Vanderbilt University Department of Economics Working Paper Series
Nashville, TN 37235-1819

Vanderbilt University Department of Economics Working Papers 16-00010

Optimal Nonlinear Taxation of Income and Savings Without Commitment

Craig Brett
Mount Allison University

John A Weymark
Vanderbilt University

Abstract

When a government is unable to commit to its future tax policies, information about taxpayers' characteristics revealed by their behavior may be used to extract more taxes from them in the future. We examine the implications of this ratchet effect for the design of redistributive income and savings tax policies in a two-period model with two types of individuals who only differ in their skill levels. When commitment is not possible, it may be optimal to separate, pool, or partially pool different types in period one. The nature of the distortions to labor supplies and savings are investigated for each of these three regimes. Novel rationales for savings distortions are identified.

Citation: Craig Brett and John A Weymark, (2016) "Optimal Nonlinear Taxation of Income and Savings Without Commitment", *Vanderbilt University Department of Economics Working Papers*, VUECON-16-00010.
Contact: Craig Brett - cbrett@mta.ca, John A Weymark - john.weymark@vanderbilt.edu.
Submitted: March 08, 2016. **Published:** March 13, 2016.
URL:http://www.accessecon.com/Pubs/VUECON/VUECON-16-00010.pdf

Optimal Nonlinear Taxation of Income and Savings Without Commitment

Craig Brett[a], **John A. Weymark**[b]

[a]*Department of Economics, Mount Allison University, 144 Main Street, Sackville NB E4L 1A7, Canada. E-mail:* `cbrett@mta.ca`

[b]*Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA. E-mail:* `john.weymark@vanderbilt.edu`

March 2016

Abstract. When a government is unable to commit to its future tax policies, information about taxpayers' characteristics revealed by their behavior may be used to extract more taxes from them in the future. We examine the implications of this ratchet effect for the design of redistributive income and savings tax policies in a two-period model with two types of individuals who only differ in their skill levels. When commitment is not possible, it may be optimal to separate, pool, or partially pool different types in period one. The nature of the distortions to labor supplies and savings are investigated for each of these three regimes. Novel rationales for savings distortions are identified.

Journal of Economic Literature classification numbers. D82, H21

Keywords. asymmetric information; commitment; optimal income taxation; ratchet effect; savings taxation

---

1. Introduction

Ever since the pathbreaking work of Mirrlees (1971), a government's lack of full information about the tax-relevant characteristics of those subject to its taxation authority has been viewed as a fundamental constraint on the design of tax policy. In the context of redistributive nonlinear income taxation, a government's egalitarian intentions may be hampered by its inability to identify the labor productivities of different taxpayers (their "types").[1] The asymmetric information approach to taxation was originally developed for atemporal environments. A major impediment to extending the Mirrlees model to dynamic settings is that information revealed by taxpayers in one period can be used by the government in subsequent periods. Aware of this possibility, rational taxpayers may modify their behavior in early periods in an attempt to better conceal their characteristics. In particular, more able taxpayers might fear the Weitzman (1980) ratchet effect, whereby the government may use any knowledge revealed by their behavior to extract more taxes from them in the future. The ratchet effect would not arise if the government could commit to forgetting any information it learns at the beginning of each new tax year. However, such a commitment is not credible.

In this article, we investigate the implications for redistributive tax policy of a government's inability to commit to its future actions. We consider an economy in which a continuum of individuals of two productivity types work and consume in each of two periods. In order to focus on the interaction between information revelation and redistribution, we set aside the social insurance motive for taxation by assuming that productivities do not change over time.[2] Individuals may also transfer resources forward in time through saving. A utilitarian government optimally chooses nonlinear taxes on income and savings in a dynamically consistent way taking into account the budget and incentive constraints that it faces. Thus, the government is unable to commit to its second period tax policy in advance, and so any information about individual productivities revealed in the first period can be used when designing second period taxes. Equivalently, the government sets a joint tax on income and savings based on all observable variables up to the time period in question.[3] The dependence of second period taxes on information

[1]See Tuomala (2016) for an introduction to optimal nonlinear income taxation using the Mirrlees framework.

[2]The role that taxation plays in providing social insurance has been a major focus of the "new dynamic public finance" literature surveyed by Golosov and Tsyvinski (2015), Golosov, Tsyvinski, and Werning (2007), and Kocherlakota (2006, 2010). With some exceptions, this literature assumes that governments can commit in advance to their future tax policies. When there is no government commitment and individual productivities are stochastic over time, types may never be known with certainty, thereby attenuating the ratchet effects that are exhibited when types are unchanged over time. See, for example, Battaglini and Coate (2008) and Golosov and Iovino (2015). If there is savings taxation, as Bisin and Rampini (2006) show, the power of the government to take advantage of information revealed can be somewhat mitigated if individuals have access to capital markets that prevent the government from observing their total savings.

[3]Conditioning taxes on both current and past values of the variables in the tax base is a feature of actual tax practice. For example, as Kocherlakota (2006, p. 269) notes, the alternative minimum tax and

revealed through behavior in the first period is rationally anticipated by the taxpayers. As is the case with the government, individuals cannot make commitments about second period behavior in the first period.

We assume that the preferences of individuals are additively separable both across time and between consumption and leisure. These assumptions on preferences guarantee that everybody's optimal marginal tax rate on savings is zero in the benchmark case in which the government can commit to ignore information revealed in the first period when designing second period tax policy. We show that in the absence of commitment, when it is optimal to induce all individuals to reveal their types in the first period, supplementing income taxation with a subsidy on the savings of the low skilled is welfare enhancing because it relaxes an incentive compatibility constraint. However, if there is no type revelation in period one, then there are zero marginal taxes on savings on average, and this necessitates subsidizing some individuals' savings and taxing the savings of others.[4] Thus, our analysis identifies novel rationales for savings distortions even when preferences are separable across time.

There are three possible kinds of optimal tax regimes. In a separating tax regime, type information is revealed in the first period, whereas in a pooling tax regime, type information remains hidden for all individuals after the first period. In a semi-pooling regime, some individuals reveal their type information in the first period, while others do not. Whether the optimal regime is separating, pooling, or semi-pooling depends on a discrete comparison among the best tax policies for each regime. Such a comparison requires additional assumptions about the functional form of the utility functions or about the values of the parameters that appear in our model. We do not explore the issue of identifying the globally optimal regime.

When types are separated in the first period, we show that it is optimal to subsidize the savings of the low skilled. By distorting their savings upward in the first period, the government makes additional resources available in the second period. The availability of these additional resources induces changes in the second period tax scheme that are more favorable to the high skilled. This, in turn, reduces the incentive for the high skilled to conceal their productivity in the first period. As in atemporal models, in the first period, it is optimal for high-skilled individuals to face a zero marginal income tax rate, whereas low-skilled individuals face a positive marginal income tax rate. Because there is complete revelation of types in the first period, personalized lump-sum taxes and transfers are optimal in the second period.

When types are pooled in the first period, it is optimal for low-skilled individuals to face a positive marginal rate of savings taxation, while the savings of high-skilled

---

the carry-forward of certain deductions introduce history dependence in the U.S. tax code. Likewise, in Canada, information about behavior in previous tax years is used in setting limits for contributions to tax-deferred savings schemes.

[4]While this result is reminiscent of the zero average capital tax result of Kocherlakota (2005), the mechanism underlying it is quite different. An aggregate distortion is not needed in Kocherlakota's model because the optimal intertemporal wedge is achieved using incomplete social insurance.

individuals are subsidized at the margin. However, on average, there is no marginal distortion on savings. With pooling in the first period, the second period is a standard one-period optimal income tax problem, and so it is optimal for high-skilled individuals to face a zero marginal income tax rate and for this rate to be positive for low-skilled individuals. In the first period, both types of individuals are distorted in the labor market, with low-skilled individuals facing a negative marginal income tax rate and high-skilled individuals facing a positive marginal income tax rate. As with savings, on average, there is no distortion in the first period labor market.

When some, but not all, of the high-skilled individuals are separated in period one, the high skilled who reveal their type face no marginal distortion in the labor market in either period. The high skilled who conceal their type face a positive marginal income tax rate in the first period, but are undistorted at the margin in the second. The low skilled face a positive marginal income tax rate in period two. In the first period, they are taxed less on the margin than the high skilled who are pooled with them, but the sign of their marginal income tax rate in this period is indeterminate. High-skilled individuals who conceal their type in the first period have their savings taxed more heavily or subsidized less at the margin than do any other individuals, but it is impossible to sign the distortions in intertemporal consumption without making further assumptions.

The rest of this article is organized as follows. In Section 2, we discuss some of the related literature. Section 3 describes the economy. In order to provide a benchmark for our analysis of optimal taxation without commitment, in Section 4, we identify the qualitative properties of the solution to the optimal tax design problem under the assumption that the government can commit to a second period tax schedule before type information is revealed. Sections 5–8 consider the optimal tax design problem without commitment. Section 5 provides an introduction to the incentive issues that arise when there is no commitment. In Section 6, we determine the properties of the solution to the optimal tax design problem when it is optimal to separate the two types in the first period. In Section 7, we analyze the second period tax design problem when some or all of the high-skilled individuals do not truthfully reveal their type in period one. The first period of this problem is analyzed in Section 8. We offer concluding remarks in Section 9. The proofs of our results are gathered in Appendix A.

2. Related Literature

The question of whether it is optimal to separate or pool individuals in nonlinear income tax models with a ratchet effect has been considered by Roberts (1984) and Berliant and Ledyard (2014). Like us, they employ a deterministic framework to analyze dynamic optimal nonlinear income taxation when the government cannot commit to ignore information gathered in earlier periods. They do not consider savings taxation as a possible instrument. Roberts shows that types are never separated in an infinite horizon economy with a finite number of types provided that the government revenue requirement is not so large as to bankrupt any individual, whereas Berliant and Ledyard identify sufficient conditions for type information to be revealed in the first period of a two-period economy with a continuum of types.[5]

The work closest to our own is that of Apps and Rees (2009).[6] As is the case here, they consider a two-type, two-period model with a continuum of individuals of each type. However, they do not allow for savings taxation. The tax distortions on labor earnings are the same in our model as in that of Apps and Rees.

An implication of the Atkinson–Stiglitz Theorem (Atkinson and Stiglitz, 1976) is that when consumption is separable from labor supply in individual preferences, as is the case here, there is no value to supplementing an optimal nonlinear income tax with savings taxation when the government can commit to its future tax policies. A number of rationales for savings taxation have been offered when we depart from this scenario.

A number of these rationales have been developed using models in which there is government commitment. Browning and Burbidge (1990) show that when the government has a different rate of time preference than does the private sector, there is a case for distortionary savings taxation. In an overlapping generation model in which individuals only work when young, Ordover and Phelps (1979) show that there is a role for savings taxation whenever the marginal rate of substitution between consumption when young and consumption when old depends on labor supply. In a similar model, Pirttilä and Tuomala (2001) argue that distorting savings decisions can be optimal even when preferences are separable across time if future relative wages are sensitive to current savings via their effect on capital accumulation. Golosov, Kocherlakota, and Tsyvinski (2003) show that savings taxation enhances work incentives so as to counteract excessive insurance against future consumption shocks.

Further rationales for savings taxation have been proposed when governments cannot commit to future policies. In an infinite-horizon representative agent model, Benhabib and Rustichini (1997) show that the government can use the capital stock as a device to make some tax policies more credible. Savings distortions can influence the evolution of the the capital stock, and so help to partially overcome its commitment problems. Other studies have highlighted commitment problems in political economy settings. Farhi, Sleet, Werning, and Yeltekin (2012) provide a justification for a progressive tax on the returns to savings (a capital income tax) in a model with repeated elections in which tax policies play a role in the probability of a political party being elected. Scheuer and Wolitzky (2016) consider a political economy model in which a sustainable tax policy must be immune to the formation of coalitions who would reform it. They show that savings distortions attenuate this threat.

In our model, savings taxation help to mitigate the distortions introduced because of the government's lack of commitment. Other policy instruments can also serve this

[5]Gaube (2010) investigates the welfare ranking of the full commitment and the no commitment regimes with a partial commitment regime in which the second period tax schedule cannot depend on first period incomes. Dillén and Lundholm (1996) also analyze when it is optimal to separate or pool types in a two-period model, but restrict attention to linear income taxation.

[6]The first version of our article was completed before we learned of their research.

purpose. For example, Boadway, Marceau, and Marchand (1996) show that mandating a minimum amount of time spent in publicly-observable education is such an instrument. For further discussion of optimal education policies, see Boadway (2012) and Golosov and Tsyvinski (2015).[7]

3. The Model

The economy lasts for two time periods. There is a continuum of individuals with unit measure. There are two types of individuals, $i = 1, 2$, who differ in labor productivity, with the skill level of an individual of type i given by the parameter w_i, where $w_1 < w_2$. An individual's skill is the same in both periods. Thus, the low skilled are of type 1 and the high skilled are of type 2. An individual is low skilled with probability α, where $0 < \alpha < 1$. An individual of type i supplies l_i^t units of labor and consumes c_i^t units of a single consumption good in period t, $t = 1, 2$.[8] Individuals may transfer wealth between the two periods by saving the amount s_i of the consumption good. In keeping with the literature on optimal nonlinear income taxation, skill is interpreted as an enhancement to effective labor, so that effective labor in period t is $y_i^t = w_i l_i^t$. The production technology exhibits constant returns to scale. In each period, one unit of effective labor is required to produce one unit of the consumption good. Each unit of the consumption good stored in the first period produces $1+r$ units of the consumption good in the second period, where $r > 0$. As in Boadway, Marceau, and Marchand (1996), individuals may not borrow against future income. The labor market is perfectly competitive in each period, so that an individual's effective labor supply equals his labor income before taxes.

