TAMPERE ECONOMIC WORKING PAPERS

NET SERIES

ON THE DESIGN OF AN OPTIMAL NON-LINEAR PENSION SYSTEM WITH HABIT FORMATION

Sanna Tenhunen
Matti Tuomala

Working Paper 68
September 2008
http://tampub.uta.fi/econet/wp68-2008.pdf

DEPARTMENT OF ECONOMICS AND ACCOUNTING

FI-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191
ISBN 978-951-44-7497-2

12.09.2008

On the design of an optimal non-linear pension system with habit formation

Sanna Tenhunen*

Finnish Centre for Pensions

and

Matti Tuomala**

University of Tampere

*E-mail: sanna.tenhunen@etk.fi
** E-mail: matti.tuomala@uta.fi

Abstract

We study optimal lifetime redistributive policy in a two-period model where individuals differ in both productivity and future discounting and where their retirement period utility is dependent of the earlier standard of living, i.e. there is a habit formation effect. We consider both welfarist government respecting individual utilities as well as the case where the government aims at correcting the short-sightedness of the individuals.

We assume that both gross income and saving are observable. Hence the pension benefits schemes can be based on both gross income and saving of each type. Analytical results depend on the pattern of incentive constraints. The two sources of imperfect information, wage rate and time discounting, interact, so the effects cannot be separated. To consider closer the tax and pension policy and the effect of the degree of habit formation the problem is also solved numerically. We found support for non-linear pension program.

1. Introduction

Publicly provided retirement program plays many different roles. It is a forced savings program, ensuring that individuals put aside sufficient amounts for old days. The public pension systems are also income redistribution and insurance programs. The rationale of forced saving for public pension system has become more motivating as the recent research in behavioural economics has demonstrated that an individual decision-making often suffers from various biases. The reason for the inadequate saving may result for example from irrationality or lack of financial skills, imperfect information or missing markets. In these situations when there is a possible conflict between an individual's long term preference and his short term behaviour the government may want to intervene.

The normative analysis of such individual decision failures in the context of the design of a pension system has received some attention. There is some literature on pension policies that attempts to take into account possible "undersaving" by households. Diamond (1977) discussed the case where individuals may undersave due to mistakes. Sheshinski (2003) proposed a general model with faulty individual decision-making, where restricting individuals' choices leads to welfare improvements. Feldstein (1985) studied the optimal pay-as-you-go system in an OLG setting in a case where individuals have higher discount rates than the government. İmrohoroğlu, İmrohoroğlu and Joines (2003) made numerical simulations in a pay-as-you-go model when individuals have hyperbolic discounting preferences. They conclude that a pension system provides additional welfare for myopic individuals. Diamond and Köszegi (2003) also employ this kind of model with hyperbolic discounting to study the policy effects of endogenous retirement choices using a public pension system as a commitment device. Diamond (2003) considers optimal distortions of savings behaviour in the case with myopia, i.e. when individuals totally ignore the effect their decisions have on future periods. Tenhunen and Tuomala (2007) extend Diamond's analysis to the case where myopia and skill are imperfectly correlated. Cremer et al. (2008) also examines a non-linear social security scheme when the government has a paternalistic view and wants to help to overcome individuals' myopia problems. They all find some support for a public intervention in terms of a pension policy or a non-linear taxation of savings.

Utility from consumption is usually assumed to be non-positional. Duesenberry (1949) in turn proposed an idea that the utility from consumption can be affected by the level of past consumption. The idea was formalized by Pollak (1970) and Ryder and Heal (1973). There is now a growing body of empirical work to support that past or habitual consumption is a determinant of current utility. In recent years, habit formation models have been used to analyze a variety of phenomena both in macro and microeconomics[1].

Rationally people with habit formation should choose an increasing consumption profile to always consume more than the level they are accustomed to. However, there is evidence that people tend to mispredict the changes in their future tastes. All models considering undersaving due to high discounting or myopia can be seen as special cases of this more general phenomenon, projection bias.[2] Projection bias implies that people, while understanding the direction of the changes in tastes, systematically make mistakes in predicting the magnitude of the changes. When projection bias is combined with habit formation, there exist two mechanisms that affect the savings behaviour. First, projection bias leads to underestimation of the effect the current consumption has on future utility, and hence encourages to consume excessively when young. The second effect arises in a dynamic setting: as one gets used to higher consumption levels, which induces even higher consumption, and less savings, than what was earlier planned. Thus the both effects tend to leave savings below the optimal level (Loewenstein, O'Donoghue and Rabin, 2003). As the importance of maintaining the consumption possibilities in the retirement period is furthermore emphasized when people exhibit habit formation, there might occur an additional rationale for public pension policy.

Our intention is to consider optimal lifetime redistribution of consumption across individuals. As in Diamond (2003) we also concentrate on within-a-cohort effects to avoid the dynamic complications that arise from intergenerational redistribution. We investigate both analytically and, to gain a better understanding, numerically the properties of an optimal lifetime redistributive policy in a two-period variant of the Mirrlees (1971) income tax problem, where all individuals work in the first period and then retire. We assume that the government can tax both labour income and the returns from savings on a nonlinear scale, which enables a consideration of a publicly provided pension program as a part of tax policy.

---

[1] Habit formation models have been applied to monetary policy questions (see e.g. Fuhrer, 2000), to research on consumption growth (see e.g. Sommer, 2007), in asset pricing literature to consider equity pricing puzzle (see e.g. Abel, 1990 and Constantinides, 1990) and to explain consumption and savings behavior (see e.g. Alessi and Lusardi, 1997, Smith, 2002).

[2] For a review of the empirical evidence on projection bias see Loewenstein and Schkade (1999) and Loewenstein O'Donoghue and Rabin (2003).

Unlike in the original Mirrlees model, we assume that individuals do not differ only in productivity, but also in time disocunting.[3] Assuming a specific correlation simplifies the analysis but is also a non-realistic assumption. Multidimensional heterogeneity in individuals' characteristics complicates the analysis notably, as discussed e.g. in Boadway et al. (2002). There are some analytical studies in a discrete case with two-dimensional heterogeneity, but they are simplified further to a three-type case (Cuff, 2000; Blomquist and Christiansen, 2004). Cremer, Lozachmeur and Pestieau (2004) also consider social security and retirement decision in a static model with individuals differing in two dimensions, productivity and health status. Moreover, due to the nature of heterogeneities they use, the pattern of the constraints for optimal tax system is simpler than in the general case. A case with imperfect correlation between the heterogeneities, i.e. a four-type case, is considered in Tenhunen and Tuomala (2007) and Cremer et al. (2008a), who consider optimal tax and pension policy in a case where people differ by their productivities and the level of myopia, but their model does not include habit formation. Whereas Loewenstein, O'Donoghue and Rabin (2003) discuss biased savings decision with a habit-forming representative consumer, models with habit formation have not been applied to redistribution policy questions. There is at least one exception. Cremer et al. (2008b) study labour supply after retirement and tax policy implications in the model where they combine myopia and habit formation.

Our purpose in this paper is to incorporate the habit formation effect into the lifetime redistribution model where individuals differ in two respect; productivity and time discounting. We explore and compare the properties of the optimal lifetime redistributive policy that would be selected by a welfarist government and the one that would be chosen by a paternalistic government. The structure of the paper is as follows. Section 2 considers a four-type version of the Mirrlees model where individuals differ in productivity skills and time discounting or discount factors with both paternalistic and welfarist governments. The optimal policy is considered both with analytical and numerical tools. A case with myopic individuals is also examined. In Section 3 we consider the numerical solution for the problem. Section 4 concludes.

2. A general four-type model

As a workhorse model we employ the much used two-type model first applied by Stern (1982) and Stiglitz (1982), where each individual has a skill level reflecting his wage rate, denoted by *n*.

[3] Sandmo (1993) considers a case where people differ in preferences, but are endowed with the same resources. Tarkiainen and Tuomala (1999; 2007) also analyze a continuum of taxpayers simultaneously distributed by skill and preferences for leisure and income.

Unlike the original Mirrlees model we assume that individuals do not differ only in productivity, but also in discount factor or time preference, denoted by δ. The difference between individual discount factors can be due to pure variation in the level of projection bias. Or, alternatively following O'Donoghue and Rabin (2001), we can interpret the disparity to result from the fact that a part of the individuals are naïve, and underestimate their level of short-sightedness, while the others are sophisticated and observe their own time discount preference.Without any assumptions on the interdependence of these two characteristics, we end up having a four-type economy, with types denoted according to table 1. The proportion of individuals of type i in the population is $N^i > 0$, with $\sum N^i = 1$.

		Wage	
		low	**high**
Discount factor	**low**	type 1	type 3
	high	type 2	type 4

TABLE 1: Individual types

Individuals supply labour in the first period, denoted by *y*, (y can also think of as the retirement age) and receive labour income *ny*, which is taxed with a non-linear scheme *T(ny)*. In the second period they retire. Individuals are free to divide their first period income between consumption, *c* and savings, *s*. Each unit of savings yields a consumer an additional 1+θ units of consumption in the second period after tax income, *x*. As a further simplification we assume that the return to savings, θ, is fixed, 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 $c = ny - T(ny) - s$ and $x = (1+\theta)s$.

Individuals get utility from consumption *c* and *x*, leisure, 1-*y*. The utility of the retirement period consumption depends on past levels of consumption through a habit formation or standard of living channel. Parameter ρ describes the magnitude of habit formation: the closer to zero ρ is, the smaller the effect of the habit is to the retirement period utility. Suppose that preferences are not additive over time.[4] Preferences could then be given by

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

[4] Non-separability of utilities would already alone imply habit formation, but we have chosen to use parameter ρ to bring the effect of habit formation clearer.

The utility function u and ψ are increasing and strictly concave. In addition, v is increasing and strictly concave with respect to x, whereas with respect to c it is decreasing and strictly convex. We also assume that all goods, both consumption goods and leisure, are normal.

Government optimization problem

The government wishes to design a lifetime tax system that may redistribute income between individuals in the same cohort. The lifetime version of the Mirrlees model can be interpreted either such that the government controls first and second period consumption and labour supply or the retirement age directly, subject to the self-selection constraints or, alternatively, if we can interpret this so that there is no private savings we have a model of labour income taxation in the first period and public provision of pension in the second period.

