



THE UNIVERSITY *of York*

*Discussion Papers in Economics*

**No. 12/13**

**Optimal Savings Taxation when Individuals have  
Different CRRA Utility Functions**

**Alan Krause, University of York**

Department of Economics and Related Studies  
University of York  
Heslington  
York, YO10 5DD



# Optimal Savings Taxation when Individuals have Different CRRA Utility Functions

Alan Krause\*  
University of York

7 May 2012

## Abstract

Recent empirical research has found that high-skill individuals tend to be less risk averse than low-skill individuals, which implies that their respective constant relative risk aversion (CRRA) utility functions have different curvature. This paper examines the effects of this form of preference heterogeneity on the classic question of whether taxing savings is desirable when the government also implements optimal nonlinear income taxation. It is shown that taxing or subsidising savings may be optimal, even if labour is separable from consumption in the utility function. Specifically, if the individuals' discount rate is lower (resp. higher) than the market interest rate, it is optimal to tax (resp. subsidise) savings. If the individuals' discount rate is equal to the market interest rate, zero taxation of savings is optimal. This basic relationship holds under both linear and nonlinear taxation of savings.

*Keywords:* Savings taxation; nonlinear income taxation; preference heterogeneity.

*JEL Classifications:* H21; H24.

---

\*Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, U.K. E-mail: alan.krause@york.ac.uk.

# 1 Introduction

In a pioneering study, Atkinson and Stiglitz (1976) show that commodity taxation is redundant if individuals share the same preferences, labour is separable from commodities in the utility function, and the government can implement optimal nonlinear income taxation.<sup>1</sup> If different commodities are interpreted as consumption in different time periods, the Atkinson-Stiglitz result implies that savings should not be taxed under their assumptions. It follows, however, that savings taxation may be desirable if preferences are heterogeneous, and the purpose of this paper is to explore this possibility in greater detail. The type of preference heterogeneity we consider is that in which high-skill individuals are less risk averse than low-skill individuals, which implies greater curvature of the latter type's utility function. Specifically, we assume that high-skill and low-skill individuals have constant relative risk aversion (CRRA) utility functions, with high-skill individuals having a lower coefficient of relative risk aversion. Our analysis is motivated by the empirical work of Dohmen, et al. (2010), who find a strong and significant inverse relationship between cognitive ability and risk aversion. Allowing for this form of preference heterogeneity in an otherwise standard optimal tax model,<sup>2</sup> we show that taxing or subsidising savings may be optimal, while maintaining the assumptions that labour is separable from consumption in the utility function and that the government uses optimal nonlinear income taxation. Thus without preference heterogeneity in our model, the Atkinson-Stiglitz result of zero savings taxation would apply.

Although our optimal tax results are generally ambiguous, we show that savings should be taxed or subsidised depending upon a simple relationship between the individuals' discount rate and the market interest rate: if the individuals' discount rate is lower (resp. higher) than the market interest rate, it is optimal to tax (resp. subsidise) savings. If the individuals' discount rate is equal to the market interest rate, zero taxation of savings is optimal. This relationship is shown to hold under both linear and

---

<sup>1</sup>See also Mirrlees (1976) and Christiansen (1984) which are other early papers that examine optimal commodity and income taxation.

<sup>2</sup>That is, we use the workhorse nonlinear income tax model of Mirrlees (1971) and Stiglitz (1982), extended to two periods to allow for savings taxation.

nonlinear taxation of savings. The intuition for the cases of linear and nonlinear savings taxation is different and is explained in detail below, but the basic reasoning common to both tax systems can be summarised as follows. If the individuals' discount rate is lower than the market interest rate, then both high-skill and low-skill individuals have a relatively strong preference for savings. However, since high-skill individuals have less curvature in their utility function, their marginal utility of savings is higher, which translates into a stronger preference for savings. Then by the arguments in the existing literature for taxing commodities preferred by the higher-skilled, taxing savings is optimal. By a reverse argument, subsidising savings is optimal if the individuals' discount rate is higher than the market interest rate, and zero taxation of savings is optimal if the individuals' discount rate and the market interest rate are equal.

To the best of our knowledge, no other papers consider the implications of heterogeneous CRRA preferences for using savings taxation alongside nonlinear income taxation. Golosov, et al. (2010) and Diamond and Spinnewijn (2011) examine the desirability of using savings taxation with nonlinear income taxation when high-skill individuals have higher discount factors, or equivalently a stronger preference for savings, than low-skill individuals. Their analyses reflect empirical evidence that suggests high-skill individuals are more patient than low-skill individuals. Tenhunen and Tuomala (2010) also examine savings taxation in a dynamic nonlinear income tax model when individuals are distinguished by their skills and discount factors. These papers, as well as ours, build on the literature which examines how commodity taxation can usefully complement income taxation under general differences in preferences (see, e.g., Saez (2002), Blomquist and Christiansen (2008), and Kaplow (2008), among others). Our paper is also related to the literature that incorporates behavioural information on preferences (such as positional concerns) into optimal nonlinear income tax models. Examples of this literature include Oswald (1983), Ireland (2001), Aronsson and Johansson-Stenman (2008, 2010), and Guo and Krause (2011).

The remainder of the paper is organised as follows. Section 2 presents our model and assumptions. Section 3 examines optimal linear taxation of savings, while Section 4 examines optimal nonlinear taxation of savings. Section 5 conducts some numerical

simulations, while Section 6 discusses the effects of extending our model to also incorporate heterogeneity in the individuals' discount factors. Section 7 concludes, while proofs and some other mathematical details are contained in an appendix.

## 2 Preliminaries

We consider a two-period model in which individuals work and save in period 1, and then live-off their first-period savings in period 2. Individuals do not work in period 2, which can be thought of as the retirement period. The size of the population is normalised to one, with a proportion  $\phi \in (0, 1)$  of the population being high-skill individuals and the remaining  $(1 - \phi)$  being low-skill individuals. The wages of the high-skill and low-skill individuals are denoted by  $w_H$  and  $w_L$  respectively, where  $w_H > w_L$ . Under the common assumption of a linear production technology, wages are fixed.<sup>3</sup>