The government designs a redistributive tax system. It cannot observe an individual's labor supply or skill level, but it can observe before-tax income and savings. Moreover, it knows the distribution of types. The total tax paid by an individual in either period can be made contingent on the amount he saves, his current income, and, in the case of period 2, his income in the previous period. Thus, the after-tax income x_i^t of a person of type i in period t, the difference between effective labor supply and the total tax paid, depends on the values of his current and past incomes and of his savings. Individuals are free to divide their first period after-tax incomes between consumption and savings. Each unit of savings affords a consumer an additional $1+r$ units of consumption in the second period over and above his second period after-tax income.[9] Therefore, consumption in each period is given by

$$c_i^1 = x_i^1 - s_i, \quad c_i^2 = x_i^2 + (1+r)s_i, \quad i = 1, 2. \tag{3.1}$$

[7]Using a model similar to ours, Krause (2009) investigates skill formation when there is learning-by-doing.

[8]Subscripts index individuals, while superscripts index time periods.

[9]Although strictly speaking, in period 2, person i's before- and after-tax incomes are $y_i^2 + rs_i$ and $x_i^2 + rs_i$, respectively, henceforth when we speak of before- and after tax income, we shall not include the interest income.

Individuals have identical preferences over consumption and labor supply, additive in all goods and across time, and represented by the utility function

$$U(c_i^1, l_i^1, c_i^2, l_i^2) = u(c_i^1) - g(l_i^1) + v(c_i^2) - h(l_i^2), \quad i = 1, 2. \tag{3.2}$$

The functions $u(\cdot)$ and $v(\cdot)$ are increasing, strictly concave, and twice continuously differentiable, while the functions $g(\cdot)$ and $h(\cdot)$ are increasing, strictly convex, and twice continuously differentiable. An allocation for an individual of type i is a vector $(x_i^1, y_i^1, x_i^2, y_i^2, s_i)$. These are the variables that the government can observe. Preference over allocations are given by

$$u(x_i^1 - s_i) - g\left(\frac{y_i^1}{w_i}\right) + v(x_i^2 + (1+r)s_i) - h\left(\frac{y_i^2}{w_i}\right), \quad i = 1, 2. \tag{3.3}$$

The marginal rate of substitution between before-tax income and after-tax income in the first period for a person of type i is

$$\text{MRS}_{y_i^1, x_i^1} = \frac{g'\left(\frac{y_i^1}{w_i}\right)}{w_i u'(c_i^1)}, \tag{3.4}$$

while that person's marginal rate of substitution between before-tax income and after-tax income in the second period is

$$\text{MRS}_{y_i^2, x_i^2} = \frac{h'\left(\frac{y_i^2}{w_i}\right)}{w_i v'(c_i^2)}. \tag{3.5}$$

Holding incomes and consumption levels constant, the marginal rates of substitution between before-tax and after-tax income are decreasing in the skill level because high-skilled individuals must work fewer additional hours for each additional unit of before-tax income than do low-skilled individuals. Thus, it takes a smaller increase in after-tax income to compensate a high-skilled individual for increases in before-tax income than it does to compensate a low-skilled individual.

The marginal rate of substitution between after-tax income in period 1 and after-tax income in period 2 for a person of type i is

$$\text{MRS}_{x_i^1, x_i^2} = -\frac{u'(c_i^1)}{v'(c_i^2)}. \tag{3.6}$$

This intertemporal marginal rate of substitution does not depend explicitly upon the skill level. Because of their common preferences over consumption and labor supply, all individuals have the same willingness to trade consumption across time periods. The additive nature of preferences implies that the marginal rate of substitution between period 1 consumption and period 2 consumption does not depend on the amount of labor supplied in either period.

The government may also engage in saving by storing an amount s_G of the consumption good. The storage technology available to the government is exactly the same as the storage technology for the private sector.

We assume that the government has a utilitarian objective function. Thus, it evaluates outcomes using the social welfare function

$$\mathcal{W}(x_1^1, x_2^1, y_1^1, y_2^1, x_1^2, x_2^2, y_1^2, y_2^2, s_1, s_2) = \alpha \left[u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_1}\right) + v(x_1^2 + (1+r)s_1) - h\left(\frac{y_1^2}{w_1}\right) \right] + (1-\alpha)\left[(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right) + v(x_2^2 + (1+r)s_2) - h\left(\frac{y_2^2}{w_2}\right) \right]. \tag{3.7}$$

4. Optimal Taxation with Commitment

First-best taxation is infeasible in this economy because the government cannot distinguish *ex ante* between the two types of individuals. In order to provide a benchmark for our analysis of the tax design problem without commitment, in this section, we assume that the government can commit to an anonymous tax schedule that specifies an individual's after-tax incomes in the two time periods as a function of his before-tax incomes in both periods and of his savings. An implication of this assumption is that the government is able to credibly commit not to use information about the skill levels of the individuals revealed in the first period to adjust taxes in the second period. By the Taxation Principle of Hammond (1979) and Guesnerie (1995) for continuum economies, having each individual choose a utility-maximizing allocation given such a tax schedule is equivalent to having the government choose the allocations directly using an incentive-compatible allocation mechanism, the same for everyone, that specifies each individual's before- and after-tax incomes and savings as a function of his type. With full commitment, incentive compatibility is the requirement that each individual weakly prefers the entire allocation (over both periods) designed for him to the allocations designed for the other individuals. Formally,

$$u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_1}\right) + v(x_1^2 + (1+r)s_1) - h\left(\frac{y_1^2}{w_1}\right) \geq u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_1}\right) + v(x_2^2 + (1+r)s_2) - h\left(\frac{y_2^2}{w_1}\right) \tag{4.1}$$

and

$$u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right) + v(x_2^2 + (1+r)s_2) - h\left(\frac{y_2^2}{w_2}\right) \geq u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_2}\right) + v(x_1^2 + (1+r)s_1) - h\left(\frac{y_1^2}{w_2}\right). \tag{4.2}$$

We assume that only the incentive compatibility condition (4.2) might potentially bind, what Stiglitz (1982) calls the normal case.[10] Given its utilitarian objective, the government wishes to redistribute income from the high skilled to the low skilled. The natural limit on this redistribution is that, if taken too far, such redistribution might induce a high-skilled individual to pretend to be low-skilled. Imposing (4.2) prevents this type of mimicking.[11]

Given the storage technology available to both individuals and the government, the materials balance constraints for the economy are

$$\alpha x_1^1 + (1-\alpha)x_2^1 + s_G \leq \alpha y_1^1 + (1-\alpha)y_2^1 \tag{4.3}$$

and

$$\alpha x_1^2 + (1-\alpha)x_2^2 \leq \alpha y_1^2 + (1-\alpha)y_2^2 + (1+r)s_G. \tag{4.4}$$

Thus, the problem faced by the government can be specified as follows:

Second-Best Tax Design Problem with Commitment. The government chooses an allocation $(x_1^1, x_2^1, y_1^1, y_2^1, x_1^2, x_2^2, y_1^2, y_2^2, s_1, s_2, s_G)$ to maximize the social welfare function (3.7) subject to the materials balance constraints (4.3) and (4.4) and the two-period incentive compatibility constraint (4.2).[12]

The second-best tax design problem with commitment is a standard one-dimensional screening problem. Because there are five components to each individual's allocation, the government has more instruments than the minimum required to achieve separation.[13] Given the adverse selection problem faced by the government, some distortions to behavior are inevitable. Proposition 1 describes the pattern of distortions at a solution to the government's full-commitment problem.

Proposition 1. *At a solution to the second-best tax design problem with commitment:*

(i) $\mathrm{MRS}_{y_2^1,x_2^1} = 1$, $\mathrm{MRS}_{y_2^2,x_2^2} = 1$ *and* $\mathrm{MRS}_{x_2^1,x_2^2} = -(1+r)$.

(ii) $\mathrm{MRS}_{y_1^1,x_1^1} < 1$, $\mathrm{MRS}_{y_1^2,x_1^2} < 1$ *and* $\mathrm{MRS}_{x_1^1,x_1^2} = -(1+r)$.

In order to interpret Proposition 1, it is useful to determine its implications for optimal marginal tax rates. Optimal income tax schedules may be nondifferentiable. The implicit marginal income tax rate in period t for an individual of type i is $1 - \mathrm{MRS}_{y_i^t,x_i^t}$. Because

[10] We also assume that low-skilled individuals do not face a binding incentive compatibility condition when we consider the no commitment case in the next section.

[11] Indeed, at a solution to the first-best taxation problem for this economy, the government wishes to equalize the consumption of all individuals in each time period and to require the high-skilled individuals to work more. Thus, (4.2) is violated at the first-best allocation, while (4.1) is slack.

[12] In all of our tax design problems, we assume that the omitted nonnegativity constraints do not bind. We also assume that each of these problems has a solution.

[13] Separation is possible in two-good worlds when there is asymmetric information in one dimension.

$\text{MRS}_{y_i^t,x_i^t} > 0$, marginal income tax rates are bounded above by 1. Similarly, because $\text{MRS}_{x_i^1,x_i^2} < 0$, the implicit marginal tax rate on first period consumption for an individual of type i is $(1+r) - |\text{MRS}_{x_i^1,x_i^2}| = (1+r) + \text{MRS}_{x_i^1,x_i^2}$. Hence, his implict marginal tax rate on savings is $-(1+r) - \text{MRS}_{x_i^1,x_i^2}$.

The first two equalities in part (i) of Proposition 1 show that the high skilled are not distorted at the margin in the labor market. That is, in both periods, they face a zero marginal income tax rate. The inequalities in part (ii) describe the distortions to the labor supply decisions of the low skilled caused by the asymmetric information. Because the first-best solution is not incentive compatible, constraint (4.2) must bind at a solution to the second-best problem with commitment. It follows from Brito, Hamilton, Slutsky, and Stiglitz (1990, Proposition 5) that the marginal rate of substitution for low-skilled individuals is distorted only for those pairs of goods for which the two types of individuals have a different marginal rate of substitution at the low-skilled allocation. Because the marginal rates of substitution between before-tax income and after-tax income vary by skill level, the effective labor-consumption margin is distorted for the low skilled. They face a positive marginal income tax rate in both periods so as to relax the high-skilled incentive constraint. On the other hand, all individuals have the same intertemporal preferences. In particular, both (a) the low skilled and (b) the high skilled when they mimic the low skilled are willing to trade consumption across time at the same implicit prices. Thus, there is no informational advantage to be had by changing the intertemporal relative price of consumption. Therefore, as shown by the last equalities in parts (i) and (ii) of Proposition 1, savings decisions are not distorted, and hence not taxed, at the margin.

5. Optimal Taxation Without Commitment

The government's ability to commit in the first period to the second-period tax schedule is not credible, nor is it credible for individuals to commit to their second-period labor supply decisions in period 1. The optimal two-period schedule with commitment offers different allocations to the two types of individuals in the first period. With full knowledge of the structure of the economy, this allows the government to infer the identities of all individuals at the end of the first period. The information asymmetry between the government and the private sector disappears, and there is no need to distort behavior in the second period. Because the optimal second-best schedule with commitment features a distortion in the period 2 labor supply of low-skilled individuals, it would not be chosen by a government that has the ability to re-optimize after the first period. Furthermore, because savings decisions have already been fixed in the first period, the government has an incentive to increase the tax on savings of high-skilled individuals beyond what is optimal with commitment in order to further its redistributional goals. Hence, the optimal tax schedule with commitment is time inconsistent.

Individuals are aware that the government is able to use information gleaned in the first period when setting second-period taxes. In particular, a high-skilled individual understands that if his type is revealed in the first period, then it will be easier to redistribute income from him to low-skilled individuals in the second period because it is no longer necessary to worry about incentive compatibility constraints. This can be accomplished by transferring more of a high-skilled individual's savings and interest income to the low-skilled or by providing an incentive for him to work more so that there is more of his labor income available to redistribute. Thus, there is an increased incentive for high-skilled individuals to conceal information in the first period.

For its part, the government is aware of the added incentive to hide information in the first period. It realizes that the full-commitment tax schedule may need to be modified in order to induce information revelation in the first period. As pointed out by Freixas, Guesnerie, and Tirole (1985) in a more general planning context and by Dillén and Lundholm (1996) for linear income taxes, such modifications may be sufficiently costly to lead the government to prefer not to separate types in the first period. In order to determine whether first-period separation is optimal, the government needs to compare the gains accruing from the use of first-best taxation in the second period to the costs incurred in the first period of extracting the information it needs to implement the second period first-best allocation.