There is asymmetric information in the sense that the tax authority is informed neither about individual skill levels, labour supply nor discount rates. It can only observe before–tax income, *ny* and savings, *s*. Thus the design of a tax policy is restricted by self-selection constraints requiring that each type should rather choose the bundle of consumption and labour supply designed to them instead of mimicking any other type. When the self-selection constraints bind depends now on how the labour supply and the distributional preferences of the government hinge on the time disocunting and skill level. Without any assumptions of the mimicking behaviour there are twelve possible self-selection constraints given by

$$u(c^i)+\delta^i v(\rho c^i, x^i)+\psi(1-y^i)\geq u(c^j)+\delta^i v(\rho c^j, x^j)+\psi\left(1-\frac{n^j y^j}{n^i}\right) \tag{2}$$

for i,j = 1,…4 and i ≠ j.

The choice of the social welfare function that the planner maximizes when designing the tax policy naturally affects the results on the optimal taxation. A welfarist case where the social utility function reflects individual's preferences is commonly used. However, it has been argued that individuals' own assessments of their utility may not be correct, even in their own opinion. This might justify a paternalist social welfare function, where the form of the utility used in the social welfare function differs from that used by individuals. The importance of a forced saving program becomes clear in the case where individuals are not saving enough, or in the case of myopic individuals, at all, to be able to survive retirement period. In these kinds of cases a public pension program representing a paternalistic view is desirable. To bring light on the differences of these two options we consider both welfarist and paternalistic case respectively.

We consider a case, where the government valuates the future with a higher discount factor δ^g than the individuals, i.e. $\delta^g > \delta^i$ for all i=1,…,4. In the welfarist case the government maximizes the sum of individuals' utilities

$$\sum N^i \left[u\left(c^i\right)+\delta^i v\left(\rho c^i, x^i\right)+\psi\left(1-y^i\right)\right] \tag{3a}$$

whereas in the paternalistic case the retirement period utility gets a higher weight as a result of a higher discount factor

$$\sum N^i \left[u\left(c^i\right)+\delta^g v\left(\rho c^i, x^i\right)+\psi\left(1-y^i\right)\right] \tag{3b}$$

The optimization is done subject to the self-selection constraints given in (2) and the production constraint

$$\sum N^i \left(n^i y^i - c^i - r x^i\right) - R \geq 0 \tag{4}$$

where $r = \dfrac{1}{1+\theta}$ is the discount factor for production.

Welfarist case

The first order conditions with respect to c, x, and y for type i, i=1,…,4 are

$$N^i\left(u_c + \delta^i \rho v_c\right) - \lambda N^i + \sum_j \mu^{ij}\left(u_c + \delta^i \rho v_c\right) - \sum_j \mu^{ji}\left(u_c + \delta^j \rho v_c\right) = 0 \tag{5}$$

$$N^i \delta^i v_x - \lambda N^i r + \sum_j \mu^{ij} \delta^i v_x - \sum_j \mu^{ji} \delta^j v_x = 0 \tag{6}$$

$$-N^i \psi_y + \lambda N^i n^i - \sum_j \left(\mu^{ij} - \frac{n^i}{n^j}\mu^{ji}\right)\psi_y = 0 \tag{7}$$

where λ and μ^{ij} are the Lagrange multipliers for the production constraint and self-selection constraint binding type i from mimicking type j.

Individual i's marginal rate of substitution between consumption in period one and consumption in period two is $\dfrac{u_c + \delta^i \rho v_c}{v_x}$. To shorten the expressions, we denote it by σ^i. With the help of the first order conditions this can be given in a form $\sigma^i = \dfrac{\delta^i}{r}\left(1-d^i\right)$, where d^i captures the implicit tax on savings. A positive d reflects marginal tax on savings, and a negative *d* implies a marginal subsidy. With the help of the first order conditions (5) and (6) we can rewrite the marginal rate of substitution as

$$\sigma^i = \frac{\delta^i}{r}\left[1 - \frac{u_c}{\lambda N^i}\sum_j \mu^{ji}\Delta^{ji}\right] \tag{8}$$

where $\Delta^{ji} = \dfrac{\delta^j - \delta^i}{\delta^i}$.

Equation (8) implies that the marginal distortion on type i's saving behaviour depends crucially on the pattern of binding self-selection constraints. When type i is mimicked by a type with a higher discount factor, the distortion tends to increase (tax). And similarly, when type i is mimicked by a type with a lower discount factor, the distortion decreases (subsidy). The overall effect might be positive, negative or zero, depending on the binding self-selection constraints.

It is also worth noticing, that there is nothing that explicitly reveals the effect induced by habit formation in the right hand side of (8). Indeed, when the same problem is solved without habit formation (i.e. δv_c=0), we would have $\sigma^i_{no\ habit} = \dfrac{u_c}{v_x} = \dfrac{\delta^i}{r}\left[1 - \dfrac{u_c}{\lambda N^i}\sum_j \mu^{ji}\Delta^{ji}\right]$. This does not imply that the implicit tax rates are not affected by habit formation, the numerical values are still likely to differ in the cases with and without habit formation.

From the first order conditions (5) and (7) we can solve the marginal tax of labour income in terms of the marginal rate of substitution, $T^{i\prime} = 1 - \dfrac{\psi_y}{n^i\left(u_c + \delta^i \rho v_c\right)}$. This can be interpreted the implicit tax (or subsidy) on prolonged activity. In general form the marginal tax of labour income for type i is given by

$$T^{i\prime} = 1 - \frac{\lambda N^i}{\lambda N^i + \rho\sum_j \mu^{ji}\left(\delta^j - \delta^i\right)v_c}\,\frac{N^i + \sum_j\left(\mu^{ij} - \mu^{ji}\right)}{N^i + \sum_j\left(\mu^{ij} - \dfrac{n^i}{n^j}\mu^{ji}\right)} \tag{9}$$

The marginal labour income tax rates also crucially depend on the pattern of binding self-selection constraints. Whereas the distortion on savings of type i depended only on the pattern of those constraints deterring others from mimicking type i, T’ depends also on those self-selection constraints stopping type i from mimicking others.

The habit effect appears now also explicitly in the marginal tax rate, implied by term $\frac{\lambda N^i}{\lambda N^i + \rho \sum_j \mu^{ji} \left(\delta^j - \delta^i\right) v_c}$. This term implies that the marginal labour income tax of type i is increased when type i is mimicked by those with a lower discount factor, and decreased if mimicked by those with higher discount factor. However, the two sources of distortion, imperfect information on both skill level and time discounting, interact. The effect of mimicking also comes from the coefficients μ^{ji} in the last part of the term, so the effect is not separable.

It is also worth noting that the famous result of Mirrlees (1971) requiring that marginal labour income tax rates must be non-negative does not necessarily hold here. This can be seen by considering the last term in (9). There is a possibility, that at least one of parts of the last term in the right hand side is high enough to make the term greater than one. The first part, $\frac{\lambda N^i}{\lambda N^i + \rho \sum_j \mu^{ji} \left(\delta^j - \delta^i\right) v_c}$, becomes greater than one when , $\sum_j \mu^{ji} \left(\delta^j - \delta^i\right) v_c < 0$, i.e. when the sum of the discount factors of the mimickers are greater than those of the mimicked (for the pairs for which the incentive constraint is binding), and the last part, $\frac{N^i + \sum_j \left(\mu^{ij} - \mu^{ji}\right)}{N^i + \sum_j \left(\mu^{ij} - \frac{n^i}{n^j} \mu^{ji}\right)}$, when $\frac{n^i}{n^j} > 1$ mimicker has lower wage rate than the mimicked (for the pairs for which the incentive constraints are binding).

The intuition behind this result is that in the Mirrlees case only the ones with lower wage rate were in danger to be mimicked, but in two-dimensional case the pattern of the binding constraint can also be other way around. The additional disutility obtained from mimicking a type with higher wage rate by having to work more than the mimicked type, can be more than compensated by the additional utility of the higher second period consumption if the difference in the discount factors is big enough. In terms of support for prolonged activity, this result means that it might be desirable to encourage prolonged activity of some types with a wage subsidy. However, with the assumptions we have made on the direction of binding incentive constraints it is not clear, whether this hypothetical outcome ever occurs. We will return to this question in the numerical part of the paper.

Paternalistic case

Next we turn to a case where the social planner uses a higher discount factor in valuating future period. The empirical findings that the consumption tends to drop more than expected[5] supports the fact that individuals may not be doing their savings decisions (fully) rationally. There are also several other studies suggesting that people do mistakes in savings decisions and fail to sufficiently for their retirement.[6] When this is the case individuals may actually wish that the government would intervene.

The paternalistic case can be interpreted so that all types, as their discount factor is below the governments' view of the optimal level, are suffering from projection bias. The government aims at correcting the projection bias individuals have on predicting the utility from retirement consumption. In addition that agents undervalue the effect of retirement consumption level, they also fail in predicting the level of change in the tastes resulting from habit formation. Thus there are now two mechanisms affecting savings behaviour. First is the projection bias concerning the utility from savings in the future. Undervaluation of the consumption level x indicates that savings tend to remain at too low level compared to the optimum. The other effect comes from underestimating the effect of habit formation.[7] As increased current consumption decreases the utility level from retirement consumption, projection bias hence encourages to consume excessively when young. Herrnstein et al. (1993) call this negative internality. Furthermore, as one gets used to higher consumption levels, which induces even higher consumption, and less savings, than what was earlier planned. Thus the both effects tend to leave savings below the optimal level

Now the government, using its own discount factor in the social welfare function, maximizes (3b), subject to production constraint (4) and self-selection constraints (2). The first order condition with respect to y is the same as in welfarist case, given in (7). The conditions with respect to c and x for i=1,…,4 are

$$N^i\left(u_c+\delta^g\rho v_c\right)-\lambda N^i+\sum_j \mu^{ij}\left(u_c+\delta^i\rho v_c\right)-\sum_j \mu^{ji}\left(u_c+\delta^j\rho v_c\right) \tag{10}$$

$$N^i\delta^g v_x-\lambda N^i r+\sum_j \mu^{ij}\delta^i v_x-\sum_j \mu^{ji}\delta^j v_x \tag{11}$$

5 See e.g. Banks, Blundell and Tanner (1998).

6 See e.g. Laibson, Repetto and Tobacman (1999), Choi et el. (2003), Haveman et al. (2007), Benartzi and Thaler (2007).

7 Loewenstein, O'Donoghue and Rabin (2003) discusses the choice of a representative consumer in a first-best world without taxation.

