 is one half of the whole population.

fraction of individuals in each group	$N^1 = 0.5, N^3 = 0.25, N^4 = 0.25$
discount factors	$\delta^L = 0.6, \delta^H = 0.8, r = 0.95$
productivities (wages)	$n^L = 2, n^H = 3$

TABLE 7: Parameter values

Numerical solution in Table 8 implies that savings of type 3 is taxed at the margin. It 8 also gives the replacement rates in terms of second period consumption relative to first period gross income. The replacement rates are be U-shaped; the replacement rate for type 3 is lower than those of types 1 and 4. The result holds for both types of utility functions.

The Gini coefficients for inequality in the first period consumption and retirement consumption in table 8 show that in the first period consumption is less dispersed than in the second period. This seems plausible, as welfarist government does not try to correct the time preference for the second period consumption. Allowing greater inequality in consumption level is a result of respecting the sovereignty of consumers to choose smaller consumption in the second period. In this case the distribution of the first period consumption is again the Lorenz dominant distribution that has the higher mean.

		U	T'	α	x/ny	Consumption dispersion
CD	type 1	-1.4763	0.0827	0	44.39	means: $\bar{c} = 0.9745$ $\bar{x} = 0.6807$ Lorenz dominance: $L_c > L_x$ Gini coefficients: $G_c = 0.081$ $G_x = 0.114$
	type 3	-1.1151	0	0.0259	36.21	
	type 4	-1.1911	0	0	43.14	
CES	type 1	-4.7301	0.1548	0	51.00	means: $\bar{c} = 0.7570$ $\bar{x} = 0.6292$ Lorenz dominance: $L_c > L_x$ Gini coefficients: $G_c = 0.061$ $G_x = 0.077$
	type 3	-4.2220	0	0.0452	41.26	
	type 4	-4.5323	0	0	44.97	

TABLE 8: Numerical solution in a welfarist case. Binding self-selection constraint are **(3,1)** and **(4,3)**.

4.2 Rawlsian case

Another possible solution to problem of adding utilities is to consider the Rawlsian social welfare function, where the government maximises the welfare of the worst-off group. In our 3 type case it is less controversial than in the 4 type case to determine the worst off individual. Social planner maximises now $N^1[u(c^1) + \delta^1 v(x^1) + \psi(1 - y^1)]$ subject to budget constraint and self-selection constraints $u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \geq \hat{u}^{ij}(c^j) + \delta^i \hat{v}^{ij}(x^j) + \hat{\psi}^{ij}\left(1 - \frac{n^j y^j}{n^i}\right)$ for $i, j = 1, 3, 4$ and $i \neq j$. The solution for numerical simulation is given in table 9. Table C2 in appendix C provides further information on the solution²¹. The results turn out to follow utilitarian case: only the savings decision of type 3 is distorted, whereas the savings of types 1 and 4 are left untaxed. Numerical solution shows that analytically unambiguous sign of the distortion α^3 is positive implying a tax at the margin. The Lorenz dominance holds for CD case. For CES case we have to apply Shorrocks' result..

		U	T'	α	x/ny	Consumption dispersion
CD	type 1	-1.4227	0.4144	0	50.97	means: $\bar{c} = 0.9016$ $\bar{x} = 0.6050$ Lorenz dominance: $L_c > L_x$ Gini coefficients: $G_c = 0.104$ $G_x = 0.141$
	type 3	-1.2237	0.0925	0.2796	31.01	
	type 4	-1.3381	0.0043	0	40.56	
CES	type 1	-4.6687	0.5074	0	57.44	means: $\bar{c} = 0.7282$ $\bar{x} = 0.5775$ Lorenz dominance: $GL_c > GL_x$ Gini coefficients: $G_c = 0.084$ $G_x = 0.084$
	type 3	-4.3546	0	0.2738	36.05	
	type 4	-4.7100	0	0	41.92	

TABLE 9: Numerical solution in a Rawlsian case. Binding self-selection constraint are **(3,1)** and **(4,3)**.

4.3 Government's and individuals' discount rates differ

Next we consider a case, where government has a discount factor δ^g and use the same separable life-time utility function given in (1). In the first best case, where the sole objective of the government is to correct the level of savings, there is a subsidy on savings equal to

²¹ Also analytical results are presented in Appendix C.

$d_{jb}^i = -\Delta^{gi}$. In the second best case with non-linear income tax the social planner maximises sum of $\sum N^i [u(c^i) + \delta^g v(x^i) + \psi(1 - y^i)]$ subject to subject to the revenue constraint and self-selection constraints $u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \geq \hat{u}^{ij}(c^j) + \delta^i \hat{v}^{ij}(x^j) + \hat{\psi}^{ij} \left(1 - \frac{n^j y^j}{n^i}\right)$ for $i, j = 1, 3, 4$ and $i \neq j$.

Numerical simulation shows that in the optimum only (8a) and (8b) are binding and the economy ends up in separating equilibrium. To consider the problem analytically, the Lagrange function is now given by

$$\begin{aligned} L = & \sum N^i [u(c^i) + \delta^g v(x^i) + \psi(1 - y^i)] + \lambda [\sum N^i (n^i y^i - c^i - r x^i) - R] \\ & + \mu^{43} [u(c^4) + \delta^H v(x^4) + \psi(1 - y^4) - \hat{u}^{43}(c^3) - \delta^H \hat{v}^{43}(x^3) - \hat{\psi}^{43}(1 - y^3)] \\ & + \mu^{31} \left[u(c^3) + \delta^L v(x^3) + \psi(1 - y^3) - \hat{u}^{31}(c^1) - \delta^L \hat{v}^{31}(x^1) - \hat{\psi}^{31} \left(1 - \frac{n^1 y^1}{n^3}\right) \right] \end{aligned} \quad (11)$$

The first order conditions with respect to c , x and y are given in Appendix A. We can express the first order conditions from government's perspective as $\left(\frac{u_c}{v_x}\right)^i = \frac{\delta^g}{r} [1 - \alpha^i]$, $i = 1, 3, 4$. From individuals point of view the distortion is still given by d^i in $\left(\frac{u_c}{v_x}\right)^i = \frac{\delta^g}{r} [1 - d^i]$, with a similar correspondence as in the two type case. The distortions government observes are given by

$$\begin{aligned} \alpha^1 &= (\varphi^1 - 1) \Delta^{Lg} \\ \alpha^3 &= \frac{\mu^{43}}{N^3 - \mu^{43} + \mu^{31}} \Delta^{Hg} - \frac{\mu^{31}}{N^3 - \mu^{43} + \mu^{31}} \Delta^{Lg} \\ \alpha^4 &= (\varphi^4 - 1) \Delta^{Hg} \end{aligned} \quad (12)$$

where $\varphi^1 = \frac{N^1}{N^1 - \mu^{31}}$ and $\varphi^4 = \frac{N^4}{N^4 + \mu^{43}}$.

