Title

Dynamic nonlinear income taxation with quasi-hyperbolic discounting and no commitment

Authors
Guo, Jang-Ting
Krause, Alan

Publication Date
2015

DOI
10.1016/j.jebo.2014.11.002

Peer reviewed

Journal of Economic Behavior & Organization 109 (2015) 101–119

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Dynamic nonlinear income taxation with quasi-hyperbolic discounting and no commitment☆

Jang-Ting Guo[a,1], Alan Krause[b,*]

[a] Department of Economics, University of California, 3133 Sproul Hall, Riverside, CA 92521, USA
[b] Department of Economics and Related Studies, University of York, Heslington, York YO10 5DD, UK

ARTICLE INFO

Article history:
Received 7 January 2014
Received in revised form 21 August 2014
Accepted 5 November 2014
Available online 18 November 2014

JEL classification:
D91
H21
H24

Keywords:
Dynamic taxation
Quasi-hyperbolic discounting
Commitment

ABSTRACT

This paper examines a dynamic model of nonlinear income taxation in which the government cannot commit to its future tax policy, and individuals are quasi-hyperbolic discounters who cannot commit to future consumption plans. The government has both paternalistic and redistributive objectives, and therefore uses its taxation powers to maximize a utilitarian social welfare function that reflects individuals' true (long-run) preferences. Under first-best taxation, quasi-hyperbolic discounting exerts no effect on the level of social welfare attainable. Under second-best taxation, quasi-hyperbolic discounting increases (resp. decreases) the level of social welfare attainable when separating (resp. pooling) taxation is optimal. In stark contrast to previous studies, this result implies that some individuals can actually be better-off in the long run as a result of their short-run impatience.

1. Introduction

The aim of this paper is to examine the effects of incorporating quasi-hyperbolic discounting by individuals into a dynamic model of optimal nonlinear income taxation without commitment. There is by now an extensive empirical and theoretical literature on quasi-hyperbolic discounting, which captures a preference many individuals have for immediate gratification.[2] This leads agents to make short-run decisions that they later regret as not being consistent with their long-run preferences. Such behavior is often described as an individual imposing a negative "internality" on their future self, which potentially justifies corrective (or paternalistic) policy intervention.[3] There is, in effect, preference heterogeneity between individuals and the government, as the government's preferences are the same as the individuals' long-run preferences, but not their short-run counterparts. In our model economy, an individual's need for immediate gratification leads them to make labor,

☆ We thank John Hey for a very helpful discussion regarding naive versus sophisticated quasi-hyperbolic discounting. We also thank Zaifu Yang and seminar participants at the University of York for their comments. We are especially grateful to an associate editor and an anonymous referee of this journal, whose comments and suggestions led to a much improved paper. Any remaining errors are our responsibility.

* Corresponding author. Tel.: +44 01904 323 572; fax: +44 01904 323 759.
E-mail addresses: guojt@ucr.edu (J.-T. Guo), alan.krause@york.ac.uk (A. Krause).

[1] Tel.: +1 951 827 1588; fax: +1 951 827 5685.

[2] See, e.g., the survey article by Frederick et al. (2002).

[3] For example, O'Donoghue and Rabin (1999) examine optimal "sin taxes", i.e., taxes on consumption goods that individuals consume too much of, relative to their long-run preferences. See also O'Donoghue and Rabin (2003, 2006), Krusell et al. (2002, 2010), Diamond and Koszegi (2003), and Amador et al. (2006).

http://dx.doi.org/10.1016/j.jebo.2014.11.002
0167-2681/

consumption and savings decisions that are not in their long-run interest. The policy instrument available to the government to offset the effects of quasi-hyperbolic discounting is dynamic nonlinear income taxation, applicable to both labor and savings.

There is currently a great deal of interest in dynamic nonlinear income taxation, such as the "new dynamic public finance" literature that extends the static Mirrlees (1971) model of optimal nonlinear income taxation to a dynamic setting.[4] The second-best nature of the Mirrlees model stems entirely from the assumption that an individual's skill type is private information, which is what prevents the government from implementing first-best taxation based on skills as the Second Welfare Theorem would recommend. In dynamic versions of the Mirrlees model, however, taxation in earlier periods may result in skill-type information being revealed to the government, which would then enable first-best taxation in latter periods. To avoid this possibility and some associated complications, the new dynamic public finance literature typically assumes that the government can commit to its future tax policy. That is, the government continues to implement second-best (incentive-compatible) taxation even after skill-type information has been revealed. However, the commitment assumption overlooks an important feature of the Mirrlees approach to optimal taxation—that no ad hoc constraints be placed on the nature of the optimal tax function, and that the tax instruments available to the government be constrained only by the information structure. Indeed, one of the motives behind the development of the new dynamic public finance literature is to avoid the need for ad hoc constraints on the tax system, as are typically imposed in the classic representative-agent Ramsey model (see Golosov et al., 2006). Therefore, we assume that the government cannot commit to its future tax policy. This means that both individuals and the government in our model cannot commit to future plans, though both would be better-off in the long run if they were able to do so.

The main complication associated with relaxing the commitment assumption is that it may no longer be social-welfare maximizing for the government to design a (separating) nonlinear income tax system in which individuals are willing to reveal their skill types. Instead, it may be optimal to pool the individuals so that skill-type information is not revealed.[5] To minimize the problems that the possible optimality of separating or pooling taxation present, we adopt the simple two-type (high-skill and low-skill) version of the Mirrlees model introduced by Stiglitz (1982),[6] and analyze a three-period model, which is the shortest time horizon that can capture the effects of quasi-hyperbolic discounting. Individuals work and save in periods 1 and 2, and live-off their second-period savings in period 3. The government imposes nonlinear taxation on labor and savings in periods 1 and 2 such that a utilitarian social welfare function based on individuals' true (long-run) preferences is maximized. Hence, the social welfare function captures a corrective or paternalistic motive for dynamic taxation, as well as the usual redistributive motive embedded in utilitarianism.

Our main result is that quasi-hyperbolic discounting increases the level of social welfare attainable when separating taxation is optimal, but decreases social welfare when pooling is optimal. This immediately implies that, under separating taxation, at least one type of individuals are actually better-off in the long run as a result of their short-run impatience. Moreover, our numerical simulations reveal that, even under pooling taxation, one type of individual is better-off in the long run. These findings stand in stark contrast to the usual result that quasi-hyperbolic discounting makes individuals worse-off in the long run. The intuition for our results, in a nutshell, can be summarized as follows. Nonlinear income taxation gives the government the power to ensure that only two allocations, one intended for low-skill individuals and the other intended for high-skill individuals, may potentially be chosen, by making the tax burden associated with all other allocations sufficiently severe. This, in effect, means that the government can override the individuals' short-run (quasi-hyperbolic) preferences. The only challenge then that the government faces is to ensure that each type chooses the allocation intended for them. Given the government's redistributive objective, low-skill individuals will never want to choose the high-skill type's allocation, but high-skill individuals may want to mimic low-skill individuals by choosing their allocation. The government can deter mimicking behavior by making sure that the allocations offered satisfy the high-skill type's incentive-compatibility constraint. Quasi-hyperbolic discounting does, however, affect the incentive-compatibility constraint, as high-skill individuals will compare the high-skill and low-skill allocations using their short-run preferences. We then show that quasi-hyperbolic discounting relaxes the high-skill type's incentive-compatibility constraint under separating taxation, but tightens it under pooling taxation. It also follows from the preceding discussion that quasi-hyperbolic discounting exerts no effect on social welfare under first-best taxation, since in this case the allocations need not be incentive compatible.

In terms of previous studies, Aronsson and Sjogren (2009), Bassi (2010), and Aronsson and Granlund (2011) are most closely related to our work, with the primary distinction being that we analytically and quantitatively demonstrate how quasi-hyperbolic discounting can actually raise long-run utility and social welfare. In Aronsson and Sjogren's model, individuals consume an unhealthy commodity when young, which in turn leads to adverse health outcomes when old. Since individuals are quasi-hyperbolic discounters, they consume too much of the unhealthy commodity. Within this setting,

[4] A survey of the new dynamic public finance literature is provided by Golosov et al. (2006), while Kocherlakota (2010) provides a textbook treatment. Earlier papers that extend the Mirrlees model to dynamic settings include Roberts (1984) and Brito et al. (1991).

[5] The reason that either separating or pooling taxation may be optimal when the government cannot commit is explained in detail in Section 4.

[6] It does not seem feasible to consider more than two types of individuals, because the number of tax regimes that must be considered increases exponentially. For example, assuming merely three types results in five regimes: complete separation, complete pooling, and three cases of pooling two types against the remaining type. Moreover, even in the two-type model that we study, there is a third possibility of partial pooling in which some, but not all, of the high-skill individuals are pooled with the low-skill individuals. However, for the sake of analytical simplicity, we restrict attention to the "pure strategy" policies of complete separating or pooling taxation.

Aronsson and Sjogren analyze how to design a mixed tax system, involving nonlinear income taxation and linear commodity taxation without commitment, to correct the effects of quasi-hyperbolic discounting and redistribute from high-skill to low-skill individuals. Bassi examines the effects of incorporating quasi-hyperbolic discounting into a dynamic Mirrlees model, but he assumes full commitment with the focus on showing how quasi-hyperbolic discounting makes a case for taxing savings. Aronsson and Granlund explore optimal provision of a public good when individuals are quasi-hyperbolic discounters. They show that quasi-hyperbolic discounting moves the second-best rule governing provision of the public good closer to the first-best Samuelson rule, through reducing the weight placed on the incentive-compatibility constraint in the second-best rule.

Our paper also complements a recent literature that examines labor and savings/capital taxation in dynamic Mirrlees-style models, albeit without quasi-hyperbolic discounting. Brett and Weymark (2008) and Farhi et al. (2012) examine the optimality of nonlinear savings/capital taxation in dynamic Mirrlees models under no commitment.[7] Brett and Weymark examine a two-type, two-period model in which the tax instruments available to the government are constrained only by asymmetric information regarding skills, as in the standard Mirrlees framework. They highlight the possibilities that separating or pooling taxation can be social-welfare maximizing, and derive the optimal marginal tax rates applicable to savings under each regime. Farhi et al. examine a model with a continuum of types, and they consider both two-period and infinite-horizon settings. In their model, taxation is constrained by asymmetric information regarding skills and by political-economy constraints, which take the form of direct or reputational costs of implementing tax reforms. This results in a "limited commitment" setting in which full redistribution is never implemented. Their main conclusion is that the optimal taxation of capital is progressive.

