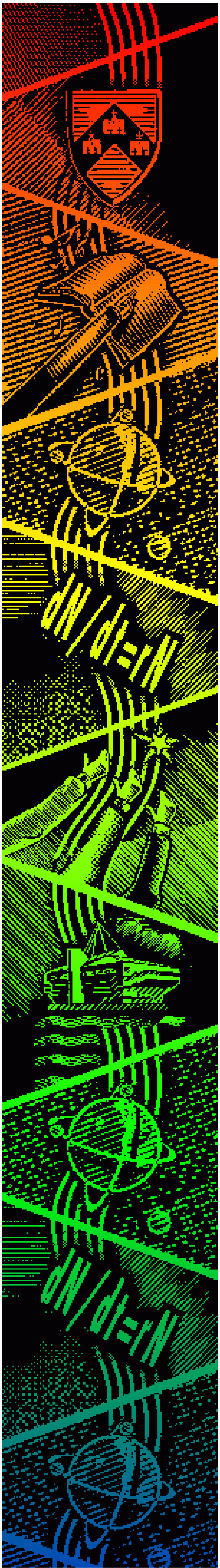


THE UNIVERSITY of York

Discussion Papers in Economics

No. 11/16

Dynamic Nonlinear Income Taxation with Quasi-Hyperbolic Discounting and No Commitment

By

Jang-Ting Guo, University of California, Riverside
Alan Krause, University of York

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

Dynamic Nonlinear Income Taxation with Quasi-Hyperbolic Discounting and No Commitment

Jang-Ting Guo*
University of California, Riverside

Alan Krause†
University of York

3 August 2011

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 uses its taxation powers to maximise a utilitarian social welfare function which reflects individuals' true (long-run) preferences. Under first-best taxation, quasi-hyperbolic discounting has 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. The effects of quasi-hyperbolic discounting on the optimal marginal tax rates applicable to labour and savings income are also explored.

Keywords: dynamic taxation; quasi-hyperbolic discounting; commitment.

JEL Classifications: D91, H21, H24.

*Department of Economics, 4123 Sproul Hall, University of California, Riverside, CA, 92521, U.S.A., Phone: 1-951-827-1588, Fax: 1-951-827-5685, E-mail: guojt@ucr.edu.

†Corresponding Author. Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, U.K., Phone: 44-(0)1904-423-572, Fax: 44-(0)1904-423-759, E-mail: alan.krause@york.ac.uk.

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.[1] This leads agents to make short-run decisions that they later regret as not being consistent with their long-run preferences. Such behaviour is often described as an individual imposing a negative "internality" on their future self, which potentially justifies corrective (or paternalistic) policy intervention.[2] There is, in effect, preference heterogeneity between individuals and the government, as the government's preferences are the same as individuals' long-run preferences, but not their short-run preferences. Policy is based on long-run preferences. In our model economy, an individual's need for immediate gratification leads them to make labour, 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 labour and savings.

There is currently a great deal of interest in dynamic nonlinear income taxation, such as the "new dynamic public finance" literature which extends the static Mirrlees [1971] model of optimal nonlinear income taxation to a dynamic setting.[3] 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 associ-

[1] See, e.g., the survey article by Frederick, et al. [2002].

[2] For example, O'Donoghue and Rabin [2006] 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].

[3] 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].

ated 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 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 the revelation principle may no longer hold. That is, it may no longer be social-welfare maximising 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.[4] To minimise the problems that the possible optimality of separating or pooling taxation present, we adopt the two-type (high-skill and low-skill) version of the Mirrlees model introduced by Stiglitz [1982],[5] and analyse 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 labour and savings in periods 1 and 2 such that a utilitarian social

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

[5]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.

welfare function based on individuals' true (long-run) preferences is maximised. Hence, the social welfare function captures a corrective motive for 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. Therefore, even though quasi-hyperbolic discounting calls for corrective taxation, it does not necessarily reduce social welfare relative to that attainable if individuals behaved according to their long-run preferences. The intuition, in a nutshell, can be summarised 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 behaviour by making sure that the allocations satisfy the high-skill type's incentive-compatibility constraint. Quasi-hyperbolic discounting does, however, affect the incentive-compatibility constraint, since 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 has no effect on social welfare under first-best taxation, since in this case the allocations need not be incentive compatible. Finally, we explore the effects of quasi-hyperbolic discounting on the optimal marginal tax rates applicable to labour and savings income, both analytically and numerically.

This paper complements a recent literature that examines labour and savings/capital taxation in dynamic Mirrlees-style optimal nonlinear income tax models. What distinguishes our work from all of these studies, however, is that none of them consider quasi-hyperbolic discounting. Brett and Weymark [2008] and Farhi, et al. [2011] examine nonlinear savings/capital taxation in dynamic Mirrlees models without commitment. 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.[6] They highlight the possibilities that separating or pooling taxation can be social-welfare maximising, 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 political-economy constraints, which take the form of direct costs 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 tax treatment of capital income is progressive.

The models examined by Golosov, et al. [2010], Tenhunen and Tuomala [2010], and Diamond and Spinnewijn [2011] assume full commitment, but individuals are distinguished by both skills and their preferences for savings. Since 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,

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

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 organised 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 concludes, while proofs are relegated to an appendix.

