The Credibility of Commitment and Optimal Nonlinear Savings Taxation*

Jang-Ting Guo†
University of California, Riverside

Alan Krause‡
University of York

September 1, 2017

Abstract

We compare optimal nonlinear savings taxation under different assumptions with regard to the government's ability to commit to its future tax policy. In particular, we incorporate the possibility that individuals may differ in their beliefs regarding the probability of commitment. When these beliefs are homogeneous, we find that optimal marginal savings tax rates always fall between those under the polar cases of full-commitment (zero marginal savings taxation) and no-commitment (progressive marginal savings taxation). However, this result no longer holds when beliefs are postulated to be heterogeneous. The effects of beliefs changing in response to past commitment or no-commitment decisions by the government are also quantitatively explored.

Keywords: Savings taxation; commitment; multi-dimensional screening.

JEL Classifications: E60; H21; H24.

*We thank seminar participants at the University of York for helpful comments and suggestions. Any errors remain our responsibility.

†Corresponding author. Department of Economics, University of California, Riverside, CA, 92521, U.S.A. E-mail: guojt@ucr.edu.

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

1 Introduction

Previous studies that examine optimal nonlinear taxation of capital/savings have typically assumed full commitment by the government, while a few studies have considered no commitment. Under full-commitment, the government announces its tax policies for the present and future, and then simply implements those polices. Importantly, it is also postulated that individuals completely believe that the government will implement its announced policies. Under no-commitment, the government re-optimizes its tax policies period by period, irrespective of any previous promises or announcements. Individuals are aware that the government will re-optimize each period. Neither full-commitment nor no-commitment is an entirely satisfactory description of how governments actually implement taxation. The problem with the full-commitment assumption is that taxation is not time consistent, as the tax system may no longer be optimal in future periods. The problem with no-commitment is that it ignores the long-term gains to be obtained by a government that can make and keep its promises. In reality, governments do reform their tax systems, though such reforms tend to be infrequent. As a result, actual practice seems to lie somewhere between what would occur under full-commitment and no-commitment.[1]

As in the related literature, we adopt the Mirrlees (1971) approach to examine optimal nonlinear taxation of labor income and savings. However, we do so in two environments that fall between the polar cases of full-commitment and no-commitment. The first is 'loose commitment', under which it is common knowledge among individuals and the government that the probability of commitment is p.[2] That is, the government does not reform the tax system with probability p, but does so with probability $(1-p)$. Loose-commitment admits full-commitment and no-commitment as special cases ($p = 1$ and

[1] Gaube (2007) assumes full commitment, but compares short-term (annual) taxation of income as practiced by governments, versus optimal long-term taxation. He shows that short-term taxation cannot implement the optimal long-term tax contract. Gaube (2010) shows that time-consistent annual taxation (so-called *partial* commitment) may yield higher welfare than long-term taxation without commitment.

[2] Guo and Krause (2014) examine optimal nonlinear income taxation under loose commitment. However, there are no savings in their model, and their focus is on the revelation and use of skill-type information. Debortoli and Nunes (2010) examine capital and labor taxation under loose commitment in a representative-agent model.

$p = 0$, respectively). While we think that loose-commitment is a better assumption than full-commitment or no-commitment, it is still not ideal. In particular, loose-commitment assumes that all individuals and the government know and agree that the probability of commitment is p. The second environment that we consider is a setting we call 'commitment without credibility'. In this setting, the government sets taxes as it would under full-commitment. The only difference is that individuals recognize that the government will be tempted to reform the tax system in the future. Individuals therefore attach some probability to re-optimization; thus, the government's promise to commit is not completely credible. Moreover, we allow individuals to differ in their beliefs regarding the probability of commitment. We think commitment-without-credibility better describes the environment in which governments actually set taxes, as opposed to full-commitment, no-commitment, or loose-commitment. Commitment-without-credibility is also a relatively modest departure from the common assumption of full-commitment.

In standard Mirrlees-style models in which individuals differ only by their skills, the existing literature has found that zero marginal savings taxation is optimal under full-commitment (based on Atkinson and Stiglitz (1976)).[3] Under no-commitment, however, savings taxation is progressive, in that optimal marginal savings tax rates are increasing in individuals' skills (see, e.g., Brett and Weymark (2008) and Farhi, et al. (2012)). It is tempting to view the results under no-commitment as providing the upper-bound on the size of the optimal marginal savings tax distortions. We show that, under loose-commitment, optimal marginal savings tax rates do always fall between those under full-commitment and no-commitment. Under commitment-without-credibility, however, circumstances are more complicated. Our main result is that some individuals may face larger marginal savings tax distortions under commitment-without-credibility than they would under no-commitment, even though all individuals believe that there is some chance of commitment under the former.

In sum, relaxing the full-commitment assumption that individuals completely believe

[3] This finding is often viewed as providing support for the Chamley (1986) and Judd (1985) result that capital should not be taxed in the long-run. Straub and Werning (2015) have recently challenged the Chamley-Judd result.

that the government can commit may have substantive effects. Specifically, the characteristics of optimal nonlinear savings taxation do not simply fall between those under full-commitment and no-commitment. The explanation for this counter-intuitive finding follows from the (we think, realistic) assumption that individuals may disagree *vis-a-vis* the perceived probability of commitment. Under full-commitment and no-commitment, all individuals agree that the probability of commitment is one and zero, respectively. Therefore, individuals differ only by their skills.[4] Under commitment-without-credibility, individuals differ by their skills as well as their beliefs. This makes the optimal tax problem more complicated (formally, it becomes a multi-dimensional screening problem), as it must now take into account this second source of heterogeneity. It will be shown that belief heterogeneity itself calls for marginal savings tax distortions; thus a further contribution of our paper is that it identifies a new rationale for taxing savings. It is this new rationale that makes it possible for some individuals to face larger marginal savings tax distortions under commitment-without-credibility than they would under no-commitment.

A government that resists the temptation to reform would hope to gain commitment credibility, making it possible to obtain higher social welfare in the future. We examine this issue in an extension of our commitment-without-credibility model, in which individuals' beliefs regarding the probability of commitment change in response to past commitment/no-commitment decisions by the government. Changing beliefs generates two opposing effects. Commitment raises probabilistic-beliefs and the level of social welfare possible in the future. But at the same time, the increase in probabilistic-beliefs makes implementing reform even more tempting, and succumbing to this temptation lowers probabilistic-beliefs and future social welfare. The net effect is, therefore, theoretically ambiguous. However, we calibrate our model using empirically-plausible parameter values, under which we find that the net effect of changing beliefs is positive, in that a higher level of social welfare in the long-run is attainable.