A high-skilled individual successfully conceals his type only if he chooses the same before-tax income, after-tax income, and savings in the first period as do low-skilled individuals. If all high-skilled individuals conceal their type information, we have a pooling outcome. A high-skilled individual reveals his type if the first-period allocation that he chooses differs in any component from that chosen by the low skilled. If every high-skilled individual reveals his type, we have a separating outcome. If high-skilled individuals are indifferent between concealing their type information and revealing it, then it is possible that some of them reveal their identities while others do not. The outcome is then semi-pooling. The tax schedule offered in the first period and the anticipated tax schedule for the second period shape the choices of the two types of individuals and implicitly determine whether there is pooling, semi-pooling, or separation in the first period. The first period revelation outcome is discrete; there is either pooling or separation or semi-pooling. Deciding which of the three configurations is best requires a comparison among the maximized values of social welfare in the three cases. In general, such a comparison depends on the exact form of the utility function and on the values of the parameters in our model. Before making this comparison, it is necessary to determine the optimal separating, pooling, and semi-pooling solutions. The properties of these solutions are the focus of our analysis.

6. The No-Commitment Tax Design Problem with First Period Separation

If the two types of individuals make different choices in the first period, then the government has sufficient information to carry out lump-sum taxation in the second period. These taxes can be paid out of an individual's second period labor income and his interest-augmented savings. Second period social welfare is

$$\begin{aligned}\mathcal{W}^2(x_1^2, x_2^2, y_1^2, y_2^2, s_1, s_2) = \alpha \left[v(x_1^2 + (1+r)s_1) - h\left(\frac{y_1^2}{w_1}\right)\right] \\ + (1-\alpha)\left[v(x_2^2 + (1+r)s_2) - h\left(\frac{y_2^2}{w_2}\right)\right].\end{aligned} \tag{6.1}$$

The problem faced by the government in the second period is:

Second Period First-Best Problem. Given (s_1, s_2, s_G), the government chooses a second period allocation $(x_1^2, x_2^2, y_1^2, y_2^2)$ to maximize the second period social welfare function (6.1) subject to the the second period materials balance constraint (4.4).

The second period first-best problem problem has a strictly concave objective function and a single linear constraint, which can easily be shown to bind at the solution to this problem. Each of the four components of this solution depends on the vector $\mathbf{s} = (s_1, s_2, s_G)$ of predetermined savings levels. Because this problem is so well-behaved, its comparative static properties with respect to each component of the savings vector can be derived using standard methods from consumer theory. The properties most pertinent to a characterization of the optimal first period distortions when types are separated in this period are collected in the following lemma.

Lemma 1. *For a given savings vector* $\mathbf{s}$*, the second period first-best problem has a unique solution. Moreover, the solution functions* $x_1^2(\mathbf{s})$*,* $x_2^2(\mathbf{s})$*,* $y_1^2(\mathbf{s})$*, and* $y_2^2(\mathbf{s})$ *are continuously differentiable and satisfy the following conditions:*

(i) $v'(x_1^2(\mathbf{s}) + (1+r)s_1) = v'(x_2^2(\mathbf{s}) + (1+r)s_2) = \dfrac{1}{w_1} h'\left(\dfrac{y_1^2(\mathbf{s})}{w_1}\right) = \dfrac{1}{w_2} h'\left(\dfrac{y_2^2(\mathbf{s})}{w_2}\right)$.

(ii) $\alpha \dfrac{\partial x_1^2(\mathbf{s})}{\partial s_i} + (1-\alpha)\dfrac{\partial x_2^2(\mathbf{s})}{\partial s_i} - \alpha \dfrac{\partial y_1^2(\mathbf{s})}{\partial s_i} - (1-\alpha)\dfrac{\partial y_2^2(\mathbf{s})}{\partial s_i} = 0, \quad i = 1, 2.$

(iii) $\dfrac{\partial y_1^2(\mathbf{s})}{\partial s_1} - \dfrac{\partial x_1^2(\mathbf{s})}{\partial s_1} > \dfrac{\partial y_2^2(\mathbf{s})}{\partial s_1} - \dfrac{\partial x_2^2(\mathbf{s})}{\partial s_1}$.

(iv) $\dfrac{\partial y_1^2(\mathbf{s})}{\partial s_2} - \dfrac{\partial x_1^2(\mathbf{s})}{\partial s_2} < \dfrac{\partial y_2^2(\mathbf{s})}{\partial s_2} - \dfrac{\partial x_2^2(\mathbf{s})}{\partial s_2}$.

With separation, in the second period, we have a full information planning problem in which the government has access to the interest-augmented savings from the first period to distribute as it wishes. Part (i) of Lemma 1 summarizes the marginal conditions for a first-best utilitarian optimum in the second period. The government wishes to equate the marginal utilities of consumption and of income for all individuals. Given identical additively separable preferences, equality of the marginal utilities of consumption implies equal consumption for all individuals. Because the marginal utility of consumption equals the marginal disutility of income, the marginal rate of substitution between labor and consumption equals the wage rate for each individual.

Because high-skilled individuals have a higher wage rate, they also have a higher marginal disutility of labor at the first-best optimum. Given identical preferences with increasing marginal disutility of labor, high-skilled individuals must work more than do low-skilled individuals. Because agreeing to work more than someone else for equal consumption is not incentive compatible, the government must make use of the skill information revealed in the first period in order to implement this scheme using person-specific lump sum taxes and transfers.

Part (ii) of Lemma 1 follows directly from the second period materials balance constraint. This does not mean that optimal second period before-tax and after-tax incomes are insensitive to individual wealth holdings. Indeed, it is feasible for the government to tax away all first-period savings. Instead, part (ii) simply says that changes in aggregate before-tax income are offset by changes in aggregate after-tax income.

Savings, regardless of who generates them, represent goods that the government may allocate as it sees fit. An increase in anyone's savings increases wealth in the second period. The government optimally responds by increasing consumption equally and decreasing before-tax incomes. The wealth generated by an increase in the savings of one of the skill types is partly redistributed to the other type and so, as as shown in parts (iii) and (iv) of Lemma 1, a marginal increase in someone's savings has a larger impact on the tax collected from him than it does from the tax collected from someone of the other skill type.

All decision makers in the economy, both private and public, recognize that the government is unable to commit to any second period taxation scheme apart from the one that is the second period optimum, given first period savings. Private individuals take this lack of commitment into account when deciding on their first period courses of action, notably when making their savings decisions. Furthermore, individuals cannot credibly commit to second period labor supply decisions that are not optimal for them when the time comes for them to supply this labor. Moreover, in order to achieve complete separation, the government must provide sufficient incentive for high-skilled individuals to reveal their type in the first period. Such an incentive is provided if the following condition is met:

$$\begin{aligned} u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right) + v(x_2^2(\mathbf{s}) + (1+r)s_2) - h\left(\frac{y_2^2(\mathbf{s})}{w_2}\right) \\ \geq u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_2}\right) + v(x_1^2(\mathbf{s}) + (1+r)s_1) - h\left(\frac{y_1^2(\mathbf{s})}{w_2}\right). \end{aligned} \tag{6.2}$$

Because there is a continuum of individuals, if a high-skilled individual lies about his type in period 1, he does not change the fraction of individuals who reveal themselves to be high skilled. Thus, he anticipates that the allocation on offer in period 2 is the one that solves the second period first-best problem. Moreover, he knows that it would be better in period 2 for him to mimic the low skilled if he has not revealed his own type in period 1 because, by Lemma 1, both types have the same second-period consumption, but the low skilled work less. Thus, the right-hand side of (6.2) is the largest two-period utility possible for a high-skilled individual if he mimics the low-skilled in period 1.

The government designs its first period tax system fully aware of how it will respond in the second period to its own first period actions and to the savings decisions of the private individuals. Its first period objective function, which includes the social welfare accruing in the second period, is

$$\mathcal{W}^{sep}(x_1^1, x_2^1, y_1^1, y_2^1, \mathbf{s}) = \alpha \left[u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_1}\right) + v(x_1^2(\mathbf{s}) + (1+r)s_1) - h\left(\frac{y_1^2(\mathbf{s})}{w_1}\right)\right] + (1-\alpha)\left[u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right) + v(x_2^2(\mathbf{s}) + (1+r)s_2) - h\left(\frac{y_2^2(\mathbf{s})}{w_2}\right)\right]. \tag{6.3}$$

Because both the incentive compatibility condition (6.2) and the objective function (6.3) include the solution functions to the second period first-best problem, the materials balance constraint in period 2 is accounted for. However, the government must take account of the first period materials balance constraint. Thus, it faces the following tax design problem in period 1.

First Period No-Commitment Tax Design Problem with Separation. The government chooses a first period allocation $(x_1^1, x_2^1, y_1^1, y_2^1, s_1, s_2, s_G)$ to maximize the objective function (6.3) subject to the first period materials balance constraint (4.3) and the incentive compatibility constraint (6.2).

The pattern of unambiguous distortions to labor supply and savings behavior arising at a solution to the first period no-commitment tax design problem with separation are given in the following proposition.

Proposition 2. *At a solution to the first period no-commitment tax design problem with separation:*

(i) $\mathrm{MRS}_{y_1^1, x_1^1} < 1$ *and* $\mathrm{MRS}_{y_2^1, x_2^1} = 1$.

(ii) $\mathrm{MRS}_{x_1^1, x_1^2} < -(1+r)$.

Part (i) of Proposition 2 indicates that, at a solution to the no-commitment tax design problem with separation, low-skilled individuals face a positive first period marginal income tax rate, whereas high-skilled individuals face a zero first period marginal income tax rate. In this respect, the qualitative features of the second-period tax schedule are the same as in a static economy. Separability of preferences across time implies that the marginal rate of substitution between first period before-tax income and first period after-tax income is independent of the second period allocation. Therefore, the existence of a future period has no effect on the type of labor supply distortions needed to induce type revelation. The magnitude of the marginal tax rate on the income of low-skilled individuals may, however, differ from the corresponding marginal tax rate in a one-period economy. Anticipated future events help to shape savings decisions, which directly affect first period consumption and an individual's marginal rate of substitution between consumption and labor supply in the first period.

Because the government has the same intertemporal preferences as the individuals, it has no desire to distort savings merely to transfer resources between periods. As we show in the proofs of Lemma 1 and Proposition 2, due to a wealth effect, an increase in anyone's savings induces the government to reduce the second-period labor supply of the low skilled. By single-crossing, this reduction is more valuable to the truly low skilled than it is to their potential mimickers. Through this channel, a marginal increase in anyone's savings relaxes the incentive-compatibility constraint. In addition, savings have an impact on tax payments, which also affect this constraint. As shown in Lemma 1, an increase in low-skilled savings reduces the taxes paid by low-skilled individuals by less than it reduces the taxes paid by the high skilled. This makes the prospect of mimicking in the second period less attractive to the high-skilled individuals and so reduces the two-period value of concealing their type in the first period. Thus, for the low skilled, the wealth and tax revenue effects of marginally increasing their savings reinforce each other. Hence, as shown in part (ii) of Proposition 2, the government's lack of commitment to a second period tax scheme results in the savings of the low skilled being subsidized. If, on the other hand, the savings of the high skilled are increased, as shown in Lemma 1, an increase the taxes paid by low-skilled individuals are reduced by *more* than the taxes paid by the high skilled. This tax revenue effect countervails the wealth effect, with the consequence that the sign of the optimal savings distortion for the high skilled is ambiguous.

In summary, it is the extra benefit of relaxing the incentive constraint (compared to the full information solution) that accounts for the upward distortion of savings for the low skilled. We thus have another instance of the observation made by Boadway (2012) that distortionary policy instruments that would not be used in the absence of asymmetric information are valuable when there is private information if these instruments can relax an incentive compatibility constraint.

7. The Second Period No-Commitment Tax Design Problem with First Period Pooling or Semi-Pooling

The government may not be able to infer the identities of all individuals after the first period. Some, or potentially all, high-skilled individuals might not reveal their type. Suppose that some positive proportion π of the high skilled mimic the low skilled in the first period. The government is then faced with a second-best problem in period 2. There is semi-pooling if $0 < \pi < 1$ and pooling if $\pi = 1$. In the semi-pooling case, there are three types of individuals. In addition to the low-skilled individuals, there are high-skilled individuals who have not revealed themselves to be high skilled in the first period, denoted by the second-period type $2p$, and there are high-skilled individuals who have revealed themselves to be high skilled in the first period, denoted by the second-period type $2s$.[14]

In this section, we consider the second period tax design problem without commitment when there is pooling or semi-pooling in the first period. To simply the exposition, we consider the pooling and semi-pooling cases simultaneously, that is, by simply assuming that $\pi > 0$. The analysis for the pooling case is obtained by setting $\pi = 1$ and omitting the variables pertaining to type $2s$.