2 Analytical Framework

The economy is assumed to last for three periods.[7] 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:

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

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

[7] We consider a finite-horizon model because the no-commitment problems faced by individuals and the government are both 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.

concave, while $v(\cdot)$ is increasing and strictly convex.

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

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

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 labour supply decisions that are not consistent with long-run utility maximisation.

2.1 Individual Behaviour 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. 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. We assume that individuals are sophisticated.[8] In this case, individual i's behaviour in period 2 can be described as follows. Choose c_i^2, l_i^2, s_i^2, and c_i^3 to maximise:

$$u(c_i^2) - v(l_i^2) + \beta\delta u(c_i^3) \tag{2.3}$$

subject to the budget constraints:

$$c_i^2 + s_i^2 \leq (1 + r)s_i^1 + w_i l_i^2 \qquad \text{and} \qquad c_i^3 \leq (1 + r)s_i^2 \tag{2.4}$$

[8] As it turns out, making the alternative assumption that individuals are naive would have no effect on our qualitative conclusions. There would, however, be some minor effects on our quantitative results. We assume that individuals are sophisticated because, as O'Donoghue and Rabin [1999] and Diamond and Koszegi [2003] point out, it is consistent with the usual assumption in economics that individuals are rational.

where s_i^t represents individual i's savings in period t, and $r > 0$ is the interest rate. Let $V_i^2(\beta, \delta, r, s_i^1, w_i)$ denote the value function associated with programme $(2.3) - (2.4)$.

In period 1, individual i knows that she will solve programme $(2.3) - (2.4)$ in period 2. Therefore, in period 1, individual i chooses c_i^1, l_i^1, and s_i^1 to maximise:

$$u(c_i^1) - v(l_i^1) + \beta\delta V_i^2(\cdot) \tag{2.5}$$

subject to the budget constraint:

$$c_i^1 + s_i^1 \leq w_i l_i^1 \tag{2.6}$$

It is shown in the Appendix that the solutions to programmes $(2.3) - (2.4)$ and $(2.5) - (2.6)$ yield the following marginal conditions between labour and consumption in period t, and between consumption in periods t and $t+1$ (for $t = 1, 2$):

$$\frac{v'(l_i^t)}{u'(c_i^t)w_i} = 1 \qquad \text{and} \qquad \frac{u'(c_i^t)}{\beta\delta(1+r)u'(c_i^{t+1})} = 1 \tag{2.7}$$

Since $\beta \neq 1$, it follows that the allocation implied by the marginal conditions in (2.7) differs from what would be optimal according to the individual's true utility function (2.1). Therefore, under laissez faire quasi-hyperbolic discounting makes each individual worse-off in the long run, which implies that any welfarist-based measure of social welfare would also be lower in the long run.

2.2 Implicit Marginal Tax Rates

As we assume that the government can impose nonlinear taxes on an individual's income from labour and savings, it may be optimal for the government to set taxes to induce violations of the marginal conditions shown in equation (2.7). Following the standard practice, one may interpret these marginal distortions as tax wedges or implicit marginal tax rates. Accordingly, we define:

$$MTRL_i^t := 1 - \frac{v'(l_i^t)}{u'(c_i^t)w_i} \qquad \text{and} \qquad MTRS_i^t := 1 - \frac{u'(c_i^t)}{\beta\delta(1+r)u'(c_i^{t+1})} \tag{2.8}$$

where $MTRL_i^t$ denotes the (implicit) marginal tax rate on labour 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 with the hypothetical case in which the government knows each individual's skill type in all three periods. In this case, the government's choice of an optimal nonlinear tax system applicable to labour 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 maximise:[9]

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

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

subject to:

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

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

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 $c_i^3 = (1+r)s_i^2$. Equation (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. Equations (3.2) and (3.3) are, respectively, the government's first- and second-period budget constraints. Implicit in equations (3.2) and (3.3) is the simplifying assumption that the government cannot save or borrow. Since it is postulated that the government can observe each individual's type from period 1

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

onwards, it does not face any incentive-compatibility constraints.

It is shown in the Appendix that the solution to programme $(3.1)-(3.3)$ yields:

Proposition 1 *Under first-best taxation*: (i) *quasi-hyperbolic discounting has no effect on the level of social welfare attainable*, (ii) $MTRL_L^1 = MTRL_H^1 = MTRL_L^2 = MTRL_H^2 = 0$, *and* (iii) $MTRS_L^1 = MTRS_H^1 = MTRS_L^2 = MTRS_H^2 = (\beta - 1)/\beta < 0$.

The result that quasi-hyperbolic discounting has 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 programme $(3.1)-(3.3)$. 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. Both types face zero marginal tax rates on their labour income, which reflects the lump-sum nature of first-best taxation. However, both types face negative marginal tax rates on their savings. 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 Second-Best Taxation

In this section, we examine nonlinear labour 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. Since 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 maximising social welfare. Accordingly, rather than design 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 maximising, depending upon the parameters of the model.[10] Therefore, we examine in turn the nature of separating and pooling nonlinear labour and savings taxation.