High-skill and low-skill individuals are also distinguished by their preferences. The utility function of every high-skill individual is:

$$\frac{(c_H^1)^{1-\alpha}}{1-\alpha} - v(l_H) + \delta \frac{(c_H^2)^{1-\alpha}}{1-\alpha} \quad (2.1)$$

while that for every low-skill individual is:

$$\frac{(c_L^1)^{1-\sigma}}{1-\sigma} - v(l_L) + \delta \frac{(c_L^2)^{1-\sigma}}{1-\sigma} \quad (2.2)$$

where  $c_i^t$  denotes type  $i$ 's consumption in period  $t$ ,  $l_i$  denotes type  $i$ 's labour supply,  $\delta = \frac{1}{1+\rho}$  where  $\delta \in (0, 1)$  is both types discount factor and  $\rho > 0$  is both types discount rate, and the function  $v(\cdot)$  is increasing and convex. Since individuals do not work in period 2,  $c_i^2 = (1 + r)s_i$  where  $s_i$  is type  $i$ 's savings and  $r > 0$  is the market interest rate, which is fixed by the assumption of a linear production technology. It can be seen that the utility functions are of the CRRA form, where  $\alpha > 0$  is the high-skill type's

---

<sup>3</sup>If wages are endogenously determined, the Atkinson-Stiglitz zero commodity taxation result may no longer hold. See, e.g., Naito (1999) and Blackorby and Brett (2004). Cremer, et al. (2001) also show that the Atkinson-Stiglitz result may no longer hold in a multi-dimensional screening setting, in which individuals are distinguished by their skills and wealth endowments. See also Cremer, et al. (2003).

coefficient of relative risk aversion, while  $\sigma > 0$  is that for the low-skill type.<sup>4</sup> Based on the empirical evidence presented in Dohmen, et al. (2010), we assume that  $\sigma > \alpha$  which reflects their finding that low-skill individuals are more risk averse than high-skill individuals.<sup>5</sup>

Finally, we assume that the parameters of the model are such that  $c_i^t > 1$  in order to avoid the problem that  $(c_i^t)^\sigma \gtrless (c_i^t)^\alpha$  depending upon whether  $c_i^t$  is greater than, equal to, or less than one. While such an assumption is unusual, it may be viewed as a slight strengthening of the standard assumption that consumption must be positive, and it ensures that our results are driven by economic factors rather than the possible choice of units of measurement.

### 3 Optimal Linear Savings Taxation

Suppose the individuals face linear taxation (or subsidisation) of their savings and non-linear income taxation. Following the standard practice, we associate a nonlinear income tax system with two tax contracts  $\langle y_L, m_L \rangle$  and  $\langle y_H, m_H \rangle$  where  $y_i = w_i l_i$  is type  $i$ 's pre-tax income and  $m_i$  is type  $i$ 's post-tax income. Therefore,  $y_i - m_i$  is equal to the income tax paid by a type  $i$  individual.

Each individual  $i$  chooses  $c_i^1$ ,  $s_i$ , and their tax contract  $\langle y_i, m_i \rangle$  (which in turn determines their labour supply) to maximise their utility:

$$\frac{(c_i^1)^{1-z}}{1-z} - v\left(\frac{y_i}{w_i}\right) + \delta \frac{((1+r)s_i)^{1-z}}{1-z} \quad (3.1)$$

subject to their budget constraint:

$$c_i^1 + qs_i \leq m_i \quad (3.2)$$

---

<sup>4</sup>If  $\alpha$  or  $\sigma$  happen to be equal to one, the utility function becomes logarithmic in consumption.

<sup>5</sup>We assume a perfect correlation between skills and preferences to avoid the difficulties involved in solving multi-dimensional screening problems. Golosov, et al. (2010) also make this simplifying assumption. Diamond and Spinnewijn (2011) allow for an imperfect correlation, but they assume that taxation is based on skill differences alone. Tenhunen and Tuomala (2010) do address the multi-dimensional screening problem, predominantly using numerical methods.

where  $z = \alpha$  for high-skill individuals and  $z = \sigma$  for low-skill individuals, and  $q = 1 + \tau$  where  $\tau$  is the tax on savings (with a negative tax being a subsidy). The solution to programme (3.1) – (3.2) yields type  $i$ 's indirect utility function  $V_i(q, m_i, \frac{y_i}{w_i})$  and their demand for savings  $s_i(q, m_i)$ .<sup>6</sup>

The government chooses the tax on savings and the nonlinear income tax system to maximise social welfare. That is, the government chooses  $\tau$ ,  $\langle y_L, m_L \rangle$ , and  $\langle y_H, m_H \rangle$  to maximise the utilitarian social welfare function:

$$(1 - \phi)V_L(q, m_L, \frac{y_L}{w_L}) + \phi V_H(q, m_H, \frac{y_H}{w_H}) \quad (3.3)$$

subject to:

$$\tau [(1 - \phi)s_L(q, m_L) + \phi s_H(q, m_H)] + (1 - \phi)(y_L - m_L) + \phi(y_H - m_H) \geq 0 \quad (3.4)$$

$$V_H(q, m_H, \frac{y_H}{w_H}) \geq V_H(q, m_L, \frac{y_L}{w_H}) \quad (3.5)$$

where equation (3.4) is the government's budget constraint, and equation (3.5) is the high-skill type's incentive-compatibility constraint. We omit the low-skill type's incentive-compatibility constraint, as we focus on the case normally studied in the literature in which the redistributive aims of the government imply that high-skill individuals have an incentive to mimic low-skill individuals, but not vice versa. Accordingly, the high-skill type's incentive-compatibility constraint will bind at an optimum, whereas the low-skill type's incentive-compatibility constraint will be slack.

It is shown in the Appendix that the solutions to programmes (3.1) – (3.2) and (3.3) – (3.5) yield the follow result:

**Proposition 1** *Under optimal linear taxation of savings and optimal nonlinear taxation of labour income when high-skill individuals are less risk averse than low-skill individuals:*

(i) if  $\rho = r$ , then  $\tau = 0$ , (ii) if  $\rho < r$ , then  $\tau > 0$ , and (iii) if  $\rho > r$ , then  $\tau < 0$ .

Therefore, the sign of the optimal linear tax applicable to savings is generally am-

---

<sup>6</sup>The assumption that labour is separable from consumption in the utility function implies that the demand for savings is independent of labour supply.



biguous, but it is determined by a simple relationship between the individuals' discount rate and the market interest rate. If  $\rho < r$  so that the individuals' discount rate is lower than the market interest rate, both types of individual are relatively patient and therefore both have a relatively strong preference for savings. But since the high-skill type's utility function has less curvature, their marginal utility of savings is higher, so they save more. Specifically, a 'mimicking' high-skill individual, i.e., a high-skill individual who chooses the tax treatment intended for low-skill individuals, saves more than the truly low-skilled even though both have the same income. In this case, a tax on savings hurts the mimicker more than the truly low-skilled. Taxing savings therefore deters mimicking behaviour or, equivalently, relaxes the incentive-compatibility constraint, which enables the government to increase social welfare. By a reverse argument, subsidising savings is optimal if  $\rho > r$ , because in this case a mimicking high-skill individual saves less than a low-skill individual. If  $\rho = r$ , zero taxation of savings is optimal, because the mimicker and low-skill individuals have the same savings; thus taxation or subsidisation of savings cannot help relax the incentive-compatibility constraint.