The model we use is admittedly simple, but simplicity is necessary for the following

[4] Under loose-commitment, individuals also differ only by their skills, because all agree that the probability of commitment is p.

reasons. First, it is well-known that optimal tax problems with two unobserved characteristics are notoriously difficult to solve, even in simple settings.[5] In our model, however, we can solve the optimal tax problem without imposing the usual restrictions needed to make the multi-dimensional screening problem tractable.[6] That said, we do restrict attention to belief heterogeneity among high-skill individuals only. Nevertheless, this restriction strengthens our main findings, as it shows that only a limited second source of heterogeneity is required. Second, we wish to compare the size (not just sign) of the optimal marginal savings tax rates across different commitment regimes. Our simple model makes such analytical comparisons possible.

The remainder of the paper is organized as follows. Section 2 outlines our modelling framework, while Sections 3, 4, and 5 present optimal nonlinear taxation under no-commitment, loose-commitment, and commitment-without-credibility, respectively. Section 6 extends our model to an infinite-horizon setting to examine commitment-without-credibility over the long-run. Section 7 concludes, while all proofs are contained in an appendix.

2 Preliminaries

We examine a two-period economy with a unit measure of individuals, with a proportion $\phi \in (0,1)$ being high-skill workers (denoted by H) and the remaining $(1-\phi)$ being low-skill workers (denoted by L). The wage rates of the high-skill and low-skill workers are w_H and w_L respectively, with $w_H > w_L > 0$. Each high-skill worker believes that the probability of commitment is high or low, denoted p^H and p^L respectively, with $1 > p^H > p^L > 0$. Let π_H^j denote the population of high-skill individuals who believe that the probability of commitment is j, where superscript $j = L$ or H for low and high probability, respectively. We then have $\pi_H^L + \pi_H^H = \phi$. Low-skill individuals have a common belief regarding the probability of commitment, but we do not specify its notation

[5] See, for example, Boadway, et al. (2002) and the discussion within.

[6] These restrictions include specifying the problem in such a way that the two characteristics can be aggregated into one (e.g., Boadway, et al. (2002) and Brett and Weymark (2003)), or assuming a perfect correlation between skills and the second source of heterogeneity (e.g., Golosov, et al. (2013) and Krause (2014)).

because it will be seen below that their common belief plays no role in the design of the optimal tax systems. The model could be extended to incorporate heterogeneous beliefs among low-skill individuals. However, the multi-dimensional screening problem would become much more complicated, making it less clear as to which incentive-compatibility constraints may bind (see Section 5).

All individuals consume, work, and save in period 1, and then live-off their savings in period 2. Thus, period 2 can be thought of as the retirement period. Preferences are represented by the quasi-linear utility function:

$$u(c_i^j) - l_i^j + \delta u((1+r)s_i^j) \tag{2.1}$$

where c_i^j is type ij's first-period consumption, l_i^j is type ij's labor supply, and s_i^j is type ij's savings. For low-skill individuals, the superscript j is redundant; thus we have c_i, l_i, and s_i, where $i = L$. The specification that the utility function is quasi-linear is a little stronger than the usual formulation in the literature whereby it is additively separable. However, quasi-linearity enables us to compare the size (not just sign) of the optimal marginal savings tax rates across different commitment regimes. The market interest rate is given by $r > 0$; therefore, second-period consumption is equal to $(1+r)s_i^j$. The function $u(\cdot)$ is increasing and strictly concave, while $\delta \in (0,1)$ denotes the discount factor. For future reference, we use $y_i^j = w_i l_i^j$ to denote the pre-tax income of a type ij individual.

In the absence of taxation, individuals would choose c_i^j, l_i^j, and s_i^j to maximize the utility function (2.1) subject to the budget constraint:

$$c_i^j + s_i^j \leq w_i l_i^j \tag{2.2}$$

which yields the marginal condition:

$$\frac{u'(c_i^j)}{\delta(1+r)u'((1+r)s_i^j)} = 1 \tag{2.3}$$

In the presence of taxation, equation (2.3) might not be satisfied. The resulting marginal distortion is commonly interpreted as a 'tax wedge' or 'implicit marginal tax rate' on savings. Thus, we define:

$$MTRS_i^j := 1 - \frac{u'(c_i^j)}{\delta(1+r)u'((1+r)s_i^j)} \tag{2.4}$$

as the (implicit) marginal tax rate on savings faced by a type ij individual.

3 No Commitment (NC)

We begin by examining optimal nonlinear taxation of labor income and savings when it is common knowledge that the government cannot commit.[7] In this case, all individuals know that the probability of commitment is zero, and therefore individuals differ only by their skills. As it is certain that the government will re-optimize in period 2, its behavior in period 2 can be described as follows. Choose consumption levels c_L and c_H for the low-skill and high-skill individuals, respectively, to maximize:

$$(1-\phi)u(c_L) + \phi u(c_H) \tag{3.1}$$

subject to:

$$(1-\phi)c_L + \phi c_H \leq (1-\phi)(1+r)s_L + \phi(1+r)s_H \tag{3.2}$$

where s_L and s_H are, respectively, the savings by low-skill and high-skill individuals in period 1. Equation (3.1) is the second-period utilitarian social welfare function, while equation (3.2) is the government's second-period budget constraint; i.e., aggregate consumption in period 2 must be less than or equal to the stock of aggregate savings.[8] It is shown in the appendix that the solution to program $(3.1)-(3.2)$ equalizes consumption

[7] In addition to Brett and Weymark (2008) and Farhi, et al. (2012), there have been a number of recent papers that examine dynamic nonlinear taxation without commitment. See, e.g., Apps and Rees (2006), Berliant and Ledyard (2014), Krause (2009, 2017), Guo and Krause (2011, 2013, 2015a, 2015b), Aronsson and Sjogren (2016), and Morita (2016). However, the focus of those papers is not on savings taxation, but rather on how the government may use skill-type information revealed in earlier periods to implement first-best taxation in latter periods.

[8] We assume throughout the paper that the government does not save or borrow and its revenue requirement is zero; thus taxation is implemented for redistributive purposes only.

($c_L = c_H$), which we denote by $c(\phi, r, s_L, s_H)$. Let $W(\phi, r, s_L, s_H)$ denote the level of social welfare attainable in period 2, which is the value function associated with program (3.1) – (3.2).

In period 1, the government chooses tax contracts, $\langle m_L, s_L, y_L\rangle$ and $\langle m_H, s_H, y_H\rangle$, for the low-skill and high-skill individuals, respectively, to maximize:

$$(1-\phi)\left[u(m_L - s_L) - \frac{y_L}{w_L}\right] + \phi\left[u(m_H - s_H) - \frac{y_H}{w_H}\right] + \delta W(\cdot) \tag{3.3}$$

subject to:

$$(1-\phi)(y_L - m_L) + \phi(y_H - m_H) \geq 0 \tag{3.4}$$

$$u(m_H - s_H) - \frac{y_H}{w_H} + \delta u(c(\cdot)) \geq u(m_L - s_L) - \frac{y_L}{w_H} + \delta u(c(\cdot)) \tag{3.5}$$

where m_i denotes type i's post-tax income; thus first-period consumption is equal to $m_i - s_i$. Equation (3.3) is a utilitarian social welfare function, which takes into account how savings decisions made in period 1 affect the level of social welfare attainable in period 2. Equation (3.4) is the government's first-period budget constraint, and equation (3.5) is the high-skill type's incentive-compatibility constraint. In period 1 the government does not know who is high-skill and who is low-skill, so the tax contracts must be incentive compatible. That is, the utility a high-skill individual obtains from choosing $\langle m_H, s_H, y_H\rangle$ must be greater than or equal to what they could obtain by choosing $\langle m_L, s_L, y_L\rangle$.[9] As all individuals receive the same level of consumption (and hence utility) in period 2, the last term on both sides of equation (3.5) is redundant. But to assist interpretation of the incentive-compatibility constraint, we do not delete this term.