The models examined by Golosov et al. (2013), Tenhunen and Tuomala (2010), and Diamond and Spinnewijn (2011) all assume full commitment, but individuals are distinguished by both skills and their preferences for savings. As the government observes neither skills nor preferences, it faces a two-dimensional screening problem. Golosov et al. avoid the complexities associated with multi-dimensional screening by assuming that preferences for savings are a function of skills. They conclude that optimal capital-income tax rates are low, and that the welfare gains from taxing capital are negligible. Diamond and Spinnewijn examine a two-period model in which individuals differ by their skills and discount factors. An individual may be high or low skill, and they may have a high or low discount factor. Although there are only four possible types in their model, Diamond and Spinnewijn note that the standard mechanism-design approach to optimal nonlinear taxation would require a highly complex tax system. Accordingly, they simplify the problem by assuming that there are only two jobs in the economy—a high-skill job and a low-skill job—and that hours worked are fixed. In this setting, their main result is that a case can be made for taxing savings. Tenhunen and Tuomala consider a similar setting to Diamond and Spinnewijn, but they do not assume jobs are skill specific. Instead, they mainly focus on numerical solutions to the multi-dimensional screening problem. They also examine the possibility that the government's discount factor may differ from that of individuals, which creates a paternalistic motive for taxation. Their analysis also provides a rationale for savings taxation.

The remainder of the paper is organized as follows. Section 2 outlines the analytical framework that we consider. Section 3 examines first-best nonlinear taxation, while Section 4 examines second-best nonlinear taxation. Section 5 presents some numerical simulations to further highlight the effects of quasi-hyperbolic discounting. Section 6 discusses the effects of assuming that the government can commit, while Section 7 discusses the effects of assuming that individuals are sophisticated agents. Section 8 concludes, while proofs are relegated to an appendix.

2. Analytical framework

The economy is assumed to last for three periods.[8] There is a continuum of individuals of unit measure who live for the three periods, with a proportion $\phi \in (0, 1)$ being high-skill workers and the remainder $(1-\phi)$ being low-skill workers. The wage rates of the high-skill and low-skill individuals are denoted by w_H and w_L respectively, where $w_H > w_L > 0$. Wages are assumed to remain constant through time.

Individuals work and save in periods 1 and 2. In period 3, which can be thought of as the retirement period, individuals do not work and must live-off their second-period savings. Therefore, savings decisions made in period 2 completely determine the outcome in period 3. Individual i's true (long-run) utility function is given by:

$$u(c_i^1) - v(l_i^1) + \delta[u(c_i^2) - v(l_i^2)] + \delta^2 u(c_i^3) \tag{2.1}$$

where c_i^t is individual i's consumption in period t, l_i^t is individual i's labor supply in period t, and $\delta \in (0, 1)$ is the discount factor. The function $u(\cdot)$ is increasing and strictly concave, while $v(\cdot)$ is increasing and strictly convex.

[7] Apps and Rees (2006), Krause (2009), and Guo and Krause (2011a, 2013) also examine two-type, two-period nonlinear income tax models without commitment, although savings do not feature in their models.

[8] We consider a finite-horizon model because the no-commitment optimal tax problem faced by the government is most easily solved by backwards induction, and as discussed earlier three periods is the shortest time horizon that can capture the effects of quasi-hyperbolic discounting.

As individuals are quasi-hyperbolic discounters, they do not act to maximize (2.1). Instead, following Laibson (1997), their objective function is better described by the utility function:

$$u(c_i^1) - v(l_i^1) + \beta\delta[u(c_i^2) - v(l_i^2)] + \beta\delta^2 u(c_i^3) \tag{2.2}$$

where $\beta \in (0, 1)$ captures the effects of quasi-hyperbolic discounting. When viewed from period 1, it can be seen that an individual's discount factor between periods 1 and 2 is $\beta\delta$, while between periods 2 and 3 it is δ. But when viewed from period 2, the discount factor between periods 2 and 3 is $\beta\delta$. Thus (2.2) captures a preference for immediate gratification that leads individuals to make short-run consumption, savings, and labor supply decisions that are not consistent with long-run utility maximization.

2.1. Individual behavior without taxation

In this subsection, we describe how individuals would behave in the absence of taxation. The literature on quasi-hyperbolic discounting has distinguished between *naive* agents and *sophisticated* agents. Naive agents are aware of their need for immediate gratification, but they (naively) think that in the future they will behave in a manner consistent with their long-run preferences. This captures the idea that individuals who act to satisfy their short-run impatience often excuse such behavior by promising themselves that they will behave more rationally in the future. On the other hand, sophisticated agents are aware of their need for immediate gratification, and they are also aware that they will feel this need again in the future and thus take this into account in their decision making. We assume that individuals are naive, mainly for simplicity but also because empirical evidence suggests that individuals are not sophisticated (see, e.g., Hey and Lotito, 2009).[9] Accordingly, individual i's behavior in period 1 can be described as follows. Choose a lifetime consumption plan c_i^1, l_i^1, s_i^1, c_i^2, l_i^2, s_i^2, and c_i^3 to maximize (2.2) subject to:

$$c_i^1 + s_i^1 \leq w_i l_i^1 \tag{2.3}$$

$$c_i^2 + s_i^2 \leq (1+r)s_i^1 + w_i l_i^2 \tag{2.4}$$

$$c_i^3 \leq (1+r)s_i^2 \tag{2.5}$$

where s_i^t is individual i's savings in period t, and $r > 0$ is the market interest rate. Eqs. (2.3)–(2.5) are the individual's first-, second-, and third-period budget constraints, respectively. It is shown in the Appendix that the solution to program (2.2)–(2.5) yields the marginal conditions:

$$\frac{v'(l_i^t)}{u'(c_i^t)w_i} = 1 \quad (\text{for } t = 1, 2), \qquad \frac{u'(c_i^1)}{\beta\delta(1+r)u'(c_i^2)} = 1 \qquad \text{and} \qquad \frac{u'(c_i^2)}{\delta(1+r)u'(c_i^3)} = 1 \tag{2.6}$$

However, individuals cannot commit themselves to their period-2 and period-3 consumption plans, so in periods 2 and 3 they will not simply implement the plans decided upon in period 1. Instead, in period 2 they will feel the need for immediate gratification again, and will therefore choose c_i^2, l_i^2, s_i^2, and c_i^3 to maximize:

$$u(c_i^2) - v(l_i^2) + \beta\delta u(c_i^3) \tag{2.7}$$

subject to their second- and third-period budget constraints, Eqs. (2.4) and (2.5). It is shown in the Appendix that the solution to this problem yields the marginal conditions:

$$\frac{v'(l_i^2)}{u'(c_i^2)w_i} = 1 \quad \text{and} \quad \frac{u'(c_i^2)}{\beta\delta(1+r)u'(c_i^3)} = 1 \tag{2.8}$$

Comparing Eqs. (2.6) and (2.8), it can be seen that the marginal condition between consumption in periods 2 and 3 that will actually be implemented differs from that which was originally planned.

2.2. Implicit marginal tax rates

As we assume that the government can impose nonlinear taxes on an individual's income from labor and savings, it may be optimal for the government to set taxes to induce violations of the marginal conditions shown above that individuals would implement in the absence of taxation. Following the standard practice, one may interpret these marginal distortions as tax wedges or implicit marginal tax rates. Thus, we define:

$$MTRL_i^t := 1 - \frac{v'(l_i^t)}{u'(c_i^t)w_i} \quad \text{and} \quad MTRS_i^t := 1 - \frac{u'(c_i^t)}{\beta\delta(1+r)u'(c_i^{t+1})} \tag{2.9}$$

[9] In any event, it will be seen in Section 7 that the assumptions of naive or sophisticated quasi-hyperbolic discounting do not matter for our main results *vis-a-vis* social welfare. All that matters is that individuals exhibit excessive short-run impatience.

where $MTRL_i^t$ denotes the (implicit) marginal tax rate on labor faced by individual i in period t, and $MTRS_i^t$ denotes the (implicit) marginal tax rate on savings faced by individual i in period t (where $t = 1, 2$).

3. First-best taxation

We begin by briefly considering the hypothetical case in which the government knows each individual's skill type in every period. In this case, the government's choice of an optimal nonlinear tax system applicable to labor income and savings is equivalent to it choosing "lifetime" allocations $\langle m_L^1, y_L^1, s_L^1, m_L^2, y_L^2, s_L^2\rangle$ and $\langle m_H^1, y_H^1, s_H^1, m_H^2, y_H^2, s_H^2\rangle$ for the low-skill and high-skill individuals, respectively, to maximize[10]:

$$(1-\phi)\left\{u(m_L^1 - s_L^1) - v\left(\frac{y_L^1}{w_L}\right) + \delta\left[u(m_L^2 + (1+r)s_L^1 - s_L^2) - v\left(\frac{y_L^2}{w_L}\right)\right] + \delta^2 u((1+r)s_L^2)\right\}$$
$$+\phi\left\{u(m_H^1 - s_H^1) - v\left(\frac{y_H^1}{w_H}\right) + \delta\left[u(m_H^2 + (1+r)s_H^1 - s_H^2) - v\left(\frac{y_H^2}{w_H}\right)\right] + \delta^2 u((1+r)s_H^2)\right\} \tag{3.1}$$

subject to:

$$(1-\phi)[y_L^1 - m_L^1] + \phi[y_H^1 - m_H^1] \geq 0 \tag{3.2}$$

$$(1-\phi)[y_L^2 - m_L^2] + \phi[y_H^2 - m_H^2] \geq 0 \tag{3.3}$$

where m_i^t is type i's post-tax income in period t, $y_i^t = w_i l_i^t$ is type i's pre-tax income in period t, and $c_i^1 = m_i^1 - s_i^1$, $c_i^2 = m_i^2 + (1+r)s_i^1 - s_i^2$, and $c_i^3 = (1+r)s_i^2$. Eq. (3.1) is a utilitarian social welfare function based on each type's true (long-run) utility function (2.1), which reflects the assumption that the government has a corrective (or paternalistic) objective in setting taxes. Eqs. (3.2) and (3.3) are, respectively, the government's first- and second-period budget constraints. Implicit in Eqs. (3.2) and (3.3) is the simplifying assumption that the government cannot save or borrow, and that its revenue requirement is zero.[11] As it is currently postulated that the government can observe each individual's type from period 1 onwards, it does not face any incentive-compatibility constraints. The solution to program (3.1)–(3.3) yields:

Remark. Under first-best taxation, quasi-hyperbolic discounting has no effect on the level of social welfare attainable.

The result that quasi-hyperbolic discounting exerts no effect on the level of social welfare attainable under first-best taxation follows simply from the fact that the quasi-hyperbolic discounting parameter, β, does not appear in program (3.1)–(3.3). This is because first-best nonlinear taxation gives the government the power to force each type i to choose $\langle m_i^1, y_i^1, s_i^1, m_i^2, y_i^2, s_i^2\rangle$ by making their tax burden associated with any other allocation sufficiently severe. Thus, the government can impose its desired allocation on each individual, irrespective of their short-run preferences.

4. Second-best taxation

In this section, we examine nonlinear labor and savings taxation when the government cannot observe each individual's skill type. Incentive-compatibility constraints must now be considered, and heterogeneity between individuals' short-run and long-run preferences now plays a role. Taxation in period 1, however, may result in skill-type information being revealed to the government, which would then enable it to implement first-best taxation in period 2. As all individuals know that if they reveal their type in period 1 they will be subjected to first-best taxation in period 2, they may have to be compensated in period 1 if they are to be willing to reveal their type. This compensation is potentially very costly from the government's perspective of maximizing social welfare. Accordingly, rather than designing a "separating" tax system in period 1 in which individuals are willing to reveal their types, it may be optimal for the government to use "pooling" taxation in which type information is not revealed, even though it is then constrained to use second-best taxation in period 2. It is theoretically possible for either the separating or pooling tax systems to be social-welfare maximizing, depending upon the parameters of the model.[12] Therefore, we examine in turn the nature of separating and pooling nonlinear labor and savings taxation.