Now there are terms resulting from both paternalist objectives (Δ^{Lg} , Δ^{Hg}) and from distributional considerations (terms including μ^{31} , μ^{43}). They cannot, however, be separated to isolate the effects of these two parts. Even when without paternalistic objectives types 1 and 3 were undistorted, the optimal distortions in case with paternalistic government the distortions

for these types depend on both effects. In (12) terms with Δ^{Lg} and Δ^{Hg} are negative as long as social planner has higher discount factor than types 3 and 4 with δ^H .

Proposition 3. *As long as $\delta^g > \delta^H (> \delta^L)$, for type 1 the marginal taxation of saving is negative (over-subsidized relative to the first best) and the marginal tax on savings for type 4 is positive (under-subsidized relative to the first best). For type 3 the sign of marginal rate is indeterminate.*

There are two distortions with opposite signs for type 3, so the overall effect on tax on savings is ambiguous. For numerical solution²², we assume that $\delta^g = 1$, otherwise the parameter values are as given in table 7. Numerical solution implies that the optimal savings tax rate for type 3 is positive, i.e. there is an implicit tax for savings (table 10). For type 1 there is an implicit subsidy and for type 4 a tax, as suggested also by the analytical results. The tax for type 3 seems to be systematically larger than that of type 4.

The replacement rate shows similar U-shape as in our welfarist case: type 3 has the lowest replacement rates. The dispersion in consumption in both periods is now reversed compared to the welfarist case: second period consumption is now more equally distributed than consumption in the first period. This implies that in a paternalist case the retirement or second period consumption is less dispersed than the first period consumption. However, in the welfarist case the result is the opposite. The Lorenz dominance holds also in our paternalistic case here.

It is clear that as long as $\delta^g > \delta^H (> \delta^L)$ and the paternalistic government values each types' second period consumption with a common discount factor the resulting consumption dispersion in the second period is more equal than with a government respecting consumers' own time preferences. This is in accordance with Diamond (2003), who analyses dependence between replacement rates (second period consumption relative to first period consumption) and risk aversion. He finds that assuming that the elderly are more risk averse than younger people the optimal lifetime redistribution tends to imply the retirement consumption should be less dispersed than the first period consumption. In our case the paternalist government internalizes the idea of risk of having low consumption in the retirement period.

²² For results of numerical simulation, see Appendix B2.

		U	T'	α	x/ny	Consumption dispersion
CD	type 1	-1.5885	0.1293	-0.1037	60.11	means: $\bar{c} = 0.8733$ $\bar{x} = 0.8957$ Lorenz dominance: $L_c < L_x$ Gini coefficients: $G_c = 0.105$ $G_x = 0.064$
	type 3	-1.1432	0	0.1035	46.92	
	type 4	-1.1545	0	0.0323	49.02	
CES	type 1	-4.8363	0.2237	-0.1763	59.58	means: $\bar{c} = 0.7186$ $\bar{x} = 0.7279$ Lorenz dominance: $L_c < L_x$ Gini coefficients: $G_c = 0.080$ $G_x = 0.048$
	type 3	-4.2316	0	0.1413	46.54	
	type 4	-4.4963	0	0.0449	47.99	

TABLE 10: Numerical solution in paternalistic case, binding self-selection constraints (3,1) and (4,3)

One possibility is, that governments' time preference and type 4's high time preference δ^H coincide, but from social planner's point of view types 1 and 3 use too low discount factor δ^L . In this case we have $\Delta^{Hg}=0$ and taxes on savings are given by distortions α^i in

$$\left(\frac{u_c}{v_x}\right)^i = \frac{\delta^g}{r} [1 - \alpha^i]:$$

$$\begin{aligned}\alpha^1 &= (\varphi^1 - 1) \Delta^{Lg} \\ \alpha^3 &= -\frac{\mu^{31}}{N^3 - \mu^{43} + \mu^{31}} \Delta^{Lg} \\ \alpha^4 &= 0\end{aligned}\tag{13}$$

Analytical consideration shows that there is no tax (or a subsidy) on the savings of type 4, a subsidy on savings of type 1 and a tax on type 3.²³

4.4 Myopic behaviour

Suppose again that myopic behaviour is illustrated by δ^L being zero, i.e. type 1 is myopic whereas types 3 and 4 are not. The social planner maximises sum of $\sum N^i [u(c^i) + \delta^g v(x^i) + \psi(1 - y^i)]$ subject to the revenue constraint and self-selection constraints $u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \geq \hat{u}^{ij}(c^j) + \delta^i \hat{v}^{ij}(x^j) + \hat{\psi}^{ij} \left(1 - \frac{n^j y^j}{n^i}\right)$ for $i, j=1, 3, 4$ and $i \neq j$. Notice now, that when mimicker is myopic, i.e. type 1 or 3, the part reflecting utility

²³ We do not report numerical simulations here, but they are available from authors.

from second period consumption disappears from both sides of the self-selection constraint. Numerical solution²⁴ suggests, that whereas any of the upwards self-selection constraint are not binding, all of the downwards constraints are.

The numerical simulation shows that in the equilibrium the second period consumption of the myopic types turns out to be at the same level, whereas the first period variables, labour supply and first period consumption are separate. The downwards self-selection constraints that are binding in the optimum are (4,3), (4,1), (3,1).

For analytical solution the Lagrange function for the problem can be written as

$$\begin{aligned}
L = & \sum N^i \left[u(c^i) + \delta^g v(x^i) + \psi(1 - y^i) \right] + \lambda \left[\sum N^i (n^i y^i - c^i - r x^i) - R \right] \\
& + \mu^{43} \left[u(c^4) + \delta^H v(x^4) + \psi(1 - y^4) - \hat{u}^{43}(c^3) - \delta^H \hat{v}^{43}(x^3) - \hat{\psi}^{43}(1 - y^3) \right] \\
& + \mu^{41} \left[u(c^4) + \delta^H v(x^4) + \psi(1 - y^4) - \hat{u}^{41}(c^1) - \delta^H \hat{v}^{41}(x^1) - \hat{\psi}^{41} \left(1 - \frac{n^L y^1}{n^H} \right) \right] \\
& + \mu^{31} \left[u(c^3) + \psi(1 - y^3) - \hat{u}^{32}(c^1) - \hat{\psi}^{31} \left(1 - \frac{n^L y^1}{n^H} \right) \right]
\end{aligned} \tag{14}$$

Writing the first order conditions in a short form as $\left(\frac{u_c}{v_x} \right)^i = \frac{\delta^g}{r} [1 - \alpha^i]$, $i = 1, 3, 4$ gives us the distortions

$$\begin{aligned}
\alpha^1 &= \frac{\mu^{41}}{N^1 - \mu^{41} - \mu^{31}} \Delta^{Hg} - \frac{\mu^{31}}{N^1 - \mu^{41} - \mu^{31}} \\
\alpha^3 &= \frac{\mu^{43}}{N^3 - \mu^{43} + \mu^{31}} \Delta^{Hg} + \frac{\mu^{31}}{N^3 - \mu^{43} + \mu^{31}} \\
\alpha^4 &= - \frac{\mu^{43} + \mu^{41}}{N^4 + \mu^{43} + \mu^{41}} \Delta^{Hg}
\end{aligned} \tag{15}$$

The analytical results follow the earlier case. For type 1 the marginal taxation of saving is negative, for type 4 positive, and indeterminate for type 3.