There are now two possible ways to consider the distortion in the savings: government's and individual's points of views. Government's view represents the actual distortion required to implement the second-best optimum, whereas what determines the behavioural effects is the distortion perceived by the individual. The marginal rate of substitution between first and second period consumption from the government's point of view can be solved from the first order conditions: $\left(\frac{u_c+\delta^g \rho v_c}{v_x}\right)^i =: \sigma^{gi}$. Rewriting this in the form $\sigma^{gi} = \frac{\delta^g}{r}\left(1-\alpha^i\right)$ gives us a possibility to consider the optimal pension program described by α^i. With help of (10) and (11) we get

$$\sigma^{gi} = \frac{\delta^g}{r}\left[1-\frac{u_c}{\lambda N^i}\left(\sum_j \mu^{ji}\Delta^{jg} - \sum_j \mu^{ij}\Delta^{ig}\right)\right] \tag{12}$$

The distortion α depends on i, which suggests that the optimal pension program should be non-linear. The sign of the distortion, whether type i should taxed, subsidized or left undistorted at the margin, depends again on the pattern of binding self-selection constraints. It is worth noting that the effect constraints of other types mimicking type i decreases the distortion (subsidy) and the effect of type i mimicking the others increases the distortion (tax).

The marginal tax rate chosen by the government, α, takes care of the negative internality, whereas individual, doing her labour supply and savings choices sees that as a pure distortion. If instead we consider a similar term as in welfarist case, $\sigma^i = \left(\frac{u_c+\delta^i \rho v_c}{v_x}\right)^i$ we get the distortions from the individuals' perspectives.

$$\sigma^{i} = \frac{\delta^i}{r}\left[1-\frac{u_c}{\lambda N^i}\left(\sum_j \mu^{ji}\Delta^{ji} - N^i\Delta^{gi}\right)\right] \tag{13}$$

Comparing this to the welfarist case, there appears now an additional term $N^i\Delta^{gi}$ in the distortion d^i that results from the difference in the discount factors between the government and individuals. Intuitively, the term tends to decrease the distortion (subsidy) to induce higher saving rate.

The marginal labour income tax in the paternalistic case can be solved with the help of the first order conditions (7) and (10): $T^{gi\prime} = 1-\frac{\psi_y}{n^i\left(u_c+\delta^g\rho v_c\right)}$. In our case this implies that

$$T^{gi'} = 1 - \frac{\lambda N^i}{\lambda N^i - \rho \sum_j \mu^{ij}\left(\delta^i - \delta^g\right)v_c + \rho \sum_j \mu^{ji}\left(\delta^j - \delta^g\right)v_c} \frac{N^i + \sum_j \left(\mu^{ij} - \mu^{ji}\right)}{N^i + \sum_j \left(\mu^{ij} - \frac{n^i}{n^j}\mu^{ji}\right)} \qquad (14)$$

There are now three sources for distortions that restricts government's aims in welfare maximization: paternalistic view of correcting projection bias, and imperfect information on both skill level and time discounting. These features interact so that the effects cannot be separated from each other, i.e. income taxation is used at the same time to take care of all distortions in the economy. Similarly, like in the welfarist case, also here the marginal labour income tax rates depend on the pattern of binding self-selection constraints. Also here the possibility of negative marginal labour tax rates remains valid. As habit formation and projection bias are likely to induce too high working period consumption, encouraging people to work might become even more desirable than in welfarist case. That way a larger share would be left for savings.

Myopia

Next we analyse the model in which some individuals are myopic. They are not able estimate at all the needs that first period consumption creates later on. Myopic behaviour may be quite common for a substantial proportion of individuals who hardly save at all and rely almost entirely on public pension benefits. Diamond (2003, Chapter 4) and Diamond and Mirrlees (2000) consider a benchmark situation with a separable utility function 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 an optimal redistributive policy for those of 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 heavy discounting of the future. If individuals have low discount factors, they will save little for their retirement consumption, but this reflects optimising behaviour. By contrast, if individuals are myopic, they get surprised by the inadequacy of their savings at the beginning of the retirement period. Thus, when myopic individuals are subjected to compulsory saving, their welfare will increase. It is also clear that with myopia the interpretation of the distortion for savings should be the compulsory pension system instead of a tax system: when individuals do not save voluntarily, there is no way to induce them to do so with any subsidy.

In the case with myopia the basic justification for a pension system is to guarantee some level of resources in the retirement period. Diamond (2003) has considered a case where all types are myopic and only differ by their productivities. Tenhunen and Tuomala (2007) extend this analysis to a case where myopia and ability are imperfectly correlated. Both of these consider only a case with a separable utility function, i.e. without habit formation.

Myopic types perceive only the apparent utility $u(c^i)+\psi(1-y^i)$. Myopic labour supply implies that the retirement consumption does not enter at all the incentive compatibility constraint of a fully myopic mimicker. Thus they don't observe either the effect of savings on retirement consumption or the effect the current consumption has on retirement utility via habit formation. 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^i being 0 for low discount factor types 1 and 3, whereas the government's discount factor δ^g presents completely rational behaviour. Thus the earlier case can be interpreted also as a case with individuals being partly myopic, some of them more than the others.[8]

The social welfare function depends on *ex post* utilities given by $u(c^i)+\delta^g v(\rho c^i, x^i)+\psi(1-y^i)$. The welfarist consideration is left here away, as the government wants to ensure some level of consumption also in the retirement period for all types, also for those with myopic preferences. Myopic types take only the apparent utility into account, also when mimicking. The other partly myopic types mimicking still perceive the second period consumption.[9]

There is now no point in deriving the distortion for savings from the point of view of the individual for the myopic types, as they will always neglect it, no matter how large subsidy or tax would be. The marginal distortion for savings, however, still exists also for the myopic types in the point of view of the government. It is given by (12). As Δ^{ig} = -1 for the fully myopic types, it can be simplified to a from

$$\sigma^{gi} = \frac{\delta^g}{r}\left[1-\frac{u_c}{\lambda N^i}\left(\sum_j \mu^{ji}\Delta^{jg} + \sum_j \mu^{ij}\right)\right] \tag{15}$$

[8] An alternative way to model would have been to assume that the utility function v(c,x) does not exist at all for the fully myopic types.

[9] This requires a discussion of the timing. We have assumed, regardless of our two-period model, the all decisions (labour supply, savings and both periods' consumption) are made at the same time in the beginning of the first period. Also the tax design is revealed at the same time. Both individuals and government are assumed to be able to commit to their decisions. Resulting of this, non-myopic consumers can count on that they have consumption on the second period even if their private savings would not be enough to provide that level of consumption.

From the government aspect taking the second period utility into account the marginal labour income tax rate is given by (14), which for the fully myopic types simplifies to form

$$T^{gi'} = 1 - \frac{\lambda N^i}{\lambda N^i + \rho \sum_j \mu^{ij} \delta^g v_c + \rho \sum_j \mu^{ji} \left(\delta^j - \delta^g\right) v_c} \frac{N^i + \sum_j \left(\mu^{ij} - \mu^{ji}\right)}{N^i + \sum_j \left(\mu^{ij} - \frac{n^i}{n^j} \mu^{ji}\right)} \tag{16}$$

Also these marginal tax rates on labour income and savings depend crucially on the pattern of binding incentive compatibility constraints. To get a closer look at the marginal tax rates and the effect of habit formation we next turn to a numerical consideration of the model.

3. Numerical solution

All the results on the marginal distortions on labour income and saving depend crucially on the pattern of binding self-selection constraints. It is typical in the multidimensional screening problems to focus on the so-called relaxed problem. In this case, the relaxed problem is defined as the problem where only the three downward incentive constraints are considered. Of course, the solution of the relaxed problem is of no intrinsic interest except insofar as it yields the solution to the fully constrained problem. In general, other self-selection constraints beside downward constraints may also be binding at the optimum. In those cases the relaxed problem is no longer relevant. Moreover, it is not easy to ascertain the direction of redistribution. In cases where the relaxed problem is not the solution of the full problem, the 'no distortion at the top' result no longer holds and we are likely to find non-monotonic pattern of replacement rates. As this pattern cannot be solved analytically without making heavy and unrealistic assumptions, we will solve the problem numerically. Numerical values also enable us to consider the consumption structure, inequality and replacement rates in the economy, which are all essential questions of a pension program.

To solve this problem numerically, we use CES utility function: $U^i = -\frac{1}{c^i} - \delta^i \frac{1}{x^i - \rho c^i} - \frac{1}{1 - y^i}$, where ρ is a parameter measuring the importance of the current consumption relative to future consumption in future utility, i.e. the degree of habit formation. If $\rho = 1$, positive utility is gained only if consumption in the retirement period increases. If $\rho = 0$, we end up back to the case with separable utility function, discussed in Tenhunen and Tuomala (2007). We do not choose the direction of redistribution a priori, i.e. the pattern of the binding self-selection constraints is not restricted a priori, and there appear 12 self-selection constraints in the optimisation problem.

The parameterization is presented in Table 2. At first we do not make any a priori assumptions on the dependence of the wage earning ability and time discounting, i.e. the size of the groups of each type is identical, 0.25. The value for ρ is chosen to be 0.3 indicating that the utility of second period consumption is mostly made up of second period consumption, and the relative consumption affects rather moderately. We consider both welfarist and paternalistic government and the case with myopia.

Fraction of each group	$N^i = 0.25$
Discount factors	$\delta^L = 0.6$, $\delta^H = 0.8$, $\delta^g = 1$, $r = 0.95$
Productivities (wages)	$n^L = 2$, $n^H = 3$
Degree of habit formation	$\rho = 0.3$

TABLE 2: Parameterisation in the two-type economy

The pattern of the binding self-selection constraints is indeed not trivially only downwards (Table 3). A reduced form consideration of the optimum, taking into account only the downward binding self-selection constraints seems to be a reasonable simplification: none of the Lagrange coefficient for the upward binding self-selection constraint is non-zero in the welfarist case.[10] In a welfarist case there are three binding constraints: one that stops patient high-skill type (type 4) from mimicking the short-sighted high-skill type (type 3), and two restricting type 3 from mimicking either of the low-skilled types. However, in the paternalistic case the same self-selection constraints bind, but in addition to that, also constraints stopping from type 1 mimicking type 2 and vice versa are binding, although the Lagrange multipliers are zero. The mutual bindingness implies here that types 1 and 2 are actually pooled in the optimum. In the case with myopia, the same three constraints are still binding, but in addition to that constraints between types 1 and 2 are now genuinely binding, and in addition that, also the constraint stopping patient high-skilled from mimicking the impatient high-skilled binds. Also here types 1 and 2 are pooled.