4.1 Separating Taxation

If the tax system was 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 behaviour in period 2 can 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 maximise:

$$(1-\phi)\left\{u(m_L^2 + (1+r)s_L^1 - s_L^2) - v\left(\frac{y_L^2}{w_L}\right) + \delta u((1+r)s_L^2)\right\}$$

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

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 programme (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 by assumption it chooses to design a separating tax system in order to obtain skill-

[10]See, e.g., the papers by Roberts [1984], Berliant and Ledyard [2005], and Guo and Krause [2011b] which examine the desirability of separating versus pooling nonlinear income taxation when the government cannot commit.

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 maximise:

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

subject to:

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

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

where:

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

and:

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

Equation (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. Equation (4.3) therefore includes the value function $Z_F^2(\cdot)$. Equation (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. Thus the 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 programmes (4.1) – (4.2) and (4.3) – (4.5) together imply:

Proposition 2 *Under second-best taxation with separation in the first period*: (i) *quasi-hyperbolic discounting increases the level of social welfare attainable*, (ii) $MTRL_L^1 > 0$ *and* $MTRL_H^1 = MTRL_L^2 = MTRL_H^2 = 0$, *and* (iii) $MTRS_L^1 < 0$, $MTRS_H^1 \gtreqless 0$ *and* $MTRS_L^2 = MTRS_H^2 = (\beta - 1)/\beta < 0$.

Interestingly, quasi-hyperbolic discounting *increases* the level of social welfare attainable, even though it calls for corrective action which requires inducing marginal distortions. 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. And 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* in 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.[11] Accordingly, high-skill individuals must be offered a relatively favourable tax treatment in period 1 if they are to reveal their type, in order to compensate them for the very unfavourable 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 enables the government to attain a higher level of social welfare.

In period 1, high-skill individuals face a zero marginal tax rate on their labour 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 labour 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.[12] 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

[11] 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.

[12] That is, the government will seek to equate second-period consumption levels. This follows from equations (A.18) and (A.21) in the Appendix.

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 made above 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 first-best taxation is used in period 2, so as in Proposition 1 there is only the corrective motive for inducing marginal savings distortions.

4.2 Pooling Taxation

If the tax system pooled 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. It 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 maximise:

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

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

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. Equation (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.[13] Since 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.

The solution to programme (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, by assumption, elects to implement pooling taxation. Therefore, it chooses a single allocation $\langle m^1, y^1, s^1 \rangle$ for all individuals to maximise:

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

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 $Z_S^2(\cdot)$. Equation (4.12) is the government's first-period budget constraint.

It is shown in the Appendix that the solutions to programmes (4.8) $-$ (4.10) and (4.11) $-$ (4.12) together imply:

Proposition 3 *Under second-best taxation with pooling in the first period*: (*i*) *quasi-hyperbolic discounting decreases the level of social welfare attainable*, (*ii*) $MTRL_L^1 < 0$, $MTRL_H^1 > 0$, $MTRL_L^2 > 0$ *and* $MTRL_H^2 = 0$, *and* (*iii*) $MTRS_L^1 \gtreqless 0$, $MTRS_H^1 < 0$ *and* $MTRS_L^2 < MTRS_H^2 < 0$.

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 con-

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

straint. 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.

In period 1, low-skill individuals face a negative marginal tax rate on their labour 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 labour supply being distorted upwards (resp. downwards) to earn y^1. In period 2 the usual pattern of optimal marginal tax rates on labour 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 Quantitative Analysis

In this section, we use numerical simulations to further explore the effects of quasi-hyperbolic discounting. To this end, we assume that:

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

where $\gamma > 0$ is a preference parameter. For our numerical simulations we set $\gamma = 2$, as this implies a Frisch labour supply elasticity of 0.5 which is in line with empirical estimates (see, e.g., Chetty, et al. [2011]).

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, i.e., the average difference in the wages of university graduates over high-school graduates, is approximately 60%. We therefore normalise 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 $0.75 - 0.95$ on each type's true (long-run) utility and on the optimal marginal tax rates,[14] holding all other parameters at their baseline levels. These effects are shown

[14] The effects of varying β on all non-zero marginal tax rates are explored, except for those equal to

in Figure 1 for separating taxation, and in Figure 2 for pooling taxation.

In Figure 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 2). 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 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. The optimal marginal tax rate applicable to the low-skill type's labour 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 labour 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 labour income. 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 Figure 2, which covers pooling taxation, social welfare is increasing in β (cf. Proposition 3). 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. The optimal marginal tax rates on labour 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 labour income. In period 2, the low-skill type's marginal tax rate on labour income is decreasing in β, because period 2 is when the

$(\beta - 1)/\beta$ because the effect in this case is obvious.

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 labour income. 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 increases, 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.

Finally, it is reported in Table 1 that for the empirically-plausible set of parameter values considered, separating taxation yields a higher level of social welfare than pooling taxation. Therefore, it can be concluded that it is optimal for the government to implement separating taxation, which implies that quasi-hyperbolic discounting raises the level of social welfare attainable using nonlinear labour and savings taxation without commitment.

6 Conclusion

In this paper we have examined, both theoretically and quantitatively, 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, as our quantitative analysis suggests is the case, quasi-hyperbolic discounting raises the level of social welfare attainable.