²⁴ See Appendix B3.

Numerical solution (table 11) also follows the same lines as paternalistic case presented earlier. The optimal savings tax rate for type 1 is negative, i.e. there is an implicit subsidy for savings. For types 3 and 4 we have a tax.

The U-shape of the replacement rates is reinforced in case with myopia. The dispersion in the first period consumption is greater than in the second period. As a matter of fact, the decrease in the Gini-coefficient from first to second period is steeper than in a case without myopia. In this case both Lorenz curves (L) and generalized Lorenz curves cross. Hence we can not make unambiguous welfare rankings between distributions.

		U	T'	α	x/ny	Consumption dispersion
CD	type 1	-1.4708	0.1477	-0.2377	64.53	means: $\bar{c} = 0.8786$ $\bar{x} = 0.8723$ Lorenz dominance: L&GL curves cross Gini coefficients: $G_c = 0.124$ $G_x = 0.031$
	type 3	-1.0169	0	0.2565	41.71	
	type 4	-1.1787	0	0.0313	48.35	
CES	type 1	-3.9647	0.2522	-0.4007	62.86	means: $\bar{c} = 0.7189$ $\bar{x} = 0.7148$ Lorenz dominance: L&GL curves cross Gini coefficients: $G_c = 0.092$ $G_x = 0.023$
	type 3	-3.3485	0	0.3385	42.90	
	type 4	-4.5236	0	0.0437	47.48	

TABLE 11: Numerical solution in paternalistic case with myopia, binding self-selection constraints (3,1), (4,1) and (4,3)

When some of the individuals are myopic, the change in the redistribution seems to be larger than with other cases considered. This holds also when a social planner is “partly welfarist”, i.e. sets its time preference equal to the time preference of the non-myopic group. It is also true in a 3-type model that when there are myopic individuals in the economy, paternalist government policy increases saving and makes saving larger than paternalist government policy without myopic individuals (Tables B3.3 and B3.5 in appendix B).

In sum, Tables 8 - 11 show that the replacement rates would be U-shaped; the replacement rate for type 3 is lower than those of types 1 and 4. The result holds for both types of utility functions and both in the cases of paternalistic and welfarist social planner. When some of the households have myopic preferences, the U-shaped pattern is emphasized: the difference in the replacement rates is bigger.

4.5 Variation of parameters

We have also checked how sensitive the results are for parameter values of discount factors, group sizes and productivity (wage) differences (see Appendix D in more detail). There are some combinations of the parameters that result in either non-solvable case or in the solution that is not appropriate for the model. However, these seem to be special cases and the failure of finding the optimum is due to technical problems in optimisation algorithm. In most cases the results found in the main text remain valid. We look at the following variations in parameters.

1. Time preferences, $\delta = (\delta^L, \delta^H)$, are allowed to get values between 0 and 1, using a step of 0.1 so that $\delta^L < \delta^H$.
2. Size of the groups $N^i = (N^1, N^2, N^3)$: to ensure the restrictions $N^i > 0$, with $\sum N^i = 1$, we allowed N^i to vary between 0.1 and 0.8, with a step of 0.1.
3. Productivity rates $n = (n^L, n^H)$ are allowed to get values between 1 and 10, with a step of one and so that $n^L < n^H$.

When the difference in time preferences is increased from (0.6, 0.8) to (0.5, 0.9), the qualitative results continue to hold. Changes in marginal tax rates for labour income and replacement rates are very small. Marginal tax rates for income from savings for types with lower discount factor are larger.

Varying the size of the groups so that the each type group i is in turn the largest group in the economy, with $N^i = 0.7$ for each $i=1,3,4$, the qualitative nature of the results does not change. When the relative number of type 1 individuals is the largest one, their marginal tax rates fall. When the largest group is the individuals of type 3 or 4 the marginal tax rates for type 1 increases. The replacement rates for types 3 and 4 are very close to each other, but the U-shape remains valid.

Increasing wage inequality marginal tax rates for both labour income tax and tax on savings increase, which is in accordance to findings in the atemporal continuous Mirrlees model (see Kanbur and Tuomala, 1994). The U-shape of replacement rate is emphasized with higher differences in wage rates.

Our sensitivity analysis shows that the same qualitative results remain with different parameterization of discount factors, group sizes and productivity rates. Interestingly the U-shaped pattern of the replacement rates does not change.

5. Conclusions

In this paper we have examined various aspects of the optimal lifetime redistribution policy within a cohort. We characterise optimal redistribution policy when society consists of individuals who do not differ only in productivity, but also in time preference or myopia. We extend Diamond's (2003) analysis on nonlinear taxation of savings into the three and four-type models. Our results provide a rationale for distortions (upward and downward) in savings behaviour in a simple two period model where high skilled and low skilled individuals have different non-observable time preferences beyond their earning capacity. If we interpret our model so that there is no private savings, but public provision of pension in period 2, then in different versions of three type model (and with different parameterizations) we find the U-shaped pattern of the replacement rates. Our numerical results suggest that the retirement consumption is less dispersed than the first period consumption in paternalist case, whereas in a welfarist case the ordering is reversed. They also show that consumption when old should be less dispersed than consumption when young when some individuals are myopic. Moreover, when there are myopic individuals in the economy, a paternalistic government policy increases saving and makes saving larger than with paternalist government policy where there are no myopic individuals. When insufficient saving is caused by myopia or low discount factor, our analytical and numerical results support the view that there is a case for a non-linear public pension program (2.best redistribution) in a world where individuals differ in skills and discount factor or myopia