[10] By downwards and upwards binding self-selection constraints we refer here to our numbering. As the numbering is chosen here randomly, it is important not to make the assumptions of the reduced form optimization before checking the actual pattern of binding constraints.

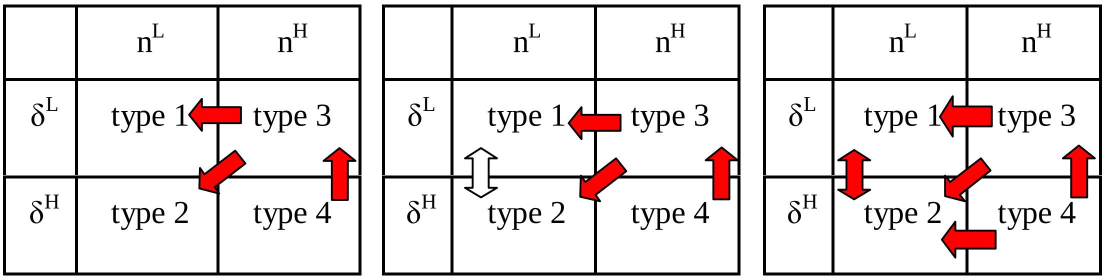

TABLE 3: Binding self-selection constraints: welfarist, paternalistic and myopia case

The results of the numerical solution are presented in Tables 4a, 4b and 4c. First column presents the utilities of each type in the optimum. It is worth noting, that impatient high-skilled type (type 3) is best of in terms of utilities in all cases. On the other hand, low-skilled patient type (type 2) gets the lowest utility. These results give further support for the fact that no clear-cut ex ante assumptions of the direction of distribution or the bindingness of the self-selection constraints should be made in a case with multidimensional heterogeneity like ours.

The marginal distortions for savings show that in the optimum the tax/pension program is non-linear: some types are subsidized at the margin while other are taxed. In the welfarist case types 1 and 4 are not distorted, whereas type 2, patient low-skilled type is subsidized at the margin and type 3, impatient high-skilled is taxed. The treatment of savings observed by the individuals in the paternalistic case shows, rather expectedly, that all types perceive a subsidy for savings (negative d^i:s), as the government with a higher discount factor wants to induce individuals to save more for their pension. The actual distortion for savings, however, shows a pattern of both subsidizing and taxing savings income (α^i:s).

There are some negative marginal labour income tax rates, even in the welfarist case. The short-sighted high-skill type (type 3) is gets a wage subsidy, while the low-skilled are both marginally taxed, or undistorted as in the welfarist case. Even if this contradicts with traditional Mirrlees results, in multidimensional optimal tax problems they are sometimes found (see e.g. Judd and Su, 2006). In our case this results from the fact that government has access to both non-linear labour income and capital income tax, which it uses for redistributional purposes. As there is imperfect information both on skill level and time discounting, the two sources of imperfect information interact. As both tax devices are used at the same time to deter mimicking also negative marginal labour income taxes are possible in the optimum. The fact that the labour supply of type 3, with a marginal wage subsidy, is not even the highest, supports the explanation that all tax devices interact.

type	U	d	α	T’	y	x/ny	x/c	Consumption dispersion
1	-5.12	0	-	0.077	0.52	63.19	120.09	$\bar{c}$ = 0.61 $\bar{x}$ = 0.78
2	-5.50	-0.140	-	0.151	0.53	67.38	139.55	Lorenz curves cross
3	-4.56	0.065	-	-0.016	0.54	51.31	117.76	Shorrocks: W(x)>W(c)
4	-4.88	0	-	0	0.56	54.06	134.02	Gini:G_c=0.071 G_x=0.068

TABLE 4a: Numerical results, welfarist case

type	U	d	α	Tg’	y	x/ny	x/c	Consumption dispersion
1	-5.23	-1.218	-0.217	0.119	0.55	69.55	155.73	$\bar{c}$ = 0.58 $\bar{x}$ = 0.86
2	-5.55	-0.592	-0.217	0.173	0.55	69.55	155.73	Lorenz: Lx > Lc
3	-4.58	-0.561	0.177	-0.066	0.56	55.43	137.99	Atkinson: W(x) > W(c)
4	-4.85	-0.251	0.057	-0.012	0.56	56.56	143.72	Gini:G_c=0.077 G_x=0.053

TABLE 4b: Numerical results, paternalistic case

type	U	d	α	Tg’	y	x/ny	x/c	Consumption dispersion
1	-4.32	-	-0.657	0.176	0.56	72.22	166.47	$\bar{c}$ =0.58 $\bar{x}$ = 0.86
2	-5.54	-	-0.174	0.234	0.56	72.22	166.47	Lorenz: Lx > Lc
3	-3.66	-	0.422	-0.217	0.55	52.57	125.32	Atkinson: W(x)>W(c)
4	-4.88	-	0.274	-0.012	0.57	56.02	143.79	Gini:G_c=0.087 G_x=0.037

TABLE 4c: Numerical results, paternalistic case with myopia

To consider the resulting pension program we consider replacement rates. The basic idea of habit formation suggests that consumer would choose an increasing consumption pattern, i.e. the replacement rate *x/c* should be greater than 100%. This seems to be happening in all cases. As usually replacement rates are given in terms of taxable income, we also consider the ratio of the second period consumption compared to gross income of the first period, *x/ny*. These replacement rates vary between 54 – 72 %, which are actually rather plausible figures. The replacement rates show a non-monotonic pattern both in the welfarist and paternalistic case, i.e. replacement rates are not increasing or decreasing in types income or consumption.

Another interesting feature in terms of tax and pension policy is to consider how equally consumption is distributed in the economy. One way to consider the overall dispersion of consumption is to calculate the Gini-coefficients[11] in both periods. It turns out that we can also make welfare rankings between the distributions of the first and the second period consumption based on the results by Atkinson (1970) and Shorrocks (1980). They determine welfare ranking between two periods with the help of Lorenz dominance. Welfare is greater if the distribution has higher mean and the Lorenz curve lies above the Lorenz curve of the comparison period at all points. Shorrocks’s result can be applied when Lorenz curves cross or when a Lorenz-dominant distribution has a lower mean. This result uses the mean-weighted cumulative distribution, called

[11] Gini-coefficients are interpreted here as descriptive measures.

generalized Lorenz curve, for comparison: welfare is higher when the generalized Lorenz curve is higher at every point of the distribution.

The numerical solution shows that in all cases the inequality, measured as dispersion in consumption, is lower in the second period. In three- and four-type models without habit formation in a case with welfarist social planner the ordering was reversed showing that in the first period inequality was lower than in the second period (Tenhunen and Tuomala, 2007). Using the information of Gini-coefficients and of Lorenz and Shorrock's results, it can be noted that tax and pension policy make retirement period consumption more equal even in a welfarist case, where the government does not have intentions to intervene with individuals projection bias.

The effect of the degree of habit formation

To consider the comparative statics with respect to the degree of habit formation, we allow ρ vary between 0 and 0.8. Every individual's replacement rates increase monotonically with the degree of habit formation (Figure 1). This is in accordance with the fact that rational decision makers pursue an increasing consumption profile when they have habit forming utilities.

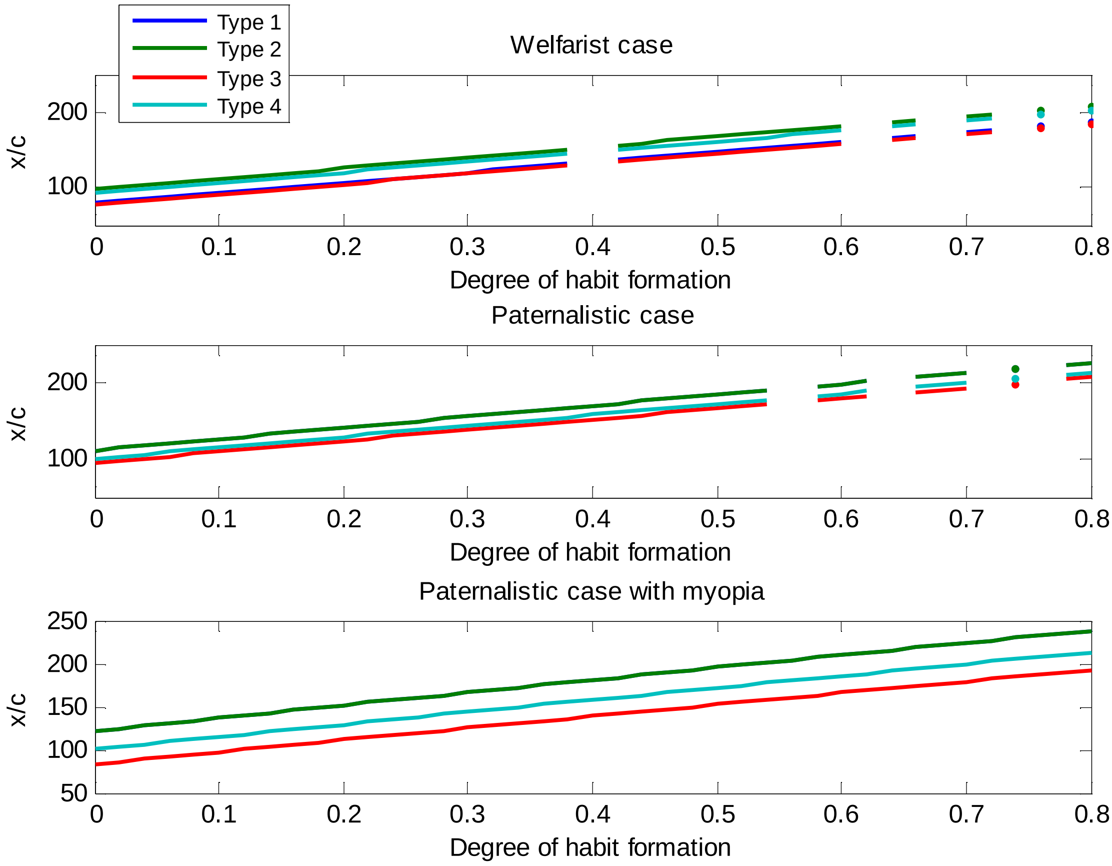

FIGURE 1: Replacement rates and the degree of habit formation

Also the labour supply or the retirement age increases with the degree of habit formation (Figure 2). These results hold for all types. When individuals have habit forming utilities and they suffer from projection bias, they tend to undersave for retirement. As the retirement period gets closer individuals might want to delay retiring to be able to continue at the consumption level they are accustomed to. A high degree of habit formation can lead to very late retiring or in an extreme case even unretiring. As in our model there is only one moment when the labour supply decision is made, the choice of retirement moment cannot be changed. However, the effect of projection bias and habit formation on the optimal labour supply is implicitly included in the government's maximization problem in paternalistic cases, where the projection bias is aimed to be corrected by using higher discounting than what (partly) myopic individuals are using.