## 4 Optimal Nonlinear Savings Taxation

Consider first individual behaviour in the absence of taxation. Individual  $i$  will choose  $c_i^1$ ,  $s_i$ ,  $l_i$ , and  $c_i^2$  to maximise utility:

$$\frac{(c_i^1)^{1-z}}{1-z} - v(l_i) + \delta \frac{(c_i^2)^{1-z}}{1-z} \quad (4.1)$$

subject to the budget constraints:

$$c_i^1 + s_i \leq w_i l_i \quad (4.2)$$

$$c_i^2 \leq (1+r)s_i \quad (4.3)$$

where, as before,  $z = \alpha$  for high-skill individuals and  $z = \sigma$  for low-skill individuals. It is shown in the Appendix that the first-order conditions corresponding to programme

(4.1) – (4.3) can be manipulated to yield the marginal condition:

$$1 = \frac{(c_i^2)^z}{\delta(1+r)(c_i^1)^z} \quad (4.4)$$

Now suppose that individuals face nonlinear taxation of their savings and income. As it may be optimal for the government to set taxes in a manner such that condition (4.4) is violated, one may define the ‘tax wedge’ or ‘implicit marginal tax rate’ applicable to type  $i$ ’s savings as:

$$MTRS_i := 1 - \frac{(c_i^2)^z}{\delta(1+r)(c_i^1)^z} \quad (4.5)$$

where  $MTRS_i$  denotes the marginal tax rate on savings faced by type  $i$  individuals.

Optimal nonlinear taxation of savings and income can be modelled as the government choosing allocations  $\langle y_L, m_L, s_L \rangle$  and  $\langle y_H, m_H, s_H \rangle$  for the low-skill and high-skill individuals, respectively, to maximise the utilitarian social welfare function:

$$(1-\phi) \left[ \frac{(m_L - s_L)^{1-\sigma}}{1-\sigma} - v\left(\frac{y_L}{w_L}\right) + \delta \frac{((1+r)s_L)^{1-\sigma}}{1-\sigma} \right] + \phi \left[ \frac{(m_H - s_H)^{1-\alpha}}{1-\alpha} - v\left(\frac{y_H}{w_H}\right) + \delta \frac{((1+r)s_H)^{1-\alpha}}{1-\alpha} \right] \quad (4.6)$$

subject to:

$$(1-\phi)(y_L - m_L) + \phi(y_H - m_H) \geq 0 \quad (4.7)$$

$$\frac{(m_H - s_H)^{1-\alpha}}{1-\alpha} - v\left(\frac{y_H}{w_H}\right) + \delta \frac{((1+r)s_H)^{1-\alpha}}{1-\alpha} \geq \frac{(m_L - s_L)^{1-\alpha}}{1-\alpha} - v\left(\frac{y_L}{w_H}\right) + \delta \frac{((1+r)s_L)^{1-\alpha}}{1-\alpha} \quad (4.8)$$

where  $c_i^1 = m_i - s_i$ , equation (4.7) is the government’s budget constraint, and equation (4.8) is the high-skill type’s incentive-compatibility constraint.<sup>7</sup>

It is shown in the Appendix that the solution to programme (4.6) – (4.8) yields:

**Proposition 2** *Under optimal nonlinear taxation of savings and income when high-skill individuals are less risk averse than low-skill individuals: (i) if  $\rho = r$ , then  $MTRS_L = 0$ , (ii) if  $\rho < r$ , then  $MTRS_L > 0$ , (iii) if  $\rho > r$ , then  $MTRS_L < 0$ , and (iv)  $MTRS_H = 0$ .*

The finding that high-skill individuals face a zero marginal tax rate on their savings is simply an example of the ‘no-distortion-at-the-top’ result that is common in optimal

---

<sup>7</sup>We again omit the low-skill type’s incentive-compatibility constraint, because by assumption it will not be binding at an optimum.

nonlinear income tax models. The sign of the marginal tax rate applicable to the low-skill type's savings is generally ambiguous, but like under linear savings taxation it is determined by a simple relationship between the individuals' discount rate and the market interest rate. However, the intuition is now a little different. If  $\rho < r$ , then all individuals have a relatively strong preference for savings. But unlike under linear savings taxation, low-skill individuals and a mimicking high-skill individual have the exact same savings, as both choose the low-skill type's allocation. However, a positive marginal tax rate on the low-skill type's savings hurts the mimicker more, because the mimicker has a higher marginal utility of savings owing to less curvature in their utility function. Thus distorting the low-skill type's savings downwards through a positive marginal tax rate deters mimicking behaviour and relaxes the incentive-compatibility constraint. Analogously, if  $\rho > r$  then distorting the low-skill type's savings upwards through a negative marginal tax rate is optimal to deter mimicking, and if  $\rho = r$  a zero marginal tax rate on the low-skill type's savings is optimal since, in this case, distorting savings cannot help deter mimicking.

## 5 Numerical Simulations

To further expose the effects of heterogeneous CRRA preferences on optimal savings taxation, we conduct some numerical simulations. The parameters of the model are chosen on the following basis. Chetty (2006) concludes that a reasonable estimate of the coefficient of relative risk aversion is one (log utility). We therefore take this as our starting point, and consider parameter values such that  $\sigma > 1 > \alpha$ , with arbitrarily chosen baseline values of  $\sigma = 1.1$  and  $\alpha = 0.9$ . For simplicity we assume that  $v(l_i) = l_i$ , so that utility is quasi-linear in labour. The OECD (2011) reports that 30% of adults across OECD countries have attained tertiary level education. Accordingly, we assume that 30% of the population are high-skill individuals, i.e.,  $\phi = 0.3$ . Fang (2006) and Goldin and Katz (2007) estimate that the college wage premium is around 60%. We therefore

assume that  $w_L = 2.5$  and  $w_H = 4.0$ .<sup>8</sup> Following common practice, we assume an annual market interest rate of 4% ( $r = 0.04$ ), and we consider values of the individuals' discount rate such that  $\rho \gtrless 0.04$ . Since we examine a two-period model, we assume that each period is 20 years in length and work with the twenty-year analogs of the annual market interest rate and the individuals' discount rate. The baseline parameter values that we use in the numerical simulations are reported in Table 1.