References

- Atkinson, A. B. (1970), "On the Measurement of Inequality", *Journal of Economic Theory* 2, 244-263.
- Atkinson, A. B. and J. E. Stiglitz (1976), "The Design of Tax Structure: direct versus Indirect Taxation". *Journal of Public Economics* 1, 55-75.
- Berliant, M. and J. O. Ledyard (2005), "Optimal dynamic nonlinear income taxation with no commitment", mimeo.
- Blomquist, S. and V. Christiansen (2004), "Taxation and Heterogeneous Preferences", mimeo.
- Cremer, H., J.-M. Lozachmeur and P. Pestieau (2004), "Social Security, retirement age and optimal income taxation". *Journal of Public Economics* 88, 2259-2281.
- Cremer, H., Ph. De Donder, D. Maldonado and P. Pestieau (2006), "Forced saving, redistribution and non linear social security scheme", mimeo.
- Cuff, K. (2000), "Optimality of Workfare with Heterogeneous Preferences", *Canadian Journal of Economics* 33, 149-174.
- Diamond, P. A. (1977), "A Framework for Social Security Analysis". *Journal of Public Economics* 8, 275-298.
- Diamond, P. A. (2003), "Taxation, Incomplete Markets, and Social Security. The 2000 Munich Lectures". MIT Press.
- Diamond, P. A. and B. Köszegi (2003), "Quasi-hyperbolic Discounting and Retirement". *Journal of Public Economics* 87, 1839-1872.
- Diamond, P. A. and J. Mirrlees (2000), "Adjusting One's Standard of Living: Two Period Models" in P. J. Hammond, and G. D. Myles (eds.) "Incentives, Organization and Public Economics, Papers in Honour of Sir James Mirrlees". 107-122. Oxford University Press.
- Feldstein, M. (1985), "The Optimal Level of Social Security Benefits", *Quarterly Journal of Economics* 100(2), 300-320.

- Gaube, T. (2005). "Optimum Taxation of Each Year's Income". Paper presented in 62nd Congress of International Institute of Public Finance, Paphos, August 2006.
- İmrohoroglu, A., S. İmrohoroglu and D. H. Joines (2003), "Time Inconsistent Preferences and Social Security". *Quarterly Journal of Economics* 118(2), 745-784.
- Kanbur, R. and Tuomala, M. (1994), "Inherent inequality and the optimal graduation of marginal tax rates". *Scandinavian Journal of Economics* 96 (2), pp. 275-282.
- Malinvaud, E. (1972), "Lectures on Microeconomic Theory". North Holland
- Mirrlees, J. (1971), "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic studies* 38, 175-208.
- Ramsey, F. P. (1928), "A Mathematical Theory of Saving", *The Economic Journal* 38, 543-559
- Rawls, J. (1971), "A Theory of Justice", Oxford, Oxford University Press.
- Roberts, K. (1984), "The Theoretical Limits to Redistribution", *Review of Economic Studies* 51, 177-195
- Saez, E. (2002), "The Desirability of Commodity Taxation Under Non-Linear Income Taxation and Heterogeneous Tastes". *Journal of Public Economics* 83, 217-230.
- Sandmo, A. (1993), "Optimal redistribution when Tastes differ", *Finanz Archiv* 50(2), 149-163.
- Sen, A.K. (1973), "On Economic Inequality", Clarendon Press, Oxford.
- Shorrocks, A. F. (1980), "The Class of Additively Decomposable Inequality Measures", *Econometrica* 48, 613-626.
- Stiglitz, Joseph E. (1982), "Self-Selection and Pareto Efficient Taxation". *Journal of Public Economics* 17, 213-240.
- Stern, N. H. (1982), "Optimum Taxation with Errors in Administration". *Journal of Public Economics* 17, 181-211.
- Tarkiainen, R and M. Tuomala (1999), "Optimal Nonlinear Income Taxation with a Two-Dimensional Population; A Computational Approach", *Computational Economics* 13, 1-16.

- Tarkiainen, R and M. Tuomala, (2007), “On optimal income taxation with heterogeneous work preferences”, forthcoming in *International Journal of Economic Theory*.
- Tuomala, M. (1990), “Optimal income tax and redistribution”, Clarendon Press, Oxford.
- Vickrey, W. (1939) ,“Averaging of Income for Income Tax purposes”, *Journal of Political Economy* 47, 379-397.

APPENDIX A: First order conditions with respect to c^i , x^i and y^i

for two-type model, welfarist case:

$$N^L u_c^L - \lambda N^L - \mu \hat{u}_c = 0 \quad (A1)$$

$$N^L \delta^L v_x^L - \lambda r N^L - \mu \delta^H \hat{v}_x = 0 \quad (A2)$$

$$-N^L \psi'^L + \lambda N^L n^L + \mu \hat{\psi}' \frac{n^L}{n^H} = 0 \quad (A3)$$

$$N^H u_c^H - \lambda N^H + \mu \hat{u}_c^H = 0 \quad (A4)$$

$$N^H \delta^H v_x^H - \lambda r N^H + \mu \delta^H v_x^H = 0 \quad (A5)$$

$$-N^H \psi'^H + \lambda N^H n^H - \mu \psi'^H = 0 \quad (A6)$$

for two-type case, paternalistic case (A1), (A3), (A4) and (A6) are the same, but instead of (A2) and (A5) we have

$$N^L \delta^g v_x^L - \lambda r N^L - \mu \delta^H \hat{v}_x = 0 \quad (A7)$$

$$N^H \delta^g v_x^H - \lambda r N^H + \mu \delta^H v_x^H = 0 \quad (A8)$$

for three-type model in a welfarist case

$$N^1 u_c^1 - \lambda N^1 - \mu^{31} \hat{u}_c^{31} = 0 \quad (A9)$$

$$N^1 \delta^L v_x^1 - \lambda r N^1 - \mu^{31} \delta^L \hat{v}_x^{31} = 0 \quad (A10)$$

$$N^1 \psi' - \lambda N^1 n^1 - \mu^{31} \hat{\psi}'^{31} \frac{n^L}{n^H} = 0 \quad (A11)$$

$$N^3 u_c^3 - \lambda N^3 - \mu^{43} \hat{u}_c^{43} + \mu^{31} u_c^3 = 0 \quad (A12)$$

$$N^3 \delta^L v_x^3 - \lambda r N^3 - \mu^{43} \delta^H \hat{v}_x^{43} + \mu^{31} \delta^L v_x^3 = 0 \quad (A13)$$

$$N^3 \psi' - \lambda N^3 n^H - \mu^{43} \hat{\psi}'^{43} + \mu^{31} \psi' = 0 \quad (A14)$$

$$N^4 u_c^4 - \lambda N^4 + \mu^{43} u_c^4 = 0 \quad (A15)$$

$$N^4 \delta^H v_x^4 - \lambda r N^4 + \mu^{43} \delta^H v_x^4 = 0 \quad (A16)$$

$$N^4 \psi' - \lambda N^4 n^H + \mu^{43} \psi' = 0 \quad (A17)$$

for a three-type model with paternalist social planner most of the first order conditions remain the same, only those with respect to future consumption x change:

$$N^1 \delta^g v_x^1 - \lambda r N^1 - \mu^{31} \delta^L \hat{v}_x^{31} = 0 \quad (\text{A18})$$

$$N^3 \delta^g v_x^3 - \lambda r N^3 - \mu^{43} \delta^H \hat{v}_x^{43} + \mu^{31} \delta^L v_x^3 = 0 \quad (\text{A19})$$