Two extensions of our work immediately 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. Nevertheless, by imposing additional structure on the model to restrict the number of potentially optimal tax regimes, it may be 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.[15] But again, imposing additional restrictions on the model may make analysis of the many-period or infinite-horizon settings feasible.

7 Appendix

Individual Behaviour Without Taxation

The Lagrangian corresponding to programme (2.3) – (2.4) is:

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

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

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

$$-v'(l_i^2) + \alpha^2 w_i = 0 \tag{A.3}$$

$$-\alpha^2 + \alpha^3(1+r) = 0 \tag{A.4}$$

$$\beta\delta u'(c_i^3) - \alpha^3 = 0 \tag{A.5}$$

The relevant first-order conditions corresponding to programme (2.5) – (2.6) are:

$$u'(c_i^1) - \alpha^1 = 0 \tag{A.6}$$

[15] 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.

$$-v'(l_i^1) + \alpha^1 w_i = 0 \tag{A.7}$$

$$\beta\delta \frac{\partial V_i^2(\cdot)}{\partial s_i^1} - \alpha^1 = 0 \tag{A.8}$$

where $\alpha^1 > 0$ is the multiplier on the first-period budget constraint (2.6). By application of the Envelope Theorem:

$$\frac{\partial V_i^2(\cdot)}{\partial s_i^1} = \frac{\partial \mathcal{L}_i^2(\cdot)}{\partial s_i^1} = \alpha^2(1+r) \tag{A.9}$$

Straightforward manipulation of (A.2) – (A.9) leads to equation (2.7). ■

Proof of Proposition 1

Part (i) follows from the fact that β does not appear in programme (3.1) – (3.3). Part (ii) is a standard result, so the proof is omitted. To prove part (iii), the first-order conditions on s_L^1, s_H^1, s_L^2, and s_H^2 can be written as, respectively:

$$-u'(c_L^1) + \delta(1+r)u'(c_L^2) = 0 \tag{A.10}$$

$$-u'(c_H^1) + \delta(1+r)u'(c_H^2) = 0 \tag{A.11}$$

$$-u'(c_L^2) + \delta(1+r)u'(c_L^3) = 0 \tag{A.12}$$

$$-u'(c_H^2) + \delta(1+r)u'(c_H^3) = 0 \tag{A.13}$$

Equation (A.10) can be rewritten as:

$$\beta\delta(1+r)u'(c_L^2) - u'(c_L^1) = \beta\delta(1+r)u'(c_L^2) - \delta(1+r)u'(c_L^2) \tag{A.14}$$

Dividing both sides of (A.14) by $\beta\delta(1+r)u'(c_L^2)$ yields:

$$1 - \frac{u'(c_L^1)}{\beta\delta(1+r)u'(c_L^2)} = \frac{\beta-1}{\beta} < 0 \tag{A.15}$$

which using equation (2.8) establishes that $MTRS_L^1 = (\beta-1)/\beta < 0$. Analogous manipulations of (A.11), (A.12) and (A.13) establish that $MTRS_H^1 = MTRS_L^2 = MTRS_H^2 = (\beta-1)/\beta < 0$. ■

Proof of Proposition 2

To prove part (i), the Lagrangian associated with programme (4.3) – (4.5) is:

$$\mathcal{L}^1(\cdot) = (1-\phi)\left\{u(m_L^1 - s_L^1) - v\left(\frac{y_L^1}{w_L}\right)\right\} + \phi\left\{u(m_H^1 - s_H^1) - v\left(\frac{y_H^1}{w_H}\right)\right\} + \delta Z_F^2(\cdot)$$
$$+ \lambda^1\left\{(1-\phi)\left[y_L^1 - m_L^1\right] + \phi\left[y_H^1 - m_H^1\right]\right\}$$
$$+ \theta_H^1\left\{u(m_H^1 - s_H^1) - v\left(\frac{y_H^1}{w_H}\right) + \beta\delta U_{HF}^2 - u(m_L^1 - s_L^1) + v\left(\frac{y_L^1}{w_H}\right) - \beta\delta U_{MF}^2\right\} \tag{A.16}$$

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

$$\frac{\partial W_S(\cdot)}{\partial\beta} = \frac{\partial\mathcal{L}^1(\cdot)}{\partial\beta} = \theta_H^1\delta\left[U_{HF}^2 + \beta\delta u(c_H^3) - U_{MF}^2 - \beta\delta u(c_L^3)\right] \tag{A.17}$$

where $W_S(\cdot)$ denotes the level of social welfare attainable under separating taxation, and use has been made of the facts that $\partial U_{HF}^2(\cdot)/\partial\beta = \delta u(c_H^3)$ and $\partial U_{MF}^2(\cdot)/\partial\beta = \delta u(c_L^3)$.