The materials balance constraint in the second period is now

$$\alpha x_1^2 + (1-\alpha)\left[\pi x_{2p}^2 + (1-\pi)x_{2s}^2\right] \leq \alpha y_1^2 + (1-\alpha)\left[\pi y_{2p}^2 + (1-\pi)y_{2s}^2\right] + (1+r)s_G. \quad (7.1)$$

Because savings is observable, any individual of type 2 can successfully conceal his type in the first period only if his before-tax income and savings are identical to those of an individual of type 1. Social welfare in the second period, which is affected by individual savings, is given by

$$\mathcal{W}^{2,pool}(x_1^2, x_{2s}^2, x_{2p}^2, y_1^2, y_{2s}^2, y_{2p}^2, s_1, s_2) = (1-\alpha)(1-\pi)\left[v(x_{2s}^2 + (1+r)s_2) - h\left(\frac{y_{2s}^2}{w_2}\right)\right] + (1-\alpha)\pi\left[v(x_{2p}^2 + (1+r)s_1) - h\left(\frac{y_{2p}^2}{w_2}\right)\right] + \alpha\left[v(x_1^2 + (1+r)s_1) - h\left(\frac{y_1^2}{w_1}\right)\right]. \quad (7.2)$$

Because the government enters the second period without being able to distinguish between individuals of types 1 and $2p$, but knowing that everybody else is of type 2 when $\pi \neq 1$, its tax design problem is constrained by the incentive compatibility requirement

$$v(x_{2p}^2 + (1+r)s_1) - h\left(\frac{y_{2p}^2}{w_2}\right) \geq v(x_1^2 + (1+r)s_1) - h\left(\frac{y_1^2}{w_2}\right). \quad (7.3)$$

The problem faced by the government in the second period is:

Second Period Tax Design Problem with Pooling or Semi-Pooling. Given $\pi > 0$ and (s_1, s_2, s_G), the government chooses a second period allocation $(x_1^2, x_{2p}^2, x_{2s}^2, y_1^2, y_{2p}^2, y_{2s}^2)$ to maximize the objective function (7.2) subject to the second period materials balance constraint (7.1) and the incentive compatibility constraint (7.3). There is semi-pooling if $0 < \pi < 1$ and pooling if $\pi = 1$.

Let $(x_1^2, x_{2p}^2, x_{2s}^2, y_1^2, y_{2p}^2, y_{2s}^2)$ denote a solution to this second period tax design problem and let $\mathcal{W}^{2,pool}(s_1, s_2, s_G, \pi)$ denote its value function. The function describing the utility level attained by an individual of type i in the solution to this problem in terms of the parameters of the problem is denoted by $\mathcal{V}^i(s_1, s_2, s_G, \pi)$, for $i = 1, 2p, 2s$.

[14]In order to avoid confusion, in the subsequent discussion, we refer to individuals as being of type 1, 2, $2p$, or $2s$, rather than as being either low or high skilled.

The second period tax design problem with semi-pooling is a hybrid of a first-best allocation problem and a standard optimal nonlinear income taxation problem. Given the utilitarian nature of the objective function, the problem is strictly redistributive in the sense of Guesnerie (1995, p. 224). Hence, the optimal second period allocation features the usual distortion for individuals of type 1, namely a positive marginal income tax rate. Type $2p$ individuals are analogous to the the highest type individuals in a standard optimal nonlinear income tax problem — potential mimickers but not potentially mimicked. Hence, the standard arguments of optimal nonlinear income tax theory can be used to show that these individuals face a zero marginal income tax rate. Because type $2s$ individuals have already revealed their tax-relevant characteristics, there is no need to distort their second period labor-leisure decisions, and their marginal income tax rate is also zero. This pattern of distortions is summarized in Proposition 3.

Proposition 3. *At a solution to the second period no-commitment tax design problem with pooling or semi-pooling,* $\text{MRS}_{y_1^2,x_1^2} < 1$*,* $\text{MRS}_{y_{2p}^2,x_{2p}^2} = 1$*, and* $\text{MRS}_{y_{2s}^2,x_{2s}^2} = 1$*.*

In addition to the within-period marginal distortions arising at a solution to the second period tax design problem with pooling or semi-pooling, the levels of consumption, before-tax income, and utility in the second period allocated to the three types of individuals are important in shaping decisions in the first period. If, for example, the government uses the information it has about individuals of type $2s$ when there is semi-pooling in order to increase their tax burden, then they must be compensated in the first period for revealing their type. Proposition 4 gives the relative magnitudes of consumption and before-tax income for individuals of the three second period types, as well as the utility of the type $2p$ individuals relative to the utilities of the other types.

Proposition 4. *At a solution to the second period no-commitment tax design problem with pooling or semi-pooling,*

(i) $x_1^2 + (1+r)s_1 < x_{2s}^2 + (1+r)s_2 < x_{2p}^2 + (1+r)s_1$.

(ii) $y_1^2 < y_{2p}^2 < y_{2s}^2$.

(iii) $\mathcal{V}^{2p}(s_1, s_2, s_G, \pi) > \mathcal{V}^1(s_1, s_2, s_G, \pi)$ *and* $\mathcal{V}^{2p}(s_1, s_2, s_G, \pi) > \mathcal{V}^{2s}(s_1, s_2, s_G, \pi)$.

Part (i) of Proposition 4 orders the consumptions of the three types of individuals, while part (ii) orders their incomes. As is the case in the standard optimal nonlinear income tax model, type 1 individuals have the lowest consumption and the lowest before-tax income. Individuals of type $2s$ have both a lower consumption and a higher before-tax income than do individuals of type $2p$. Thus, when there is semi-pooling, it is optimal for the government to use the information it gains in the first period about the type $2s$ individuals to increase their tax burdens. Because it is optimal in period 2 for the incentive compatibility constraint to bind, as in an atemporal Mirrlees model, the type $2p$ individuals must be better off in period 2 than those of type 1, as indicated by the first inequality in part (iii). Because the government taxes type $2s$ individuals more heavily than it does type $2p$ individuals, revealing one's type in the first period is costly in terms of second period utility, as indicated by the second inequality in part (iii).

8. The First Period No-Commitment Tax Design Problem with Pooling or Semi-Pooling

The government foresees the impact of second period decisions when choosing the optimal first period allocation with pooling or semi-pooling. Thus, given that a proportion $\pi > 0$ of the type 2 individuals pool in the first period, its objective function is

$$\begin{aligned}\mathcal{W}^{pool}(x_1^1, x_2^1, y_1^1, y_2^1, s_1, s_2, s_G) \\ &= \alpha \left[u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_1}\right)\right] + (1-\alpha)\pi \left[u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_2}\right)\right] \\ &\quad + (1-\alpha)(1-\pi)\left[u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right)\right] + \mathcal{W}^{2,pool}(s_1, s_2, s_G, \pi).\end{aligned} \quad (8.1)$$

The objective function of the first period no-commitment tax design problem with pooling or semi-pooling depends on first period savings in two ways. There is a dependence due to the direct effects of private savings on consumption in each period. There are also indirect effects that depend on how the components of the optimal second period allocation depend on public and private savings. While the exact comparative static responses of the optimal second period allocations to savings are difficult to determine, as we shall see, it is nevertheless possible to sign most of the marginal distortions in the first period.

Because the objective function (8.1) incorporates the second period decisions of the government, it takes account of the second period materials balance constraint (7.1) and the incentive compatibility condition (7.3). However, the first period tax design problem is constrained by a materials balance constraint for period 1, namely

$$[\alpha + (1-\alpha)\pi]x_1^1 + (1-\alpha)(1-\pi)x_2^1 + s_G \leq [\alpha + (1-\alpha)\pi]y_1^1 + (1-\alpha)(1-\pi)y_2^1. \quad (8.2)$$

If there is pooling, the variables x_2^1, y_2^1, and s_2 are omitted from (8.1) and (8.2) because there are no type $2s$ individuals.

If there is semi-pooling (i.e., when $0 < \pi < 1$), individuals of type 2 must be indifferent between revealing their type in the first period and hiding their private information by mimicking type 1 individuals in period 1. Individuals of type $2s$ rationally anticipate that revealing their type is costly to them in the next period, and so they must be compensated in period 1. Thus, the bundles offered to the type 2 individuals must satisfy

$$u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right) + \mathcal{V}^{2s}(s_1, s_2, s_G, \pi) = u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_2}\right) + \mathcal{V}^{2p}(s_1, s_2, s_G, \pi). \quad (8.3)$$

We consider the pooling and semi-pooling cases separately.

8.1. Pooling

When there is pooling, the government's first period decision problem is described as follows:

First Period No-Commitment Tax Design Problem with Pooling. The government chooses a first period allocation (x_1^1, y_1^1, s_1, s_G) to maximize the objective function (8.1) subject to the first period materials balance constraint (8.2).

Proposition 5 describes the pattern of distortions to first period labor supply and savings when there is pooling.

Proposition 5. *At a solution to the first period no-commitment tax design problem with pooling:*

(i) $\text{MRS}_{y_1^1,x_1^1} > 1 > \text{MRS}_{y_{2p}^1,x_{2p}^1}$.

(ii) $\text{MRS}_{x_1^1,x_1^2} > -(1+r) > \text{MRS}_{x_{2p}^1,x_{2p}^2}$.[15]

(iii) $\alpha\text{MRS}_{y_1^1,x_1^1} + (1-\alpha)\text{MRS}_{y_{2p}^1,x_{2p}^1} = 1$.

(iv) $\alpha\text{MRS}_{x_1^1,x_1^2} + (1-\alpha)\text{MRS}_{x_{2p}^1,x_{2p}^2} = -(1+r)$.

With pooling and no commitment, individuals of type $2p$ face a higher implicit marginal income tax rate than do type 1 individuals in period 1, which is the reverse of what occurs in the standard nonlinear income tax problem and in the full commitment case. Interestingly, individuals of type 1 face a negative marginal income tax rate and individuals of type $2p$ face a positive marginal income tax rate.[16] Nevertheless, as part (iii) shows, on average, there is no marginal distortion to first period labor supply decisions. Turning now to savings, individuals of type 1 are subsidized on the margin, whereas individuals of type $2p$ are taxed. However, on average, there is no marginal distortion to the savings decisions.

Because the utility function is additively separable in labor supply and consumption, equal consumption in the first period implies equal marginal utility of consumption in that period. Equal incomes in period 1 imply that individuals of type $2p$ have a smaller marginal disutility of labor in period 1 than do persons of type 1. The monotonicity of second period consumption in type implies that individuals of type $2p$ have a lower marginal utility of consumption in the second period than do individuals of type 1. The requirements that $\text{MRS}_{y_1^1,x_1^1} > \text{MRS}_{y_{2p}^1,x_{2p}^1}$ in part (i) of Proposition 5 and $\text{MRS}_{x_1^1,x_1^2} > \text{MRS}_{x_{2p}^1,x_{2p}^2}$

[15] Recall that marginal rates of substitution between consumption in the two periods are negative, so in absolute value, individuals of type 1 have a smaller intertemporal marginal rate of substitution than individuals of type $2p$.

[16] Dillén and Lundholm (1996) have found in their model of dynamic linear income taxation without commitment that pooling with a negative first period marginal income tax rate may be optimal.

in part (ii) then follow from the definitions of these marginal rates of substitution in (3.4) and (3.6). Because there is no incentive constraint in the first period no-commitment tax design problem with pooling, on average, there is no need to distort first period labor supply or savings decisions on the margin. Thus, it is optimal to marginally subsidize the labor supply and savings decisions of type 1 individuals and to marginally tax these decisions for type $2p$ individuals.

Further intuition for the signs of these distortions may be obtained by considering the implications of perturbing some of the decision variables. At an optimal allocation with first-period pooling, it is possible to infinitesimally decrease the common first period consumption of individuals of types 1 and $2p$ and before-tax incomes of types 1 and $2p$ by the same amount holding savings fixed without violating the materials balance constraints in either period. Because savings are held constant, this change has no effect on the second period incentive compatibility constraint (7.3). If $\mathrm{MRS}_{y^1_{2p},x^1_{2p}} \geq 1$ and, hence, $\mathrm{MRS}_{y^1_1,x^1_1} > 1$, this change is a Pareto improvement. Hence, it must be optimal to have $\mathrm{MRS}_{y^1_{2p},x^1_{2p}} < 1$.[17] By reversing the direction of change in first period consumption and before-tax income, it follows that it is also optimal to have $\mathrm{MRS}_{y^1_1,x^1_1} > 1$.

As we have seen, at an optimal allocation with first-period pooling, it is necessary for individuals of type 1 to have the smaller intertemporal marginal rate of substitution in absolute value. Suppose that $\mathrm{MRS}_{x^1_1,x^2_1} \leq -(1+r)$ and, hence, $\mathrm{MRS}_{x^1_{2p},x^2_{2p}} < -(1+r)$. Consider modifying the optimal allocation by having each individual of type 1 or $2p$ transfer a common infinitesimally small amount from savings into first period consumption and then decreasing second period consumption by $-\mathrm{MRS}_{x^1_1,x^2_1}$ for individuals of type 1 and by $-\mathrm{MRS}_{x^1_{2p},x^2_{2p}}$ for individuals of type $2p$. This composite change has no effect on the variables that appear in the first period materials balance constraint (8.2), but it relaxes the second-period materials balance constraint (7.1); that is, it is resource saving. Furthermore, this reallocation is a matter of indifference for each individual. From the definition of the intertemporal marginal rate of substitution (3.6), we see that second period utility has decreased by $u'(c^1_{2p})$ for those of type $2p$ and by $u'(c^1_1)$ for those of type 1. Because $c^1_{2p} = c^1_1$ when there is pooling in period 1, $u'(c^1_{2p}) = u'(c^1_1)$. Hence, this composite change does not violate the incentive compatibility constraint (7.3). The resource savings can now be used to increase everybody's second period consumption without violating any of the constraints of the tax design problem with pooling, contradicting the optimality of the initial allocation. Thus, it must be optimal to have $\mathrm{MRS}_{x^1_1,x^2_1} > -(1+r)$. A similar argument can be used to show that is also optimal to have $\mathrm{MRS}_{x^1_{2p},x^2_{2p}} < -(1+r)$, for otherwise it would be possible to obtain a Pareto improvement by transferring consumption from the first period to the second.