4.1. Separating taxation

If the tax system is designed to separate the high-skill individuals from the low-skill individuals in period 1, the government has enough information to implement first-best taxation in period 2. The government's behavior in period 2 can

[10] Recall that individuals do not work in period 3. Thus savings decisions made in period 2 completely determine the outcome in period 3.

[11] We maintain these simplifying assumptions throughout the paper, since assuming otherwise would not affect any of our main results.

[12] See, e.g., the papers by Roberts (1984), Berliant and Ledyard (2014), and Guo and Krause (2011b) which examine the desirability of separating versus pooling nonlinear income taxation when the government cannot commit.

then be described as follows. Choose allocations $\langle m_L^2, y_L^2, s_L^2\rangle$ and $\langle m_H^2, y_H^2, s_H^2\rangle$ for the low-skill and high-skill individuals, respectively, to maximize:

$$(1-\phi)\left\{u(m_L^2+(1+r)s_L^1-s_L^2)-v\left(\frac{y_L^2}{w_L}\right)+\delta u((1+r)s_L^2)\right\}+\phi\left\{u(m_H^2+(1+r)s_H^1-s_H^2)-v\left(\frac{y_H^2}{w_H}\right)+\delta u((1+r)s_H^2)\right\} \tag{4.1}$$

subject to:

$$(1-\phi)\left[y_L^2-m_L^2\right]+\phi\left[y_H^2-m_H^2\right]\geq 0 \tag{4.2}$$

where (4.1) is a utilitarian social welfare function which reflects each type's true utility over periods 2 and 3, while (4.2) is the government's second-period budget constraint. The solution to program (4.1)–(4.2) yields the functions $m_L^2(\phi, r, s_L^1, w_L, \delta, s_H^1, w_H), y_L^2(\cdot), s_L^2(\cdot), m_H^2(\cdot), y_H^2(\cdot)$, and $s_H^2(\cdot)$. Substituting these functions into (4.1) yields the value function $Z_F^2(\cdot)$, with the subscript F indicating that the value function is associated with a first-best taxation problem.

In period 1, the government cannot distinguish high-skill from low-skill individuals, but it designs a separating tax system in order to obtain skill-type information. Accordingly, the government chooses allocations $\langle m_L^1, y_L^1, s_L^1\rangle$ and $\langle m_H^1, y_H^1, s_H^1\rangle$ for the low-skill and high-skill individuals, respectively, to maximize:

$$(1-\phi)\left\{u(m_L^1-s_L^1)-v\left(\frac{y_L^1}{w_L}\right)\right\}+\phi\left\{u(m_H^1-s_H^1)-v\left(\frac{y_H^1}{w_H}\right)\right\}+\delta Z_F^2(\cdot) \tag{4.3}$$

subject to:

$$(1-\phi)[y_L^1-m_L^1]+\phi[y_H^1-m_H^1]\geq 0 \tag{4.4}$$

$$u(m_H^1-s_H^1)-v\left(\frac{y_H^1}{w_H}\right)+\beta\delta U_{HF}^2\geq u(m_L^1-s_L^1)-v\left(\frac{y_L^1}{w_H}\right)+\beta\delta U_{MF}^2 \tag{4.5}$$

where

$$U_{HF}^2 := u(m_H^2(\cdot)+(1+r)s_H^1-s_H^2(\cdot))-v\left(\frac{y_H^2(\cdot)}{w_H}\right)+\delta u((1+r)s_H^2(\cdot)) \tag{4.6}$$

and

$$U_{MF}^2 := u(m_L^2(\cdot)+(1+r)s_L^1-s_L^2(\cdot))-v\left(\frac{y_L^2(\cdot)}{w_H}\right)+\delta u((1+r)s_L^2(\cdot)) \tag{4.7}$$

Eq. (4.3) is a utilitarian social welfare function, which takes into account how savings decisions made in period 1 affect the level of social welfare attainable over periods 2 and 3. Eq. (4.3) therefore includes the value function $Z_F^2(\cdot)$. Eq. (4.4) is the government's first-period budget constraint, while (4.5) is the high-skill type's incentive-compatibility constraint. In order for a high-skill individual to be willing to reveal their type, the utility they obtain from choosing $\langle m_H^1, y_H^1, s_H^1\rangle$ in period 1 and thus revealing their type, plus the *quasi-hyperbolic* discounted value of the utility, U_{HF}^2, high-skill individuals obtain over periods 2 and 3 under first-best taxation, must be greater than or equal to the utility they could obtain by pretending to be low skill. A high-skill individual may pretend to be low skill by choosing $\langle m_L^1, y_L^1, s_L^1\rangle$ in period 1. They will then be treated as low-skill by the government under first-best taxation in period 2, which implies that they obtain a "mimickers" utility level, U_{MF}^2, over periods 2 and 3. Since high-skill individuals are free to choose between $\langle m_H^1, y_H^1, s_H^1\rangle$ and $\langle m_L^1, y_L^1, s_L^1\rangle$, they evaluate these allocations and the respective utility streams that follow according to their short-run (quasi-hyperbolic) preferences. As a result, the high-skill type's incentive-compatibility constraint must take quasi-hyperbolic discounting into account.

Notice that the low-skill type's incentive-compatibility constraint is omitted, because we focus on what Stiglitz (1982) calls the "normal" case and what Guesnerie (1995) calls "redistributive equilibria". Specifically, we make the standard assumption that the parameters of the model are such that the government will be seeking to redistribute from high-skill to low-skill individuals. This implies that high-skill individuals have an incentive to mimic low-skill individuals, but not vice versa. Therefore, 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 programs (4.1)–(4.2) and (4.3)–(4.5) together imply:

Proposition 1. *Under second-best taxation with separation in the first period, quasi-hyperbolic discounting increases the level of social welfare attainable.*

Interestingly, quasi-hyperbolic discounting *increases* the level of social welfare attainable, which implies that at least one type of individual must be better-off in the long run as a result of their short-run impatience. The intuition for this result is two-fold. First, nonlinear taxation gives the government the power to ensure that either the $\langle m_H^1, y_H^1, s_H^1\rangle$ or $\langle m_L^1, y_L^1, s_L^1\rangle$ allocations will be chosen in period 1, simply by making the tax burden associated with all other allocations sufficiently severe. Given the government's redistributive goals, low-skill individuals will always want to choose $\langle m_L^1, y_L^1, s_L^1\rangle$, so all the government has to worry about is making sure that high-skill individuals choose $\langle m_H^1, y_H^1, s_H^1\rangle$. This will happen provided the high-skill type's incentive-compatibility constraint (4.5) is satisfied. Second, quasi-hyperbolic discounting relaxes the high-skill type's incentive-compatibility constraint. This follows from a well-known, though somewhat strange, feature of first-best taxation, i.e., individual utility is *decreasing* with respect to the wage rate. This is because under first-best taxation, it is optimal to give both types the same level of consumption, but high-skill individuals are required to work longer.[13] Accordingly, high-skill individuals must be offered a relatively favorable tax treatment in period 1 if they are to reveal their type, in order to compensate them for the very unfavorable tax treatment they will face under first-best taxation in period 2. However, quasi-hyperbolic discounting means that in period 1 high-skill individuals care less than they should about the utility they obtain in periods 2 and 3. Therefore, high-skill individuals require less compensation in period 1 to reveal their type, which in turn enables the government to attain a higher level of social welfare.

4.2. Marginal tax rates under separating taxation

One can also compute the implicit marginal tax rates associated with separating taxation, although we hasten to add that these are either standard results or are straightforward extensions of those in Brett and Weymark (2008) once the effects of quasi-hyperbolic discounting are taken into account. This shows that quasi-hyperbolic discounting has little qualitative effect on optimal marginal tax rates, although Section 5 finds that it can have a substantive quantitative impact.

It is shown in the Appendix that the optimal marginal tax rates under separating taxation applicable to labor income are $MTRL_H^1 = MTRL_L^2 = MTRL_H^2 = 0$ and $MTRL_L^1 > 0$, while those applicable to savings are $MTRS_L^1 < 0$, $MTRS_H^1 \gtreqless 0$, and $MTRS_L^2 = MTRS_H^2 = (\beta - 1)/\beta < 0$.

In period 1, high-skill individuals face a zero marginal tax rate on their labor income, while that for low-skill individuals is positive. These are the well-known "no-distortion-at-the-top" and "downward-distortion-at-the-bottom" results that typify second-best nonlinear income taxation. Likewise, both types face zero marginal tax rates on labor income in period 2 simply because first-best taxation is used in that period.

Low-skill individuals face a negative marginal tax rate on savings in period 1 for two reasons. First, quasi-hyperbolic discounting means that they want to save less than they should; thus the government distorts their savings upwards to correct this effect. Second, distorting savings by low-skill individuals upwards relaxes the high-skill type's incentive-compatibility constraint. To see this, note that under first-best taxation in period 2, the government will choose allocations such that $u'(m_H^2 + (1+r)s_H^1 - s_H^2) = u'(m_L^2 + (1+r)s_L^1 - s_L^2)$, taking first-period savings s_H^1 and s_L^1 as given.[14] Now since $u(\cdot)$ is strictly concave, an increase in s_L^1 will reduce the low-skill type's marginal utility of consumption relative to that for the high-skill type, meaning the government can raise social welfare by transferring income from low-skill to high-skill individuals. It follows that an increase in s_L^1 makes high-skill individuals better-off under first-best taxation in period 2. This in turn makes them more willing to reveal their type in period 1, or equivalently the incentive-compatibility constraint is relaxed.

The sign of the marginal tax rate on savings faced by high-skill individuals in period 1 is ambiguous. On the one hand, the government wants to distort their savings upwards to correct the effects of quasi-hyperbolic discounting. But on the other hand, the government wants to distort their savings downwards to relax the incentive-compatibility constraint; the intuition for this follows by mirroring the argument just made for increasing the low-skill type's savings to relax the incentive-compatibility constraint. Finally, both types face negative marginal tax rates on their second-period savings, equal to $(\beta - 1)/\beta$. This is because quasi-hyperbolic discounting implies that individuals would choose to save less than they should according to their long-run preferences. Therefore, optimal nonlinear taxation distorts each type's savings upwards via negative marginal tax rates to correct the effects of quasi-hyperbolic discounting.

4.3. Pooling taxation

If the tax system pools the individuals in period 1, the government cannot distinguish high-skill from low-skill individuals in period 2. Therefore, in period 2 the government must solve a second-best (information constrained) optimal nonlinear income tax problem. As this problem is essentially a static optimal nonlinear income tax problem, separating taxation is optimal. The government therefore chooses allocations $\langle m_L^2, y_L^2, s_L^2\rangle$ and $\langle m_H^2, y_H^2, s_H^2\rangle$ for the low-skill and high-skill individuals, respectively, to maximize:

$$(1-\phi)\left\{u(m_L^2 + (1+r)s^1 - s_L^2) - v\left(\frac{y_L^2}{w_L}\right) + \delta u((1+r)s_L^2)\right\} + \phi\left\{u(m_H^2 + (1+r)s^1 - s_H^2) - v\left(\frac{y_H^2}{w_H}\right) + \delta u((1+r)s_H^2)\right\} \tag{4.8}$$

[13] This has led some to describe first-best taxation as Marxist in nature, because it takes from each individual according to their ability and gives to each individual according to their need.