To determine the sign of the expression in (A.17), the first-order conditions corresponding to programme (4.1) – (4.2) are:

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

$$-v'\left(\frac{y_L^2}{w_L}\right)\frac{1}{w_L} + \lambda^2 = 0 \tag{A.19}$$

$$-u'(m_L^2 + (1+r)s_L^1 - s_L^2) + \delta(1+r)u'((1+r)s_L^2) = 0 \tag{A.20}$$

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

$$-v'\left(\frac{y_H^2}{w_H}\right)\frac{1}{w_H} + \lambda^2 = 0 \tag{A.22}$$

$$-u'(m_H^2 + (1+r)s_H^1 - s_H^2) + \delta(1+r)u'((1+r)s_H^2) = 0 \tag{A.23}$$

$$(1-\phi)\left[y_L^2 - m_L^2\right] + \phi\left[y_H^2 - m_H^2\right] = 0 \tag{A.24}$$

where $\lambda^2 > 0$ is the multiplier on the government's second-period budget constraint (4.2). Equations (A.18) and (A.21) imply that $c_L^2 = c_H^2$, which using (A.20) and (A.23) implies that $c_L^3 = c_H^3$. Furthermore, equations (A.19) and (A.22) imply that $y_H^2 > y_L^2$.

Therefore, from (4.6) and (4.7) we obtain $U^2_{HF} < U^2_{MF}$, which using (A.17) establishes that $\partial W_S(\cdot)/\partial\beta < 0$.

The results in part (ii) that low-skill individuals face a positive marginal tax rate on labour 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 labour income in period 2 because first-best taxation is used in that period.

The proofs of the results in part (iii) that $MTRS^2_L = MTRS^2_H = (\beta - 1)/\beta < 0$ are analogous to those for Proposition 1. To prove that $MTRS^1_L < 0$, the first-order conditions on m^1_L and s^1_L from programme (4.3) – (4.5) can be written as, respectively:

$$(1 - \phi - \theta^1_H)u'(c^1_L) - \lambda^1(1 - \phi) = 0 \tag{A.25}$$

$$-(1 - \phi - \theta^1_H)u'(c^1_L) + \delta\frac{\partial Z^2_F(\cdot)}{\partial s^1_L} + \theta^1_H\beta\delta\left[\frac{\partial U^2_{HF}(\cdot)}{\partial s^1_L} - \frac{\partial U^2_{MF}(\cdot)}{\partial s^1_L}\right] = 0 \tag{A.26}$$

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

$$\mathcal{L}^2(\cdot) = (1 - \phi)\left\{u(m^2_L + (1 + r)s^1_L - s^2_L) - v\left(\frac{y^2_L}{w_L}\right) + \delta u((1 + r)s^2_L)\right\}$$

$$+ \phi\left\{u(m^2_H + (1 + r)s^1_H - s^2_H) - v\left(\frac{y^2_H}{w_H}\right) + \delta u((1 + r)s^2_H)\right\} + \lambda^2\left\{(1 - \phi)\left[y^2_L - m^2_L\right] + \phi\left[y^2_H - m^2_H\right]\right\} \tag{A.27}$$

By the Envelope Theorem:

$$\frac{\partial Z^2_F(\cdot)}{\partial s^1_L} = \frac{\partial \mathcal{L}^2(\cdot)}{\partial s^1_L} = (1 - \phi)u'(c^2_L)(1 + r) \tag{A.28}$$

Note that (A.25) implies that $1 - \phi - \theta^1_H > 0$. Using (2.8), (4.6), (4.7) and (A.28), equation (A.26) can be rewritten as:

$$MTRS^1_L := 1 - \frac{u'(c^1_L)}{\beta\delta(1 + r)u'(c^2_L)} = \frac{(1 - \phi)(\beta - 1)}{(1 - \phi - \theta^1_H)\beta} +$$

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

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.18) and (A.19) we obtain $u'(c_L^2) = v'\left(\frac{y_L^2}{w_L}\right)\frac{1}{w_L} > v'\left(\frac{y_L^2}{w_H}\right)\frac{1}{w_H}$. By applying the Implicit Function Theorem and Cramer's Rule to (A.18) – (A.24) it can be shown that:

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

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. ■

Proof of Proposition 3

To prove part (i), the Lagrangians associated with programmes (4.11) – (4.12) and (4.8) – (4.10) are, respectively:

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

$$\mathcal{L}^2(\cdot) = (1-\phi)\left\{u(m_L^2+(1+r)s^1-s_L^2)-v\left(\frac{y_L^2}{w_L}\right)+\delta u((1+r)s_L^2)\right\}$$

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

$$+\ \theta_H^2\{u(m_H^2+(1+r)s^1-s_H^2)-v\left(\frac{y_H^2}{w_H}\right)+\beta\delta u((1+r)s_H^2)$$

$$-\ u(m_L^2+(1+r)s^1-s_L^2)+v\left(\frac{y_L^2}{w_H}\right)-\beta\delta u((1+r)s_L^2)\} \tag{A.32}$$

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

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

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 programme (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 $\partial W_P(\cdot)/\partial\beta > 0$.

To prove part (ii), the first-order conditions on m^1 and y^1 from programme (4.11) – (4.12) can be written as, respectively:

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

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

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

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

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.

To prove part (iii), the first-order condition on s^1 from programme (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:

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

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

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

Equation (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.

Equation (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 equation (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 $MTRS_L^2 < MTRS_H^2 < 0$. ■

References

[1] Apps, P. and R. Rees [2006], "Repeated Optimal Nonlinear Income Taxation", *mimeo*.