Figure 1 illustrates the effects of changes in some key parameters on the optimal linear savings tax rate  $\tau$ , while holding all other parameters at their baseline levels. Simulations are conducted for  $\rho < r$  ( $\rho = 0.0395$ ),  $\rho = r$  ( $\rho = 0.04$ ), and  $\rho > r$  ( $\rho = 0.0405$ ).<sup>9</sup> The top panel of Figure 1 confirms Proposition 1: if  $\rho < r$  then  $\tau$  is positive and increasing in the degree of preference heterogeneity; if  $\rho > r$  then  $\tau$  is negative and decreasing in the degree of preference heterogeneity; and if  $\rho = r$  then  $\tau$  is always equal to zero. The middle panel of Figure 1 shows the effects of increasing the degree of wage inequality. As wage inequality increases, it becomes optimal for the government to engage in more redistribution, which strengthens the incentive high-skill individuals have to mimic low-skill individuals. Accordingly, when  $\rho \neq r$ , it is optimal to move  $\tau$  further away from zero as savings taxation/subsidisation is used more intensely to deter mimicking. The bottom panel of Figure 1 shows the effects of changing the proportion of high-skill individuals  $\phi$  in the economy. An increase in  $\phi$  has two opposing effects. On the one hand, an increase in  $\phi$  means that high-skill individuals obtain more weight in the social welfare function, which reduces the extent of redistribution that the government wants to undertake. On the other hand, an increase in  $\phi$  raises the economy's endowments and makes redistribution less costly, thus strengthening the government's desire to redistribute. For the empirically-plausible parameter values that we consider, the second effect dominates. Therefore, when  $\rho \neq r$ , the optimal level of  $\tau$  moves further away from zero as  $\phi$  increases, in order to facilitate redistribution by

---

<sup>8</sup>It would be more natural to normalise the low-skill type's wage to unity and correspondingly set the high-skill type's wage equal to 1.6. However, we choose  $w_L = 2.5$  and  $w_H = 4.0$  to ensure that  $c_i^t > 1$  in all of our numerical simulations.

<sup>9</sup>We consider small differences in the annual market interest rate and the individuals' discount rate, because these imply relatively large differences over the 20-year horizon.

detering mimicking.

Figure 2 shows the effects of parameter changes on optimal nonlinear savings taxation; specifically, on the optimal marginal tax rate applicable to the low-skill type's savings. As expected, the simulation results are qualitatively the same as those for linear savings taxation, as is the intuition underlying the results.

## 6 Extension: Heterogeneous Discount Rates

In light of the empirical evidence that suggests high-skill individuals are more patient than low-skill individuals, and the related literature which has focused on this case, we consider an extension of our model to incorporate this second source of preference heterogeneity. We assume that  $\delta_H > \delta_L$ , or equivalently  $\rho_H < \rho_L$ , so that high-skill individuals have a higher discount factor (or lower discount rate) than low-skill individuals.

With this extension, the analogues to Propositions 1 and 2 are:

**Proposition 1a** *Under optimal linear taxation of savings and optimal nonlinear taxation of labour income when high-skill individuals are less risk averse and more patient than low-skill individuals: (i) if  $\rho_L = r$ , then  $\tau > 0$ , (ii) if  $\rho_L < r$ , then  $\tau > 0$ , and (iii) if  $\rho_L > r$ , then  $\tau \geq 0$ .*

**Proposition 2a** *Under optimal nonlinear taxation of savings and income when high-skill individuals are less risk averse and more patient than low-skill individuals: (i) if  $\rho_L = r$ , then  $MTRS_L > 0$ , (ii) if  $\rho_L < r$ , then  $MTRS_L > 0$ , (iii) if  $\rho_L > r$ , then  $MTRS_L \geq 0$ , and (iv)  $MTRS_H = 0$ .*

The proofs of Propositions 1a and 2a are provided in the Appendix, but the intuition underlying these results is fairly straightforward. If high-skill individuals are more patient than low-skill individuals, then the former always have a stronger preference for savings. It therefore becomes optimal, *ceteris paribus*, to tax savings. Thus in Propositions 1a and 2a for each possible relationship between  $\rho_L$  and  $r$ ,<sup>10</sup> the optimal tax on savings is simply increased relative to that in Propositions 1 and 2 for the corresponding

---

<sup>10</sup>One could rewrite Propositions 1a and 2a in terms of the relationship between  $\rho_H$  and  $r$  by using the fact that  $\rho_H = \beta\rho_L$  for some  $\beta \in (0, 1)$ .

relationship between  $\rho$  and  $r$ .<sup>11</sup> Therefore, the intuition behind Propositions 1a and 2a is the same as that behind Propositions 1 and 2, except that the stronger preference high-skill individuals now have for savings strengthens the case for taxing savings.

## 7 Conclusion

Following the pioneering work of Atkinson and Stiglitz (1976), an extensive literature has arisen which examines the robustness of their zero commodity taxation result. This paper joins this literature by considering the effects of individuals having different CRRA utility functions, which is motivated by empirical evidence that suggests high-skill individuals are less risk averse than low-skill individuals. Our main result is that savings taxation or subsidisation may now be optimal, depending upon a simple relationship between the individuals' discount rate and the market interest rate. Interestingly, this simple relationship holds under both linear and nonlinear taxation of savings.

A number of potential extensions of our paper come to mind. As is common in the literature, we have assumed a linear production technology which fixes wages and the market interest rate. Since our results depend upon the relationship between the individuals' discount rate and the market interest rate, it may be interesting to extend our model to a setting in which the market interest rate is endogenous. Likewise, to maintain consistency with the existing literature, we have assumed that there is no uncertainty. Given our focus on heterogeneous degrees of risk aversion, an extension incorporating uncertainty might be worth pursuing. Another possible extension would be to relax the assumption that there is a perfect correlation between skills and preferences. However, optimal nonlinear taxation would then become a multi-dimensional screening problem, and such problems can be very difficult to solve for clear-cut results.

---

<sup>11</sup>The only exception is the marginal tax rate applicable to the high-skill type's savings under nonlinear savings taxation, in which the 'no-distortion-at-the-top' result remains intact.