[14] That is, the government will seek to equate second-period consumption levels. This follows from Eqs. (A.16) and (A.19) in the Appendix.

subject to:

$$(1-\phi)\left[y_L^2 - m_L^2\right] + \phi\left[y_H^2 - m_H^2\right] \geq 0 \tag{4.9}$$

$$u(m_H^2 + (1+r)s^1 - s_H^2) - v\left(\frac{y_H^2}{w_H}\right) + \beta\delta u((1+r)s_H^2) \geq u(m_L^2 + (1+r)s^1 - s_L^2) - v\left(\frac{y_L^2}{w_H}\right) + \beta\delta u((1+r)s_L^2) \tag{4.10}$$

where s^1 denotes the first-period savings of both types under pooling in period 1. Eq. (4.8) is a utilitarian social welfare function, (4.9) is the government's second-period budget constraint, and (4.10) is the high-skill type's incentive-compatibility constraint. In period 2 the government cannot distinguish high-skill from low-skill individuals, so the allocations must be incentive compatible.[15] As high-skill individuals are free to choose between $\langle m_H^2, y_H^2, s_H^2\rangle$ and $\langle m_L^2, y_L^2, s_L^2\rangle$, their short-run preference for immediate gratification must be taken into account. Hence, the quasi-hyperbolic discounting parameter, β, enters the incentive-compatibility constraint.

It should be noted that based on the individuals' second-period choices, the government can distinguish high-skill from low-skill individuals in period 3. However, as the second-period choices completely determine the period-3 allocations, the government cannot take advantage of the skill-type information it acquires to improve the period-3 allocations. This is because we assume that taxation of labor income and savings occurs in the periods when labor and savings decisions are made (periods 1 and 2), not the retirement period (period 3).

The solution to program (4.8)–(4.10) yields the functions $m_L^2(\phi, r, s^1, w_L, \delta, w_H, \beta)$, $y_L^2(\cdot)$, $s_L^2(\cdot)$, $m_H^2(\cdot)$, $y_H^2(\cdot)$, and $s_H^2(\cdot)$. Substituting these functions into (4.8) yields the value function $Z_S^2(\cdot)$, with the subscript S indicating that the value function is associated with a second-best taxation problem.

In period 1 the government implements pooling taxation. Therefore, it chooses a single allocation $\langle m^1, y^1, s^1\rangle$ for all individuals to maximize:

$$(1-\phi)\left\{u(m^1 - s^1) - v\left(\frac{y^1}{w_L}\right)\right\} + \phi\left\{u(m^1 - s^1) - v\left(\frac{y^1}{w_H}\right)\right\} + \delta Z_S^2(\cdot) \tag{4.11}$$

subject to:

$$y^1 - m^1 \geq 0 \tag{4.12}$$

where (4.11) is the first-period utilitarian social welfare function, but takes into account how the choice of first-period savings affects the level of social welfare attainable over periods 2 and 3; thus (4.11) includes the value function $Z_S^2(\cdot)$. Eq. (4.12) is the government's first-period budget constraint.

It is shown in the Appendix that the solutions to programs (4.8)–(4.10) and (4.11)–(4.12) together imply:

Proposition 2. *Under second-best taxation with pooling in the first period, quasi-hyperbolic discounting decreases the level of social welfare attainable.*

When pooling in period 1 is optimal, quasi-hyperbolic discounting reduces the level of social welfare attainable. This is because, unlike in the separating case, quasi-hyperbolic discounting under pooling *tightens* the high-skill type's incentive-compatibility constraint. The intuition is as follows. Under first-best taxation, it is optimal to equate c_H^3 and c_L^3; but under second-best taxation it is optimal to set $c_H^3 > c_L^3$ as this helps relax the high-skill type's incentive-compatibility constraint. Now from (4.10) it can be seen that as the extent of quasi-hyperbolic discounting rises, or equivalently as β falls, the government must further raise c_H^3 relative to c_L^3 in order to exert the same impact, *ceteris paribus*, on the incentive-compatibility constraint. Therefore, quasi-hyperbolic discounting requires that the government move c_H^3 and c_L^3 further away from their first-best levels, which in turn reduces social welfare.

4.4. Marginal tax rates under pooling taxation

It is shown in the Appendix that under pooling taxation the optimal marginal tax rates applicable to labor income are $MTRL_L^1 < 0$, $MTRL_H^1 > 0$, $MTRL_L^2 > 0$, and $MTRL_H^2 = 0$, while those applicable to savings are $MTRS_L^1 \gtreqless 0$, $MTRS_H^1 < 0$, and $MTRS_L^2 < MTRS_H^2 < 0$. As under separating taxation, these marginal tax rates are either standard results or are similar to those in Brett and Weymark (2008), indicating again that quasi-hyperbolic discounting has little qualitative effect on optimal marginal tax rates.

In period 1, low-skill individuals face a negative marginal tax rate on their labor income, while that for high-skill individuals is positive. To understand these results, note that in the absence of taxation high-skill individuals would choose to earn more income than low-skill individuals (as both types have the same preferences, but $w_H > w_L$). When both types are subjected to the same allocation under pooling in period 1, the government, in effect, chooses y^1 based on a weighted average of w_L and w_H. This results in the low-skill (resp. high-skill) type's labor supply being distorted upwards (resp. downwards)

[15] We again omit the low-skill type's incentive-compatibility constraint because, given the government's redistributive objective, it will not be binding.

Table 1
Baseline parameter values for numerical simulations.[a]

φ	0.25	δ	0.38	w_L	1
γ	2	r	0.05	w_H	1.6
		β	0.85		

	Separating	Pooling
Long-run utility: low-skill	−0.481	−0.579
Long-run utility: high-skill	−0.182	−0.144
Social welfare	−0.407	−0.470

[a] Each period is assumed to be 20 years in length.

to earn y^1. In period 2 the usual pattern of optimal marginal tax rates on labor income applies, because in period 2 the government essentially solves a standard second-best optimal nonlinear income tax problem.

The sign of the marginal tax rate on savings faced by low-skill individuals in period 1 is ambiguous, while that for high-skill individuals is negative. This is because in period 2 it is optimal to set $c_H^2 > c_L^2$ to relax the high-skill type's incentive-compatibility constraint, as in the standard Mirrlees model. Implicitly distorting the low-skill type's first-period savings downwards, and the high-skill type's first-period savings upwards, makes it easier to implement $c_H^2 > c_L^2$ in period 2. But concurrently, the government wants to distort both types' first-period savings upwards to offset the effects of quasi-hyperbolic discounting. Thus the two motives for marginal distortions work in opposite directions for low-skill individuals, rendering their optimal marginal tax rate on first-period savings ambiguous; whereas both motives encourage an upward distortion to high-skill individuals' first-period savings, making a negative marginal tax rate optimal. In period 2, both types face negative marginal tax rates on their savings, in order to correct the effects of quasi-hyperbolic discounting. However, the subsidy for low-skill individuals is larger, because the further upward distortion to their second-period savings makes it easier to implement $c_H^2 > c_L^2$, which again helps relax the high-skill type's incentive-compatibility constraint.

5. Numerical simulations

In this section, we explore the effects of varying the quasi-hyperbolic discounting parameter, β. Ideally, these comparative statics results could be derived analytically; however the literature on the comparative statics of optimal nonlinear income taxes has found that analytical results are generally obtainable only when the utility function is quasi-linear.[16] We therefore calibrate our model with empirically plausible parameter values, and then use numerical simulations to examine the effects of changing β. To this end, we postulate that:

$$u(c_i^t) = \ln(c_i^t) \quad \text{and} \quad v(l_i^t) = \frac{1}{1+\gamma}(l_i^t)^{1+\gamma} \tag{5.1}$$

where $\gamma > 0$. Chetty (2006) concludes that a reasonable estimate of the coefficient of relative risk aversion is one (log utility); hence our assumption that utility is logarithmic in consumption. In addition, we set $\gamma = 2$, as this implies a labor supply elasticity of 0.5 which is in line with empirical estimates (see, e.g., Chetty et al., 2011).[17]

Table 1 presents the remaining parameter values. The OECD (2010) reports that on average across OECD countries, approximately one-quarter of all adults have attained tertiary level education. We therefore assume that 25% of individuals are high-skill workers, i.e., we set $\phi = 0.25$. We assume an annual interest rate of $r = 0.05$, which is consistent with common practice, and that the long-run discount factor δ is equal to $1/(1+r)$. Since most individuals work for around 40 years of their lives, we take each period to be 20 years in length. An annual discount rate of 5% then corresponds to a 20-year discount factor of $\delta = 0.38$. Fang (2006) and Goldin and Katz (2007) estimate that the college wage premium is approximately 60%. We therefore normalize the low-skill type's wage to unity, and set the high-skill type's wage equal to 1.6. Finally, we begin with an arbitrary baseline value of $\beta = 0.85$, and then examine the effects of varying β between 0.75 and 0.95 on each type's true (long-run) utility and on the optimal marginal tax rates,[18] holding all other parameters at their baseline levels. These effects are shown in Fig. 1 for separating taxation, and in Fig. 2 for pooling taxation.

In Fig. 1, it can be seen that social welfare under separating taxation is decreasing in β or, equivalently, increasing in the degree of quasi-hyperbolic discounting (cf. Proposition 1). High-skill individuals are better-off as β increases, while low-skill individuals are worse-off. As discussed earlier, an increase in β under separating taxation tightens the high-skill

[16] See, e.g., Brett and Weymark (2011) and Simula (2010). We cannot assume quasi-linearity in our model, because the first-best optimal tax problem is then no longer uniquely determined.

[17] Micro-econometric estimates of the labor supply elasticity tend to be low (around 0.1), while macro estimates are typically much higher (close to 1). We have examined the sensitivity of our numerical simulation results with respect to the labor supply elasticity, by conducting them using elasticities of 0.1 and 1. The effects of changes in β on social welfare, utility, and the optimal marginal tax rates turn out to be qualitatively identical to those reported in this section when the labor supply elasticity is 0.5. The details of these results are available upon request.

[18] The effects of varying β on all non-zero marginal tax rates are explored, except for those equal to $(\beta-1)/\beta$ because the effect in this case is obvious.