[2] Berliant, M. and J. Ledyard [2005], "Optimal Dynamic Nonlinear Income Taxes with No Commitment", *mimeo*.

[3] Brett, C. and J. Weymark [2008], "Optimal Nonlinear Taxation of Income and Savings without Commitment", *Vanderbilt University Working Paper*, No. 08-W05.

[4] Brito, D., J. Hamilton, S. Slutsky and J. Stiglitz [1991], "Dynamic Optimal Income Taxation with Government Commitment", *Journal of Public Economics*, 44, 15-35.

[5] Chetty, R., A. Guren, D. Manoli and A. Weber [2011], "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins", *American Economic Review*, 101, 471-475.

[6] Diamond, P. and B. Koszegi [2003], "Quasi-Hyperbolic Discounting and Retirement", *Journal of Public Economics*, 87, 1839-1872.

[7] Diamond, P. and J. Spinnewijn [2011], "Capital Income Taxes with Heterogeneous Discount Rates", *American Economic Journal: Economic Policy*, forthcoming.

[8] Fang, H. [2006], "Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices", *International Economic Review*, 47, 1151-1185.

[9] Farhi, E., C. Sleet, I. Werning and S. Yeltekin [2011], "Nonlinear Capital Taxation Without Commitment", *mimeo*.

[10] Frederick, S., G. Loewenstein and T. O'Donoghue [2002], "Time Discounting and Time Preference: A Critical Review", *Journal of Economic Literature*, 40, 351-401.

[11] 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.

[12] Golosov, M., M. Troshkin, A. Tsyvinski and M. Weinzeirl [2010], "Preference Heterogeneity and Optimal Capital Income Taxation", *mimeo*.

[13] Golosov, M., A. Tsyvinski and I. Werning [2006], "New Dynamic Public Finance: A User's Guide", *NBER Macroeconomics Annual*, 2006, 317-387.

[14] Guesnerie, R. [1995], *A Contribution to the Pure Theory of Taxation*, Cambridge University Press.

[15] Guo, J-T. and A. Krause [2011a], "Optimal Nonlinear Income Taxation with Habit Formation", *Journal of Public Economic Theory*, 13, 463-480.

[16] Guo, J-T. and A. Krause [2011b], "Dynamic Income Taxation without Commitment: Comparing Alternative Tax Systems", *mimeo*.

[17] Kocherlakota, N. [2010], *The New Dynamic Public Finance*, Princeton University Press.

[18] Krause, A. [2009], "Optimal Nonlinear Income Taxation with Learning-by-Doing", *Journal of Public Economics*, 93, 1098-1110.

[19] Laibson, D. [1997], "Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics*, 112, 443-477.

[20] Mirrlees, J. [1971], "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies*, 38, 175-208.

[21] O'Donoghue, T. and M. Rabin [1999], "Doing It Now or Later", *American Economic Review*, 89, 103-124.

[22] O'Donoghue, T. and M. Rabin [2003], "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes", *American Economic Review*, 93, 186-191.

[23] O'Donoghue, T. and M. Rabin [2006], "Optimal Sin Taxes", *Journal of Public Economics*, 90, 1825-1849.

[24] OECD [2010], *Highlights from Education at a Glance*, OECD publications 2010.

[25] Roberts, K. [1984], "The Theoretical Limits to Redistribution", *Review of Economic Studies*, 51, 177-195.

[26] Stiglitz, J. [1982], "Self-Selection and Pareto Efficient Taxation", *Journal of Public Economics*, 17, 213-240.

[27] Tenhunen, S. and M. Tuomala [2010], "On Optimal Lifetime Redistribution Policy", *Journal of Public Economic Theory*, 12, 171-198.

TABLE 1

Baseline Parameter Values for Numerical Simulations*

ϕ	0.25	δ	0.38	w_L	1.00
γ	2.00	r	0.05	w_H	1.60
		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

* Each period is assumed to be 20 years in length.