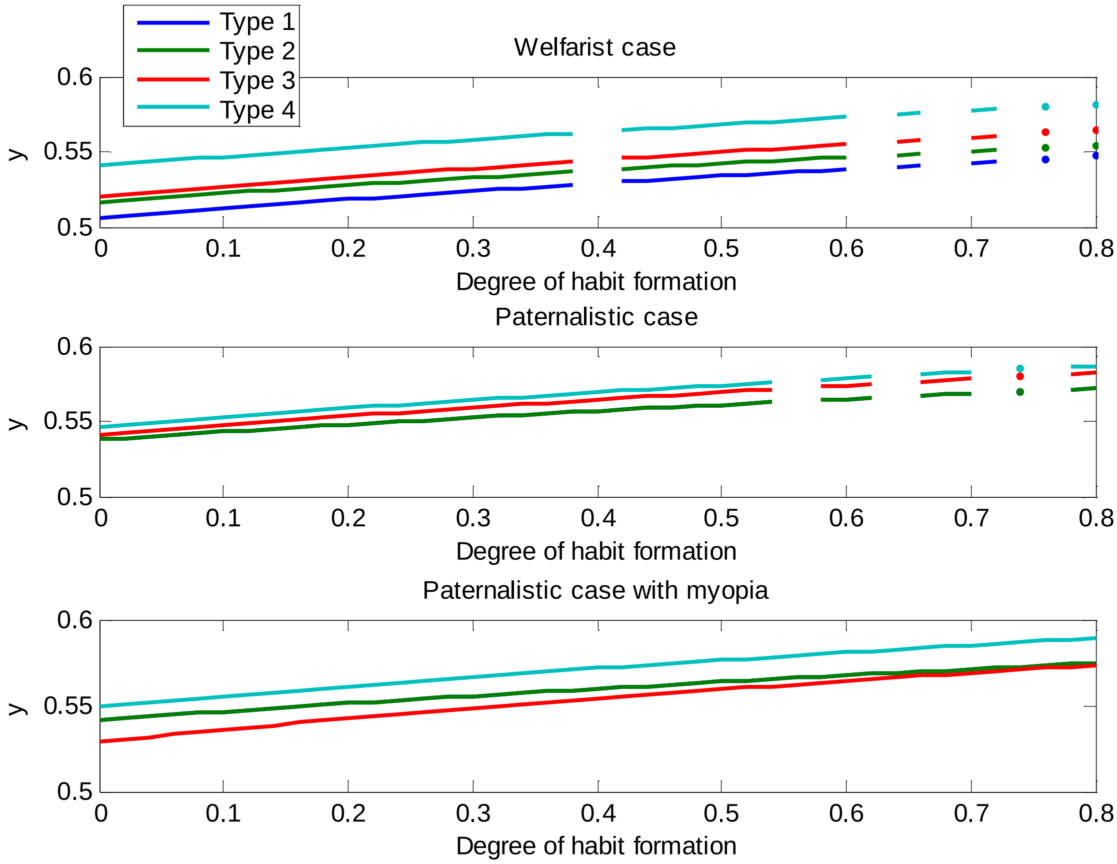

FIGURE 2: Labour supply and the degree of habit formation

The information from Gini coefficients (Figure 3) suggests that the retirement period consumption is always more equally distributed between types than the first period consumption. This can be explained by that when government is maximizing the sum of individual utilities with diminishing marginal utilities, more equal distribution yields higher overall welfare. In our case, the government has more tools to equalize retirement period consumption than in the working period, as both labour income taxation and taxation of savings (the pension system) affect retirement period consumption possibilities. In the welfarist case the difference is small, and with the habit formation parameter equal to zero retirement period consumption is actually more dispersed. In the paternalistic cases the difference between the Gini coefficients decreases as the degree of habit formation increases. It is also worth noting, that in the paternalistic cases the Gini coefficient of the retirement period slightly increases as the degree of habit formation increases. In other words, when government aims at correcting the impatience of the individuals and when the utility of the retirement period consumption is dependent on the earlier consumption, it is optimal to allow higher differences in the retirement period consumption. The consumption differences in the working period in turn decrease slightly with the degree of habit formation in all cases.

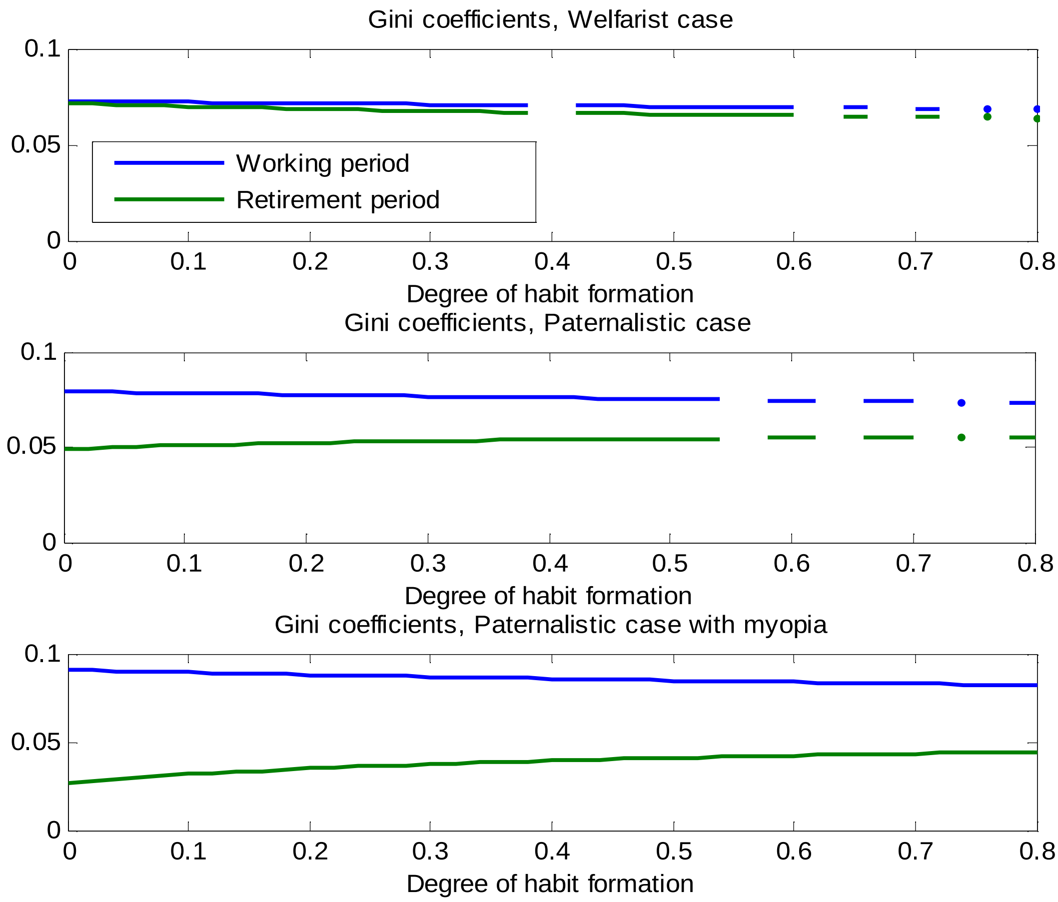

FIGURE 3: Gini coefficients and the degree of habit formation

The marginal labour income tax rates (Figure 4) are also affected by the degree of habit formation. The differences seem to spread as the degree of habit formation increases. In most cases the marginal distortions for labour income increase with the degree of habit formation. For types 1 and 2 the marginal labour income tax rate increases or remains constant with the degree of habit formation. For type 3 and type 4 for in the paternalistic cases the marginal labour income tax rates are negative, and the marginal wage subsidy increases with the degree of habit formation. The pattern looks the same with marginal tax rates on savings (Figure 5). The magnitude of distortion (subsidy of tax) in savings increases with the degree of habit formation. It can be noted that the marginal subsidy on savings perceived by type 1 is required to be more than 100 % in order to induce the second best outcome, regardless of very low levels of habit formation.

The increase in the marginal tax or subsidy rates alongside with rising level of habit formation results from the fact that with higher level of habit formation, projection bias causes bigger welfare losses to (partly and fully) myopic individuals. To maximize overall welfare bigger distortions are required to induce the optimal outcome. More surprisingly, a similar pattern happens also in the welfarist case, where, instead of projection bias, time discounting is interpreted as a pure time preference which is respected by government.