4 Loose Commitment (LC)

In this section, we examine optimal nonlinear taxation of labor income and savings under loose commitment. In this case, it is common knowledge among individuals and

[9] Following the standard practice, we omit the low-skill type's incentive-compatibility constraint, as it will not be binding. This is because the redistributive goals of the government create an incentive for high-skill individuals to mimic low-skill individuals, but not vice versa.

the government that the probability of commitment is p. Individuals therefore still differ only by their skills. LC admits NC and full-commitment (FC) as special cases, with $p = 0$ corresponding to the former and $p = 1$ corresponding to the latter.

In period 1, the government chooses tax contracts, $\langle m_L, s_L, y_L \rangle$ and $\langle m_H, s_H, y_H \rangle$, for the low-skill and high-skill individuals, respectively, to maximize:

$$(1-\phi)\left[u(m_L - s_L) - \frac{y_L}{w_L}\right] + \phi\left[u(m_H - s_H) - \frac{y_H}{w_H}\right]$$

$$+\, p\delta\left[(1-\phi)u((1+r)s_L) + \phi u((1+r)s_H)\right] + (1-p)\delta W(\cdot) \tag{4.1}$$

subject to:

$$(1-\phi)(y_L - m_L) + \phi(y_H - m_H) \geq 0 \tag{4.2}$$

$$u(m_H - s_H) - \frac{y_H}{w_H} + \delta\left[pu((1+r)s_H) + (1-p)u(c(\cdot))\right] \geq$$

$$u(m_L - s_L) - \frac{y_L}{w_H} + \delta\left[pu((1+r)s_L) + (1-p)u(c(\cdot))\right] \tag{4.3}$$

where equation (4.1) is a utilitarian social welfare function, equation (4.2) is the government's budget constraint, and equation (4.3) is the high-skill type's incentive-compatibility constraint. The social welfare function incorporates the expected level of social welfare in period 2. That is, with probability p the government can commit, and social welfare will simply equal the aggregate utility individuals obtain from consuming their first-period savings. But with probability $(1-p)$ the government will re-optimize (as under NC), and social welfare will be equal to $W(\cdot)$. Likewise, the incentive-compatibility constraint incorporates the high-skill type's expected utility in period 2. It is shown in the appendix that:

Proposition 1 *The optimal marginal tax rates applicable to savings are:* $MTRS_H^{NC} > MTRS_H^{LC} > 0 > MTRS_L^{LC} > MTRS_L^{NC}$.

The superscripts in Proposition 1 are used to distinguish the cases of no-commitment and loose-commitment. Under both NC and LC, high-skill individuals face a positive marginal savings tax rate, while that for low-skill individuals is negative.[10] However,

[10] The results under no-commitment are closely related to those obtained by Brett and Weymark

the marginal tax rates are closer to zero under LC. It is straightforward to show that if the government's ability to commit was perfectly credible (full-commitment), then all individuals would face a zero marginal tax rate on their savings. That benchmark can be viewed as an application of the Atkinson and Stiglitz (1976) result on the redundancy of commodity taxes alongside nonlinear income taxation.[11] Under NC, high-skill individuals know that the government will re-optimize the savings tax in period 2, in order to redistribute some of their savings toward the low-skill. This creates an incentive for high-skill individuals to mimic low-skill individuals in period 1. Thus, to help deter mimicking, the government brings forward consumption by high-skill individuals (by taxing their savings) and delays consumption by low-skill individuals (by subsidizing their savings). These same forces are also in effect under LC, but since LC admits some chance of commitment, the optimal marginal savings tax rates fall somewhere between those under NC and FC.

5 Commitment Without Credibility (CWC)

We now consider optimal nonlinear taxation of labor income and savings when the government has no intention of redesigning the tax system in period 2. However, individuals recognize that once savings decisions are made in period 1, the government will be tempted to re-optimize in period 2. Individuals therefore assign some probability to the possibility of re-optimization. We call this case 'commitment without credibility', and the heterogeneous beliefs (p^H and p^L) now play a role. The government chooses a tax contract $\langle m_L, s_L, y_L \rangle$ for the low-skill individuals, and $\langle m_H^j, s_H^j, y_H^j \rangle$ for each type of

(2008) and by Farhi, et al. (2012). Brett and Weymark consider a two-type, two-period, model without commitment, in which individuals work in both periods. They show that high-skill individuals face a positive marginal savings tax rate, while that for low-skill individuals is lower and may be negative. Farhi, et al. consider a two-period model with a continuum of types, in which individuals work only in the first period. They conclude that optimal marginal savings tax rates without commitment are increasing in skills, and are always positive at the top of the skill distribution and always negative at the bottom.

[11] Strictly speaking, the Atkinson-Stiglitz result depends upon the assumptions that: (i) labor is separable from consumption in the utility function, and (ii) all individuals have the same preferences. We make these assumptions in this paper, as does most of the related literature.

high-skill individual, to maximize:

$$(1-\phi)\left[u(m_L - s_L) - \frac{y_L}{w_L} + \delta u((1+r)s_L)\right] + \sum_j \pi_H^j \left[u(m_H^j - s_H^j) - \frac{y_H^j}{w_H} + \delta u((1+r)s_H^j)\right] \tag{5.1}$$

subject to:

$$(1-\phi)(y_L - m_L) + \sum_j \pi_H^j (y_H^j - m_H^j) \geq 0 \tag{5.2}$$

$$u(m_H^H - s_H^H) - \frac{y_H^H}{w_H} + \delta\left[p^H u((1+r)s_H^H) + (1-p^H)u(c(\cdot))\right] \geq$$

$$u(m_H^L - s_H^L) - \frac{y_H^L}{w_H} + \delta\left[p^H u((1+r)s_H^L) + (1-p^H)u(c(\cdot))\right] \tag{5.3}$$

$$u(m_H^H - s_H^H) - \frac{y_H^H}{w_H} + \delta\left[p^H u((1+r)s_H^H) + (1-p^H)u(c(\cdot))\right] \geq$$

$$u(m_L - s_L) - \frac{y_L}{w_H} + \delta\left[p^H u((1+r)s_L) + (1-p^H)u(c(\cdot))\right] \tag{5.4}$$

$$u(m_H^L - s_H^L) - \frac{y_H^L}{w_H} + \delta\left[p^L u((1+r)s_H^L) + (1-p^L)u(c(\cdot))\right] \geq$$

$$u(m_L - s_L) - \frac{y_L}{w_H} + \delta\left[p^L u((1+r)s_L) + (1-p^L)u(c(\cdot))\right] \tag{5.5}$$