FIGURE 1

Numerical Simulations: Separating Taxation

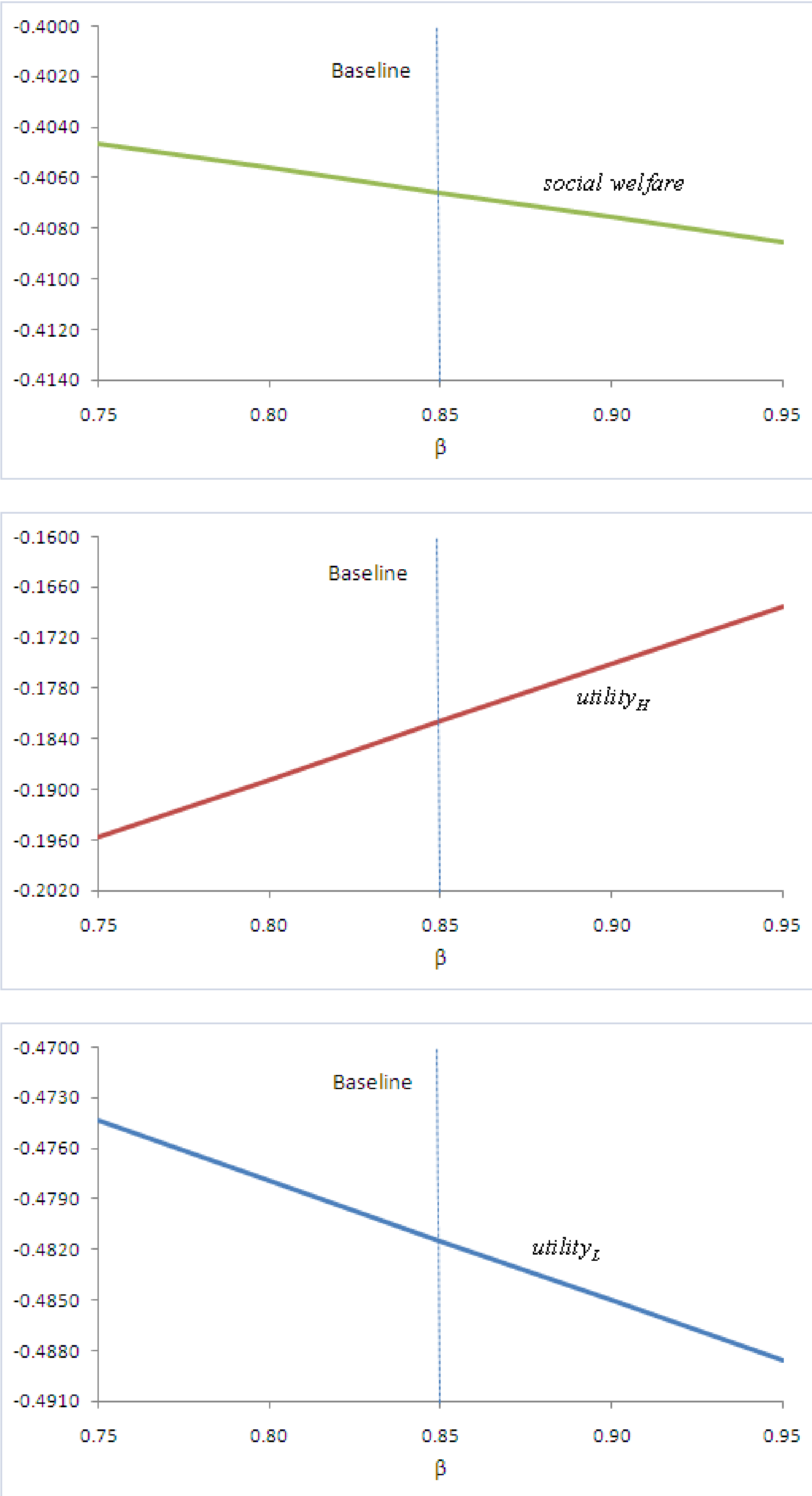
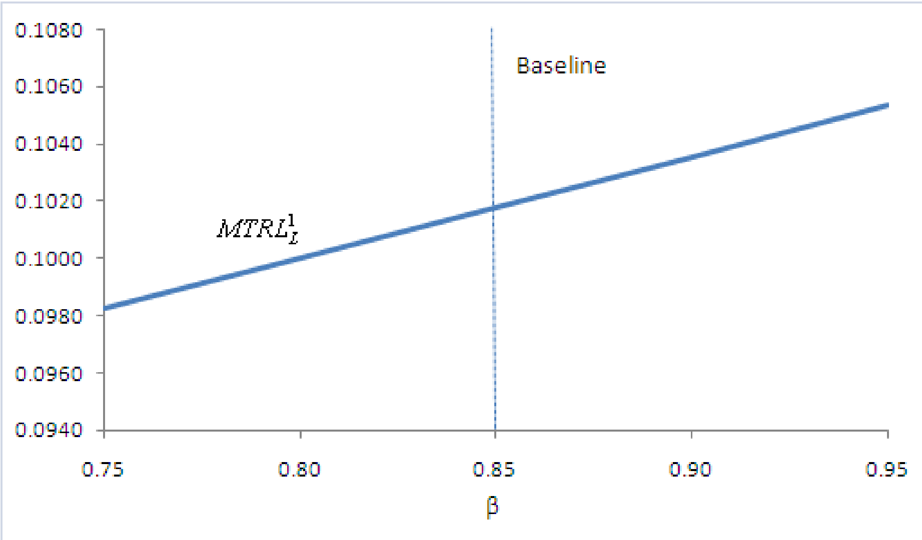


FIGURE 1 (cont.)