## 8 Appendix

### A.1 Proof of Proposition 1

The Lagrangian corresponding to programme (3.1) – (3.2) is:

$$\mathcal{L} = \frac{(c_i^1)^{1-z}}{1-z} - v\left(\frac{y_i}{w_i}\right) + \delta \frac{((1+r)s_i)^{1-z}}{1-z} + \lambda_i [m_i - c_i^1 - qs_i] \quad (\text{A.1})$$

where  $\lambda_i > 0$  is the multiplier on individual  $i$ 's budget constraint. The first-order conditions are:

$$(c_i^1)^{-z} - \lambda_i = 0 \quad (\text{A.2})$$

$$\delta((1+r)s_i)^{-z}(1+r) - \lambda_i q = 0 \quad (\text{A.3})$$

$$m_i - c_i^1 - qs_i = 0 \quad (\text{A.4})$$

These first-order conditions can be manipulated to yield:

$$s_i = \frac{m_i}{\left(\frac{q}{\delta}\right)^{\frac{1}{z}} (1+r)^{\frac{z-1}{z}} + q} \quad \text{and} \quad \hat{s}_H = \frac{m_L}{\left(\frac{q}{\delta}\right)^{\frac{1}{\alpha}} (1+r)^{\frac{\alpha-1}{\alpha}} + q} \quad (\text{A.5})$$

where  $\hat{s}_H$  denotes the savings of a mimicking high-skill individual.

For use below, note that by the Envelope Theorem we obtain:

$$\frac{\partial \hat{V}_H(\cdot)}{\partial q} = \frac{\partial \mathcal{L}(\cdot)}{\partial q} = -\hat{\lambda}_H \hat{s}_H \quad \text{and} \quad \frac{\partial \hat{V}_H(\cdot)}{\partial m_L} = \frac{\partial \mathcal{L}(\cdot)}{\partial m_L} = \hat{\lambda}_H \quad (\text{A.6})$$

where  $\hat{V}_H(q, m_L, \frac{y_L}{w_H})$  is the mimicker's indirect utility function, and  $\hat{\lambda}_H > 0$  is the multiplier on the mimicker's budget constraint. Also, equation (A.5) implies that:

$$m_L \left[ \frac{1}{s_L} - \frac{1}{\hat{s}_H} \right] = \left( \frac{q}{\delta} \right)^{\frac{1}{\sigma}} (1+r)^{\frac{\sigma-1}{\sigma}} - \left( \frac{q}{\delta} \right)^{\frac{1}{\alpha}} (1+r)^{\frac{\alpha-1}{\alpha}} \quad (\text{A.7})$$

The first-order conditions on  $\tau$ ,  $m_L$ , and  $m_H$  in programme (3.3) – (3.5) are:

$$(1-\phi) \frac{\partial V_L(\cdot)}{\partial q} + \phi \frac{\partial V_H(\cdot)}{\partial q} + \lambda [(1-\phi)s_L + \phi s_H] + \lambda \tau \left[ (1-\phi) \frac{\partial s_L(\cdot)}{\partial q} + \phi \frac{\partial s_H(\cdot)}{\partial q} \right] + \theta_H \left[ \frac{\partial V_H(\cdot)}{\partial q} - \frac{\partial \hat{V}_H(\cdot)}{\partial q} \right] = 0 \quad (\text{A.8})$$

$$(1 - \phi) \frac{\partial V_L(\cdot)}{\partial m_L} + \lambda \tau (1 - \phi) \frac{\partial s_L(\cdot)}{\partial m_L} - \lambda (1 - \phi) - \theta_H \frac{\partial \widehat{V}_H(\cdot)}{\partial m_L} = 0 \quad (\text{A.9})$$

$$\phi \frac{\partial V_H(\cdot)}{\partial m_H} + \lambda \tau \phi \frac{\partial s_H(\cdot)}{\partial m_H} - \lambda \phi + \theta_H \frac{\partial V_H(\cdot)}{\partial m_H} = 0 \quad (\text{A.10})$$

where  $\lambda > 0$  is the multiplier on the government's budget constraint (3.4), and  $\theta_H > 0$  is the multiplier on the incentive-compatibility constraint (3.5).

Adding equations (A.8), (A.9), and (A.10) after first multiplying equation (A.9) by  $s_L$  and equation (A.10) by  $s_H$  yields the following expression for the optimal linear tax on savings:

$$\tau = \frac{\theta_H \widehat{\lambda}_H [s_L - \widehat{s}_H]}{\lambda \left[ (1 - \phi) \frac{\partial h_L(\cdot)}{\partial q} + \phi \frac{\partial h_H(\cdot)}{\partial q} \right]} \quad (\text{A.11})$$

where use has been made of equation (A.6), Roy's identity, and the Slutsky equation in deriving equation (A.11), where  $h_i(q, u_i)$  is type  $i$ 's Hicksian (or compensated) demand for savings, with  $u_i$  denoting type  $i$ 's utility. Standard results in consumer theory ensure that  $\partial h_i(\cdot)/\partial q < 0$ , which implies that the denominator in equation (A.11) is negative.

To show that  $\rho = r \implies \tau = 0$ , we proceed by contradiction. Suppose that  $\rho = r$  and  $\tau > 0$ . This is equivalent to having  $\delta(1 + r) = 1$  and  $q > 1$ , which using equation (A.7) implies:

$$m_L \left[ \frac{1}{s_L} - \frac{1}{\widehat{s}_H} \right] = (1 + r) \left[ q^{\frac{1}{\sigma}} - q^{\frac{1}{\alpha}} \right] < 0 \quad (\text{A.12})$$

and therefore  $s_L > \widehat{s}_H$ . Using equation (A.11),  $s_L > \widehat{s}_H$  implies  $\tau < 0$ , yielding a contradiction. Analogously, suppose that  $\rho = r$  and  $\tau < 0$ . This is equivalent to having  $\delta(1 + r) = 1$  and  $q < 1$ , which using equation (A.7) implies:

$$m_L \left[ \frac{1}{s_L} - \frac{1}{\widehat{s}_H} \right] = (1 + r) \left[ q^{\frac{1}{\sigma}} - q^{\frac{1}{\alpha}} \right] > 0 \quad (\text{A.13})$$

and therefore  $s_L < \widehat{s}_H$ . Using equation (A.11),  $s_L < \widehat{s}_H$  implies  $\tau > 0$ , yielding a contradiction.

To show that  $\rho < r \implies \tau > 0$ , we proceed by contradiction. Suppose that  $\rho < r$  and  $\tau \leq 0$ . This is equivalent to having  $\delta(1 + r)a = 1$ , where  $a \in (0, 1)$ , and  $q \leq 1$ , which



using equation (A.7) implies:

$$m_L \left[ \frac{1}{s_L} - \frac{1}{\widehat{s}_H} \right] = (1+r) \left[ (qa)^{\frac{1}{\sigma}} - (qa)^{\frac{1}{\alpha}} \right] > 0 \quad (\text{A.14})$$

and therefore  $s_L < \widehat{s}_H$ . Using equation (A.11),  $s_L < \widehat{s}_H$  implies  $\tau > 0$ , yielding a contradiction.