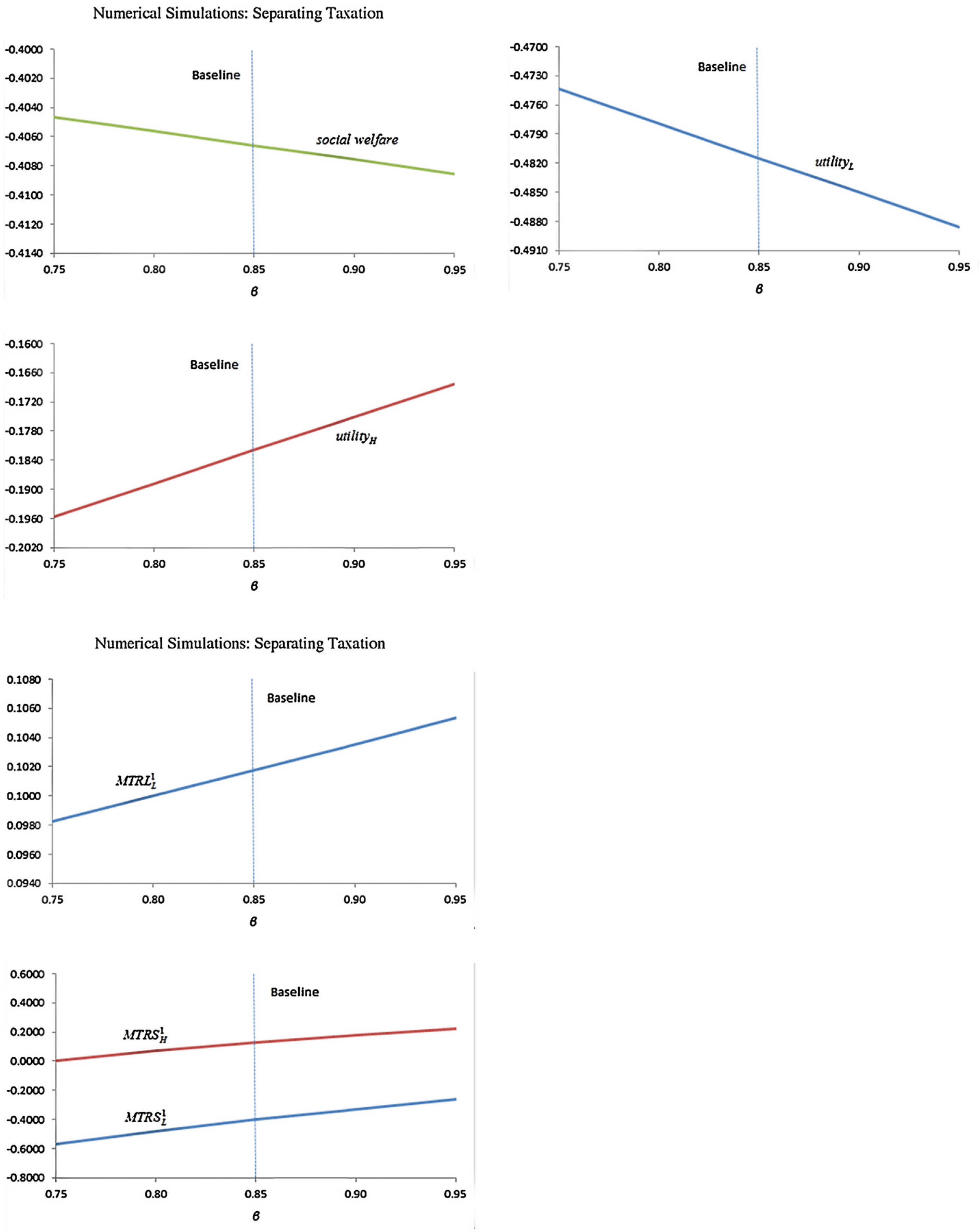

Fig. 1. Numerical simulations: separating taxation.

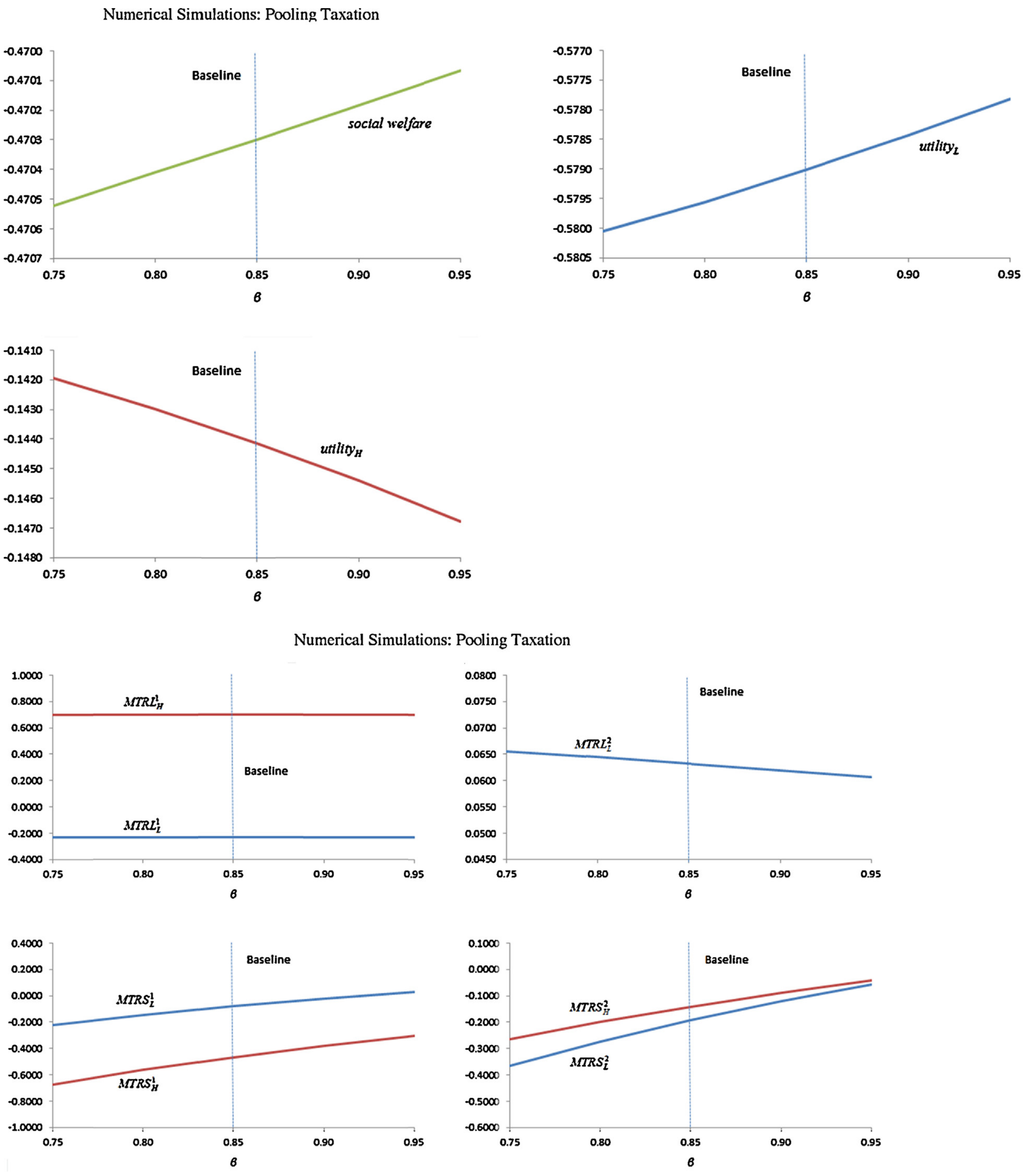

Fig. 2. Numerical simulations: pooling taxation.

type's incentive-compatibility constraint. Thus high-skill individuals must be offered a more attractive tax treatment, which comes at the expense of low-skill individuals. Next, the optimal marginal tax rate applicable to the low-skill type's labor income in period 1 is increasing in β. As in the standard Mirrlees model, low-skill individuals face a positive marginal tax rate on their labor income to relax the high-skill type's incentive-compatibility constraint. Since in our model an increase in β tightens the high-skill type's incentive-compatibility constraint, low-skill individuals must face a higher marginal tax rate on their labor income. Furthermore, the marginal tax rates for both types on their first-period savings are increasing in β, simply because the need to correct the effects of quasi-hyperbolic discounting is attenuated. While the sign of the high-skill type's optimal marginal tax rate on first-period savings is theoretically ambiguous, in our numerical simulations it is positive. This indicates that the motive the government has to distort their savings downwards to relax the incentive-compatibility constraint outweighs the motive it has to distort their savings upwards to offset the effects of quasi-hyperbolic discounting.

In Fig. 2, which covers pooling taxation, social welfare is increasing in β (cf. Proposition 2). Mirroring the results under separating taxation, high-skill individuals are worse-off as β increases, while low-skill individuals are better-off, because an increase in β under pooling taxation relaxes the high-skill type's incentive-compatibility constraint. Moreover, the optimal marginal tax rates on labor income in period 1 for both types are independent of β. Since pooling takes place in period 1, there are no incentive-compatibility constraints in the first period; hence changes in β have no effect on the first-period marginal tax rates applicable to labor income. In period 2, the low-skill type's marginal tax rate on labor income is decreasing in β, because period 2 is when the government faces the high-skill type's incentive-compatibility constraint. As increases in β relax the high-skill type's incentive-compatibility constraint, the government can reduce the marginal tax rate applicable to the low-skill type's labor income. Finally, the marginal tax rates on savings for both types in both periods are increasing in β, again simply because the need to correct the effects of quasi-hyperbolic discounting is attenuated. Although the sign of the low-skill type's first-period marginal tax rate on savings is theoretically ambiguous, in our numerical simulations it is negative, indicating that the corrective motive *vis-a-vis* quasi-hyperbolic discounting for marginal savings distortions dominates the redistributive motive *vis-a-vis* the high-skill type's incentive-compatibility constraint. The marginal tax rates on savings for both types in period 2 approach zero as β rises, because there is only the corrective motive for inducing marginal savings distortions. However, for lower values of β the savings subsidy for low-skill individuals increases relative to that for high-skill individuals, because this makes it easier for the government to satisfy the high-skill type's incentive-compatibility constraint.

6. Discussion: commitment by the government

We have assumed that the government cannot commit to its future tax policy, because as discussed earlier we think that the no-commitment case is of significant interest in itself, and is also more consistent with the spirit of the Mirrlees approach to nonlinear income taxation. Nevertheless, since much of the dynamic nonlinear income tax literature has assumed commitment by the government, one may wonder how our main results change if the government can commit.

If the government can commit in our model, it will solve program (3.1)–(3.3) but subject to the additional constraint:

$$u(m_H^1 - s_H^1) - v\left(\frac{y_H^1}{w_H}\right) + \beta\delta\left[u(m_H^2 + (1+r)s_H^1 - s_H^2) - v\left(\frac{y_H^2}{w_H}\right)\right] + \beta\delta^2 u((1+r)s_H^2) \geq u(m_L^1 - s_L^1) - v\left(\frac{y_L^1}{w_H}\right)$$
$$+\beta\delta\left[u(m_L^2 + (1+r)s_L^1 - s_L^2) - v\left(\frac{y_L^2}{w_H}\right)\right] + \beta\delta^2 u((1+r)s_L^2) \tag{6.1}$$

which is the high-skill type's incentive-compatibility constraint. The effect of quasi-hyperbolic discounting on social welfare when the government can commit can be found by applying the Envelope Theorem to program (3.1)–(3.3) and (6.1):

$$\frac{\partial W}{\partial \beta} = \theta_H \left\{\delta\left(u(c_H^2) - v\left(\frac{y_H^2}{w_H}\right)\right) + \delta^2 u(c_H^3) - \delta\left(u(c_L^2) - v\left(\frac{y_L^2}{w_H}\right)\right) - \delta^2 u(c_L^3)\right\} \gtreqless 0 \tag{6.2}$$

where W denotes the level of social welfare attainable from program (3.1)–(3.3) and (6.1), and $\theta_H > 0$ is the Lagrange multiplier on constraint (6.1).