Equation (5.1) is a utilitarian social welfare function, which is analogous to the social welfare function that the government would maximize under FC. This reflects the government's promise that it will not re-optimize in period 2. Equation (5.2) is the government's budget constraint, while equations (5.3) – (5.5) are incentive-compatibility constraints.[12] At this point we encounter the usual challenge associated with multi-dimensional screening problems, in that there are numerous (in our case, six) incentive-compatibility constraints that must be considered.[13] However, we conjecture with some

[12] There are some similarities between the CWC program (5.1) – (5.5) and optimal taxation when individuals differ by their skills and preferences for savings. This is because the CWC probabilities could be interpreted as applying different weights to second-period consumption. However, there are also some key differences, such as that the social welfare function (5.1) does not reflect any preference heterogeneity. Optimal taxation when individuals differ by their skills and preferences for savings has been examined by Saez (2002), Tenhunen and Tuomala (2010), Diamond and Spinnewijn (2011), and Golosov, et al. (2013).

[13] If low-skill individuals were also distinguished by their beliefs regarding the probability of commitment, there would be twelve incentive-compatibility constraints to consider.

confidence that only (5.3) $-$ (5.5) may be binding for the following reasons.

Consider incentive-compatibility constraint (5.3), which links the HH and HL types. Recall that HH individuals are high-skill individuals who believe that there is a high probability of commitment. These individuals think that there is a good chance that the government will not re-optimize in period 2, and therefore their second-period consumption will be $(1+r)s_H^H$. Recall that if the government were to re-optimize in period 2, it would equalize consumption across all individuals. Accordingly, high-skill individuals would have their savings taxed to subsidize second-period consumption by low-skill individuals. As before, let $c(\cdot)$ denote the equalized level of consumption in period 2. Type HL individuals are high-skill individuals who believe that the probability of commitment is low. They think it is relatively more likely that the government in period 2 will re-optimize, and their second-period consumption will be only $c(\cdot)$. Accordingly, HL individuals will be relatively more reluctant to reveal their type, and they must be offered a generous tax contract to do so. This means that HH individuals will be attracted to the tax contract intended for HL individuals, but not vice versa. Incentive-compatibility constraint (5.3) prevents such mimicking behavior, by ensuring that the utility a HH individual can expect from choosing $\langle m_H^H, s_H^H, y_H^H \rangle$ is greater than or equal to what they could obtain by choosing $\langle m_H^L, s_H^L, y_H^L \rangle$. Specifically, HH individuals believe that the probability of commitment is p^H, so with this probability they expect to be able to consume their savings in period 2, but with probability $(1-p^H)$ they expect that the government will re-optimize and their second-period consumption will be $c(\cdot)$. Finally, incentive-compatibility constraints (5.4) and (5.5) reflect the standard assumption that the government seeks to redistribute from high-skill to low-skill individuals, which creates an incentive for each type of high-skill individual to mimic the low-skill, but not vice versa. We now obtain the following results:

Proposition 2 *Under CWC, the optimal marginal tax rates applicable to savings are*: $MTRS_H^L > MTRS_H^H > 0 > MTRS_L$.

High-skill individuals face positive marginal savings tax rates, while low-skill individuals face a negative marginal savings tax rate. This is qualitatively the same as under NC and LC, and the intuition for the progressive pattern of optimal marginal savings taxation is analogous. Those high-skill individuals who believe that the probability of commitment is low (the *HL* type) face the largest marginal savings tax distortion. This is because *HL* individuals think it is relatively more likely that the government will re-optimize the savings tax. Accordingly, they must face the highest marginal savings tax rate to deter mimicking.

Proposition 3 *HH individuals and low-skill individuals always face lower marginal savings tax distortions under CWC than under NC, i.e.,* $MTRS_H^H < MTRS_H^{NC}$ *and* $MTRS_L^{NC} < MTRS_L$. *However, HL individuals may face a greater marginal savings tax distortion under CWC than under NC, i.e.,* $MTRS_H^L > MTRS_H^{NC}$ *is possible.*

The *HH* individuals and low-skill individuals always face lower marginal savings tax distortions under CWC than under NC, for the same reason that optimal marginal savings tax distortions are lower under LC than under NC. However, it is possible that *HL* individuals face a larger marginal savings tax distortion under CWC than under NC. This is possible even though the only difference between FC and CWC is that individuals under CWC attach some (differing) probability to re-optimization. The intuition is as follows. Under NC individuals are distinguished only by their skills, as it is common knowledge that the government cannot commit. Under CWC, however, individuals are distinguished by their skills and their beliefs regarding the probability of commitment. Belief heterogeneity itself calls for marginal savings tax rate distortions. An increase in the degree of belief heterogeneity makes the tax contract intended for *HL* individuals more attractive to *HH* individuals. Accordingly, a higher marginal tax rate on savings by *HL* individuals is required to deter mimicking. When belief heterogeneity is sufficiently severe, *HL* individuals face a higher marginal savings tax rate under CWC than they would under NC.

Next, we provide a numerical example to demonstrate the possibility stated in Proposition 3 that $MTRS_H^L > MTRS_H^{NC}$. This possibility is demonstrated using plausible parameter values, thus illustrating that it cannot be considered an unlikely occurrence. For comparison purposes, the results under FC and LC are also shown.

We first postulate that the utility function (2.1) takes the logarithmic form:

$$\ln(c_i^j) - l_i^j + \delta \ln((1+r)s_i^j) \tag{5.6}$$

which follows from Chetty (2006) who concludes that a reasonable estimate of the coefficient of relative risk aversion is one (which implies log utility). Numerical values of the parameters are chosen on the following basis. The OECD (2014) reports that approximately one-third of working-age individuals have attained tertiary level education. We assume that such individuals are high-skill and the remainder are low-skill, i.e., $\phi = 1/3$. We then assume that equal numbers of high-skill individuals believe that the probability of commitment is high or low, i.e., $\pi_H^H = \pi_H^L = 1/6$. Fang (2006) and Goldin and Katz (2007) estimate that the college wage premium is around 60%. We therefore normalize the low-skill type's wage to unity and set the high-skill type's wage equal to 1.6. Following common practice, we assume an annual market interest rate of 4% and that each period of our two-period economy is 20-years in length. Therefore, $\delta = 1/(1+r)^{20} = 0.456$. Finally, we assume that the LC probability p is 0.5, while the CWC probabilities are $p^H = 0.75$ and $p^L = 0.25$.