To show that  $\rho > r \implies \tau < 0$ , we proceed by contradiction. Suppose that  $\rho > r$  and  $\tau \geq 0$ . This is equivalent to having  $\delta(1+r)b = 1$ , where  $b > 1$ , and  $q \geq 1$ , which using equation (A.7) implies:

$$m_L \left[ \frac{1}{s_L} - \frac{1}{\widehat{s}_H} \right] = (1+r) \left[ (qb)^{\frac{1}{\sigma}} - (qb)^{\frac{1}{\alpha}} \right] < 0 \quad (\text{A.15})$$

and therefore  $s_L > \widehat{s}_H$ . Using equation (A.11),  $s_L > \widehat{s}_H$  implies  $\tau < 0$ , yielding a contradiction. ■

## A.2 Derivation of Equation (4.4)

The relevant first-order conditions corresponding to programme (4.1) – (4.3) are:

$$(c_i^1)^{-z} - \lambda_i^1 = 0 \quad (\text{A.16})$$

$$-\lambda_i^1 + \lambda_i^2(1+r) = 0 \quad (\text{A.17})$$

$$\delta(c_i^2)^{-z} - \lambda_i^2 = 0 \quad (\text{A.18})$$

where  $\lambda_i^1 > 0$  and  $\lambda_i^2 > 0$  are the multipliers on constraints (4.2) and (4.3), respectively.

Straightforward manipulation of equations (A.16) – (A.18) yields equation (4.4). ■

## A.3 Proof of Proposition 2

The first-order condition on  $s_L$  in programme (4.6) – (4.8) is:

$$(1-\phi) \left[ \delta(1+r)(c_L^2)^{-\sigma} - (c_L^1)^{-\sigma} \right] - \theta_H \left[ \delta(1+r)(c_L^2)^{-\alpha} - (c_L^1)^{-\alpha} \right] = 0 \quad (\text{A.19})$$

where  $\theta_H > 0$  is the multiplier on the incentive-compatibility constraint (4.8).

Equation (A.19) can be manipulated to yield:

$$\frac{(c_L^2)^\sigma}{(c_L^1)^\sigma} = \delta(1+r) \times \frac{1 - \phi - \theta_H(c_L^2)^{\sigma-\alpha}}{1 - \phi - \theta_H(c_L^1)^{\sigma-\alpha}} \quad (\text{A.20})$$

or using equation (4.5):

$$MTRS_L := 1 - \frac{(c_L^2)^\sigma}{\delta(1+r)(c_L^1)^\sigma} = \frac{(c_L^2)^\sigma}{\delta(1+r)(c_L^1)^\sigma} \left[ \frac{1 - \phi - \theta_H(c_L^1)^{\sigma-\alpha}}{1 - \phi - \theta_H(c_L^2)^{\sigma-\alpha}} - 1 \right] \quad (\text{A.21})$$

which implies that  $MTRS_L \geq 0$  iff  $c_L^2 \geq c_L^1$ .

To show that  $\rho = r \implies MTRS_L = 0$ , we proceed by contradiction. Suppose that  $\rho = r$  and  $MTRS_L > 0$ . This is equivalent to having  $\delta(1+r) = 1$  and  $c_L^2 > c_L^1$ , which using equation (A.20) implies that  $1 - \phi - \theta_H(c_L^2)^{\sigma-\alpha} > 1 - \phi - \theta_H(c_L^1)^{\sigma-\alpha}$  or  $c_L^2 < c_L^1$  which yields a contradiction. Analogously, suppose that  $\rho = r$  and  $MTRS_L < 0$ . This is equivalent to having  $\delta(1+r) = 1$  and  $c_L^2 < c_L^1$ , which using equation (A.20) implies that  $1 - \phi - \theta_H(c_L^2)^{\sigma-\alpha} < 1 - \phi - \theta_H(c_L^1)^{\sigma-\alpha}$  or  $c_L^2 > c_L^1$  which yields a contradiction.

To show that  $\rho < r \implies MTRS_L > 0$ , we proceed by contradiction. Suppose that  $\rho < r$  and  $MTRS_L \leq 0$ . This is equivalent to having  $\delta(1+r) > 1$  and  $c_L^2 \leq c_L^1$ , which using equation (A.20) implies that  $1 - \phi - \theta_H(c_L^2)^{\sigma-\alpha} < 1 - \phi - \theta_H(c_L^1)^{\sigma-\alpha}$  or  $c_L^2 > c_L^1$  which yields a contradiction.

To show that  $\rho > r \implies MTRS_L < 0$ , we proceed by contradiction. Suppose that  $\rho > r$  and  $MTRS_L \geq 0$ . This is equivalent to having  $\delta(1+r) < 1$  and  $c_L^2 \geq c_L^1$ , which using equation (A.20) implies that  $1 - \phi - \theta_H(c_L^2)^{\sigma-\alpha} > 1 - \phi - \theta_H(c_L^1)^{\sigma-\alpha}$  or  $c_L^2 < c_L^1$  which yields a contradiction.

Finally, the first-order condition on  $s_H$  in programme (4.6) – (4.8) is:

$$(\phi + \theta_H) [\delta(1+r)(c_H^2)^{-\alpha} - (c_H^1)^{-\alpha}] = 0 \quad (\text{A.22})$$

which using equation (4.5) implies that  $MTRS_H = 0$ . ■

#### A.4 Proof of Proposition 1a

With heterogeneous discount factors, equation (A.7) becomes:

$$m_L \left[ \frac{1}{s_L} - \frac{1}{\widehat{s}_H} \right] = \left( \frac{q}{\delta_L} \right)^{\frac{1}{\sigma}} (1+r)^{\frac{\sigma-1}{\sigma}} - \left( \frac{q}{\delta_H} \right)^{\frac{1}{\alpha}} (1+r)^{\frac{\alpha-1}{\alpha}} \quad (\text{A.23})$$

To show that  $\rho_L = r \implies \tau > 0$ , we proceed by contradiction. Suppose that  $\rho_L = r$  and  $\tau \leq 0$ . This is equivalent to having  $\delta_L(1+r) = 1$  and  $q \leq 1$ , which using equation (A.23) implies:

$$m_L \left[ \frac{1}{s_L} - \frac{1}{\widehat{s}_H} \right] = (1+r) \left[ q^{\frac{1}{\sigma}} - (qx)^{\frac{1}{\alpha}} \right] > 0 \quad (\text{A.24})$$

where use has been made of the fact that  $\delta_L = x\delta_H$  for some  $x \in (0, 1)$ . Therefore,  $s_L < \widehat{s}_H$  which using equation (A.11) implies that  $\tau > 0$ , yielding a contradiction.