When the government can commit, the effect of quasi-hyperbolic discounting on social welfare is ambiguous. This is because satisfaction of the incentive-compatibility constraint (6.1) requires only that high-skill individuals be indifferent between choosing the lifetime tax treatments $\langle m_H^1, y_H^1, s_H^1, m_H^2, y_H^2, s_H^2\rangle$ and $\langle m_L^1, y_L^1, s_L^1, m_L^2, y_L^2, s_L^2\rangle$. There is no need for them to be indifferent between the sub-parts of these tax treatments that relate to periods 2 and 3, i.e., the parts directly affected by quasi-hyperbolic discounting. Accordingly, constraint (6.1) can potentially be satisfied with high-skill individuals strictly preferring their own tax treatment over periods 2 and 3 to that intended for low-skill individuals, or vice versa, depending upon the specific details regarding preferences and the parameters of the model. This renders the effects of quasi-hyperbolic discounting on social welfare theoretically ambiguous when the government can commit.

7. Discussion: sophisticated agents

We have assumed that individuals are naive, rather than sophisticated, quasi-hyperbolic discounters, but this assumption has no bearing on our main results regarding social welfare. Indeed, we think one of the strengths of our results are their generality, in that *any* behavioral assumption in which individuals discount the near-future too much relative to their long-run preferences would lead to the same conclusions. We examine quasi-hyperbolic discounting simply because it is the most common behavioral assumption in which individuals exhibit excessive short-run impatience.

Likewise, as our discussion of the effects of quasi-hyperbolic discounting on the optimal marginal tax rates suggests, these tend to be straightforward extensions of existing results on optimal savings taxation. That is, quasi-hyperbolic discounting makes a case for subsidizing savings, *ceteris paribus*, to offset its effects, and this applies regardless of whether agents are naive or sophisticated. However, the expressions for the marginal tax rate equations are derived from how individuals would behave in the absence of taxation (see Section 2), and such behavior does depend upon whether individuals are naive or sophisticated. Sophisticated agents feel the need for immediate gratification, but they are also aware that they will feel this need again in the future. Therefore, in the absence of taxation, their behavior can be modeled as a dynamic game played by their "present-self" and their "future-self". That is, in period 2 the individual chooses c_i^2, l_i^2, s_i^2, and c_i^3 to maximize:

$$u(c_i^2) - v(l_i^2) + \beta\delta u(c_i^3) \tag{7.1}$$

subject to:

$$c_i^2 + s_i^2 \le (1+r)s_i^1 + w_i l_i^2 \tag{7.2}$$

$$c_i^3 \le (1+r)s_i^2 \tag{7.3}$$

which yields solutions for the choice variables as functions of the parameters of the program and first-period savings, i.e., $c_i^2(\beta, \delta, r, s_i^1, w_i)$, $l_i^2(\cdot)$, $s_i^2(\cdot)$, and $c_i^3(\cdot)$.

In period 1, the individual knows that he/she will solve program (7.1)–(7.3) in period 2. Therefore, in period 1 the individual chooses c_i^1, l_i^1, and s_i^1 to maximize:

$$u(c_i^1) - v(l_i^1) + \beta\delta[u(c_i^2(\cdot)) - v(l_i^2(\cdot))] + \beta\delta^2 u(c_i^3(\cdot)) \tag{7.4}$$

subject to:

$$c_i^1 + s_i^1 \le w_i l_i^1 \tag{7.5}$$

It is shown in the Appendix that the solutions to programs (7.1)–(7.3) and (7.4)–(7.5) yield the following marginal conditions that would hold in the absence of taxation:

$$\frac{v'(l_i^t)}{u'(c_i^t)w_i} = 1\,(\text{for}\, t = 1, 2), \quad \frac{u'(c_i^1)}{\beta\delta(1+r)u'(c_i^2)} > 1 \quad \text{and} \quad \frac{u'(c_i^2)}{\beta\delta(1+r)u'(c_i^3)} = 1 \tag{7.6}$$

Comparing these with those obtained when agents are naive (see Eqs. (2.6) and (2.8)), it can be seen that the only difference is the marginal condition relating first- and second-period consumption. Specifically, in the absence of taxation, sophisticated agents save more in period 1 than do naive agents. The intuition is fairly straightforward. Naive agents feel the need for immediate gratification in period 1, and therefore save less than they should, but they falsely believe that they will behave rationally in the future and save optimally in period 2. On the other hand, sophisticated agents know that they will feel the need for immediate gratification again in the future, and therefore their first-period under-saving will not be alleviated by saving appropriately in period 2. Accordingly, sophisticated agents save more than naive agents, and it then follows that the government has less need to correct first-period under-saving by individuals when they are sophisticated, so the implicit marginal subsidy will be lower.

8. Conclusion

In this paper we have examined, both theoretically and numerically, the effects of incorporating quasi-hyperbolic discounting by individuals into a dynamic model of optimal nonlinear income taxation without commitment. Although quasi-hyperbolic discounting calls for marginal tax distortions to correct its effects, social welfare is not necessarily reduced. In fact, when separating taxation is optimal, quasi-hyperbolic discounting raises the level of social welfare attainable and therefore makes some individuals better-off in the long run. Furthermore, our numerical simulations show that even under pooling taxation some individuals are better-off in the long run as a result of quasi-hyperbolic discounting. We view these findings as being of both theoretical interest and of demonstrating the power of Mirrlees-style nonlinear taxation. At the very least, our main results stand in stark contrast to the usual result that quasi-hyperbolic discounting, or any form of excessive short-run impatience, makes individuals worse-off in the long run.

Although our model is relatively simple, the intuition driving our main results suggests that they would hold-up in more complex settings. Nevertheless, two potential extensions of our work come to mind. The first would be to extend the model to more than two types, but as discussed earlier this does not seem feasible as the number of possible tax regimes increases exponentially. However, imposing additional structure on the model to restrict the number of potentially optimal tax regimes may make it possible to examine the many-type case. The second extension would be to move beyond three periods, and possibly to an infinite-horizon setting. One does, however, run into the same problem as going to more than two types, in that the number of possible tax regimes increases.[19] But again, imposing additional restrictions on the model may make analysis of the many-period or infinite-horizon settings feasible.

Appendix A.

A.1. Individual behavior without taxation

The Lagrangian corresponding to program (2.2)–(2.5) is:

$$\mathcal{L}_i^1(\cdot) = u(c_i^1) - v(l_i^1) + \beta\delta[u(c_i^2) - v(l_i^2)] + \beta\delta^2 u(c_i^3) + \alpha^1\{w_i l_i^1 - c_i^1 - s_i^1\} + \alpha^2\{(1+r)s_i^1 + w_i l_i^2 - c_i^2 - s_i^2\} + \alpha^3\{(1+r)s_i^2 - c_i^3\} \tag{A.1}$$

where $\alpha^1 > 0$, $\alpha^2 > 0$, and $\alpha^3 > 0$ are Lagrange multipliers. The relevant first-order conditions can be written as:

$$u'(c_i^1) - \alpha^1 = 0 \tag{A.2}$$

$$-v'(l_i^1) + \alpha^1 w_i = 0 \tag{A.3}$$

$$-\alpha^1 + \alpha^2(1+r) = 0 \tag{A.4}$$

$$\beta\delta u'(c_i^2) - \alpha^2 = 0 \tag{A.5}$$

$$-\beta\delta v'(l_i^2) + \alpha^2 w_i = 0 \tag{A.6}$$

$$-\alpha^2 + \alpha^3(1+r) = 0 \tag{A.7}$$

$$\beta\delta^2 u'(c_i^3) - \alpha^3 = 0 \tag{A.8}$$

Simple algebraic manipulation of these first-order conditions yields Eq. (2.6).

The Lagrangian corresponding to the maximization of Eq. (2.7) subject to Eqs. (2.4) and (2.5) is:

$$\mathcal{L}_i^2(\cdot) = u(c_i^2) - v(l_i^2) + \beta\delta u(c_i^3) + \alpha^2\{(1+r)s_i^1 + w_i l_i^2 - c_i^2 - s_i^2\} + \alpha^3\{(1+r)s_i^2 - c_i^3\} \tag{A.9}$$

where $\alpha^2 > 0$ and $\alpha^3 > 0$ are Lagrange multipliers. The relevant first-order conditions are:

$$u'(c_i^2) - \alpha^2 = 0 \tag{A.10}$$

$$-v'(l_i^2) + \alpha^2 w_i = 0 \tag{A.11}$$

$$-\alpha^2 + \alpha^3(1+r) = 0 \tag{A.12}$$

$$\beta\delta u'(c_i^3) - \alpha^3 = 0 \tag{A.13}$$

Simple algebraic manipulation of these first-order conditions yields Eq. (2.8).

A.2. Proof of Proposition 1

The Lagrangian associated with program (4.3)–(4.5) is:

$$\mathcal{L}^1(\cdot) = (1-\phi)\left\{u(m_L^1 - s_L^1) - v\left(\frac{y_L^1}{w_L}\right)\right\} + \phi\left\{u(m_H^1 - s_H^1) - v\left(\frac{y_H^1}{w_H}\right)\right\} + \delta Z_F^2(\cdot) + \lambda^1\{(1-\phi)[y_L^1 - m_L^1] + \phi[y_H^1 - m_H^1]\} + \theta_H^1\left\{u(m_H^1 - s_H^1) - v(\frac{y_H^1}{w_H}) + \beta\delta U_{HF}^2 - u(m_L^1 - s_L^1) + v\left(\frac{y_L^1}{w_H}\right) - \beta\delta U_{MF}^2\right\} \tag{A.14}$$

where $\lambda^1 > 0$ and $\theta_H^1 > 0$ are Lagrange multipliers. By the Envelope Theorem:

$$\frac{\partial W_S(\cdot)}{\partial \beta} = \frac{\partial \mathcal{L}^1(\cdot)}{\partial \beta} = \theta_H^1 \delta[U_{HF}^2 - U_{MF}^2] \tag{A.15}$$

where $W_S(\cdot)$ denotes the level of social welfare attainable under separating taxation.

[19] For example, going to a four-period model, with three periods of taxation, generates three possibilities: (i) separate in period 1, and use first-best taxation in periods 2 and 3, (ii) pool in periods 1 and 2, and use second-best taxation in period 3, and (iii) pool in period 1, separate in period 2, and use first-best taxation in period 3.