The results are shown in Table 1. All individuals face zero marginal savings taxation under FC (Atkinson and Stiglitz (1976)). Under NC, optimal marginal savings taxation is progressive (Brett and Weymark (2008) and Farhi, et al. (2012)). The optimal marginal savings tax rates under LC fall between those under FC and NC (Proposition 1). Under CWC, we have $MTRS_H^L > MTRS_H^H > 0 > MTRS_L$ (Proposition 2), and HH individuals and low-skill individuals face lower marginal savings tax distortions under CWC than under NC (Proposition 3). We also have $MTRS_H^L > MTRS_H^{NC}$, thus demonstrating the counter-intuitive possibility identified in Proposition 3.

6 CWC over the Long-Run

The government's objective is to maximize social welfare. Given this objective, the government faces a trade-off when deciding whether or not to reform the tax system.

Reforming the tax system yields an immediate increase in social welfare, but also reduces the level of social welfare attainable in the future if individuals revise down their beliefs regarding the probability of commitment. On the other hand, a government that resists the temptation to reform forgoes an immediate increase in social welfare, but may gain commitment credibility (and higher social welfare in the future) if individuals revise up their probabilistic-beliefs. To explore these issues, we examine an infinitely-repeated version of the two-period CWC economy described in Section 5.

Before proceeding with our analysis, note that Farhi, et al. (2012) consider an extension of their two-period model to an infinite-horizon setting, in order to incorporate the benefits versus costs of reform. They examine a dynamic game in which a deviation (reform) triggers a bad equilibrium, hence no government chooses to reform in the trigger-strategy equilibrium. The advantage of their approach is that it incorporates strategic decision-making. The disadvantages are its complexity and that reforms never occur, which seems unrealistic. Our approach is to assume that each government chooses to reform or not-reform with some probability. This approach is relatively simple and consistent with some plausible assumptions regarding government behavior. For example, a government that expects to be in power for some time may resist reform to raise long-run social welfare, while a government that does not expect to be re-elected may do the opposite. Likewise, governments with different political ideologies, e.g., left-wing versus right-wing, may be more or less inclined to succumb to the temptation to reform.

6.1 The Infinitely-Repeated CWC Economy

In our model, individuals' probabilistic-beliefs and the actual probability of commitment are assumed to follow a Markov process. The initial probabilistic-beliefs, in economy 1 which covers periods 1 and 2, are as previously denoted, i.e., p^H and p^L. Subsequently, if the government did not reform (commitment) in economy T, then individuals in economy $T+1$ revise up their beliefs to $p^{\overline{H}}$ and $p^{\overline{L}}$, where $p^{\overline{H}} > p^H$ and $p^{\overline{L}} > p^L$. For example, if the government did not reform in economy $T = 3$ (which covers periods 5 and 6), then beliefs in economy $T+1 = 4$ (which covers periods 7 and 8) will be $p^{\overline{H}}$ and $p^{\overline{L}}$. Analogously, if the government did reform (no-commitment) in economy T, the revised beliefs in economy $T+1$ are $p^{\underline{H}}$ and $p^{\underline{L}}$, where $p^{\underline{H}} < p^H$ and $p^{\underline{L}} < p^L$. We also require an actual probability of commitment. In economy 1, this probability is denoted by q. Subsequently, q_C denotes the probability of commitment in economy $T+1$ following commitment in economy T, and q_N denotes the probability of commitment in economy $T+1$ following no-commitment in economy T.

From economy $T+1$ onward, continuation social welfare can be represented by the following recursive equations:

$$Z_{CC}^{T+1} = W_{CC}^{T+1} + q_C \beta Z_{CC}^{T+2} + (1-q_C)\beta Z_{NC}^{T+2} \tag{6.1}$$

$$Z_{NC}^{T+1} = W_{NC}^{T+1} + q_N \beta Z_{CN}^{T+2} + (1-q_N)\beta Z_{NN}^{T+2} \tag{6.2}$$

$$Z_{CN}^{T+1} = W_{CN}^{T+1} + q_C \beta Z_{CC}^{T+2} + (1-q_C)\beta Z_{NC}^{T+2} \tag{6.3}$$

$$Z_{NN}^{T+1} = W_{NN}^{T+1} + q_N \beta Z_{CN}^{T+2} + (1-q_N)\beta Z_{NN}^{T+2} \tag{6.4}$$

where $\beta = \delta^2$ is the discount factor between any two economies (recall that each economy has two periods). Social welfare in economy $T+1$ under commitment when commitment also occurred in economy T is denoted by W_{CC}^{T+1}. Social welfare in economy $T+1$ under no-commitment when commitment occurred in economy T is denoted by W_{NC}^{T+1}, while W_{CN}^{T+1} and W_{NN}^{T+1} can be similarly interpreted. To interpret the recursive equations, consider for example equation (6.1). If commitment occurs in economy $T+1$ following commitment in economy T, yielding social welfare of W_{CC}^{T+1} in economy $T+1$, there is then q_C probability that commitment will occur again in economy $T+2$. But with probability $(1-q_C)$ no-commitment will occur in economy $T+2$, yielding social welfare W_{NC}^{T+2} and activating recursive equation (6.2). The remaining recursive equations can be similarly interpreted.

Aggregate social welfare over the infinite horizon can now be written as:

$$SW = qW_C^1 + (1-q)W_N^1 + q\beta\left[q_C Z_{CC}^2 + (1-q_C)Z_{NC}^2\right] + (1-q)\beta\left[q_N Z_{CN}^2 + (1-q_N)Z_{NN}^2\right] \tag{6.5}$$

where W_C^1 and W_N^1 denote social welfare in economy 1 under commitment and no-commitment, respectively.

6.2 Numerical Simulations

To explore how changing beliefs affect the level of social welfare attainable, we conduct numerical simulations. First, we calibrate the model using plausible parameter values. These are the same as discussed in Section 5, but with the additional parameters needed for the infinite-horizon extension. We set $q = 0.5$, together with $q_C = 0.55$ and $q_N = 0.45$ to capture the idea that commitment in economy T is more likely to be followed by commitment again in economy $T+1$, and vice versa. Likewise, the baseline parameters chosen for $p^{\overline{H}}$ and $p^{\overline{L}}$ reflect a 10 percent upward revision to the probabilistic-beliefs following commitment, while $p^{\underline{H}}$ and $p^{\underline{L}}$ reflect a 10 percent downward revision following no-commitment. The full set of baseline parameter values is shown in Table 2. We also report the optimal marginal tax rates applicable to savings in economy 1, in economy $T+1$ following commitment, and in economy $T+1$ following no-commitment. It can be seen that the qualitative relationships stated in Propositions 2 and 3 continue to hold.

Figure 1 shows the effects of increasing the sensitivity of beliefs in response to commitment and no-commitment decisions by the government, holding all other parameters (including the actual probabilities of commitment) at their baseline levels. For example, a 15 percent sensitivity represents a 15 percent rise in probabilistic-beliefs following commitment ($p^{\overline{H}} = 0.8625$ and $p^{\overline{L}} = 0.2875$), but also a 15 percent decline in probabilistic-beliefs following no-commitment ($p^{\underline{H}} = 0.6375$ and $p^{\underline{L}} = 0.2125$). An increase in belief sensitivity has two opposing effects. On the one hand, it rewards a government who resists the temptation to reform, by making higher social welfare attainable in the future. On the other hand, a government that chooses to reform to obtain the short-run increase in social welfare is penalized by the decline in beliefs, making it harder to raise future social welfare. It turns out that the first effect quantitatively dominates in our model, with an increase in belief sensitivity raising the level of social welfare attainable in the long-run.