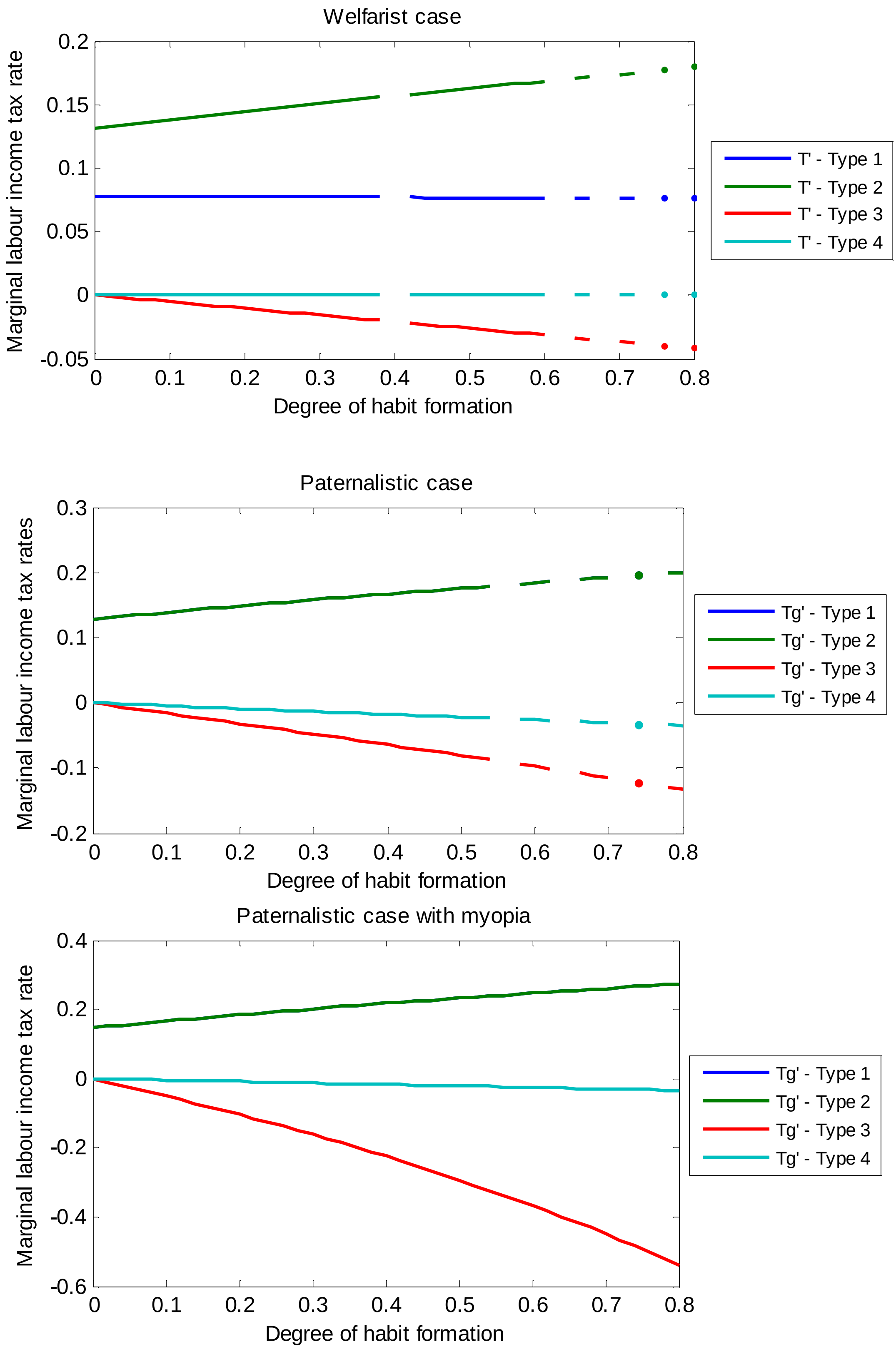

FIGURE 4: Marginal labour income tax rates and the degree of habit formation

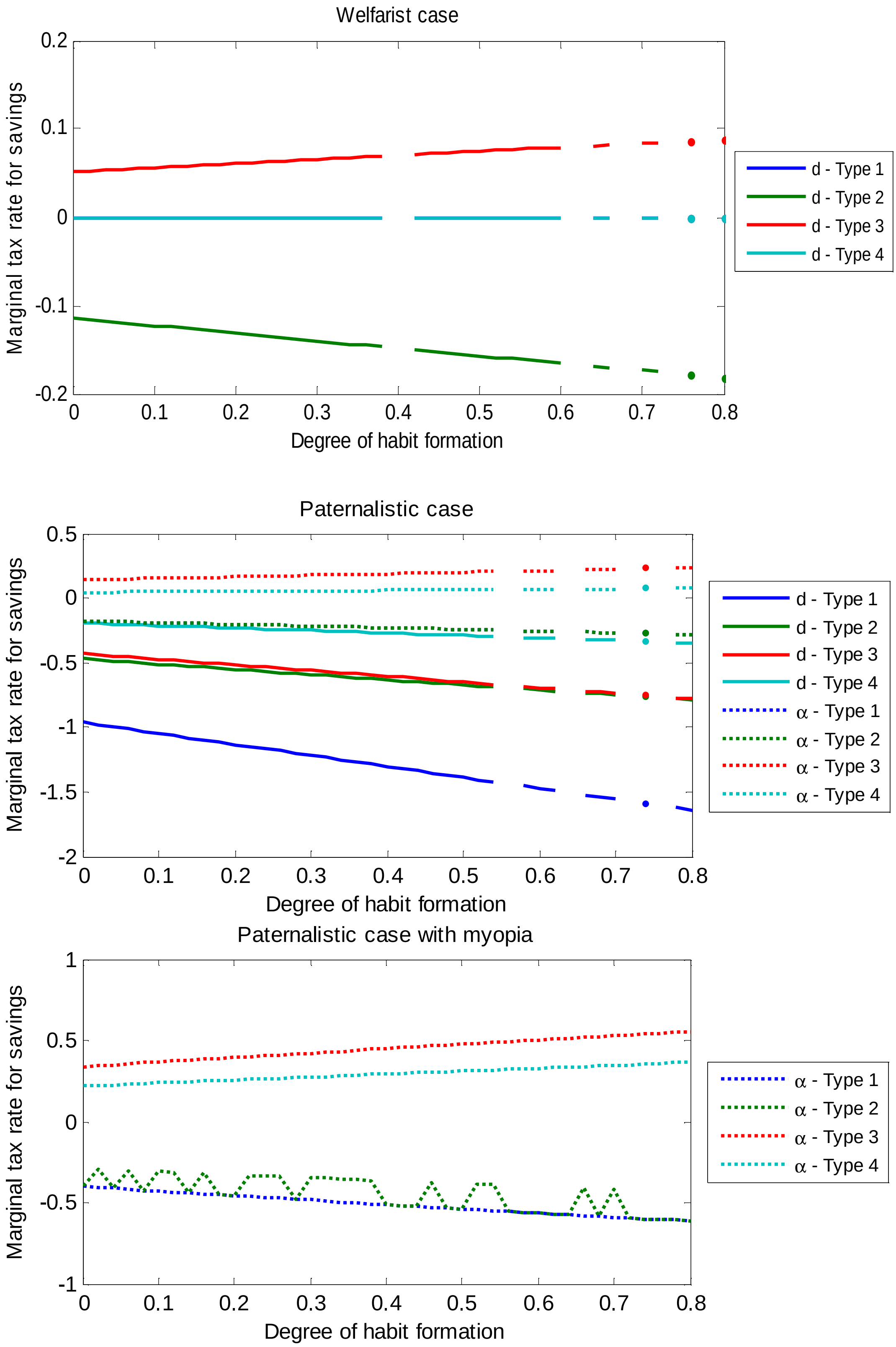

FIGURE 5: Marginal tax rates for savings and the degree of habit formation

Correlation between skill level and time discounting

The effect of the correlation can be investigated by changing the structure of the economy. Assuming a perfect correlation would return the analysis to a two-type economy.[12] However, the two characteristics may also be imperfectly correlated. We have solved the case with some other structures of the economy. At the same time, changing the assumption of the correlation allows us to consider the robustness of our results with respect to the distribution of types. To maintain the tractability of the results, we have made a simplifying assumption, following Cremer et al. (2008a), and fixed the fractions so that a half of each wage group has low and other half has high discount factor, i.e. $N^1=N^4$ and $N^2=N^3$.

The structure of the economy does not seem to affect the fact that retirement period consumption is less dispersed than working period consumption (Figure 6). In the welfarist case the difference is small especially when skill level and time discounting are perfectly correlated. In the paternalistic case Gini coefficients remain rather robust regardless of the case when the two charasteristics are strongly negatively correlated. In the case with myopia the difference between Gini coefficients in each period increases as the correlation coefficient increases from -1 to 1. It can also be concluded that the more positively correlated skill level and time discounting is the less dispersed the retirement period consumption is.

[12] In our case perfect positive correlation means that the economy consists of types 1 and 4 only, whereas perfect negative correlation means that there are only types 2 and 3 in the economy.

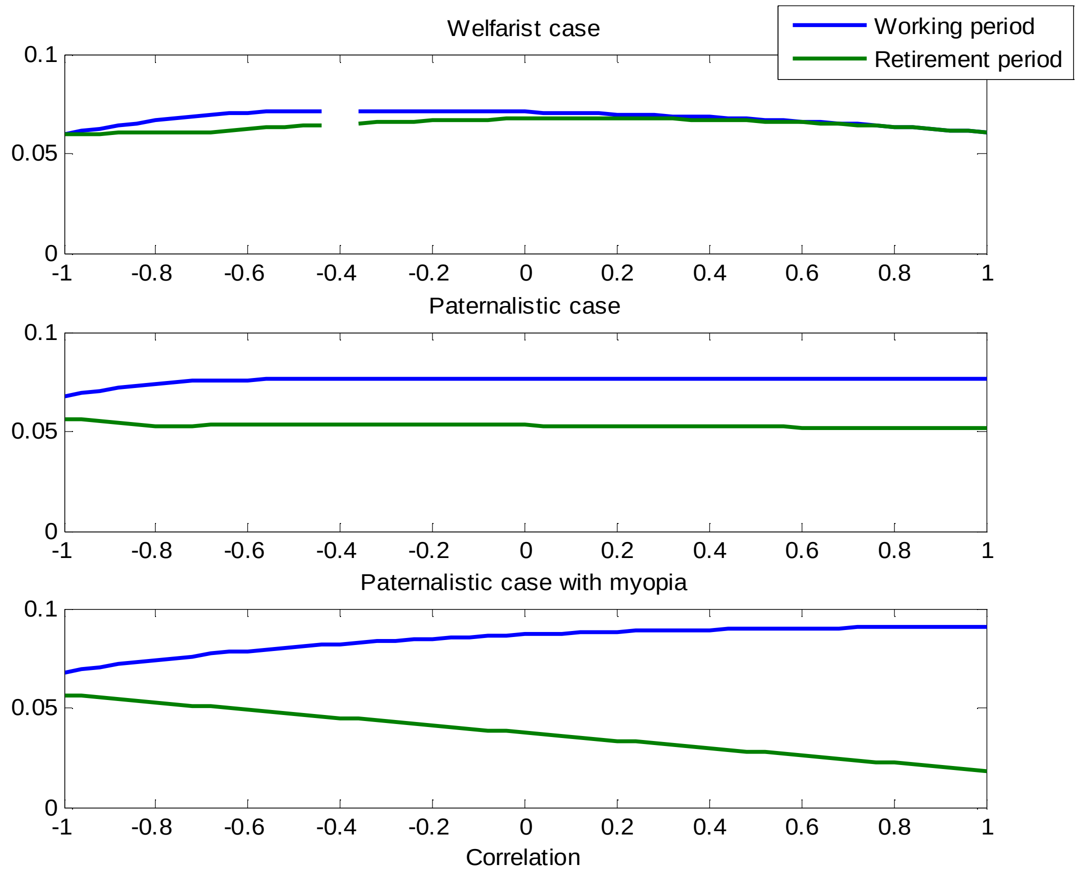

FIGURE 6: Gini coefficients and correlation between skill level and time discounting