[17]Except in the borderline case in which $\mathrm{MRS}_{y^1_2,x^1_2} = 1$, both types benefit from this change. In the borderline case, this change is a matter of indifference to individuals of type $2p$, but strictly benefits individuals of type 1.

8.2. Semi-Pooling

When there is semi-pooling, the government's first period decision problem is described as follows:

First Period No-Commitment Tax Design Problem with Semi-Pooling. The government chooses a first period allocation $(x_1^1, x_2^1, y_1^1, y_2^1, s_1, s_2, s_G)$ to maximize the objective function (8.1) subject to the first period materials balance constraint (8.2) and the indifference condition (8.3).

The pattern of the first period distortions are summarized in Proposition 6.

Proposition 6. *At a solution to the first period no-commitment tax design problem with semi-pooling:*

(i) $\mathrm{MRS}_{y_1^1,x_1^1} > \mathrm{MRS}_{y_{2p}^1,x_{2p}^1}$, $1 > \mathrm{MRS}_{y_{2p}^1,x_{2p}^1}$, *and* $\mathrm{MRS}_{y_{2s}^1,x_{2s}^1} = 1$.

(ii) $\mathrm{MRS}_{x_1^1,x_1^2} > \mathrm{MRS}_{x_{2p}^1,x_{2p}^2}$ *and* $\mathrm{MRS}_{x_{2s}^1,x_{2s}^2} > \mathrm{MRS}_{x_{2p}^1,x_{2p}^2}$.

As in the pooling case, with semi-pooling, the type 1 individuals have their labor supply and savings decisions taxed less at the margin than the type $2p$ individuals. The intuition for these findings is exactly the same as when there is pooling. As in the standard optimal nonlinear income tax problem, there is no need to distort the labor supply decisions of the type $2s$ individuals because there is no higher type from which information rents can be extracted. Individuals of type $2s$ have a higher consumption in the first period and lower consumption in the second period than do individuals of type $2p$. The requirement that $\mathrm{MRS}_{x_{2s}^1,x_{2s}^2} > \mathrm{MRS}_{x_{2p}^1,x_{2p}^2}$ in part (ii) follows directly from these observations. Thus, the marginal tax rate on savings for individuals of type $2s$ is smaller than that for type $2p$. In combination with a positive first period marginal income tax rate for type $2p$ individuals, this is what allows the government to separate the type $2s$ individuals from the rest of the type 2 population. In general, it is not possible to sign any of the marginal distortions on savings.

The formulae in Proposition 7 help provide insight into why it is not possible to sign some of the tax distortions when there is semi-pooling.

Proposition 7. *At a solution to the first period no-commitment tax design problem with semi-pooling, there exist* $\gamma \in (0, \alpha + (1-\alpha)\pi)$ *and* $\phi \in (0, \alpha)$ *such that:*

(i) $\left[\frac{\alpha}{\alpha+(1-\alpha)\pi-\gamma}\right] \mathrm{MRS}_{y_1^1,x_1^1} + \left[\frac{(1-\alpha)\pi-\gamma}{\alpha+(1-\alpha)\pi-\gamma}\right] \mathrm{MRS}_{y_{2p}^1,x_{2p}^1} = 1.$

(ii) $\left[\frac{\alpha}{\alpha+(1-\alpha)\pi}\right] \mathrm{MRS}_{y_1^1,x_1^1} + \left[\frac{(1-\alpha)\pi}{\alpha+(1-\alpha)\pi}\right] \mathrm{MRS}_{y_{2p}^1,x_{2p}^1} < 1.$

(iii) $$\mathrm{MRS}_{x_1^1,x_1^2} = -(1+r)\left(\frac{\alpha-\phi}{\alpha}\right) - \frac{\gamma}{[\alpha+(1-\alpha)\pi]v'(c_1^2)}\left[u'(x_1^1 - s_1) + \frac{\partial \mathcal{V}^{2s}}{\partial s_1} - \frac{\partial \mathcal{V}^{2p}}{\partial s_1}\right].$$

$$\text{(iv)}\ \ \text{MRS}_{x^1_{2p},x^2_{2p}} = -(1+r)\left(\frac{(1-\alpha)\pi+\phi}{(1-\alpha)\pi}\right) - \frac{\gamma}{[\alpha+(1-\alpha)\pi]v'(c^2_{2p})}\left[u'(x^1_1-s_1)+\frac{\partial\mathcal{V}^{2s}}{\partial s_1}-\frac{\partial\mathcal{V}^{2p}}{\partial s_1}\right].$$

$$\text{(v)}\ \ \text{MRS}_{x^1_{2s},x^2_{2s}} = -(1+r)+\frac{\gamma}{(1-\alpha)(1-\pi)v'(c^2_{2s})}\left[u'(x^1_2-s_2)+\frac{\partial\mathcal{V}^{2p}}{\partial s_2}-\frac{\partial\mathcal{V}^{2s}}{\partial s_2}\right].$$

The parameter γ is the multiplier on the indifference condition (8.3) in the Lagrangian for the first period no-commitment tax design problem with semi-pooling. *A priori*, γ can be of either sign, but at a solution to the first period problem, it must be in the interval $(0, \alpha+(1-\alpha)\pi)$. The parameter ϕ is the multiplier on the incentive compatibility constraint (7.3) in the Lagrangian for the second period no-commitment tax design problem with semi-pooling (or pooling). At the solution to this problem, ϕ is in the interval $(0, \alpha)$.

If $\gamma < (1-\alpha)\pi$, then the weights on the marginal rates of substitution in part (i) of Proposition 7 are both positive and less than 1. When this inequality is satisfied, it follows from part (i) that $\text{MRS}_{y^1_1,x^1_1} > 1 > \text{MRS}_{y^1_{2p},x^1_{2p}}$, as in the case of pooling. That is, in period 1, the marginal income tax rate for type 1 individuals is negative, whereas it is positive for individuals of type $2p$. However, if $\gamma \geq (1-\alpha)\pi$, then the weight on $\text{MRS}_{y^1_1,x^1_1}$ is at least 1 and the weight on $\text{MRS}_{y^1_{2p},x^1_{2p}}$ is nonpositive. Consequently, it is possible to have $\text{MRS}_{y^1_1,x^1_1} \leq 1$ (i.e., a nonnegative marginal income tax rate for type 1 individuals) even though $\text{MRS}_{y^1_{2p},x^1_{2p}} < 1$. Hence, if the resource cost of satisfying the indifference condition (8.3) is sufficiently high, it may be desirable to tax the labor supply of the type 1 individuals at the margin.

Part (ii) of Proposition 7 shows that on average the individuals that are pooled in the first period face a positive marginal tax rate on their labor incomes. When considering changes in the consumptions and incomes in the first period, the second period terms in the indifference condition (8.3) play no role. Hence, this condition has the same implications as a binding incentive compatibility constraint in a static Mirrlees model. Thus, the argument establishing an optimal downward distortion of labor supply in the static nonlinear income tax model applies for the pooled individuals here.

Let Δ_1 denote the last term in square brackets on the right-hand sides of the equations in parts (iii) and (iv) of Proposition 7. Δ_1 is the change in the utility over both periods of a type $2p$ individual if the savings of those individuals who are pooled together is marginally decreased. To preserve the indifference condition (8.3) following such a change, the utility of type $2s$ individuals must also change by Δ_1. In general, the sign of Δ_1 is indeterminate. If $\Delta_1 = 0$, the terms relating to the indifference condition in parts (iii) and (iv) vanish. Hence, because $0 < \phi < \alpha$, $\text{MRS}_{x^1_1,x^2_1} > -(1+r) > \text{MRS}_{x^1_{2p},x^2_{2p}}$ and, therefore, on the margin, type 1 individuals have their savings subsidized and type $2p$ individuals have their savings taxed, as in the pooling case. For $i = 1, 2p$, the value of $\text{MRS}_{x^1_i,x^2_i}$ is decreasing (i.e., increasing in absolute value) in the value of Δ_1. Thus, if $\Delta_1 < 0$ (increasing his savings is detrimental to a type $2p$ individual), it is optimal to decrease the marginal tax rate on savings for type $2p$ individuals and to increase the marginal subsidy on savings for type 1 individuals compared to when $\Delta_1 = 0$. If this effect is sufficiently strong, type $2p$ individuals could also have their savings subsidized on the margin. Similar reasoning shows that if $\Delta_1 > 0$, then it is possible for both type 1 and type $2p$ individuals to face a positive marginal tax on savings. Analogous reasoning using part (v) shows why the marginal distortion on the savings of type $2s$ individuals may differ from 0 (as is the case when there is complete separation in period 1) and why the sign of this tax rate is indeterminate when there is semi-pooling.[18]

9. Conclusion

Our analysis suggests that a government's inability to commit to its future tax policy provides a rationale for distortions in savings behavior. Extending our analysis to more than two types of individuals is not straightforward. It is easy to construct models of static nonlinear income taxation that exhibit considerable pooling of types (see, for example, Weymark, 1986). Dynamic extensions of such models would invariably uncover cases of pooling, semi-pooling, and separation, each with its own distinct pattern of savings distortions. Our fundamental insight—that the inability to commit to future tax schedules necessitates the use of savings taxation when taxes are set optimally—is likely to carry over to economies with any number of types. This article can be seen as a step toward incorporating dynamic consistency constraints into normative approaches to tax theory alongside the familiar budget and incentive constraints.

Acknowledgements

This research has been generously supported by the Canada Research Chair Programme Grant No. 950-203857 and the Social Sciences and Humanities Research Council of Canada Grant No. 410-2008-2308. We have benefited from comments received from Gadi Barlevy, Marcus Berliant, Pierre Boyer, Francesco Cohen, Stéphane Gauthier, Jonathan Hamilton, Florian Scheuer, and Johannes Spinnewijn. This article was previously circulated as "Optimal Nonlinear Taxation of Income and Savings in a Two Class Economy."

[18]Let Δ_2 denote the last term in square brackets on the right-hand side of the equation in part (v) of Proposition 7. Note that $\mathrm{MRS}_{x^1_{2s},x^2_{2s}}$ is increasing, not decreasing, in Δ_2.

Appendix A. Proofs

Proof of Proposition 1. The Lagrangian associated with the second-best tax design problem with commitment is

$$\begin{aligned}
&\alpha\left[u(x_1^1-s_1)-g\left(\frac{y_1^1}{w_1}\right)+v(x_1^2+(1+r)s_1)-h\left(\frac{y_1^2}{w_1}\right)\right]\\
&+(1-\alpha)\left[u(x_2^1-s_2)-g\left(\frac{y_2^1}{w_2}\right)+v(x_2^2+(1+r)s_2)-h\left(\frac{y_2^2}{w_2}\right)\right]\\
&+\lambda^1\left[\alpha(y_1^1-x_1^1)+(1-\alpha)(y_2^1-x_2^1)-s_G\right]\\
&+\lambda^2\left[\alpha(y_1^2-x_1^2)+(1-\alpha)(y_2^2-x_2^2)+(1+r)s_G\right]\\
&+\mu\left[u(x_2^1-s_2)-g\left(\frac{y_2^1}{w_2}\right)+v(x_2^2+(1+r)s_2)-h\left(\frac{y_2^2}{w_2}\right)\right.\\
&\left.-u(x_1^1-s_1)+g\left(\frac{y_1^1}{w_2}\right)-v(x_1^2+(1+r)s_1)+h\left(\frac{y_1^2}{w_2}\right)\right].
\end{aligned} \tag{A.1}$$

The associated first-order conditions for an interior solution are:

$$x_1^1:\ \alpha u'(c_1^1)-\lambda^1\alpha-\mu u'(c_1^1)=0; \tag{A.2}$$
$$x_2^1:\ (1-\alpha)u'(c_2^1)-\lambda^1(1-\alpha)+\mu u'(c_2^1)=0; \tag{A.3}$$
$$y_1^1:\ -\frac{\alpha}{w_1}g'\left(\frac{y_1^1}{w_1}\right)+\lambda^1\alpha+\frac{\mu}{w_2}g'\left(\frac{y_1^1}{w_2}\right)=0; \tag{A.4}$$
$$y_2^1:\ -\frac{(1-\alpha)}{w_2}g'\left(\frac{y_2^1}{w_2}\right)+\lambda^1(1-\alpha)-\frac{\mu}{w_2}g'\left(\frac{y_2^1}{w_2}\right)=0; \tag{A.5}$$
$$x_1^2:\ \alpha v'(c_1^2)-\lambda^2\alpha-\mu v'(c_1^2)=0; \tag{A.6}$$
$$x_2^2:\ (1-\alpha)v'(c_2^2)-\lambda^2(1-\alpha)+\mu v'(c_2^2)=0; \tag{A.7}$$
$$y_1^2:\ -\frac{\alpha}{w_1}h'\left(\frac{y_1^2}{w_1}\right)+\lambda^2\alpha+\frac{\mu}{w_2}h'\left(\frac{y_1^2}{w_2}\right)=0; \tag{A.8}$$
$$y_2^2:\ -\frac{(1-\alpha)}{w_2}h'\left(\frac{y_2^2}{w_2}\right)+\lambda^2(1-\alpha)-\frac{\mu}{w_2}h'\left(\frac{y_2^2}{w_2}\right)=0; \tag{A.9}$$
$$s_1:\ \alpha\left[-u'(c_1^1)+(1+r)v'(c_1^2)\right]+\mu u'(c_1^1)-(1+r)\mu v'(c_1^2)=0; \tag{A.10}$$
$$s_2:\ (1-\alpha)\left[-u'(c_2^1)+(1+r)v'(c_2^2)\right]-\mu u'(c_2^1)+(1+r)\mu v'(c_2^2)=0; \tag{A.11}$$
$$s_G:\ -\lambda^1+(1+r)\lambda^2=0. \tag{A.12}$$

The first equality of part (i) follows from solving each of (A.3) and (A.5) for $\lambda^1(1-\alpha)$ and rearranging the resulting equality. Similar algebra applied to (A.7) and (A.9) yields the second equality. From (A.11),

$$(1-\alpha+\mu)u'(c_2^1)=(1-\alpha+\mu)(1+r)v'(c_2^2), \tag{A.13}$$

from which the final equality of part (i) follows.