$$N^4 \delta^g v_x^4 - \lambda r N^4 + \mu^{43} \delta^H v_x^4 = 0 \quad (\text{A20})$$

for a three-type model with myopia

$$N^1 u_c^1 - \lambda N^1 - \mu^{41} \hat{u}_c^{41} - \mu^{31} \hat{u}_c^{31} = 0 \quad (\text{A21})$$

$$N^1 \delta^g v_x^1 - \lambda r N^1 - \mu^{41} \delta^H \hat{u}_c^{41} = 0 \quad (\text{A22})$$

$$N^1 \psi' - \lambda N^1 n^1 - \mu^{41} \hat{\psi}'^{41} \frac{n^L}{n^H} - \mu^{31} \hat{\psi}'^{31} \frac{n^L}{n^H} = 0 \quad (\text{A23})$$

$$N^3 u_c^3 - \lambda N^3 - \mu^{43} \hat{u}_c^{43} + \mu^{31} u_c^3 = 0 \quad (\text{A24})$$

$$N^3 \delta^g v_x^3 - \lambda r N^3 - \mu^{43} \delta^H \hat{v}_x^{43} = 0 \quad (\text{A25})$$

$$N^3 \psi' - \lambda N^3 n^H - \mu^{43} \hat{\psi}'^{43} + \mu^{31} \psi' = 0 \quad (\text{A26})$$

$$N^4 u_c^4 - \lambda N^4 + \mu^{43} u_c^4 + \mu^{41} u_c^4 = 0 \quad (\text{A27})$$

$$N^4 \delta^g v_x^4 - \lambda r N^4 + \mu^{43} \delta^H v_x^4 + \mu^{41} \delta^H v_x^4 = 0 \quad (\text{A28})$$

$$N^4 \psi' - \lambda N^4 n^H + \mu^{43} \psi' + \mu^{41} \psi' = 0 \quad (\text{A29})$$

APPENDIX B: Numerical simulations

B1: 2-type cases

parameterization:

$$n^L = 2, n^H = 3, N^i = 0.5, i=L,H,$$

$$\delta^L = 0.6, \delta^H = 0.8, r = 0.95$$

welfarist case

		c	x	y
CD	type L	0.8256	0.5022	0.5629
	type H	1.0082	0.8490	0.6639
CES	type L	0.6667	0.5118	0.4997
	type H	0.7899	0.7248	0.5440

TABLE B1.1: Equilibrium consumption and labour supply

	λ	μ^{LH}	μ^{HL}
CD	1.091	0 (-4.1493)	0.050
CES	1.872	0 (-2.8564)	0.084

TABLE B1.2: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^j - U^i$) in the optimum is given in parenthesis.

paternalistic case

		c	x	y
CD	type L	0.6716	0.7383	0.6221
	type H	0.9668	0.9870	0.6777
CES	type L	0.5911	0.6287	0.5340
	type H	0.7813	0.7843	0.5489

TABLE B1.3: Equilibrium consumption and labour supply

	λ	μ^{LH}	μ^{HL}
CD	1.221	0 (*)	0.090
CES	2.084	0 (-2.9152)	0.136

TABLE B1.4: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.

We denote by (*) the case where utility doesn't get a real value. This is due to the fact that low skilled type should supply labour more than the total time endowment is. This is, of course, not in the interest of any consumer. Thus, even if the difference cannot be determined accurately, the self-selection constraint in question is not binding.

Rawlsian case²⁵

		c	x	y
CD	type L	0.8169	0.3195	0.3780
	type H	0.9437	0.7897	0.6861
CES	type L	0.6720	0.4078	0.3842
	type H	0.7517	0.6898	0.5660

TABLE B1.5. Equilibrium consumption, labour supply, utility level at the optimum, replacement rates and Gini-coefficients

	λ	μ^{LH}	μ^{HL}
CD	0.566	0 (*)	0.267
CES	0.984	0 (-4.2401)	0.278

TABLE B1.6: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.

²⁵ Optimisation with Cobb-Douglas utility function fails when size of the both groups is exactly 0.5. Thus we have solved problem in Cobb-Douglas case with closest possible combination, where the size of group 1 is 0.4986. In the case of CES utility function there occurs no problem due to size of the groups.

B2: 4-type cases

parameterization:

$$n^L = 2, n^H = 3, N^i = 0.25, i=1,2,3,4,$$

$$\delta^L = 0.6, \delta^H = 0.8, r = 0.95$$

welfarist case

	CD			CES		
	c	x	y	c	x	y
type 1	0.7998	0.5051	0.5704	0.6483	0.5152	0.5063
type 2	0.7092	0.6396	0.5902	0.6018	0.5828	0.5167
type 3	1.1146	0.6814	0.6285	0.8318	0.6438	0.5198
type 4	1.0172	0.8566	0.6609	0.7960	0.7305	0.5404

TABLE B2.1: Equilibrium consumption and labour supply

	λ	μ^{12}	μ^{13}	μ^{14}	μ^{21}	μ^{23}	μ^{24}
CD	1.10	0 (-0.0256)	0 (-1.5027)	0 (-3.3553)	0 (-0.0216)	0 (-1.4644)	0 (-3.2713)
CES	1.90	0 (-0.0275)	0 (-1.9404)	0 (-2.6254)	0 (-0.0175)	0 (-1.8804)	0 (-2.5286)
		μ^{31}	μ^{32}	μ^{34}	μ^{41}	μ^{42}	μ^{43}
CD		0.03	0.06	0 (-0.0457)	0 (-0.0599)	0 (-0.0127)	0.03
CES		0.05	0.08	0 (-0.0369)	0 (-0.0775)	0 (-0.0325)	0.05

TABLE B2.2: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.

Paternalistic case:

	CD			CES		
	c	x	y	c	x	y
type 1	0.6532	0.7590	0.6317	0.5759	0.6408	0.5378
type 2	0.6532	0.7590	0.6317	0.5759	0.6408	0.5378
type 3	1.0034	0.9392	0.6698	0.7948	0.7556	0.5411
type 4	0.9850	0.9883	0.6773	0.7851	0.7872	0.5467

TABLE B2.3: Equilibrium consumption and labour supply

	λ	μ^{12}	μ^{13}	μ^{14}	μ^{21}	μ^{23}	μ^{24}
CD	1.22	0 (0)	0 (*)	0 (*)	0 (0)	0 (*)	0 (*)
CES	2.09	0 (0)	0 (-2.5271)	0 (-2.7585)	0 (0)	0 (-2.4797)	0 (-2.7005)
		μ^{31}	μ^{32}	μ^{34}	μ^{41}	μ^{42}	μ^{43}
		0.05	0.05	0 (-0.0111)	0 (-0.0375)	0 (-0.0375)	0.05
		0.08	0.08	0 (-0.0106)	0 (-0.0474)	0 (-0.474)	0.07

TABLE B2.4: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.