To determine the sign of the expression in (A.15), the first-order conditions corresponding to program (4.1)–(4.2) are:

$$u'(m_L^2+(1+r)s_L^1-s_L^2)-\lambda^2=0 \tag{A.16}$$

$$-v'\left(\frac{y_L^2}{w_L}\right)\frac{1}{w_L}+\lambda^2=0 \tag{A.17}$$

$$-u'(m_L^2+(1+r)s_L^1-s_L^2)+\delta(1+r)u'((1+r)s_L^2)=0 \tag{A.18}$$

$$u'(m_H^2+(1+r)s_H^1-s_H^2)-\lambda^2=0 \tag{A.19}$$

$$-v'\left(\frac{y_H^2}{w_H}\right)\frac{1}{w_H}+\lambda^2=0 \tag{A.20}$$

$$-u'(m_H^2+(1+r)s_H^1-s_H^2)+\delta(1+r)u'((1+r)s_H^2)=0 \tag{A.21}$$

$$(1-\phi)[y_L^2-m_L^2]+\phi[y_H^2-m_H^2]=0 \tag{A.22}$$

where $\lambda^2>0$ is the multiplier on the government's second-period budget constraint (4.2). Eqs. (A.16) and (A.19) imply that $c_L^2=c_H^2$, which using (A.18) and (A.21) implies that $c_L^3=c_H^3$. Furthermore, Eqs. (A.17) and (A.20) imply that $y_H^2>y_L^2$. Therefore, from (4.6) and (4.7) we obtain $U_{HF}^2<U_{MF}^2$, which using (A.15) establishes that $\partial W_S(\cdot)/\partial\beta<0$. □

A.3. Marginal tax rates under separating taxation

The results that low-skill individuals face a positive marginal tax rate on their labor income in period 1, while high-skill individuals face a zero marginal tax rate, are standard results so the proofs are omitted. Similarly, both types face zero marginal tax rates on labor income in period 2 because first-best taxation is used in that period.

To prove that $MTRS_L^2=(\beta-1)/\beta<0$, Eq. (A.18) can be rewritten as:

$$\beta\delta(1+r)u'(c_L^3)-u'(c_L^2)=\beta\delta(1+r)u'(c_L^3)-\delta(1+r)u'(c_L^3) \tag{A.23}$$

Dividing both sides of (A.23) by $\beta\delta(1+r)u'(c_L^3)$ yields:

$$1-\frac{u'(c_L^2)}{\beta\delta(1+r)u'(c_L^3)}=\frac{\beta-1}{\beta}<0 \tag{A.24}$$

which using Eq. (2.9) establishes that $MTRS_L^2=(\beta-1)/\beta<0$. Analogous manipulations of (A.21) establish that $MTRS_H^2=(\beta-1)/\beta<0$.

To prove that $MTRS_L^1<0$, the first-order conditions on m_L^1 and s_L^1 from program (4.3)–(4.5) can be written as, respectively:

$$(1-\phi-\theta_H^1)u'(c_L^1)-\lambda^1(1-\phi)=0 \tag{A.25}$$

$$-(1-\phi-\theta_H^1)u'(c_L^1)+\delta\frac{\partial Z_F^2(\cdot)}{\partial s_L^1}+\theta_H^1\beta\delta\left[\frac{\partial U_{HF}^2(\cdot)}{\partial s_L^1}-\frac{\partial U_{MF}^2(\cdot)}{\partial s_L^1}\right]=0 \tag{A.26}$$

The Lagrangian for program (4.1)–(4.2) is:

$$\begin{aligned}\mathcal{L}^2(\cdot)=&(1-\phi)\left\{u(m_L^2+(1+r)s_L^1-s_L^2)-v\left(\frac{y_L^2}{w_L}\right)+\delta u((1+r)s_L^2)\right\}\\&+\phi\left\{u(m_H^2+(1+r)s_H^1-s_H^2)-v\left(\frac{y_H^2}{w_H}\right)+\delta u((1+r)s_H^2)\right\}+\lambda^2\left\{(1-\phi)\left[y_L^2-m_L^2\right]+\phi\left[y_H^2-m_H^2\right]\right\}\end{aligned} \tag{A.27}$$

By the Envelope Theorem:

$$\frac{\partial Z_F^2(\cdot)}{\partial s_L^1}=\frac{\partial\mathcal{L}^2(\cdot)}{\partial s_L^1}=(1-\phi)u'(c_L^2)(1+r) \tag{A.28}$$

Note that (A.25) implies that $1-\phi-\theta_H^1>0$. Using (2.9), (4.6), (4.7) and (A.28), Eq. (A.26) can be rewritten as:

$$MTRS_L^1 := 1-\frac{u'(c_L^1)}{\beta\delta(1+r)u'(c_L^2)} = \frac{(1-\phi)(\beta-1)}{(1-\phi-\theta_H^1)\beta} + \frac{\theta_H^1}{(1-\phi-\theta_H^1)(1+r)u'(c_L^2)} \times \left[u'(c_L^2)\frac{\partial m_L^2(\cdot)}{\partial s_L^1} - \frac{v'(y_L^2/w_H)}{w_H}\frac{\partial y_L^2(\cdot)}{\partial s_L^1} - u'(c_H^2)\frac{\partial m_H^2(\cdot)}{\partial s_L^1} + \frac{v'(y_H^2/w_H)}{w_H}\frac{\partial y_H^2(\cdot)}{\partial s_L^1}\right] \tag{A.29}$$

where use has also been made of the facts that $c_L^2=c_H^2$ and $c_L^3=c_H^3$. Using (A.16) and (A.17) we obtain $u'(c_L^2)=v'(y_L^2/w_L)1/w_L > v'(y_L^2/w_H)\frac{1}{w_H}$. By applying the Implicit Function Theorem and Cramer's Rule to (A.16)–(A.22) it can be shown that:

$$\frac{\partial m_L^2(\cdot)}{\partial s_L^1} < \frac{\partial y_L^2(\cdot)}{\partial s_L^1} < 0, \quad \frac{\partial m_H^2(\cdot)}{\partial s_L^1} > 0 \quad \text{and} \quad \frac{\partial y_H^2(\cdot)}{\partial s_L^1} < 0 \tag{A.30}$$

It now follows from (A.29) that $MTRS_L^1<0$.

Finally, the result that $MTRS_H^1 \gtreqless 0$ has been established using numerical examples, details of which are available upon request.

A.4. Proof of Proposition 2

The Lagrangians associated with programs (4.11)–(4.12) and (4.8)–(4.10) are, respectively:

$$\mathcal{L}^1(\cdot) = (1-\phi)\left\{u(m^1-s^1)-v\left(\frac{y^1}{w_L}\right)\right\} + \phi\left\{u(m^1-s^1)-v\left(\frac{y^1}{w_H}\right)\right\} + \delta Z_S^2(\cdot) + \lambda^1\left\{y^1-m^1\right\} \tag{A.31}$$

$$\begin{aligned}\mathcal{L}^2(\cdot) = {} & (1-\phi)\left\{u(m_L^2+(1+r)s^1-s_L^2)-v\left(\frac{y_L^2}{w_L}\right)+\delta u((1+r)s_L^2)\right\} \\ & +\phi\left\{u(m_H^2+(1+r)s^1-s_H^2)-v\left(\frac{y_H^2}{w_H}\right)+\delta u((1+r)s_H^2)\right\} + \lambda^2\left\{(1-\phi)\left[y_L^2-m_L^2\right]+\phi\left[y_H^2-m_H^2\right]\right\} \\ & +\theta_H^2\{u(m_H^2+(1+r)s^1-s_H^2)-v\left(\frac{y_H^2}{w_H}\right)+\beta\delta u((1+r)s_H^2)-u(m_L^2+(1+r)s^1-s_L^2)+v\left(\frac{y_L^2}{w_H}\right)-\beta\delta u((1+r)s_L^2)\}\end{aligned} \tag{A.32}$$

where $\lambda^1>0$, $\lambda^2>0$, and $\theta_H^2>0$ are Lagrange multipliers. By repeated application of the Envelope Theorem:

$$\frac{\partial W_P(\cdot)}{\partial\beta} = \frac{\partial\mathcal{L}^1(\cdot)}{\partial\beta} = \delta\frac{\partial Z_S^2(\cdot)}{\partial\beta} = \delta\frac{\partial\mathcal{L}^2(\cdot)}{\partial\beta} = \delta^2\theta_H^2[u(c_H^3)-u(c_L^3)] \tag{A.33}$$

where $W_P(\cdot)$ denotes the level of social welfare attainable under pooling taxation.

The first-order conditions on m_L^2, s_L^2, m_H^2, and s_H^2 from program (4.8)–(4.10) can be written as, respectively:

$$(1-\phi-\theta_H^2)u'(c_L^2)-\lambda^2(1-\phi)=0 \tag{A.34}$$

$$-(1-\phi-\theta_H^2)u'(c_L^2)+(1-\phi-\beta\theta_H^2)\delta(1+r)u'(c_L^3)=0 \tag{A.35}$$

$$(\phi+\theta_H^2)u'(c_H^2)-\lambda^2\phi=0 \tag{A.36}$$

$$-(\phi+\theta_H^2)u'(c_H^2)+(\phi+\beta\theta_H^2)\delta(1+r)u'(c_H^3)=0 \tag{A.37}$$

Adding (A.34) and (A.35), and adding (A.36) and (A.37), and then combining the results of these additions yields:

$$\frac{u'(c_L^3)}{u'(c_H^3)} = \frac{(\phi+\beta\theta_H^2)(1-\phi)}{(1-\phi-\beta\theta_H^2)\phi} = \frac{\phi(1-\phi)+\beta\theta_H^2(1-\phi)}{\phi(1-\phi)-\phi\beta\theta_H^2} > 1 \tag{A.38}$$

which implies that $c_H^3>c_L^3$. It now follows from (A.33) that $\partial W_P(\cdot)/\partial\beta>0$. □

A.5. Marginal tax rates under pooling taxation

The first-order conditions on m^1 and y^1 from program (4.11)–(4.12) can be written as, respectively:

$$u'(c^1)-\lambda^1=0 \tag{A.39}$$

$$-(1-\phi)v'\left(\frac{y^1}{w_L}\right)\frac{1}{w_L}-\phi v'\left(\frac{y^1}{w_H}\right)\frac{1}{w_H}+\lambda^1=0 \tag{A.40}$$

Eqs. (A.39) and (A.40) can be manipulated to obtain:

$$MTRL_L^1 := 1-\frac{v'(y^1/w_L)}{u'(c^1)w_L}=\frac{\phi}{u'(c^1)}\left[v'\left(\frac{y^1}{w_H}\right)\frac{1}{w_H}-v'\left(\frac{y^1}{w_L}\right)\frac{1}{w_L}\right]<0 \tag{A.41}$$

which is negative because $w_H > w_L$ and $v(\cdot)$ is strictly convex. Similarly, (A.39) and (A.40) can be manipulated to obtain:

$$MTRL_H^1 := 1-\frac{v'\left(\frac{y^1}{w_H}\right)}{u'(c^1)w_H}=\frac{(1-\phi)}{u'(c^1)}\left[v'\left(\frac{y^1}{w_L}\right)\frac{1}{w_L}-v'\left(\frac{y^1}{w_H}\right)\frac{1}{w_H}\right]>0 \tag{A.42}$$

The results that $MTRL_L^2 > 0$ and $MTRL_H^2 = 0$ are standard results for second-best nonlinear income taxation, so the proofs are omitted.