The marginal labour income tax rates for types 1, 2 and 3 increase as the fraction of types 1 and 4 increases and the fraction of types 2 and 3 decreases (Figure 7). For type 4 the marginal labour income tax rate either remains constant (welfarist case) or decreases slightly (paternalistic cases) as the correlation coefficient increases. The marginal labour income tax for type 3, the impatient high wage type, implies wage subsidy, when skill level and time discounting are strongly negatively correlated, whereas the marginal subsidy decreases as the negative correlation decreases and turns to positive. The marginal distortions on savings in turn move towards subsidizing savings as the correlation coefficient increases (Figure 8). In the welfarist case the marginal tax on type 3 decreases and the marginal subsidy on types 1 and 2 increase with correlation coefficient. In the paternalistic cases the decrease in tax (or increase in subsidy) is more modest. For type 1 in all cases and for types 1 and 2 in the paternalistic cases the marginal subsidy on savings perceived by herself is again more than 100%. In the paternalistic case with myopia the marginal subsidy for the low skill types increases as the correlation moves from negative to positive.

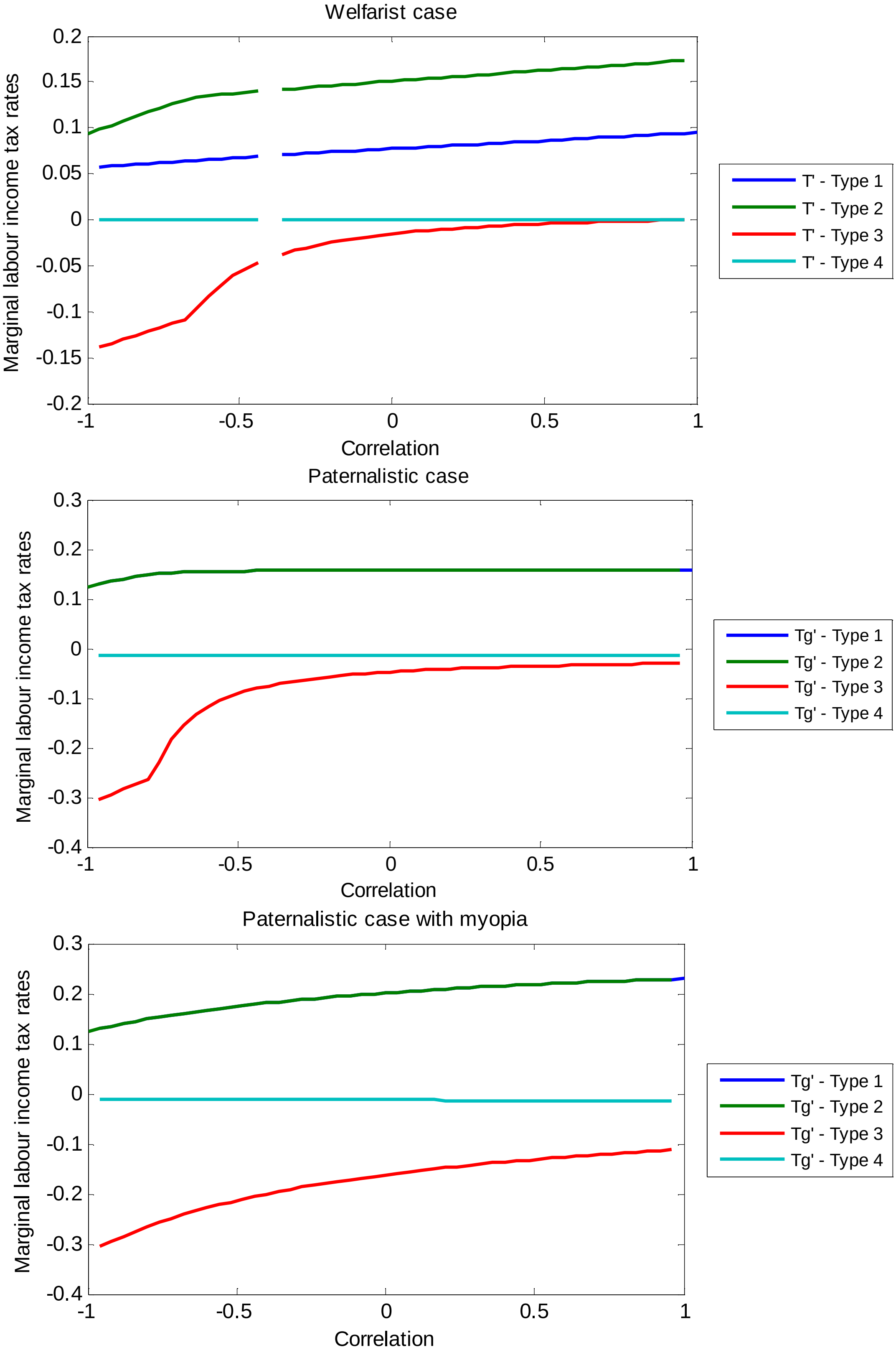

FIGURE 7: Marginal labour income tax rates and correlation between skill level and time discounting

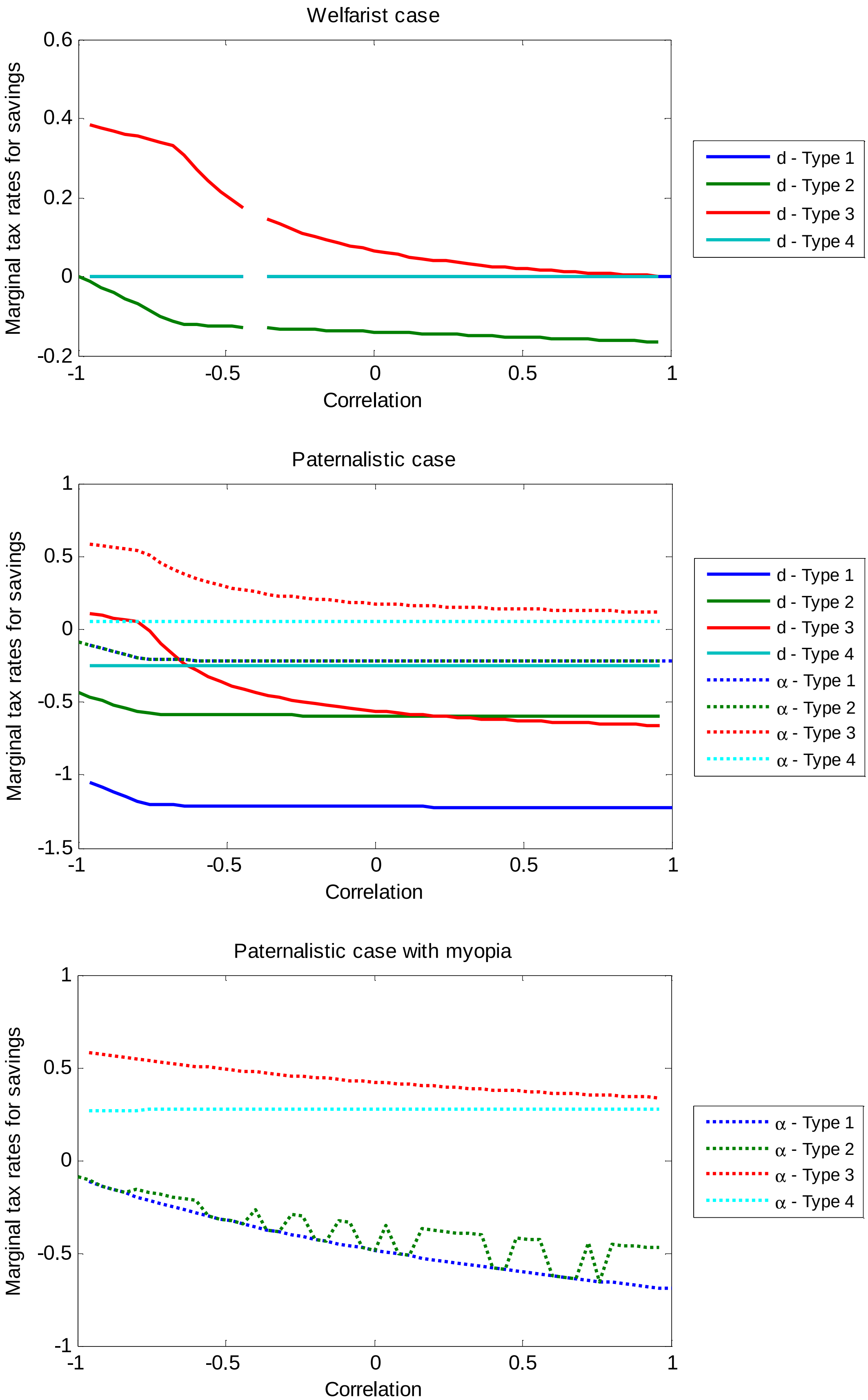

FIGURE 8: Marginal distortion for savings and correlation between skill level and time discounting

4. Conclusions

We have studied the optimal lifetime redistribution policy in a two-period model where individuals differ in both productivity and time discount factor and where their utility of the retirement period consumption is not independent of the earlier standard of living, i.e. there is a habit formation effect. We have considered both welfarist government maximizing the sum of individual utilities as well as the case where the government acts aiming at correcting the short-sightedness of the individuals by using higher discount factor for retirement period utility. We have also considered a case where a part of the individuals in the economy totally neglect retirement when making their labour supply and savings decisions.