Paternalistic case with myopia ($\delta^L=0$):

	CD			CES		
	c	x	y	c	x	y
type 1	0.6295	0.8209	0.6371	0.5607	0.6808	0.5415
type 2	0.6295	0.8209	0.6371	0.5607	0.6808	0.5415
type 3	1.0414	0.8209	0.6533	0.8159	0.6808	0.5289
type 4	0.9776	0.9820	0.6808	0.7802	0.7828	0.5495

TABLE B2.5: Equilibrium consumption and labour supply

	λ	μ^{12}	μ^{13}	μ^{14}	μ^{21}	μ^{23}	μ^{24}
CD	1.227	0 (0)	0 (-2.3893)	0 (*)	0 (0)	0 (-2.3893)	0 (*)
CES	2.102	0 (0)	0 (-2.1014)	0 (-3.0086)	0 (0)	0 (-2.1014)	0 (-2.8555)
		μ^{31}	μ^{32}	μ^{34}	μ^{41}	μ^{42}	μ^{43}
		0.043	0.043	0 (-0.1459)	0.016	0.016	0.016
		0.062	0.062	0 (-0.1531)	0.023	0.023	0.023

TABLE B2.6: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.

B3: 3-type cases**parameterization:**

$n^L = 2, n^H = 3, N^1 = 0.5, N^3 = N^4 = 0.25$ $\delta^L = 0.6, \delta^H = 0.8, r = 0.95$

Welfarist case

		c	x	y
CD	type 1	0.7958	0.5026	0.5662
	type 3	1.1116	0.6839	0.6295
	type 4	1.0162	0.8558	0.6613
CES	type 1	0.6459	0.5133	0.5032
	type 3	0.8301	0.6446	0.5207
	type 4	0.7952	0.7297	0.5409

TABLE B3.1: Equilibrium consumption and labour supply

	λ	μ^{13}	μ^{14}	μ^{31}	μ^{34}	μ^{41}	μ^{43}
CD	1.0754	0 (-1.5323)	0 (-3.4165)	0.0721	0 (-0.0448)	0 (-0.0616)	0.0232
CES	1.8556	0 (-1.9739)	0 (-2.6514)	0.1130	0 (-0.0362)	0 (-0.0794)	0.0433

TABLE B3.2: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.**Paternalistic case**

		c	x	y
CD	type 1	0.6491	0.7541	0.6273
	type 3	0.9962	0.9402	0.6679
	type 4	0.9747	0.9929	0.6751
CES	type 1	0.5759	0.6408	0.5378
	type 3	0.7948	0.7556	0.5411
	type 4	0.7851	0.7872	0.5467

TABLE B3.3: Equilibrium consumption and labour supply

	λ	μ^{13}	μ^{14}	μ^{31}	μ^{34}	μ^{41}	μ^{43}
CD	1.2236	0 (*)	0 (*)	0.1029	0 (-0.0109)	0 (-0.0441)	0.0482
CES	2.0928	0 (-2.5271)	0 (-2.7585)	0.1529	0 (-0.0106)	0 (-0.0474)	0.0725

TABLE B3.4: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.

Paternalistic case with myopia

		c	x	y
CD	type 1	0.6258	0.8168	0.6329
	type 3	1.0419	0.8168	0.6528
	type 4	0.9682	0.9834	0.6779
CES	type 1	0.5607	0.6808	0.5415
	type 3	0.8159	0.6808	0.5289
	type 4	0.7802	0.7828	0.5495

TABLE B3.5: Equilibrium consumption and labour supply

	λ	μ^{13}	μ^{14}	μ^{31}	μ^{34}	μ^{41}	μ^{43}
CD	1.2267	0 (-2.3626)	0 (*)	0.0851	0 (-0.1484)	0.0309	0.0155
CES	2.1017	0 (-2.1014)	0 (-3.0086)	0.1231	0 (-0.1531)	0.0466	0.0233

TABLE B3.6: Lagrange multipliers for each constraint, bolded are binding in the optimum. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.

APPENDIX C: Case with the Rawlsian social welfare function

Social planner maximises now $N^1[u(c^1) + \delta^1 v(x^1) + \psi(1 - y^1)]$ subject to budget constraint and self-selection constraints $u(c^i) + \delta^i v(x^i) + \psi(1 - y^i) \geq \hat{u}^{ij}(c^j) + \delta^i \hat{v}^{ij}(x^j) + \hat{\psi}^{ij}\left(1 - \frac{n^j y^j}{n^i}\right)$ for $i, j = 1, 3, 4$ and $i \neq j$. The solution for numerical simulation²⁶ is given in table C.1. Table C.2 provides information of the binding constraints and the Lagrange multipliers attached to each constraint.

		c	x	y
CD	type 1	0.6898	0.4190	0.4110
	type 3	1.0604	0.5680	0.6105
	type 4	0.9457	0.8280	0.6804
CES	type 1	0.5885	0.4677	0.4071
	type 3	0.8309	0.5627	0.5203
	type 4	0.76651	0.7021	0.5582

TABLE C.1: Equilibrium consumption and labour supply

	λ	μ^{13}	μ^{14}	μ^{31}	μ^{34}	μ^{41}	μ^{43}
CD	0.5760	0 (-1.3324)	0 (*)	0.3037	0 (-0.0767)	0 (-0.0493)	0.1385
CES	1.0152	0 (-2.1562)	0 (-3.6415)	0.3242	0 (-0.0706)	0 (-0.0722)	0.1462

TABLE C.2: Lagrange multipliers. For the nonbinding constraints the value of the constraint ($U^{ij} - U^i$) in the optimum is given in parenthesis.

Using the information of the binding self-selection constraints provided by numerical solution, the Lagrange function can be written as

²⁶ Due to solvability problem the results with CES utility functions are given with parametric values $N^1=0.5$, $N^2=0.254$ and $N^3=0.246$.

$$\begin{aligned}
L = & \sum N^1 \left[u(c^1) + \delta^i v(x^1) + \psi(1 - y^1) \right] + \lambda \left[\sum N^i (n^i y^i - c^i - r x^i) - R \right] \\
& + \mu^{43} \left[u(c^4) + \delta^H v(x^4) + \psi(1 - y^3) - \hat{u}^{43}(c^3) - \delta^H \hat{v}^{43}(x^3) - \hat{\psi}^{43}(1 - y^3) \right] \\
& + \mu^{31} \left[u(c^3) + \delta^L v(x^3) + \psi(1 - y^3) - \hat{u}^{31}(c^1) - \delta^L \hat{v}^{31}(x^1) - \hat{\psi}^{31} \left(1 - \frac{n^1 y^1}{n^2} \right) \right]
\end{aligned} \tag{C1}$$