Figure 2 shows the effects of greater persistence in commitment and no-commitment decisions, holding all other parameters (including probabilistic-beliefs) at their baseline levels. For example, a 15 percent sensitivity represents a 15 percent rise in the actual probability of commitment following commitment ($q_C = 0.575$), but also a 15 percent decline in the actual probability of commitment following no-commitment ($q_N = 0.425$). As with changes in belief sensitivity, there are two opposing effects. An increase in persistence means that a commitment decision is more likely to be repeated, thus keeping the economy in the 'good state'. But a no-commitment decision is also more likely to be repeated, keeping the economy in the 'bad state'. In our model the second effect dominates, with greater persistence in commitment and no-commitment decisions reducing long-run social welfare.

Figure 3 shows the effect on long-run social welfare when individuals' beliefs and the actual probability of commitment move together. For example, a 15 percent sensitivity means that individuals revise up their probabilistic-beliefs by 15 percent following commitment ($p^{\overline{H}} = 0.8625$ and $p^{\overline{L}} = 0.2875$), while the actual probability of commitment also rises by 15 percent ($q_C = 0.575$). Analogously, individuals' beliefs and the actual probability of commitment fall by 15 percent following no-commitment ($p^{\underline{H}} = 0.6375$, $p^{\underline{L}} = 0.2125$, and $q_N = 0.425$). It can be seen that more persistence in commitment/no-commitment decisions, together with corresponding changes in beliefs, increases long-run social welfare.

7 Concluding Comments

This paper compares optimal nonlinear savings taxation under different commitment assumptions, and introduces a setting we call 'commitment without credibility'. We think CWC better describes the environment in which governments actually implement their tax policies. CWC is a relatively modest departure from the common FC assumption, as compared to NC and LC. Our specification of CWC captures the observation that when setting taxes governments do, at least implicitly, indicate that they currently have no intention of reforming the tax system in the future. However, individuals recognize that once the future arrives, the government will face the temptation to re-optimize the tax system. Individuals therefore attach some probability to re-optimization, and we think it is likely that they will differ in their beliefs regarding these subjective probabilities.

We show that belief heterogeneity itself calls for savings tax distortions, and as such optimal marginal savings tax rates under CWC do not necessarily fall between those under the polar cases of FC and NC. This counter-intuitive result is not possible under LC, in which all individuals and the government know and agree on the probability of commitment. However, we think CWC is a more realistic assumption than LC. In a repeated CWC setting in which beliefs may change, past commitment decisions by the government are rewarded as individuals revise up their beliefs regarding the probability of commitment. However, past no-commitment decisions are penalized as beliefs are then revised downwards. The net effect of changing beliefs on long-run social welfare could therefore be positive or negative, but our numerical simulations suggest that it is positive.

When examining CWC over the long-run, we have assumed that each government chooses to reform or not-reform with some probability. The temptation to reform the tax system follows from the immediate increase in social welfare that it provides. This increase in social welfare comes from more redistribution; thus low-skill individuals will be in favor of reform, *ceteris paribus*, while high-skill individuals will not. Accordingly, one could argue that the political power of low-skill versus high-skill individuals may influence the government's decision. In an interesting paper, Scheuer and Wolitzky (2016) examine 'sustainable' capital taxation, in that a policy is sustainable if the government will not face sufficient pressure to force reform. Sustainable capital taxation essentially buys commitment credibility, the price paid being the tax distortions required to deter the gathering of popular support for reform. While sustainable capital taxation captures an important positive (political economy) aspect of tax policy, it is not necessarily optimal from the normative perspective of maximizing social welfare.[14] In future research, it would be interesting to try to model the government's commitment decision, in a manner that allows the government to sometimes succumb to the temptation to reform.

[14]Scheuer and Wolitzky (2016) note that sometimes it may be optimal to let the reform occur, but they suggest that the no-reform constraint might be justified for some unmodeled reason. These include the government's loss of political power or future reputation.

8 Appendix

A.1 Proof of Proposition 1

The first-order conditions on y_L and y_H in program (3.3) – (3.5) and in program (4.1) – (4.3) are, respectively:

$$\frac{-(1-\phi)}{w_L} + \lambda(1-\phi) + \frac{\theta_H}{w_H} = 0 \tag{A.1}$$

$$\frac{-(\phi+\theta_H)}{w_H} + \lambda\phi = 0 \tag{A.2}$$

where $\lambda \geq 0$ is the multiplier on the budget constraint (equation (3.4) and (4.2)), and $\theta_H \geq 0$ is the multiplier on the incentive-compatibility constraint (equation (3.5) and (4.3)). Solving (A.1) and (A.2) yields:

$$\lambda = \frac{(1-\phi)}{w_L} + \frac{\phi}{w_H} > 0 \text{ and } \theta_H = \phi(1-\phi)\left(\frac{w_H}{w_L} - 1\right) > 0 \tag{A.3}$$

Note that the values of the multipliers are the same in both programs (3.3) – (3.5) and (4.1) – (4.3).

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

$$\mathcal{L} = (1-\phi)u(c_L)+\phi u(c_H)+\lambda^2\left[(1-\phi)(1+r)s_L + \phi(1+r)s_H - (1-\phi)c_L - \phi c_H\right] \tag{A.4}$$

where $\lambda^2 \geq 0$ is the Lagrange multiplier. The first-order conditions on c_L and c_H are, respectively:

$$(1-\phi)u'(c_L) - \lambda^2(1-\phi) = 0 \tag{A.5}$$

$$\phi u'(c_H) - \lambda^2\phi = 0 \tag{A.6}$$

which implies that $c_L = c_H$. Let c denote this equalized level of consumption. By the Envelope Theorem:

$$\frac{\partial W(\cdot)}{\partial s_L} = \frac{\partial \mathcal{L}(\cdot)}{\partial s_L} = u'(c)(1-\phi)(1+r) \text{ and } \frac{\partial W(\cdot)}{\partial s_H} = \frac{\partial \mathcal{L}(\cdot)}{\partial s_H} = u'(c)\phi(1+r) \tag{A.7}$$

The first-order condition on s_L in program $(3.3)-(3.5)$ can be written as:

$$-(1-\phi-\theta_H)u'(m_L-s_L)+\delta\frac{\partial W(\cdot)}{\partial s_L}=0 \tag{A.8}$$

Note that (A.8) implies that $1-\phi-\theta_H>0$. Using (2.4) and (A.7), equation (A.8) can be manipulated to yield:

$$MTRS_L^{NC}=\frac{-\theta_H}{1-\phi-\theta_H}<0 \tag{A.9}$$

The first-order condition on s_H in program $(3.3)-(3.5)$ can be written as:

$$-(\phi+\theta_H)u'(m_H-s_H)+\delta\frac{\partial W(\cdot)}{\partial s_H}=0 \tag{A.10}$$