To show that  $\rho_L < r \implies \tau > 0$ , we proceed by contradiction. Suppose that  $\rho_L < r$  and  $\tau \leq 0$ . This is equivalent to having  $\delta_L(1+r)a = 1$ , where  $a \in (0, 1)$ , and  $q \leq 1$ , which using equation (A.23) implies:

$$m_L \left[ \frac{1}{s_L} - \frac{1}{\widehat{s}_H} \right] = (1+r) \left[ (qa)^{\frac{1}{\sigma}} - (qax)^{\frac{1}{\alpha}} \right] > 0 \quad (\text{A.25})$$

where use has been made of the fact that  $\delta_L = x\delta_H$  for some  $x \in (0, 1)$ . Therefore,  $s_L < \widehat{s}_H$  which using equation (A.11) implies that  $\tau > 0$ , yielding a contradiction.

Finally,  $\rho_L > r \implies \tau \gtrless 0$  has been confirmed using numerical examples, details of which are available upon request. ■

## A.5 Proof of Proposition 2a

With heterogeneous discount factors, equations (A.20) and (A.21) become:

$$\frac{(c_L^2)^\sigma}{(c_L^1)^\sigma} = \delta_L(1+r) \times \frac{1 - \phi - \theta_H b (c_L^2)^{\sigma-\alpha}}{1 - \phi - \theta_H (c_L^1)^{\sigma-\alpha}} \quad (\text{A.26})$$

$$MTRS_L := 1 - \frac{(c_L^2)^\sigma}{\delta_L(1+r)(c_L^1)^\sigma} = \frac{(c_L^2)^\sigma}{\delta_L(1+r)(c_L^1)^\sigma} \left[ \frac{1 - \phi - \theta_H (c_L^1)^{\sigma-\alpha}}{1 - \phi - \theta_H b (c_L^2)^{\sigma-\alpha}} - 1 \right] \quad (\text{A.27})$$

where  $\delta_H = b\delta_L$  for some  $b > 1$ .

To show that  $\rho_L = r \implies MTRS_L > 0$ , we proceed by contradiction. Suppose that  $\rho_L = r$  and  $MTRS_L \leq 0$ . This is equivalent to having  $\delta_L(1+r) = 1$  and  $1 -$

$\phi - \theta_H b(c_L^2)^{\sigma-\alpha} \geq 1 - \phi - \theta_H (c_L^1)^{\sigma-\alpha}$ , which implies  $c_L^2 < c_L^1$ . However, using (A.26),  $1 - \phi - \theta_H b(c_L^2)^{\sigma-\alpha} \geq 1 - \phi - \theta_H (c_L^1)^{\sigma-\alpha}$  implies  $c_L^2 \geq c_L^1$ , yielding a contradiction.

To show that  $\rho_L < r \implies MTRS_L > 0$ , we proceed by contradiction. Suppose that  $\rho_L < r$  and  $MTRS_L \leq 0$ . This is equivalent to having  $\delta_L(1+r) > 1$  and  $1 - \phi - \theta_H b(c_L^2)^{\sigma-\alpha} \geq 1 - \phi - \theta_H (c_L^1)^{\sigma-\alpha}$ , which implies  $c_L^2 < c_L^1$ . However, using (A.26),  $1 - \phi - \theta_H b(c_L^2)^{\sigma-\alpha} \geq 1 - \phi - \theta_H (c_L^1)^{\sigma-\alpha}$  implies  $c_L^2 > c_L^1$ , yielding a contradiction.

The result that  $\rho_L > r \implies MTRS_L \geq 0$  has been confirmed using numerical examples, details of which are available upon request.

Finally, with heterogeneous discount factors, the first-order condition on  $s_H$  in programme (4.6) – (4.8) becomes:

$$(\phi + \theta_H) [\delta_H(1+r)(c_H^2)^{-\alpha} - (c_H^1)^{-\alpha}] = 0 \quad (\text{A.28})$$

and equation (4.5) becomes:

$$MTRS_i := 1 - \frac{(c_i^2)^z}{\delta_i(1+r)(c_i^1)^z} \quad (\text{A.29})$$

which implies that  $MTRS_H = 0$ . ■

## References

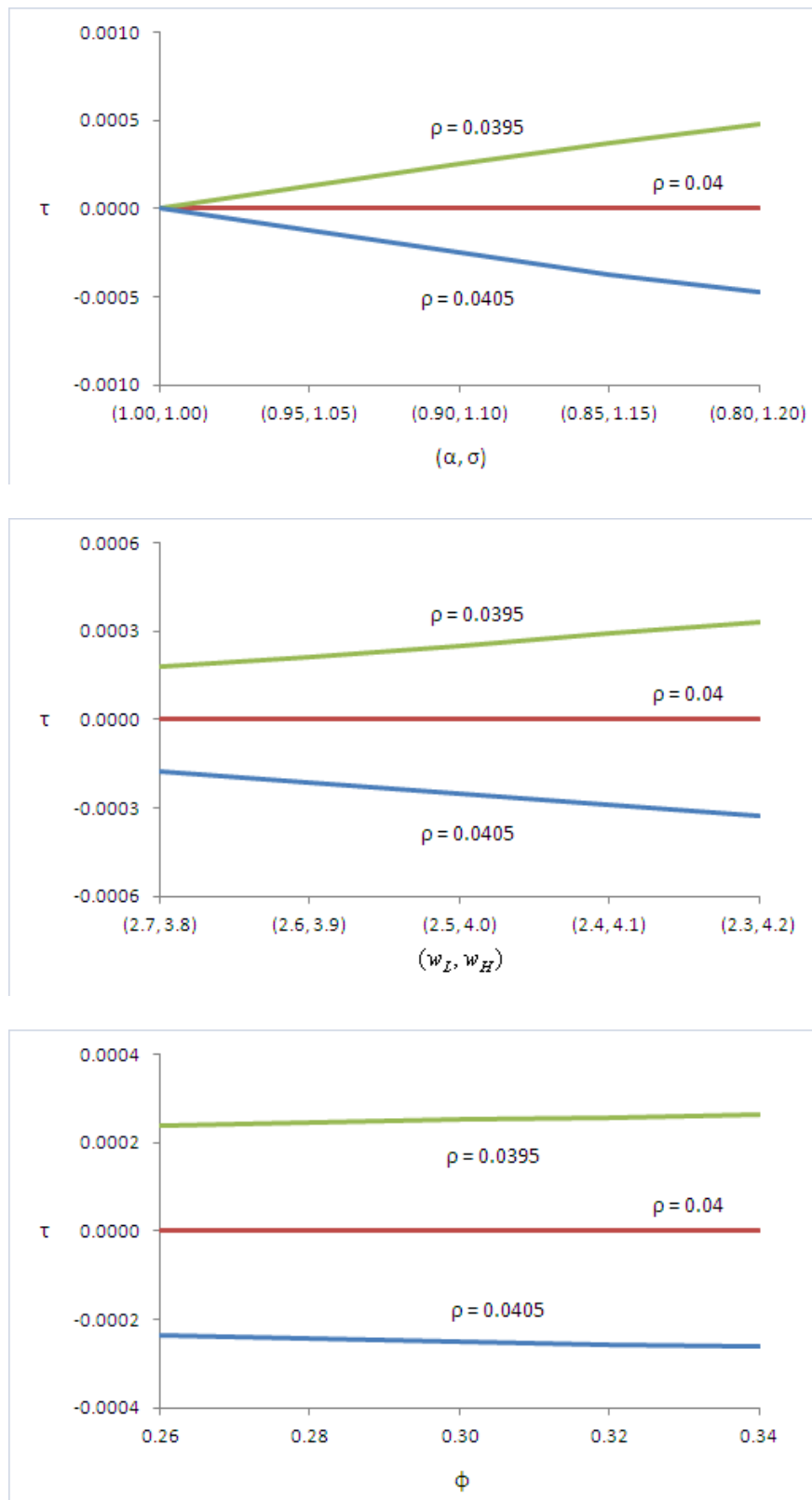
- [1] Aronsson, T. and O. Johansson-Stenman (2008), “When the Joneses’ Consumption Hurts: Optimal Public Good Provision and Nonlinear Income Taxation”, *Journal of Public Economics*, 92, 986-997.
- [2] Aronsson, T. and O. Johansson-Stenman (2010), “Positional Concerns in an OLG Model: Optimal Labor and Capital Income Taxation”, *International Economic Review*, 51, 1071-1095.
- [3] Atkinson, A. and J. Stiglitz (1976), “The Design of Tax Structure: Direct Versus Indirect Taxation”, *Journal of Public Economics*, 6, 55-75.
- [4] Blackorby, C. and C. Brett (2004), “Production Efficiency and the Direct-Indirect Tax Mix”, *Journal of Public Economic Theory*, 6, 165-180.
- [5] Blomquist, S. and V. Christiansen (2008), “Taxation and Heterogeneous Preferences”, *FinanzArchiv*, 64, 218-244.
- [6] Chetty, R. (2006), “A New Method of Estimating Risk Aversion”, *American Economic Review*, 96, 1821-1834.
- [7] Christiansen, V. (1984), “Which Commodity Taxes Should Supplement the Income Tax?”, *Journal of Public Economics*, 24, 195-220.
- [8] Cremer, H., P. Pestieau and J. Rochet (2001), “Direct versus Indirect Taxation: The Design of the Tax Structure Revisited”, *International Economic Review*, 42, 781-799.
- [9] Cremer, H., P. Pestieau and J. Rochet (2003), “Capital Income Taxation when Inherited Wealth is not Observable”, *Journal of Public Economics*, 87, 2475-2490.
- [10] Diamond, P. and J. Spinnewijn (2011), “Capital Income Taxes with Heterogeneous Discount Rates”, *American Economic Journal: Economic Policy*, 3, 52-76.
- [11] Dohmen, T., A. Falk, D. Huffman and U. Sunde (2010), “Are Risk Aversion and Patience Related to Cognitive Ability?”, *American Economic Review*, 100, 1238-1260.
- [12] Fang, H. (2006), “Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices”, *International Economic Review*, 47, 1151-1185.
- [13] Goldin, C. and L. Katz (2007), “The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890-2005”, *NBER Working Paper*, No. 12984.
- [14] Golosov, M., M. Troshkin, A. Tsyvinski and M. Weinzeirl (2010), “Preference Heterogeneity and Optimal Capital Income Taxation”, *mimeo*.

- [15] Guo, J-T. and A. Krause (2011), “Optimal Nonlinear Income Taxation with Habit Formation”, *Journal of Public Economic Theory*, 13, 463-480.
- [16] Ireland, N. (2001), “Optimal Income Tax in the Presence of Status Effects”, *Journal of Public Economics*, 81, 193-212.
- [17] Kaplow, L. (2008), “Optimal Policy with Heterogeneous Preferences”, *B.E. Journal of Economic Analysis and Policy*, 8, Article 40.
- [18] Mirrlees, J. (1971), “An Exploration in the Theory of Optimum Income Taxation”, *Review of Economic Studies*, 38, 175-208.
- [19] Mirrlees, J. (1976), “Optimal Tax Theory: A Synthesis”, *Journal of Public Economics*, 6, 327-358.
- [20] Naito, H. (1999), “Re-examination of Uniform Commodity Taxes Under a Non-linear Income Tax System and its Implication for Production Efficiency”, *Journal of Public Economics*, 71, 165-188.
- [21] OECD (2011), *Highlights from Education at a Glance*, OECD publications.
- [22] Oswald, A. (1983), “Altruism, Jealousy and the Theory of Optimal Non-Linear Taxation”, *Journal of Public Economics*, 20, 77-87.
- [23] Saez, E. (2002), “The Desirability of Commodity Taxation under Non-linear Income Taxation and Heterogeneous Tastes”, *Journal of Public Economics*, 83, 217-230.
- [24] Stiglitz, J. (1982), “Self-Selection and Pareto Efficient Taxation”, *Journal of Public Economics*, 17, 213-240.
- [25] Tenhunen, S. and M. Tuomala (2010), “On Optimal Lifetime Redistribution Policy”, *Journal of Public Economic Theory*, 12, 171-198.

**TABLE 1**  
**Baseline Parameter Values**

$\sigma$	1.10	$\phi$	0.30	$w_L$	2.50
$\alpha$	0.90	$r$	0.04	$w_H$	4.00

**FIGURE 1**  
**Optimal Linear Savings Taxation**





**FIGURE 2**  
**Optimal Nonlinear Savings Taxation**