By (A.2) and (A.4),

$$(\alpha-\mu)u'(c_1^1)=\frac{\alpha}{w_1}g'\left(\frac{y_1^1}{w_1}\right)-\frac{\mu}{w_2}g'\left(\frac{y_1^1}{w_2}\right)=\lambda^1\alpha. \tag{A.14}$$

Because $w_1<w_2$ and $g(\cdot)$ is strictly convex,

$$\frac{\alpha}{w_1}g'\left(\frac{y_1^1}{w_1}\right)-\frac{\mu}{w_2}g'\left(\frac{y_1^1}{w_2}\right)>\frac{(\alpha-\mu)}{w_1}g'\left(\frac{y_1^1}{w_1}\right). \tag{A.15}$$

Combining (A.14) and (A.15) yields

$$(\alpha-\mu)u'(c_1^1)>\frac{(\alpha-\mu)}{w_1}g'\left(\frac{y_1^1}{w_1}\right). \tag{A.16}$$

Because the multiplier on the resource constraint, λ^1, is positive, (A.14) implies that $(\alpha-\mu)$ is positive. Dividing both sides of (A.16) by $(\alpha-\mu)u'(c_1^1)$ and rearranging yields the first inequality of part (ii). The second inequality follows from a similar argument applied to (A.6) and (A.8). From (A.10),

$$(\alpha-\mu)u'(c_1^1)=(\alpha-\mu)(1+r)v'(c_1^2), \tag{A.17}$$

from which the final equality of part (ii) follows. □

Proof of Lemma 1. The objective function of the second period first-best problem is strictly concave and the constraint set is convex. Hence, by Sundaram (1996, Theorem 7.14), the problem has a unique solution. The associated Lagrangian is

$$\begin{aligned}\alpha\left[v(x_1^2+(1+r)s_1)\ -h\left(\frac{y_1^2}{w_1}\right)\right]+(1-\alpha)\left[v(x_2^2+(1+r)s_2)-h\left(\frac{y_2^2}{w_2}\right)\right]\\ +\lambda\left[\alpha(y_1^2-x_1^2)+(1-\alpha)(y_2^2-x_2^2)+(1+r)s_G\right].\end{aligned} \tag{A.18}$$

The first-order conditions for an optimum are:

$$x_1^2:\ \alpha\left[v'(c_1^2)-\lambda\right]=0; \tag{A.19}$$

$$x_2^2:\ (1-\alpha)\left[v'(c_2^2)-\lambda\right]=0; \tag{A.20}$$

$$y_1^2:\ \alpha\left[-\frac{1}{w_1}h'\left(\frac{y_1^2}{w_1}\right)+\lambda\right]=0; \tag{A.21}$$

$$y_2^2:\ (1-\alpha)\left[-\frac{1}{w_2}h'\left(\frac{y_2^2}{w_2}\right)+\lambda\right]=0; \tag{A.22}$$

$$\lambda:\ \alpha(y_1^2-x_1^2)+(1-\alpha)(y_2^2-x_2^2)+(1+r)s_G=0. \tag{A.23}$$

Part (i) of the lemma follows from solving each of (A.19)–(A.22) for λ.

The bordered Hessian matrix for this problem is

$$A = \begin{bmatrix} \alpha v''(c_1^2) & 0 & 0 & 0 & -\alpha \\ 0 & (1-\alpha)v''(c_2^2) & 0 & 0 & -(1-\alpha) \\ 0 & 0 & -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ -\alpha & -(1-\alpha) & \alpha & (1-\alpha) & 0 \end{bmatrix}. \tag{A.24}$$

Its determinant is

$$\begin{aligned} |A| = \alpha^2(1-\alpha)^2 v''(c_1^2)v''(c_2^2)\left[(1-\alpha)\frac{h''(l_1^2)}{(w_1)^2} + \alpha\frac{h''(l_2^2)}{(w_2)^2}\right] \\ - \alpha^2(1-\alpha)^2\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2}\left[(1-\alpha)v''(c_1^2) + \alpha v''(c_2^2)\right]. \end{aligned} \tag{A.25}$$

Because $0 < \alpha < 1$, the strict convexity of $h(\cdot)$ implies that the term inside the first square bracket in (A.25) is positive. It then follows from the strict concavity of $v(\cdot)$ that the first term on the right-hand side of (A.25) is positive. On the other hand, the strict concavity of $v(\cdot)$ implies that the sum inside the second square bracket is negative. Because $h(\cdot)$ is strictly convex, this sum is multiplied by a positive number. Thus, (A.25) expresses $|A|$ as a positive quantity minus a negative quantity. Hence, $|A| > 0$ and A is invertible. It then follows from the Implicit Function Theorem (see Sundaram, 1996, Theorem 1.77) that the solution functions are continuously differentiable.

Part (ii) of the lemma follows directly from differentiating the materials balance constraint, which is also the first-order condition (A.23), with respect to s_1 and s_2.

We now prove part (iii). Implicitly differentiating (A.19)–(A.23) with respect to the endogenous variables $(x_1^2, x_2^2, y_1^2, y_2^2, \lambda)$ and the parameter s_1, we obtain

$$A\begin{bmatrix} dx_1^2 \\ dx_2^2 \\ dy_1^2 \\ dy_2^2 \\ d\lambda \end{bmatrix} = \begin{bmatrix} -\alpha(1+r)v''(c_1^2) \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} ds_1. \tag{A.26}$$

Because A is invertible, we can now use Cramer's Rule to compute the derivatives of the allocation functions with respect to s_1. For x_1^2, we have

$$|A|\frac{\partial x_1^2}{\partial s_1} = \begin{vmatrix} -\alpha(1+r)v''(c_1^2) & 0 & 0 & 0 & -\alpha \\ 0 & (1-\alpha)v''(c_2^2) & 0 & 0 & -(1-\alpha) \\ 0 & 0 & -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ 0 & -(1-\alpha) & \alpha & (1-\alpha) & 0 \end{vmatrix}. \tag{A.27}$$

Computing the determinant by expanding along the first column,

$$|A|\frac{\partial x_1^2}{\partial s_1} = -\alpha(1+r)v''(c_1^2)\left\{(1-\alpha)v''(c_2^2)\begin{vmatrix} -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ \alpha & (1-\alpha) & 0 \end{vmatrix}\right. \\ \left. + (1-\alpha)\begin{vmatrix} 0 & 0 & -(1-\alpha) \\ -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \end{vmatrix}\right\} \tag{A.28}$$

or, equivalently,

$$|A|\frac{\partial x_1^2}{\partial s_1} = -\alpha^2(1-\alpha)^2(1+r)v''(c_1^2) \\ \times\left[(1-\alpha)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2} + \alpha v''(c_2^2)\frac{h''(l_2^2)}{(w_2)^2} - (1-\alpha)\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2}\right]. \tag{A.29}$$

For x_2^2, we have

$$|A|\frac{\partial x_2^2}{\partial s_1} = \begin{vmatrix} \alpha v''(c_1^2) & -\alpha(1+r)v''(c_1^2) & 0 & 0 & -\alpha \\ 0 & 0 & 0 & 0 & -(1-\alpha) \\ 0 & 0 & -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ -\alpha & 0 & \alpha & (1-\alpha) & 0 \end{vmatrix}. \tag{A.30}$$

Computing the determinant by expanding along the second column,

$$|A|\frac{\partial x_2^2}{\partial s_1} = \alpha(1+r)v''(c_1^2)\alpha\begin{vmatrix} 0 & 0 & -(1-\alpha) \\ -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \end{vmatrix} \tag{A.31}$$

or, equivalently,

$$= -\alpha^3(1-\alpha)^2(1+r)v''(c_1^2)\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2}. \tag{A.32}$$

For y_1^2, we have

$$|A|\frac{\partial y_1^2}{\partial s_1} = \begin{vmatrix} \alpha v''(c_1^2) & 0 & -\alpha(1+r)v''(c_1^2) & 0 & -\alpha \\ 0 & (1-\alpha)v''(c_2^2) & 0 & 0 & -(1-\alpha) \\ 0 & 0 & 0 & 0 & \alpha \\ 0 & 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ -\alpha & -(1-\alpha) & 0 & (1-\alpha) & 0 \end{vmatrix}. \tag{A.33}$$

Computing the determinant by expanding along the third column,

$$|A|\frac{\partial y_1^2}{\partial s_1} = -\alpha(1+r)v''(c_1^2)\alpha \begin{vmatrix} 0 & (1-\alpha)v''(c_2^2) & 0 \\ 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} \\ -\alpha & -(1-\alpha) & (1-\alpha) \end{vmatrix} \tag{A.34}$$

or, equivalently,

$$= -\alpha^3(1-\alpha)^2(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_2^2)}{(w_2)^2}. \tag{A.35}$$

Finally, for y_2^2, we have

$$|A|\frac{\partial y_2^2}{\partial s_1} = \begin{vmatrix} \alpha v''(c_1^2) & 0 & 0 & -\alpha(1+r)v''(c_1^2) & -\alpha \\ 0 & (1-\alpha)v''(c_2^2) & 0 & 0 & -(1-\alpha) \\ 0 & 0 & -\frac{\alpha h''(l_1^2)}{(w_2)^1} & 0 & \alpha \\ 0 & 0 & 0 & 0 & (1-\alpha) \\ -\alpha & -(1-\alpha) & \alpha & 0 & 0 \end{vmatrix}. \tag{A.36}$$

Computing the determinant by expanding along the fourth column,

$$|A|\frac{\partial y_2^2}{\partial s_1} = \alpha(1+r)v''(c_1^2)(1-\alpha) \begin{vmatrix} 0 & (1-\alpha)v''(c_2^2) & 0 \\ 0 & 0 & -\frac{\alpha)h''(l_1^2)}{(w_1)^2} \\ -\alpha & -(1-\alpha) & \alpha \end{vmatrix} \tag{A.37}$$

or, equivalently,

$$|A|\frac{\partial y_2^2}{\partial s_1} = -\alpha^3(1-\alpha)^2(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2}. \tag{A.38}$$

Combining (A.29)–(A.38) yields

$$\begin{aligned} |A|\left[\frac{\partial x_1^2}{\partial s_1} - \frac{\partial x_2^2}{\partial s_1} - \frac{\partial y_1^2}{\partial s_1} + \frac{\partial y_2^2}{\partial s_1}\right] &= -\alpha^2(1-\alpha)^3(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2} \\ - \alpha^3(1-\alpha)^2(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_2^2)}{(w_2)^2} &+ \alpha^2(1-\alpha)^3(1+r)v''(c_1^2)\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2} \\ + \alpha^3(1-\alpha)^2(1+r)v''(c_1^2)\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2} &+ \alpha^3(1-\alpha)^2(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_2^2)}{(w_2)^2} \\ - \alpha^3(1-\alpha)^2(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2}. \end{aligned} \tag{A.39}$$

The second and fifth terms on the right-hand side of (A.39) cancel. Factoring common elements, grouping the first term with the sixth, and the third with the fourth yields

$$\frac{\partial x_1^2}{\partial s_1} - \frac{\partial x_2^2}{\partial s_1} - \frac{\partial y_1^2}{\partial s_1} + \frac{\partial y_2^2}{\partial s_1} = \frac{\alpha^2(1-\alpha)^2(1+r)v''(c_1^2)}{|A|}\left[\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2} - v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2}\right]. \tag{A.40}$$

We have already established that $|A|$ is positive. Thus, curvature properties of $v(\cdot)$ and $h(\cdot)$ imply that ther right-hand side of (A.40) is negative. Therefore, the left-hand side of (A.40) is also negative, which establishes part (iii) of the lemma.