Using (2.4) and (A.7), equation (A.10) can be manipulated to yield:

$$MTRS_H^{NC}=\frac{\theta_H}{\phi+\theta_H}>0 \tag{A.11}$$

The first-order condition on s_L in program $(4.1)-(4.3)$ can be written as:

$$-(1-\phi-\theta_H)u'(m_L-s_L)+p(1-\phi-\theta_H)\delta(1+r)u'((1+r)s_L)+(1-p)\delta\frac{\partial W(\cdot)}{\partial s_L}=0 \tag{A.12}$$

Using (2.4) and (A.7), equation (A.12) can be manipulated to yield:

$$MTRS_L^{LC}=\frac{-\theta_H(1-p)u'(c)}{(1-\phi-\theta_H)\left[pu'((1+r)s_L)+(1-p)u'(c)\right]}<0 \tag{A.13}$$

The first-order condition on s_H in program $(4.1)-(4.3)$ can be written as:

$$-(\phi+\theta_H)u'(m_H-s_H)+p(\phi+\theta_H)\delta(1+r)u'((1+r)s_H)+(1-p)\delta\frac{\partial W(\cdot)}{\partial s_H}=0 \tag{A.14}$$

Using (2.4) and (A.7), equation (A.14) can be manipulated to yield:

$$MTRS_H^{LC}=\frac{\theta_H(1-p)u'(c)}{(\phi+\theta_H)\left[pu'((1+r)s_H)+(1-p)u'(c)\right]}>0 \tag{A.15}$$

Finally, using (A.9) and (A.13) it can be shown that $MTRS_L^{NC}<MTRS_L^{LC}$, and

using (A.11) and (A.15) it can be shown that $MTRS_H^{NC} > MTRS_H^{LC}$. ■

A.2 Proof of Proposition 2

The first-order conditions on y_L, y_H^L, and y_H^H from program (5.1) – (5.5) can be written as, respectively:

$$\frac{-(1-\phi)}{w_L} + \lambda(1-\phi) + \frac{\gamma_H^H + \gamma_H^L}{w_H} = 0 \tag{A.16}$$

$$\frac{-(\pi_H^L - \theta_H^H + \gamma_H^L)}{w_H} + \lambda\pi_H^L = 0 \tag{A.17}$$

$$\frac{-(\pi_H^H + \theta_H^H + \gamma_H^H)}{w_H} + \lambda\pi_H^H = 0 \tag{A.18}$$

where $\lambda \geq 0$, $\theta_H^H \geq 0$, $\gamma_H^H \geq 0$, and $\gamma_H^L \geq 0$ are the multipliers on constraints (5.2) – (5.5), respectively. Solving (A.16) – (A.18) yields:

$$\lambda = \frac{(1-\phi)}{w_L} + \frac{\phi}{w_H} > 0,\ \theta_H^H + \gamma_H^H = \pi_H^H(1-\phi)\left(\frac{w_H}{w_L} - 1\right) > 0, \text{ and } \gamma_H^L - \theta_H^H = \pi_H^L(1-\phi)\left(\frac{w_H}{w_L} - 1\right) > 0 \tag{A.19}$$

Note that the multiplier on the budget constraint, λ, is the same as in programs (3.3) – (3.5) and (4.1) – (4.3), and that $\gamma_H^L > 0$.

The first-order condition on s_L in program (5.1) – (5.5) can be written as:

$$-(1-\phi-\gamma_H^H-\gamma_H^L)u'(m_L-s_L)+\delta(1+r)u'((1+r)s_L)\left[1-\phi-p^H\gamma_H^H-p^L\gamma_H^L\right] = 0 \tag{A.20}$$

The first-order condition on m_L in program (5.1) – (5.5) can be written as:

$$(1-\phi-\gamma_H^H-\gamma_H^L)u'(m_L-s_L) - \lambda(1-\phi) = 0 \tag{A.21}$$

which implies that $1-\phi-\gamma_H^H-\gamma_H^L > 0$. Using (2.4) and (A.19), equation (A.20) can be manipulated to yield:

$$MTRS_L = \frac{-\left[\gamma_H^H(1-p^H) + \gamma_H^L(1-p^L\right]}{1-\phi-\gamma_H^H-\gamma_H^L} < 0 \tag{A.22}$$

The first-order condition on s_H^H in program (5.1) – (5.5) can be written as:

$$-(\pi_H^H+\theta_H^H+\gamma_H^H)u'(m_H^H-s_H^H)+\delta(1+r)u'((1+r)s_H^H)\left[\pi_H^H+p^H(\theta_H^H+\gamma_H^H)\right]=0 \quad \text{(A.23)}$$

Using (2.4) and (A.19), equation (A.23) can be manipulated to yield:

$$MTRS_H^H=\frac{(1-p^H)(\theta_H^H+\gamma_H^H)}{\pi_H^H+\theta_H^H+\gamma_H^H}>0 \quad \text{(A.24)}$$

The first-order condition on s_H^L in program (5.1) – (5.5) can be written as:

$$-(\pi_H^L-\theta_H^H+\gamma_H^L)u'(m_H^L-s_H^L)+\delta(1+r)u'((1+r)s_H^L)\left[\pi_H^L-p^H\theta_H^H+p^L\gamma_H^L)\right]=0 \quad \text{(A.25)}$$

Using (2.4) and (A.19), equation (A.25) can be manipulated to yield:

$$MTRS_H^L=\frac{\gamma_H^L(1-p^L)-\theta_H^H(1-p^H)}{\pi_H^L-\theta_H^H+\gamma_H^L}>0 \quad \text{(A.26)}$$

Finally, using (A.24) and (A.26) it can be shown that $MTRS_H^L>MTRS_H^H$. ■

A.3 Proof of Proposition 3

Using (A.11) and (A.24) it can be shown that $MTRS_H^{NC}>MTRS_H^H$, and using (A.9) and (A.22) it can be shown that $MTRS_L>MTRS_L^{NC}$. Using (A.11) and (A.26) it can be shown that $MTRS_H^L>MTRS_H^{NC}$ if and only if $(p^L/p^H)<(\theta_H^H/\gamma_H^L)$. Table 1 provides an example where $MTRS_H^L>MTRS_H^{NC}$. ■

References

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

[2] Aronsson, T. and T. Sjogren (2016), "Quasi-hyperbolic Discounting, Paternalism and Optimal Mixed Taxation", *Mathematical Social Sciences*, 84, 24-36.

[3] Atkinson, A. and J. Stiglitz (1976), "The Design of Tax Structure: Direct Versus Indirect Taxation", *Journal of Public Economics*, 6, 55-75.

[4] Berliant, M. and J. Ledyard (2014), "Optimal Dynamic Nonlinear Income Taxes with No Commitment", *Journal of Public Economic Theory*, 16, 196-221.