The first order condition with respect to c^i , x^i and y^i , $i=1,3,4$ are given by

$$N^1 u_c^1 - \lambda N^1 - \mu^{31} \hat{u}_c^{31} = 0 \tag{C2}$$

$$N^1 \delta^L v_x^1 - \lambda r N^1 - \mu^{31} \delta^L \hat{v}_x^{31} = 0 \tag{C3}$$

$$N^1 \psi' - \lambda N^1 n^1 - \mu^{31} \hat{\psi}'^{31} \frac{n^1}{n^2} = 0 \tag{C4}$$

$$-\lambda N^3 - \mu^{43} \hat{u}_c^{43} + \mu^{31} u_c^3 = 0 \tag{C5}$$

$$-\lambda r N^3 - \mu^{43} \delta^H \hat{v}_x^{43} + \mu^{31} \delta^L v_x^3 = 0 \tag{C6}$$

$$-\lambda N^3 n^2 - \mu^{43} \hat{\psi}'^{43} + \mu^{31} \psi' = 0 \tag{C7}$$

$$-\lambda N^4 + \mu^{43} u_c^4 = 0 \tag{C8}$$

$$-\lambda r N^4 + \mu^{43} \delta^H v_x^4 = 0 \tag{C9}$$

$$-\lambda N^4 n^2 + \mu^{43} \psi' = 0 \tag{C10}$$

To solve the distortions for savings rewrite the first order condition in form

$$\left(\frac{u_c}{v_x} \right)^i = \frac{\delta^g}{r} [1 - \alpha^i], i = 1, 3, 4, \text{ where the distortions are given by}$$

$$\begin{aligned}
\alpha^1 &= 0 \\
\alpha^2 &= - \frac{\mu^{43}}{-\mu^{43} + \mu^{31}} \Delta^{HL} \\
\alpha^3 &= 0
\end{aligned} \tag{C11}$$

APPENDIX D: Sensitivity analysis

We now check how sensitive the results are for parameter values of discount factors, group sizes and productivity rates.

Variation of parameters

- Time preferences, $\delta = (\delta^L, \delta^H)$, are allowed to get values between 0 and 1, using a step of 0.1 so that $\delta^L < \delta^H$.
- Size of the groups $N^i = (N^1, N^2, N^3)$: to ensure the restrictions $N^i > 0$, with $\sum_i N^i = 1$, we allowed N^i to vary between 0.1 and 0.8, with a step of 0.1.
- Productivity rates $n = (n^L, n^H)$ are allowed to get values between 1 and 10, with a step of one and so that $n^L < n^H$.

Solvability of the problem

When using Cobb-Douglas utility function there seems to be some combinations of (δ^L, δ^H) ²⁷ where the optimization algorithm fails to solve the problem at all. However, solution is found with δ^L and δ^H close to these points (± 0.02). The problem referred by the program relates to definiteness of Hessian matrix in search direction. Upwards self-selection constraints do not become binding at any solvable combination of δ^L and δ^H . Ignoring the upwards self-selection constraints removes this problem. With CES utility functions this problem does not occur. Thus we believe this does not pose a significant threat to the generality of our results.

The same way as with time preferences, there were a few combination of N^i with which the problem was not solvable²⁸ with Cobb-Douglas utility. The problem referred by the program relates again to definiteness of Hessian matrix in search direction. However, changing the parameters a little, by ± 0.03 , made the problem solvable again. Also when upwards self-selection constraints were excluded the solution was found for all compositions of the economy.

Changing productivity rates from $n^L, n^H = (1, 10)$, with a step of one and so that $n^L < n^H$ produces some combinations of productivities that lead to a non-solvable case. With the Cobb-Douglas utility function these are pairs (3,7), (4,9) and (5,8). With the CES utility function there are more

²⁷ These combinations are (0.3, 0.7), (0.4, 0.5), (0.5, 0.7) and (0.8, 0.9).

²⁸ These combinations (N^1, N^2, N^3) are (0.1, 0.1, 0.8), (0.1, 0.5, 0.4), (0.2, 0.6, 0.2), (0.4, 0.5, 0.1) and (0.6, 0.3, 0.1).

pairs when the optimisation fails. This happens when the difference between wage rates is big²⁹. In addition to these, the program fails to solve the problem also when $n^H=2*n^L$ with both utility functions.

In welfarist case there are fewer cases when the program totally fails to solve the problem. Variation of the composition of the population, N , does not result in any non-solvable cases. Variation in discount factor δ , with the Cobb-Douglas utility produces two combination when the solution is not solved: $\delta = (0.5, 1)$ and $(0.8, 1)$. As in paternalistic case, also here the solution is not found when $n^H=2*n^L$. There are also some cases when the optimum program finds is not appropriate, as the condition for non-negativity of labour supply becomes binding. This occurs mainly while the wage rates are varied: $n^L=1$ and $n^H \geq 3$, $n^L=2$ and $n^H \geq 5$, $n^L=3$ and $n^H \geq 7$.

Type of equilibrium and the bindingness of the self-selection constraints

There are also some compositions of the economy, where when using Cobb-Douglas utility function also upwards self-selection constraint become binding. Table D1 presents the combinations, the binding self-selection constraints and a possible explanation for these cases. However, as in most cases with the Cobb-Douglas utility function and in all cases with the CES utility function the upwards self-selection constraints are not binding, we will concentrate on the case where only downwards self-selection constraints are taken into account.

N^1	N^2	N^3	binding constraints ij , $i = \text{mimicker}, j =$ mimicked	explanation
0.1	0.3	0.6	31, 23	?
0.2	0.2	0.6	31, 23	?
0.2	0.3	0.5	23	Lagrange multiplier $\mu^{23} \approx 0$, all types choose the same amount of consumption on both periods.
0.2	0.4	0.4	32, 23	Lagrange multipliers $\mu^{32} \approx 0$ and $\mu^{23} \approx 0$, all

²⁹ Optimisation fails or gives spurious results when $n^L=1$, $n^L=2$ and $n^H \geq 5$, $n^L=3$ and $n^H \geq 5$, $n^L=4$ and $n^H \geq 6$ and $n^L=5$ and $n^H=8$.

				types choose the same amount of consumption on both periods.
0.2	0.5	0.3	32, 23	Lagrange multipliers $\mu^{32} \approx 0$ and $\mu^{23} \approx 0$, all types choose the same amount of consumption on both periods.

TABLE D1: Cases, when not all upwards self-selection constraints are redundant, Cobb-Douglas utility function.