The first-order condition on s^1 from program (4.11)–(4.12) is:

$$-u'(c^1)+\delta\frac{\partial Z_S^2(\cdot)}{\partial s^1}=0 \tag{A.43}$$

By the Envelope Theorem:

$$\frac{\partial Z_S^2(\cdot)}{\partial s^1}=\frac{\partial \mathcal{L}^2(\cdot)}{\partial s^1}=(1+r)[(1-\phi-\theta_H^2)u'(c_L^2)+(\phi+\theta_H^2)u'(c_H^2)] \tag{A.44}$$

Eqs. (A.43) and (A.44) can be manipulated to yield:

$$MTRS_H^1 := 1-\frac{u'(c^1)}{\beta\delta(1+r)u'(c_H^2)}=\frac{\beta-\phi-\theta_H^2}{\beta}-\frac{(1-\phi-\theta_H^2)u'(c_L^2)}{\beta u'(c_H^2)} \tag{A.45}$$

Eq. (A.34) implies that $1-\phi-\theta_H^2 > 0$, while straightforward manipulation of (A.34) and (A.36) establishes that $u'(c_L^2) > u'(c_H^2)$. It now follows from (A.45) that $MTRS_H^1 < 0$. The result that $MTRS_L^1 \gtreqless 0$ has been established using numerical examples, details of which are available upon request.

Eq. (A.35) can be manipulated to yield:

$$MTRS_L^2 := 1-\frac{u'(c_L^2)}{\beta\delta(1+r)u'(c_L^3)}=\frac{(1-\phi)(\beta-1)}{(1-\phi-\theta_H^2)\beta}<0 \tag{A.46}$$

and Eq. (A.37) can be manipulated to yield:

$$MTRS_H^2 := 1-\frac{u'(c_H^2)}{\beta\delta(1+r)u'(c_H^3)}=\frac{\phi(\beta-1)}{(\phi+\theta_H^2)\beta}<0 \tag{A.47}$$

which establishes that $MTRS_L^2 < MTRS_H^2 < 0$.

A.6. Sophisticated agents

The first-order conditions from program (7.1)–(7.3) are:

$$u'(c_i^2)-\alpha^2=0 \tag{A.48}$$

$$-v'(l_i^2)+\alpha^2 w_i=0 \tag{A.49}$$

$$-\alpha^2+\alpha^3(1+r)=0 \tag{A.50}$$

$$\beta\delta u'(c_i^3)-\alpha^3=0 \tag{A.51}$$

$$(1+r)s_i^1+w_i l_i^2-c_i^2-s_i^2=0 \tag{A.52}$$

$$(1+r)s_i^2-c_i^3=0 \tag{A.53}$$

where $\alpha^2 > 0$ and $\alpha^3 > 0$ are the Lagrange multipliers on Eqs. (7.2) and (7.3), respectively. Straightforward manipulation of the above equations yields the marginal conditions:

$$\frac{v'(l_i^2)}{u'(c_i^2)w_i} = 1 \quad \text{and} \quad \frac{u'(c_i^2)}{\beta\delta(1+r)u'(c_i^3)} = 1 \tag{A.54}$$

The relevant first-order conditions from program (7.4)–(7.5) are:

$$u'(c_i^1) - \alpha^1 = 0 \tag{A.55}$$

$$-v'(l_i^1) + \alpha^1 w_i = 0 \tag{A.56}$$

$$\beta\delta u'(c_i^2)\frac{\partial c_i^2(\cdot)}{\partial s_i^1} - \beta\delta v'(l_i^2)\frac{\partial l_i^2(\cdot)}{\partial s_i^1} + \beta\delta^2 u'(c_i^3)\frac{\partial c_i^3(\cdot)}{\partial s_i^1} - \alpha^1 = 0 \tag{A.57}$$

where $\alpha^1 > 0$ is the Lagrange multiplier on Eq. (7.5). Straightforward manipulation of (A.55) and (A.56) yields:

$$\frac{v'(l_i^1)}{u'(c_i^1)w_i} = 1 \tag{A.58}$$

By applying the Implicit Function Theorem and Cramer's Rule to (A.48)–(A.53) it can be shown that:

$$\frac{\partial c_i^2(\cdot)}{\partial s_i^1} = \frac{-v''(l_i^2)\beta\delta(1+r)^3 u''(c_i^3)}{-u''(c_i^2)v''(l_i^2) + \beta\delta(1+r)^2 u''(c_i^3)[u''(c_i^2)w_i^2 - v''(l_i^2)]} > 0 \tag{A.59}$$

$$\frac{\partial l_i^2(\cdot)}{\partial s_i^1} = \frac{-u''(c_i^2)w_i\beta\delta(1+r)^3 u''(c_i^3)}{-u''(c_i^2)v''(l_i^2) + \beta\delta(1+r)^2 u''(c_i^3)[u''(c_i^2)w_i^2 - v''(l_i^2)]} < 0 \tag{A.60}$$

$$\frac{\partial c_i^3(\cdot)}{\partial s_i^1} = \frac{-v''(l_i^2)u''(c_i^2)(1+r)^2}{-u''(c_i^2)v''(l_i^2) + \beta\delta(1+r)^2 u''(c_i^3)[u''(c_i^2)w_i^2 - v''(l_i^2)]} > 0 \tag{A.61}$$

Substituting (A.59), (A.60), and (A.61) into Eq. (A.57) and undertaking some algebraic manipulation yields:

$$u'(c_i^1) = \beta\delta(1+r)u'(c_i^2)\left[\frac{-u''(c_i^2)v''(l_i^2) + \beta^2\delta(1+r)^2 u''(c_i^3)[u''(c_i^2)w_i^2 - v''(l_i^2)]}{-\beta u''(c_i^2)v''(l_i^2) + \beta^2\delta(1+r)^2 u''(c_i^3)[u''(c_i^2)w_i^2 - v''(l_i^2)]}\right] \tag{A.62}$$

which implies that $u'(c_i^1) > \beta\delta(1+r)u'(c_i^2)$.

References

Amador, M., Werning, I., Angeletos, G., 2006. Commitment versus flexibility. Econometrica 74, 365–396.
Apps, P., Rees, R., 2006. Repeated Optimal Nonlinear Income Taxation (mimeo).
Aronsson, T., Granlund, D., 2011. Public goods and optimal paternalism under present-biased preferences. Econ. Lett. 113, 54–57.
Aronsson, T., Sjogren, T., 2009. Quasi-hyperbolic discounting and mixed taxation. In: Umea Economic Studies Working Paper, No. 778.
Bassi, M., 2010. Mirrlees meets Laibson: optimal income taxation with bounded rationality. In: CSEF Working Paper, No. 266.
Berliant, M., Ledyard, J., 2014. Optimal dynamic nonlinear income taxes with no commitment. J. Publ. Econ. Theory 16, 196–221.
Brett, C., Weymark, J., 2008. Optimal nonlinear taxation of income and savings without commitment. In: Vanderbilt University Working Paper, No. 08-W05.
Brett, C., Weymark, J., 2011. How optimal nonlinear income taxes change when the distribution of the population changes. J. Publ. Econ. 95, 1239–1247.
Brito, D., Hamilton, J., Slutsky, S., Stiglitz, J., 1991. Dynamic optimal income taxation with government commitment. J. Publ. Econ. 44, 15–35.
Chetty, R., 2006. A new method of estimating risk aversion. Am. Econ. Rev. 96, 1821–1834.
Chetty, R., Guren, A., Manoli, D., Weber, A., 2011. Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins. Am. Econ. Rev. Pap. Proc. 101, 471–475.
Diamond, P., Koszegi, B., 2003. Quasi-hyperbolic discounting and retirement. J. Publ. Econ. 87, 1839–1872.
Diamond, P., Spinnewijn, J., 2011. Capital income taxes with heterogeneous discount rates. Am. Econ. J.: Econ. Policy 3, 52–76.
Fang, H., 2006. Disentangling the college wage premium: estimating a model with endogenous education choices. Int. Econ. Rev. 47, 1151–1185.
Farhi, E., Sleet, C., Werning, I., Yeltekin, S., 2012. Nonlinear capital taxation without commitment. Rev. Econ. Stud. 79, 1469–1493.
Frederick, S., Loewenstein, G., O'Donoghue, T., 2002. Time discounting and time preference: a critical review. J. Econ. Lit. 40, 351–401.
Goldin, C., Katz, L., 2007. The race between education and technology: the evolution of U.S. educational wage differentials, 1890–2005. In: NBER Working Paper, No. 12984.
Golosov, M., Troshkin, M., Tsyvinski, A., Weinzeirl, M., 2013. Preference heterogeneity and optimal capital income taxation. J. Publ. Econ. 97, 160–175.
Golosov, M., Tsyvinski, A., Werning, I., 2006. New dynamic public finance: a user's guide. In: NBER Macroeconomics Annual., pp. 317–387.
Guesnerie, R., 1995. A Contribution to the Pure Theory of Taxation. Cambridge University Press.
Guo, J.-T., Krause, A., 2011a. Optimal nonlinear income taxation with habit formation. J. Publ. Econ. Theory 13, 463–480.
Guo, J.-T., Krause, A., 2011b. Dynamic Income Taxation without Commitment: Comparing Alternative Tax Systems (mimeo).
Guo, J.-T., Krause, A., 2013. Optimal nonlinear taxation of income and education expenditures. Oxf. Econ. Pap. 65, 74–95.
Hey, J., Lotito, G., 2009. Naive, resolute or sophisticated? A study of dynamic decision making. J. Risk Uncertain. 38, 1–25.
Kocherlakota, N., 2010. The New Dynamic Public Finance. Princeton University Press.
Krause, A., 2009. Optimal nonlinear income taxation with learning-by-doing. J. Publ. Econ. 93, 1098–1110.
Krusell, P., Kuruscu, B., Smith, A., 2002. Equilibrium welfare and government policy with quasi-geometric discounting. J. Econ. Theory 105, 42–72.
Krusell, P., Kuruscu, B., Smith, A., 2010. Temptation and taxation. Econometrica 78, 2063–2084.
Laibson, D., 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112, 443–477.

Mirrlees, J., 1971. An exploration in the theory of optimum income taxation. Rev. Econ. Stud. 38, 175–208.
O'Donoghue, T., Rabin, M., 1999. Doing it now or later. Am. Econ. Rev. 89, 103–124.
O'Donoghue, T., Rabin, M., 2003. Studying optimal paternalism, illustrated by a model of sin taxes. Am. Econ. Rev. 93, 186–191.
O'Donoghue, T., Rabin, M., 2006. Optimal sin taxes. J. Publ. Econ. 90, 1825–1849.
OECD, 2010. Highlights from Education at a Glance. OECD Publications.
Roberts, K., 1984. The theoretical limits to redistribution. Rev. Econ. Stud. 51, 177–195.
Simula, L., 2010. Optimal nonlinear income tax and nonlinear pricing: optimality conditions and comparative static properties. Soc. Choice Welfare 35, 199–220.
Stiglitz, J., 1982. Self-selection and pareto efficient taxation. J. Publ. Econ. 17, 213–240.
Tenhunen, S., Tuomala, M., 2010. On optimal lifetime redistribution policy. J. Publ. Econ. Theory 12, 171–198.