Numerical Simulations: Separating Taxation



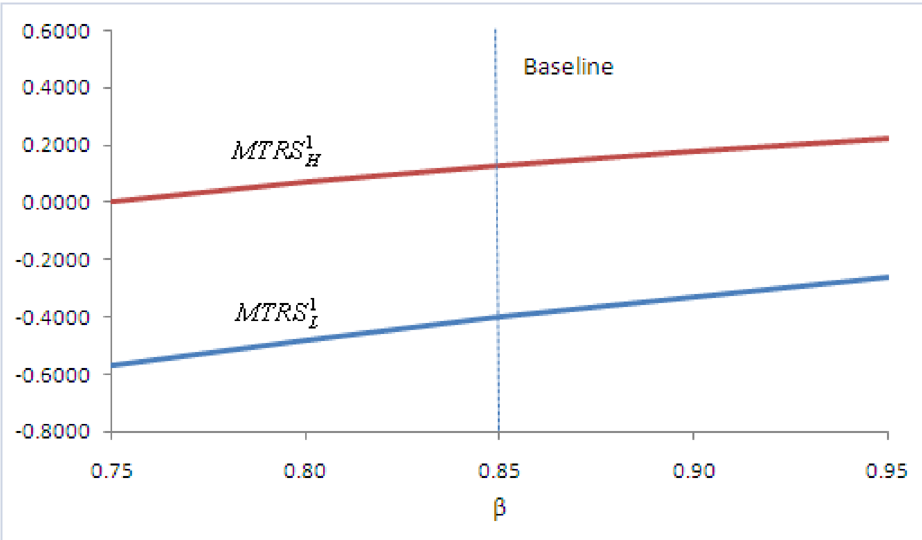


FIGURE 2

Numerical Simulations: Pooling Taxation

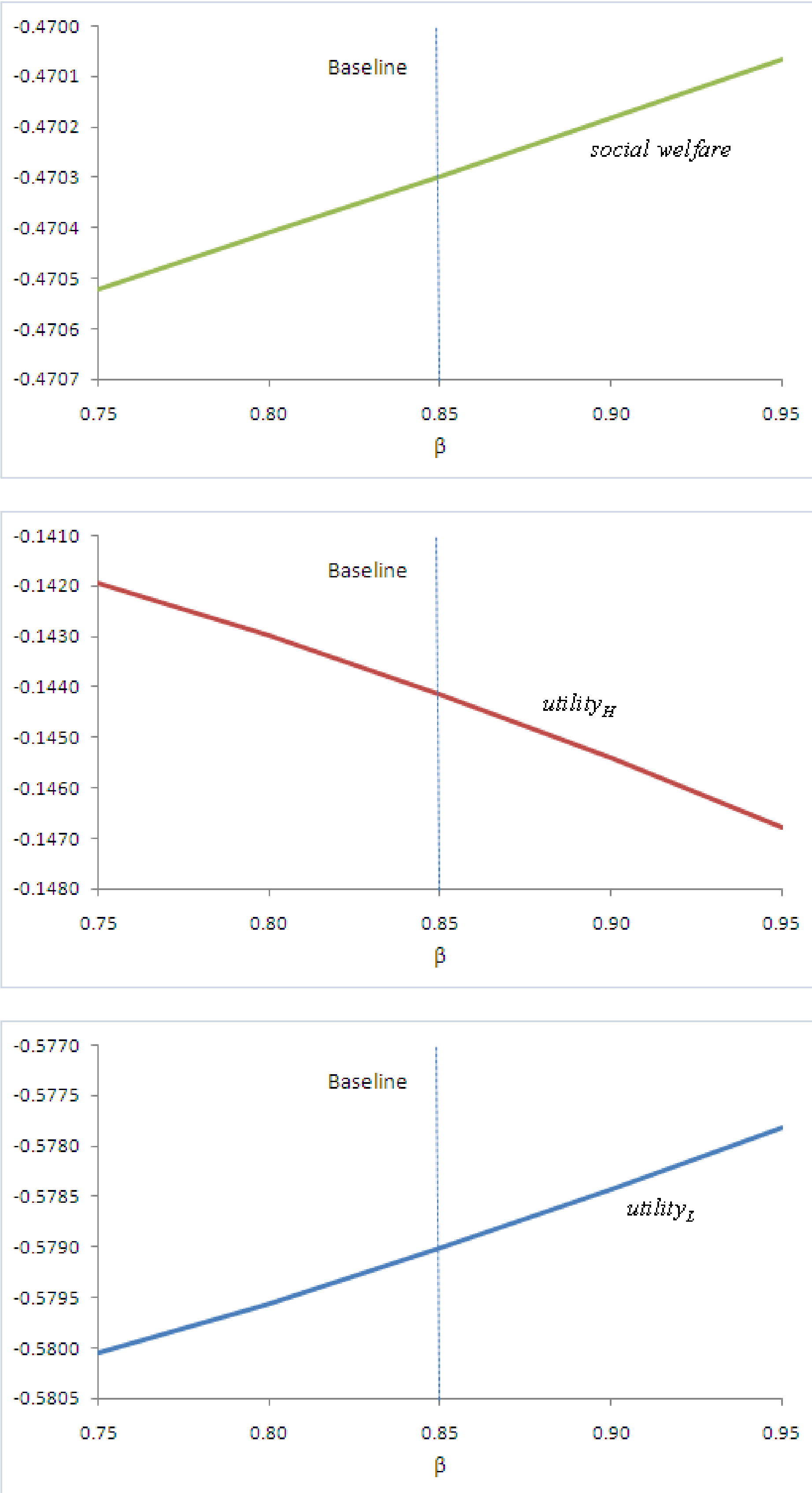


FIGURE 2 (cont.)

Numerical Simulations: Pooling Taxation