Turning now to part (iv) for $i = 2$, in Appendix B, we show that the analog to (A.40) is

$$\frac{\partial x_1^2}{\partial s_1} - \frac{\partial x_2^2}{\partial s_1} - \frac{\partial y_1^2}{\partial s_1} + \frac{\partial y_2^2}{\partial s_1} = -\frac{\alpha^2(1-\alpha)^2(1+r)v''(c_2^2)}{|A|}\left[\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2} - v''(c_1^2)\frac{h''(l_2^2)}{(w_2)^2}\right]. \tag{A.41}$$

This expression is positive, from which part (iv) of the lemma follows. □

Proof of Proposition 2. The Lagrangian associated with the first-period no-commitment tax design problem with separation is

$$\begin{aligned}
&\alpha\left[u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_1}\right) + v(x_1^2(\mathbf{s}) + (1+r)s_1) - h\left(\frac{y_1^2(\mathbf{s})}{w_1}\right)\right] \\
&\quad + (1-\alpha)\left[u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right) + v(x_2^2(\mathbf{s}) + (1+r)s_2) - h\left(\frac{y_2^2(\mathbf{s})}{w_2}\right)\right] \\
&\quad + \eta\big[\alpha(y_1^1 - x_1^1) + (1-\alpha)(y_2^1 - x_2^1) - s_G\big] \\
&\quad + \psi\left[u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right) + v(x_2^2(\mathbf{s}) + (1+r)s_2) - h\left(\frac{y_2^2(\mathbf{s})}{w_2}\right)\right. \\
&\qquad \left. - u(x_1^1 - s_1) + g\left(\frac{y_1^1}{w_2}\right) - v(x_1^2(\mathbf{s}) + (1+r)s_1) + h\left(\frac{y_1^2(\mathbf{s})}{w_2}\right)\right].
\end{aligned} \tag{A.42}$$

The associated first-order equations include:

$$x_1^1:\ \alpha u'(c_1^1) - \eta\alpha - \psi u'(c_1^1) = 0; \tag{A.43}$$

$$x_2^1:\ (1-\alpha)u'(c_2^1) - \eta(1-\alpha) + \psi u'(c_2^1) = 0; \tag{A.44}$$

$$y_1^1:\ -\frac{\alpha}{w_1}g'\left(\frac{y_1^1}{w_1}\right) + \eta\alpha + \frac{\psi}{w_2}g'\left(\frac{y_1^1}{w_2}\right) = 0; \tag{A.45}$$

$$y_2^1:\ -\frac{(1-\alpha)}{w_2}g'\left(\frac{y_2^1}{w_2}\right) + \eta(1-\alpha) - \frac{\psi}{w_2}g'\left(\frac{y_1^2}{w_2}\right) = 0; \tag{A.46}$$

$$\begin{aligned}
s_1:\ & -\alpha u'(c_1^1) + \alpha v'(c_1^2)\left[\frac{\partial x_1^2}{\partial s_1} + (1+r)\right] - \frac{\alpha}{w_1}h'\left(\frac{y_1^2}{w_1}\right)\frac{\partial y_1^2}{\partial s_1} \\
&\quad + (1-\alpha)v'(c_2^2)\frac{\partial x_2^2}{\partial s_1} - \frac{(1-\alpha)}{w_2}h'\left(\frac{y_2^2}{w_2}\right)\frac{\partial y_2^2}{\partial s_1} \\
&\quad + \psi\left[u'(c_1^1) + v'(c_2^2)\frac{\partial x_2^2}{\partial s_1} - \frac{1}{w_2}h'\left(\frac{y_2^2}{w_2}\right)\frac{\partial y_2^2}{\partial s_1}\right. \\
&\qquad \left. - v'(c_1^2)\left[\frac{\partial x_1^2}{\partial s_1} + (1+r)\right] + \frac{1}{w_2}h'\left(\frac{y_1^2}{w_2}\right)\frac{\partial y_1^2}{\partial s_1}\right] = 0;
\end{aligned} \tag{A.47}$$

$$
\begin{aligned}
s_2\colon\ & \alpha v'(c_1^2)\frac{\partial x_1^2}{\partial s_2} - \frac{\alpha}{w_1}h'\left(\frac{y_1^2}{w_1}\right)\frac{\partial y_1^2}{\partial s_2} - (1-\alpha)u'(c_2^1) \\
& + (1-\alpha)v'(c_2^2)\left[\frac{\partial x_2^2}{\partial s_2} + (1+r)\right] - \frac{(1-\alpha)}{w_2}h'\left(\frac{y_2^2}{w_2}\right)\frac{\partial y_2^2}{\partial s_2} \\
& + \psi\left[-u'(c_2^1) + v'(c_2^2)\left[\frac{\partial x_2^2}{\partial s_2} + (1+r)\right] - \frac{1}{w_2}h'\left(\frac{y_2^2}{w_2}\right)\frac{\partial y_2^2}{\partial s_2}\right. \\
& \left. - v'(c_1^2)\frac{\partial x_1^2}{\partial s_2} + \frac{1}{w_2}h'\left(\frac{y_1^2}{w_2}\right)\frac{\partial y_1^2}{\partial s_2}\right] = 0.
\end{aligned}
\tag{A.48}
$$

Equations (A.43)–(A.46) are identical to equations (A.2)–(A.5), except that λ^1 is replaced by η and μ is replaced by ψ. Thus, the arguments used in the proof of Proposition 1 may be repeated to prove part (i) of the proposition.

By part (i) of Lemma 1, (A.47) is equivalent to

$$
\begin{aligned}
-\alpha u'(c_1^1) + \alpha(1+r)v'(c_1^2) + v'(c_1^2)\left[\alpha\frac{\partial x_1^2}{\partial s_1} + (1-\alpha)\frac{\partial x_2^2}{\partial s_1} - \alpha\frac{\partial y_1^2}{\partial s_1} - (1-\alpha)\frac{\partial y_2^2}{\partial s_1}\right] \\
+\psi u'(c_1^1) + \psi v'(c_1^2)\left[-\frac{\partial x_1^2}{\partial s_1} + \frac{\partial x_2^2}{\partial s_1} + \frac{\partial y_1^2}{\partial s_1} - \frac{\partial y_2^2}{\partial s_1}\right] - \psi v'(c_1^2)(1+r) \\
+ \psi\left[\frac{1}{w_2}h'\left(\frac{y_1^2}{w_2}\right) - \frac{1}{w_1}h'\left(\frac{y_1^2}{w_1}\right)\right]\frac{\partial y_1^2}{\partial s_1} = 0.
\end{aligned}
\tag{A.49}
$$

By part (ii) of Lemma 1, the first term in square brackets on the left-hand side of (A.49) is zero, so that

$$
\begin{aligned}
-\alpha u'(c_1^1) + (\alpha-\psi)(1+r)v'(c_1^2) + \psi u'(c_1^1) \\
+ \psi v'(c_1^2)\left[-\frac{\partial x_1^2}{\partial s_1} + \frac{\partial x_2^2}{\partial s_1} + \frac{\partial y_1^2}{\partial s_1} - \frac{\partial y_2^2}{\partial s_1}\right] \\
+ \psi\left[\frac{1}{w_2}h'\left(\frac{y_1^2}{w_2}\right) - \frac{1}{w_1}h'\left(\frac{y_1^2}{w_1}\right)\right]\frac{\partial y_1^2}{\partial s_1} = 0,
\end{aligned}
\tag{A.50}
$$

which is equivalent to

$$
\begin{aligned}
(\alpha-\psi)u'(c_1^1) + \psi v'(c_1^2)\left[\frac{\partial x_1^2}{\partial s_1} - \frac{\partial x_2^2}{\partial s_1} - \frac{\partial y_1^2}{\partial s_1} + \frac{\partial y_2^2}{\partial s_1}\right] \\
+ \psi\left[\frac{1}{w_1}h'\left(\frac{y_1^2}{w_1}\right) - \frac{1}{w_2}h'\left(\frac{y_1^2}{w_2}\right)\right]\frac{\partial y_1^2}{\partial s_1} = (\alpha-\psi)(1+r)v'(c_1^2).
\end{aligned}
\tag{A.51}
$$

Because both $u(\cdot)$ and $v(\cdot)$ are increasing, it follows from (A.43) that $(\alpha - \psi) > 0$. Rearranging (A.51) yields

$$
\begin{aligned}
\frac{u'(c_1^1)}{v'(c_1^2)} + \left(\frac{\psi}{\alpha-\psi}\right)\left[\frac{\partial x_1^2}{\partial s_1} - \frac{\partial x_2^2}{\partial s_1} - \frac{\partial y_1^2}{\partial s_1} + \frac{\partial y_2^2}{\partial s_1}\right] \\
+ \left(\frac{\psi}{\alpha-\psi}\right)\frac{1}{v'(c_1^2)}\left[\frac{1}{w_1}h'\left(\frac{y_1^2}{w_1}\right) - \frac{1}{w_2}h'\left(\frac{y_1^2}{w_2}\right)\right]\frac{\partial y_1^2}{\partial s_1} = (1+r).
\end{aligned}
\tag{A.52}
$$

By part (iii) of Lemma 1, the second term on the left-hand side of (A.52) is negative. The final term on the left-hand side of (A.52) is also negative. The part in square brackets is positive because $w_1 < w_2$ and $h(\cdot)$ is convex; the derivative is negative by (A.35). The inequality in part (ii) of the proposition follows from these observations.

In Appendix B, we show that the sign of the left-hand side of the analog of (A.52) for s_2 is indeterminate. □

Proof of Proposition 3. The Lagrangian associated with the second period no-commitment tax design problem with pooling or semi-pooling is

$$\begin{aligned}
&\alpha\left[v(x_1^2+(1+r)s_1)-h\left(\frac{y_1^2}{w_1}\right)\right]+(1-\alpha)\pi\left[v(x_{2p}^2+(1+r)s_1)-h\left(\frac{y_{2p}^2}{w_2}\right)\right]\\
&\quad+(1-\alpha)(1-\pi)\left[v(x_{2s}^2+(1+r)s_2)-h\left(\frac{y_{2s}^2}{w_2}\right)\right]\\
&\quad\quad+\zeta\left[\alpha(y_1^2-x_1^2)+(1-\alpha)\pi(y_{2p}^2-x_{2p}^2)\right.\\
&\quad\quad\quad\left.+(1-\alpha)(1-\pi)(y_{2s}^2-x_{2s}^2)+(1+r)s_G\right]\\
&\quad\quad\quad\quad+\phi\left[v(x_{2p}^2+(1+r)s_1)-h\left(\frac{y_{2p}^2}{w_2}\right)-v(x_1^2+(1+r)s_1)+h\left(\frac{y_1^2}{w_2}\right)\right].
\end{aligned}\tag{A.53}$$

The associated first-order conditions include:

$$x_1^2\colon\ \alpha v'(c_1^2)-\alpha\zeta-\phi v'(c_1^2)=0;\tag{A.54}$$

$$y_1^2\colon\ -\frac{\alpha}{w_1}h'\left(\frac{y_1^2}{w_1}\right)+\alpha\zeta+\frac{\phi}{w_2}h'\left(\frac{y_1^2}{w_2}\right)=0;\tag{A.55}$$

$$x_{2p}^2\colon\ (1-\alpha)\pi v'(c_{2p}^2)-\zeta(1-\alpha)\pi+\phi v'(c_{2p}^2)=0;\tag{A.56}$$

$$y_{2p}^2\colon\ -\frac{(1-\alpha)\pi}{w_2}h'\left(\frac{y_{2p}^2}{w_2}\right)+\zeta(1-\alpha)\pi-\frac{\phi}{w_2}h'\left(\frac{y_{2s}^2}{w_2}\right)=0;\tag{A.57}$$

$$x_{2s}^2\colon\ (1-\alpha)(1-\pi)v'(c_{2s}^2)-\zeta(1-\alpha)(1-\pi)=0;\tag{A.58}$$

$$y_{2s}^2\colon\ -\frac{(1-\alpha)(1-\pi)}{w_2}h'\left(\frac{y_{2s}^2}{w_2}\right)+\zeta(1-\alpha)(1-\pi)=0.\tag{A.59}$$

The three relations in the statement of the proposition follow from combining (A.54) with (A.55), (A.56) with (A.57), and (A.58) with (A.59) and using definition (3.5). □

Proof of Proposition 4. Rearranging equations (A.54), (A.56), and (A.58) yields:

$$\left(\frac{\alpha-\phi}{\alpha}\right)v'(c_1^2)=v'(c_{2s}^2)=\left(\frac{(1-\alpha)\pi+\phi}{(1-\alpha)\pi}\right)v'(c_{2p}^2)=\zeta.\tag{A.60}$$

The multiplier ϕ is positive, so

$$v'(c_1^2)>v'(c_{2s}^2)>v'(c_{2p}^2).\tag{A.61}$$

But v is strictly concave. Thus,

$$c_1^2 < c_{2s}^2 < c_{2p}^2, \tag{A.62}$$

and part (i) of the proposition then follows. Note that (A.60) implies that $\phi < \alpha$.

By standard arguments from the optimal nonlinear income tax literature, $y_{2p}^2 > y_1^2$. Equations (A.57) and (A.59) imply

$$\frac{1}{w_2} h'\left(\frac{y_{2s}^2}{w_2}\right) = \left(\frac{(1-\alpha)\pi + \phi}{(1-\alpha)\pi}\right) \frac{1}{w_2} h'\left(\frac{y_{2p}^2}{w_2}\right) = \zeta. \tag{A.63}$$

Because $\phi > 0$,

$$h'\left(\frac{y_{2s}^2}{w_2}\right) > h'\left(\frac{y_{2p}^2}{w_2}\right). \tag{A.64}$$

But h is strictly convex. Thus,

$$y_{2s}^2 > y_{2p}^2. \tag{A.65}$$

Part (ii) then follows.