With imperfect information and two-dimensional heterogeneity of the individuals the optimal redistribution policy depends crucially on the pattern of binding self-selection constraints. We assumed that the government can observe both gross income and saving. Hence the pension benefits schemes can be based on both gross income and saving of each type. As there is imperfect information in two dimensions, the two sources of imperfect information interact. As both tax devices are used at the same time to deter mimicking the effects of each source of heterogeneity cannot be separated. Indeed, with habit forming utilities and individuals suffering of projection bias, both labour supply and consumption decisions are suboptimal without government intervening wit tax and pension policy. To get a closer look at the tax and pension policy and the effect of the degree of habit formation on it we also solved the problem numerically.

The numerical analysis shows that, as we might expect in a case of multidimensional heterogeneity, the pattern of the binding self-selection constraints is non-trivial. The marginal distortions for savings show that in the optimum the tax/ pension program is non-linear: some types are subsidized at the margin while other are taxed. Also negative marginal labour income taxes are possible in the optimum. In most cases the marginal distortions for labour income and savings increase with the degree of habit formation. For the high skill types in the paternalistic cases the marginal labour income tax rates are negative, and the marginal wage subsidy increases with the degree of habit formation. The pattern looks the same with marginal tax rates on savings: the magnitude of distortion (subsidy of tax) in savings increases with the degree of habit formation. With a government aiming at correcting the impatience of the individuals, the difference between the dispersion of consumption in each period measured by Gini coefficients decreases. The dispersion in the retirement period increases as the degree of habit formation increases. Our numerical result also show that retirement age increases with the degree of habit formation.

Finally, although the present paper in several respects generalizes the literature on optimal redistributive policy there is still many important aspects left. To keep the mechanism behind imperfect information on the agents' characteristics tractable, we have limited the period wise decision making to one. If dynamic decision making at the beginning of each period was allowed, our setting would give opportunity to consider retirement decisions. As in aging societies the demand for labour is becoming an important issue, a consideration of tax and pension policy in a world where individuals have habit forming utilities but suffer from projection bias would be valuable information. Furthermore, we assumed that individuals do not differ in the degree of habit formation. This may crucially affect individuals' choices of retirement age, and the optimality of tax and pension policies. These questions are left for future research.

References

Abel, A. (1990), "Asset prices under habit formation and catching up with the Jones", *American Economic Review* 40, 38-42.

Alessi, R. and A. Lusardi (1997), "Consumption, Saving and Habit Formation", *Economics Letters* 55, 103-108.

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.

Banks, J., R. Blundell and S. Tanner (1998), "Is There a Retirement-Savings Puzzle?", *American Economic Review* 88, 769-788.

Benartzi, S., and R. Thaler (2007), "Heuristics and Biases in Retirement Savings Behaviour", *Journal of Economic Perspectives* 21, 81-104.

Boadway, R., M. Marchand, P. Pestieau and M. del Mar Racionero (2002), "Optimal Redistribution with Heterogeneous Preferences for Leisure", *Journal of Public Economic Theory* 4, 475-498.

Blomquist, S. and V. Christiansen (2004), "Taxation and Heterogeneous Preferences", CESifo Working Paper 1244.

Choi J., D. Laibson, B. Madriand and A. Metrick (2003), "Optimal Defaults", *American Economic Review Papers and Proceedings* 93, 180-185.

Constantinides, G. (1990), "Habit Formation: A resolution of the equity premium puzzle", *Journal of Political Economy* 98, 519-543.

Cremer, H., P. De Donder, D. Maldonado and P. Pestieau (2008a), "Forced saving, redistribution and non-linear social security scheme", CESifo Working Paper No. 2325.

Cremer, H., P. De Donder, D. Maldonado and P. Pestieau (2008b), "Habit Formation and Labor Supply", CESifo Working Paper No. 2351.

Cremer, H., J.-M. Lozachmeur and P. Pestieau (2004), "Social Security, retirement age and optimal income taxation", *Journal of Public Economics* 88, 2259-2281.

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. Kőszegi (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.

Duesenberry, J. S. (1949), "Income, Saving and the Theory of Consumer Behaviour". Harvard University Press, Cambridge, MA.

Feldstein, M. (1985), "The Optimal Level of Social Security Benefits", *Quarterly Journal of Economics* 100(2), 300-320.

Fuhrer, J. (2000), "An Optimizing Model for Monetary Policy: Can Habit Formation Help?", *American Economic Review* 90, 367-390.

Haveman, R., K. Holden, A. Romanov and B. Wolfe (2007), "Assessing the Maintenance of Savings Sufficiency over the Firsts Decade of Retirement", *International Tax and Public Finance* 14(4), 481-502.

Herrnstein R., G. Loewenstein, D. Prelec and W. Vaughan (1993), "Utility Maximization and Melioration: Internalities in Individual Choice", *Journal of Behavioral Decision Making* 6, 149-185.

İmrohoroğlu, A., S. İmrohoroğlu and D. H. Joines (2003), "Time Inconsistent Preferences and Social Security". *Quarterly Journal of Economics* 118(2), 745-784.

Judd, K. L. and C.-L. Su (2006), "Optimal Income Taxation with Multidimensional Taxpayers Types", Computing in Economics and Finance 2006, 471, Society for Computational Economics.

Laibson, D, A. Repetto and J. Tobacman (1998), "Self-Control and Saving for Retirement", *Brookings Papers on Economic Activity* 1, 91-196.

Loewenstein, G., T. O'Donoghue and M. Rabin (2003), "Projection Bias in Predicting Future Utility", *Quarterly Journal of Economics* 118(4), 1209-1248.

Loewenstein, G. and D. Schkade (1999), "Wouldn't It Be Nice? Predicting Future Feelings" in D. Kahneman, E. Diener and N. Schwartz (eds.) "*Well-Being: The Foundations of Hedonic Psychology*", Russel Sage Foundation, New York.

Mirrlees, J. (1971), "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic studies* 38, 175-208.

O'Donoghue, T. and M. Rabin (2001), "Choice and Procrastination", *Quarterly Journal of Economics* 116, 121-160.

Pollak, R. (1970), "Habit Formation and Dynamic Demand Functions", *Journal of Political Economy* 78(4), 745-763.

Ryder, H. and G. Heal (1973), "Optimal Growth with Intertemporally Dependent Preferences", *Review of Economic Studies* 40, 1-33.

Sandmo, A. (1993), "Optimal Redistribution when Tastes Differ", *Finanz Archiv* 50(2), 149-163.

Sheshinski, E. (2003), "Optimal Policy to Influence Individual Choice Probabilities", University of Hebrew, mimeo.

Shorrocks, A. F. (1980), "The Class of Additively Decomposable Inequality Measures", *Econometrica* 48, 613-626.

Smith, W. (2002), "Consumption and saving with habit formation and durability", *Economics Letters* 75, 369-375.

Sommer, M. (2007), "Habit Formation and Aggregate Consumption Dynamics", *Journal of Macroeconomics* 7:1, Article 21.

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", *International Journal of Economic Theory* 3, 35-46.

Tenhunen, S. and M. Tuomala (2007), "On Optimal Lifetime Redistribution Policy" in Tenhunen S. "Essays on the Theory of Optimal Taxation". Academic Dissertation, Acta Universitatis Tamperensis 1276. Tampere University Press.

APPENDIX: Numerical simulations

Procedure

Numerical simulation was carried out with the Matlab programme. The function used (*fmincon*) solves the optimum of a multivariable function with equality and inequality constraints that may be linear or nonlinear. It also determines which of the constraints are binding. The same procedure is applied to all cases considered in this paper.

Note that as the optimization function also allowed slack constraints, we were not restricted to a priori assumptions on the binding self-selection constraints. Thus, we included all possible constraints in the optimization procedure and simply determined the binding constraints with the help of numerical solutions.

Multidimensional screening problems are numerically challenging to solve, as discussed in Judd and Su (2006). We also encountered some difficulties with the solvability of the problem, as the matrix including the constraints with opposite self-selection constraints is very close to being singular. These difficulties were met during the sensitivity analysis; with some combinations of parameters the problem was not solvable, or gave irrational values.

Parameterization

The distribution of the economy was chosen to be uniform merely for the simplicity and comparability for the following cases. The discount factors, δ^L and δ^H are set at 0.6 and 0.8, while δ^g is set equal to 1. These values are in line with e.g. Cremer et al. (2006) doing similar numerical calculations. The myopic cases were modelled by assuming that $\delta^L = 0$.[13] The wage rates reflecting the productivities were chosen to be 2 and 3 respectively. The magnitude of the wage rates was determined by the solvability of the problem; with these wage rates, the labour supply decision, restricted to lie between 0 and 1 were at reasonable levels.

Sensitivity analysis

We have considered the sensitivity of the analysis with help of varying i) the degree of habit formation, which allows us to compare the cases with and without habit effect, and ii) the correlation between the characteristics of the individuals, skill level and discount factor. With some combinations of the parameters there is a failure to find the optimum, but it is due to technical problems in the optimisation algorithm.

[13] Note that myopic behaviour does not necessarily imply 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.

Numerical tables

WELFARIST CASE

	c	x
type 1	0.55	0.66
type 2	0.51	0.72
type 3	0.70	0.83
type 4	0.68	0.91

λ	μ^{12}	μ^{13}	μ^{14}	μ^{21}	μ^{23}	μ^{24}
2.0469	0	0	0	0	0	0
	μ^{31}	μ^{32}	μ^{34}	μ^{41}	μ^{42}	μ^{43}
	0.0502	**0.0759**	0	0	0	**0.0498**

PATERNALISTIC CASE

	c	x
type 1	0.49	0.77
type 2	0.49	0.77
type 3	0.67	0.93
type 4	0.67	0.96

λ	μ^{12}	μ^{13}	μ^{14}	μ^{21}	μ^{23}	μ^{24}
2.2483	**0**	0	0	**0**	0	0
	μ^{31}	μ^{32}	μ^{34}	μ^{41}	μ^{42}	μ^{43}
	0.0741	**0.0741**	0	0	0	**0.0704**

PATERNALISTIC CASE WITH MYOPIA

	c	x
type 1	0.48	0.80
type 2	0.48	0.80
type 3	0.69	0.86
type 4	0.66	0.95

λ	μ^{12}	μ^{13}	μ^{14}	μ^{21}	μ^{23}	μ^{24}
2.2567	**0.0226**	0	0	**0.0226**	0	0
	μ^{31}	μ^{32}	μ^{34}	μ^{41}	μ^{42}	μ^{43}
	0.0816	**0.0363**	0	**0**	**0.0453**	**0.0226**