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

[6] Brett, C. and J. Weymark (2003), "Financing Education Using Optimal Redistributive Taxation", *Journal of Public Economics*, 87, 2549-2569.

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

[8] Chamley, C. (1986), "Optimal Taxation of Capital Income in General Equilibrium with Infinite lives", *Econometrica*, 54, 607-622.

[9] Chetty, R. (2006), "A New Method of Estimating Risk Aversion", *American Economic Review*, 96, 1821-1834.

[10] Debortoli, D. and R. Nunes (2010), "Fiscal Policy under Loose Commitment", *Journal of Economic Theory*, 145, 1005-1032.

[11] Diamond, P. and J. Spinnewijn (2011), "Capital Income Taxes with Heterogeneous Discount Rates", *American Economic Journal: Economic Policy*, 3, 52-76.

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

[13] Farhi, E., C. Sleet, I. Werning and S. Yeltekin (2012), "Nonlinear Capital Taxation Without Commitment", *Review of Economic Studies*, 79, 1469-1493.

[14] Gaube, T. (2007), "Optimum Taxation of Each Year's Income", *Journal of Public Economic Theory*, 9, 127-150.

[15] Gaube, T. (2010), "Taxation of Annual Income as a Commitment Device", *mimeo.*

[16] 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.

[17] Golosov, M., M. Troshkin, A. Tsyvinski and M. Weinzeirl (2013), "Preference Heterogeneity and Optimal Capital Income Taxation", *Journal of Public Economics*, 97, 160-175.

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

[19] Guo, J-T. and A. Krause (2013), "Optimal Nonlinear Taxation of Income and Education Expenditures", *Oxford Economic Papers*, 65, 74-95.

[20] Guo, J-T. and A. Krause (2014), "Optimal Dynamic Nonlinear Income Taxation under Loose Commitment", *Macroeconomic Dynamics*, 18, 1403-1427.

[21] Guo, J-T. and A. Krause (2015a), "Dynamic Nonlinear Income Taxation with Quasi-Hyperbolic Discounting and No Commitment", *Journal of Economic Behavior and Organization*, 109, 101-119.

[22] Guo, J-T. and A. Krause (2015b), "Dynamic Income Taxation without Commitment: Comparing Alternative Tax Systems", *Economic Modelling*, 47, 319-326.

[23] Judd, K. (1985), "Redistributive Taxation in a Simple Perfect Foresight Model", *Journal of Public Economics*, 28, 59-83.

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

[25] Krause, A. (2014), "Optimal Savings Taxation when Individuals have Different CRRA Utility Functions", *Annals of Economics and Statistics*, 113/114, 207-223.

[26] Krause, A. (2017), "On Redistributive Taxation Under the Threat of High-Skill Emigration", *Social Choice and Welfare*, 48, 845-856.

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

[28] Morita, S. (2016), "The Time-Consistent Public Goods Provision", *Journal of Public Economic Theory*, 18, 923-941.

[29] OECD (2014), *Highlights from Education at a Glance*, OECD publications.

[30] Saez, E. (2002), "The Desirability of Commodity Taxation under Non-linear Income Taxation and Heterogeneous Tastes", *Journal of Public Economics*, 83, 217-230.

[31] Scheuer, F. and A. Wolitzky (2016), "Capital Taxation under Political Constraints", *American Economic Review*, 106, 2304-2328.

[32] Straub, L. and I. Werning (2015), "Positive Long Run Capital Taxation: Chamley-Judd Revisited", *mimeo*.

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

TABLE 1

A Numerical Example

Baseline parameter values

$\phi = 1/3$	$\delta = 0.456$	$w_L = 1.0$	$p^L = 0.25$
$\pi_H^H = 1/6$	$r = 0.04$	$w_H = 1.6$	$p^H = 0.75$
$\pi_H^L = 1/6$			$p = 0.50$

	FC	*CWC*	*LC*	*NC*
$MTRS_H^H$	0.000	0.072	0.167	0.286
$MTRS_H^L$	0.000	0.329	0.167	0.286
$MTRS_L$	0.000	–0.174	–0.110	–0.249

Memo item: Under CWC, incentive-compatibility constraints (5.3), (5.4), and (5.5) are all binding. All omitted incentive-compatibility constraints are slack.

TABLE 2

CWC over the Long-Run

Baseline parameter values

$\phi = 1/3$	$\delta = 0.456$	$w_L = 1.0$	$p^L = 0.25$	$p^{\bar{L}} = 0.275$	$p^{\underline{L}} = 0.225$
$\pi_H^H = 1/6$	$r = 0.04$	$w_H = 1.6$	$p^H = 0.75$	$p^{\bar{H}} = 0.825$	$p^{\underline{H}} = 0.675$
$\pi_H^L = 1/6$	$\beta = 0.208$		$q = 0.50$	$q_C = 0.55$	$q_N = 0.45$

	Economy 1	*Economy T+1 following commit*	*Economy T+1 following no-commit*
$MTRS_H^H$	0.193	0.174	0.211
$MTRS_H^L$	0.308	0.313	0.304
$MTRS_L$	–0.210	–0.202	–0.218

FIGURE 1

CWC over the long-run: Belief sensitivity

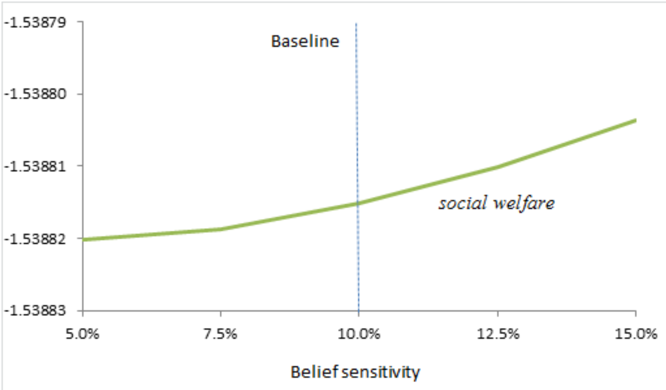

FIGURE 2

CWC over the long-run: Commitment sensitivity

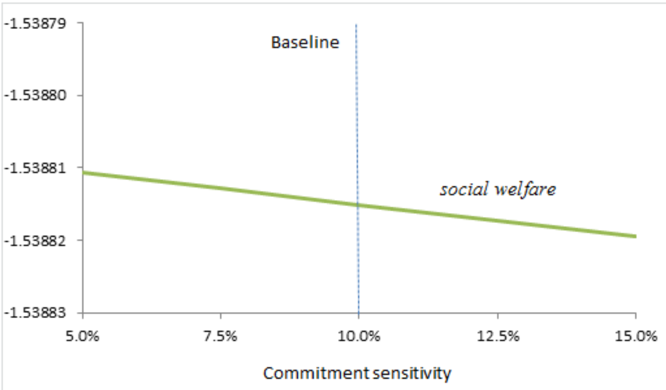

FIGURE 3

CWC over the long-run: Equal belief and commitment sensitivity