In welfarist case upwards binding self-selection constraints are usually redundant. There are, however, some exceptions. They occur when we vary

- wage rate n : With Cobb-Douglas utility function when wage rates are (7,8), (8,9) or (9,10) the binding self-selection constraints are 21, 12 and 23. With CES utility function simulations with wage combinations (1,3), (1,4) and (3,5) do not work properly (labour supply of type 1 aims to be negative, which have to be prohibited artificially, so the optimum the programme finds is not a good one).
- discount factor δ : None of the upwards self-selection constraints are binding either with Cobb-Douglas or CES utility function.
- composition of the economy, N : In the case of CES utility with $N = (0.3, 0.4, 0.3)$ self-selection constraints 12 and 23 are binding. With very low fraction of type 1, ($N^1 = 0.1$ with Cobb-Douglas utility function and $N^1 = 0.1-0.2$ with CES utility function) self-selection constraint 23 becomes binding.

When all three downwards self-selection constraints are binding, the partial bunching equilibrium occurs. Bunching takes place in the second period, whereas in the first period each group has separate bundle. In other cases, only two of the downwards self-selection constraints are binding and the economy ends up in separating equilibrium. Partial bunching equilibrium occurs when

- variation of N does not affect the type of equilibrium either with Cobb-Douglas or CES utility function
- δ^L is sufficiently low and the difference between time preferences is sufficiently large. With the CES utility function this occurs when $\delta^L=0.1$ and

$\delta^H \in (0.5, 1)$, whereas with Cobb-Douglas there are more cases, presented in Figure D1.

FIGURE D1: Pairs of time preference factors when all downwards self-selection constraints are binding and partial bunching equilibrium occurs, the Cobb-Douglas utility function

- With the CES utility function bunching at the second period occurs with following wage rates: (1,2), (2,5-9), (3,5-6), (4,6), (4,8) and (5,10). Also with the Cobb-Douglas utility function wage rates seem to affect; partial bunching seem to be more common when wage difference is small and when wage rates are high (figure D2).

FIGURE D2: Pairs of productivity rates when all downwards self-selection constraints are binding and partial bunching equilibrium occurs.

Marginal implicit tax rate on savings

With both Cobb-Douglas and CES utility functions, paternalistic government and only downwards self-selection constraints binding, with all possible combinations of $\delta^L, \delta^H \in (0,1)$ with a step of 0.1 and $\delta^L < \delta^H$, $N^i \in (0.1, 0.8)$ so that $\sum N^i = 1$ and $n^L, n^H \in (1, 10)$ so that $n^L < n^H$ with a step of 1 the signs of α :s are the same as in the case presented in the paper: type 1 faces a positive marginal distortion and types 2 and 3 face a negative marginal distortion.

Varying each parameter at the time and allowing all self-selection constraints to bind, in all solvable combinations of $\delta^L, \delta^H \in (0,1)$ with both Cobb-Douglas and CES utility functions the signs of α :s remain as before; a marginal tax for type 1 and a marginal subsidy for type 2 and 3, except in cases where $\delta^H=1$ when α^3 is zero. In case of Cobb-Douglas utility function when going through all solvable combinations of N^i we found 3 special cases, where $\alpha^i=0$ for all types. These

special cases happen with combinations (0.1, 0.3, 0.6), (0.2, 0.2, 0.6) and (0.2, 0.3, 0.5). In the two first cases binding self-selection constraints turn out to be 31 and 23, whereas in the last case all types choose exactly the same amount of labour income and consumption in both periods. These are clearly only special cases.

Finally, variation of productivity rates shows that also here the results are rather robust. In most solvable cases signs of α :s remain the same. There are some exceptions, where all α :s are either zero or very close to it. This happens with Cobb-Douglas utility function with productivity pairs (1,3), (2,5) and some other pairs where the difference is very large³⁰.

In welfarist case there are some parameter values, where the distortions α are zero or almost zero. These cases occur with Cobb-Douglas utility function when composition of the economy N is (0.1, 0.8, 0.1) or when discount rates are (0.4, 0.9), (0.3, 1), (0.1, 0.8) or (0.1, 0.4). With CES utility function similar case occurs when wages are given by (1, 3), or when composition of the economy N is (0.1, 0.5, 0.4) or discount rates are (0.3, 0.9-1), (0.2, 0.7-0.8) or (0.1, 0.4).

Overall, the result on the signs of marginal implicit tax rate on savings seems to be very robust to the chosen parameters.

U-shapeness of replacement rates and consumption dispersion

With both Cobb-Douglas and CES utility functions, paternalistic government and only downwards self-selection constraints binding, U-shapeness remains valid with all possible combinations of $\delta^L, \delta^H \in (0,1)$ with a step of 0.1 and $\delta^L < \delta^H$, $N^i \in (0.1, 0.8)$ so that $\sum N^i = 1$ and $n^L, n^H \in (1, 10)$ so that $n^L < n^H$ with a step of 1.

Also consumption dispersion follows the results presented earlier: consumption in the retirement period is more equally distributed than in the first period when government is paternalistic. With welfarist government first period consumption is less dispersed.

With welfarist government there are exceptions for the generality of U-shapeness.

- Varying discount factor, δ : With CES utility functions replacement rate measured by x/c is non-U-shaped when discount factors are (0.3, 0.8), (0.3,

³⁰ These pairs are (1,7), (1,8), (1,9), (1,10), (2,8), (2,9), (2,10), (3,9), (3,10) and (4,10), i.e. when the difference is more than 6.

0.9), (0.3, 1), (0.1, 0.8), (0.1, 1), (0, 0.1) and (0,0.2). In these cases x/c for types 1 and 2 is equal (with accuracy of 4 digits). With Cobb-Douglas utility function there are similar cases³¹. There is also three exceptions, where replacement rate x/c is increasing: this occurs when discount factors are (0.6, 0.9), (0.6, 1) and (0.7, 0.9). In all these cases, however, replacement rate measured as a ratio between second period consumption and first period net income, x/ny remain U-shaped.

- Varying composition of the economy, N : With CES utility function there are three exceptions to U-shapeness of x/c : they occur with composition (0.2, 0.4, 0.5) where replacement rate is increasing and (0.1, 0.6, 0.3) and (0.1, 0.5, 0.4) when replacement rate of types 1 and 2 are equal (or almost equal). With Cobb-Douglas utility function there are only two cases with composition (0.2, 0.7, 0.1) and (0.1, 0.6, 0.3) when the replacement rates of types 1 and 2 are equal or almost equal.
- Varying productivities, n : With CES utility function in all solvable cases U-shapeness remains valid. With Cobb-Douglas utility there are three cases, where replacement rate x/c is increasing: (3,5), (1,4) and (1,3).

In sum: There are some combinations of the parameters that result in either non-solvable case or in the solution that is not appropriate for the model. However, these seem to be special cases and the failure of finding the optimum is due to technical problems in optimisation algorithm. In most cases the results found in the main text remain valid.

³¹ These occur when discount rates are (0.1, 0.4 – 0.6), (0.1, 1), (0.2, 0.6-0.8), (0.2, 1), (0.3, 0.7-0.9) and (0.4, 0.9).