Because the incentive constraint (7.3) binds, $w_2 > w_1$, and h is strictly convex, the first inequality in part (iii) of the proposition holds. The second inequality in this part follows directly from parts (i) and (ii). $\square$

Proof of Propositions 5, 6, and 7. The Lagrangian associated with the first period no-commitment tax design problem with semi-pooling is

$$\begin{aligned}
&\alpha \left[u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_1}\right)\right] + (1-\alpha)\pi \left[u(x_1^1 - s_1) - g\left(\frac{y_1^1}{w_2}\right)\right] \\
&\quad + (1-\alpha)(1-\pi)\left[u(x_2^1 - s_2) - g\left(\frac{y_2^1}{w_2}\right)\right] + \mathcal{W}^{2,pool}(s_1, s_2, s_g, \pi) \\
&\quad + \gamma \left[u(x_2^1 - s_2^1) - g\left(\frac{y_2^1}{w_2}\right) + \mathcal{V}^{2s}(s_1, s_2, s_G, \pi)\right. \\
&\qquad \left. - u(x_1^1 - s_1) + g\left(\frac{y_1^1}{w_2}\right) - \mathcal{V}^{2p}(s_1, s_2, s_G, \pi)\right] \\
&\qquad + \sigma\left[(\alpha + (1-\alpha)\pi)(y_1^1 - x_1^1) + (1-\alpha)(1-\pi)(y_2^1 - x_2^1) - s_G\right].
\end{aligned} \tag{A.66}$$

If there is pooling, the third and fifth terms in this sum are omitted. The multiplier σ is non-negative, but the multiplier γ is *a priori* of indeterminate sign because it is associated with an equality constraint.

The associated first-order conditions include:

$$x_1^1\colon\ (\alpha + (1-\alpha)\pi)u'(c_1^1) - (\alpha + (1-\alpha)\pi)\sigma - \gamma u'(c_1^1) = 0; \tag{A.67}$$

$$y_1^1\colon\ -\frac{\alpha}{w_1} g'\left(\frac{y_1^1}{w_1}\right) - \frac{(1-\alpha)\pi}{w_2} g'\left(\frac{y_1^1}{w_2}\right) + (\alpha + (1-\alpha)\pi)\sigma + \frac{\gamma}{w_2} g'\left(\frac{y_1^1}{w_2}\right) = 0; \tag{A.68}$$

$$x_2^1\colon\ (1-\alpha)(1-\pi)u'(c_2^1) - (1-\alpha)(1-\pi)\sigma + \gamma u'(c_2^1) = 0; \tag{A.69}$$

$$y_2^1\colon \quad -\frac{(1-\alpha)(1-\pi)}{w_2} g'\left(\frac{y_2^1}{w_2}\right) + (1-\alpha)(1-\pi)\sigma - \frac{\gamma}{w_2} g'\left(\frac{y_2^1}{w_2}\right) = 0. \tag{A.70}$$

$$s_1\colon \quad -[\alpha + (1-\alpha)\pi] u'(c_1^1) + \frac{\partial \mathcal{W}^{2,pool}}{\partial s_1} + \gamma \left[u'(c_1^1) + \frac{\partial \mathcal{V}^{2s}}{\partial s_1} - \frac{\partial \mathcal{V}^{2p}}{\partial s_1}\right] = 0. \tag{A.71}$$

$$s_2\colon \quad -(1-\alpha)(1-\pi) u'(c_2^1) + \frac{\partial \mathcal{W}^{2,pool}}{\partial s_2} + \gamma \left[-u'(c_2^1) + \frac{\partial \mathcal{V}^{2s}}{\partial s_2} - \frac{\partial \mathcal{V}^{2p}}{\partial s_2}\right] = 0. \tag{A.72}$$

Single crossing implies that

$$\mathrm{MRS}_{y_1^1, x_1^1} > \mathrm{MRS}_{y_{2p}^1, x_{2p}^1}, \tag{A.73}$$

which is the first inequality in part (i) of Proposition 6. The equality in this part follows directly from (A.69) and (A.70).

Using either (A.67) or (A.69), it is easy to show that $\sigma > 0$. Equations (A.67) and (A.68) imply

$$[\alpha + (1-\alpha)\pi - \gamma] u'(c_1^1) = [\alpha + (1-\alpha)\pi]\sigma = \frac{\alpha}{w_1} g'\left(\frac{y_1^1}{w_1}\right) + \frac{(1-\alpha)\pi - \gamma}{w_2} g'\left(\frac{y_1^1}{w_2}\right). \tag{A.74}$$

Because $\sigma > 0$ and utility is increasing in consumption, $[\alpha + (1-\alpha)\pi - \gamma] > 0$. Dividing the extreme right-hand side of (A.74) by its extreme left-hand side and using (3.4) yields

$$\frac{\alpha}{\alpha + (1-\alpha)\pi - \gamma} \mathrm{MRS}_{y_1^1, x_1^1} + \frac{(1-\alpha)\pi - \gamma}{\alpha + (1-\alpha)\pi - \gamma} \mathrm{MRS}_{y_{2p}^1, x_{2p}^1} = 1, \tag{A.75}$$

which is the equation in part (i) of Proposition 7. Setting $\gamma = 0$ and $\pi = 1$ in (A.75) shows that part (iii) of Proposition 5 holds when there is pooling. Part (i) of this proposition then follows from (A.73).

By (A.73), the left-hand side of (A.75) is strictly larger than $\mathrm{MRS}_{y_{2p}^1, x_{2p}^1}$. Hence, $1 > \mathrm{MRS}_{y_{2p}^1, x_{2p}^1}$, which is the second inequality in part (i) of Proposition 6.

Part (i) of Proposition 4 and definition (3.6) imply that

$$\mathrm{MRS}_{x_1^1, x_1^2} > \mathrm{MRS}_{x_{2p}^1, x_{2p}^2}, \tag{A.76}$$

which is the first inequality in part (ii) of Proposition 6.

Equations (A.67) and (A.69) imply that

$$\sigma = \left[1 - \frac{\gamma}{\alpha + (1-\alpha)\pi}\right] u'(c_1^1) = \left[1 + \frac{\gamma}{(1-\alpha)(1-\pi)}\right] u'(c_2^1), \tag{A.77}$$

so that

$$c_2^1 \leq c_1^1 \leftrightarrow \gamma \leq 0. \tag{A.78}$$

We shall now show that $\gamma > 0$ when there is semi-pooling. Suppose, by way of contradiction, that $\gamma \leq 0$. Then $c_2^1 \leq c_1^1$. But type $2s$ individuals are worse off in the second period than are their type $2p$ counterparts, so they must be better off in period 1. Hence, it must be the case that $y_2^1 < y_1^1$. Now, (A.68) and (A.70) imply

$$\begin{aligned} &\frac{1}{w_2} g'\left(\frac{y_2^1}{w_2}\right)\left[\frac{(1-\alpha)(1-\pi)+\gamma}{(1-\alpha)(1-\pi)}\right] \\ &\qquad = \frac{(1-\alpha)\pi-\gamma}{[\alpha+(1-\alpha)\pi]w_2} g'\left(\frac{y_1^1}{w_2}\right) + \frac{\alpha}{[\alpha+(1-\alpha)\pi]w_1} g'\left(\frac{y_1^1}{w_1}\right). \end{aligned} \tag{A.79}$$

Because $w_2 > w_1$, (A.79) implies

$$\frac{1}{w_2} g'\left(\frac{y_2^1}{w_2}\right)\left[\frac{(1-\alpha)(1-\pi)+\gamma}{(1-\alpha)(1-\pi)}\right] > \frac{1}{w_2} g'\left(\frac{y_1^1}{w_2}\right)\left[\frac{\alpha+(1-\alpha)\pi-\gamma}{\alpha+(1-\alpha)\pi}\right]. \tag{A.80}$$

Because $[\alpha + (1-\alpha)\pi - \gamma] > 0$, the right-hand side of (A.80) is positive. Therefore, the left-hand side is also positive. Because $y_2^1 < y_1^1$ and g is strictly convex, a necessary condition for (A.80) to hold is

$$\frac{(1-\alpha)(1-\pi)+\gamma}{(1-\alpha)(1-\pi)} > \frac{\alpha+(1-\alpha)\pi-\gamma}{\alpha+(1-\alpha)\pi}. \tag{A.81}$$

But (A.81) holds if and only if

$$\frac{\gamma}{(1-\alpha)(1-\pi)} > -\frac{\gamma}{\alpha+(1-\alpha)\pi} \tag{A.82}$$

or, equivalently,

$$\gamma > 0, \tag{A.83}$$

which contradicts the initial assumption that $\gamma \leq 0$. Thus, it must be the case that $\gamma > 0$. We have already shown that $[\alpha + (1-\alpha)\pi - \gamma] > 0$. Thus, $\gamma \in (0, \alpha + (1-\alpha)\pi)$. We have also shown in the proof of Proposition 3 that $\phi < \alpha$. Thus, the parameter restrictions for γ and ϕ in Proposition 7 are satisfied.

Because $\gamma > 0$, part (ii) of Proposition 7 follows from part (i) of this proposition and the first inequality in part (i) of Proposition 6.

It follows from (A.78) and (A.83) that $c_2^1 > c_1^1$. In light of the strict concavity of both u and v, combining this inequality with part (i) of Propostion 4 implies

$$\frac{u'(c_2^1)}{v'(c_{2s}^2)} < \frac{u'(c_1^1)}{v'(c_{2p}^2)}. \tag{A.84}$$

The second inequality of part (ii) of Proposition 6 now follows from (3.6).

Applying the Envelope Theorem to (A.53) yields

$$\frac{\partial \mathcal{W}^{2,pool}}{\partial s_1} = (\alpha-\phi)(1+r)v'(c_1^2) + [(1-\alpha)\pi+\phi](1+r)v'(c_{2p}^2). \tag{A.85}$$

Substituting (A.54) and (A.56) into (A.85) yields

$$\frac{\partial \mathcal{W}^{2,pool}}{\partial s_1} = (1+r)\zeta[\alpha + (1-\alpha)\pi]. \tag{A.86}$$

Substituting (A.86) into (A.71) and rearranging yields

$$u'(c_1^1) = (1+r)\zeta + \frac{\gamma}{\alpha + (1-\alpha)\pi}\left[u'(c_1^1) + \frac{\partial \mathcal{V}^{2s}}{\partial s_1} - \frac{\partial \mathcal{V}^{2p}}{\partial s_1}\right]. \tag{A.87}$$

Solving (A.54) for ζ and substituting the resulting expression into (A.87) gives

$$u'(c_1^1) = \frac{(1+r)(\alpha - \phi)}{\alpha} v'(c_1^2) + \frac{\gamma}{\alpha + (1-\alpha)\pi}\left[u'(c_1^1) + \frac{\partial \mathcal{V}^{2s}}{\partial s_1} - \frac{\partial \mathcal{V}^{2p}}{\partial s_1}\right]. \tag{A.88}$$

Part (iii) of Proposition 7 follows from dividing both sides of (A.88) by $-v'(c_1^2)$.

Solving (A.56) for ζ and substituting the resulting expression into (A.87) gives

$$u'(c_1^1) = (1+r)\left(\frac{(1-\alpha)\pi + \phi}{(1-\alpha)\pi}\right) v'(c_{2p}^2) + \frac{\gamma}{\alpha + (1-\alpha)\pi}\left[u'(c_1^1) + \frac{\partial \mathcal{V}^{2s}}{\partial s_1} - \frac{\partial \mathcal{V}^{2p}}{\partial s_1}\right]. \tag{A.89}$$

Part (iv) of Proposition 7 follows from dividing both sides of (7) by $-v'(c_{2p}^2)$.

When there is pooling, the expressions corresponding to parts (iii) and (iv) of Proposition 7 are obtained by setting $\gamma = 0$. Taking population weighted sums of these marginal rates of substitution yields the equation in part (iv) of Proposition 5. Part (ii) of this proposition then follows from (A.76).

Applying the Envelope Theorem to (A.53) yields

$$\frac{\partial \mathcal{W}^{2,pool}}{\partial s_2} = (1-\alpha)(1-\pi)(1+r)v'(c_{2s}^2). \tag{A.90}$$

Substituting (A.90) into (A.72) and rearranging yields

$$u'(c_2^1) = (1+r)v'(c_{2s}^2) + \frac{\gamma}{(1-\alpha)(1-\pi)}\left[-u'(c_2^1) + \frac{\partial \mathcal{V}^{2s}}{\partial s_2} - \frac{\partial \mathcal{V}^{2p}}{\partial s_2}\right]. \tag{A.91}$$

Part (v) of Proposition 7 follows from dividing both sides of (A.91) by $-v'(c_{2s}^2)$. □

References

Apps, P., Rees, R., 2009. Two extensions to the theory of optimal taxation, unpublished manuscript, University of Sydney and University of Munich.

Atkinson, A. B., Stiglitz, J. E., 1976. The design of tax structure: Direct versus indirect taxation. Journal of Public Economics 6, 55–75.

Battaglini, M., Coate, S., 2008. Pareto efficient income taxation with stochastic abilities. Journal of Public Economics 92, 844–868.

Benhabib, J., Rustichini, A., 1997. Optimal taxes without commitment. Journal of Economic Theory 77, 231–259.

Berliant, M., Ledyard, J. O., 2014. Optimal dynamic nonlinear income taxes with no commitment. Journal of Public Economic Theory 16, 196–221.

Bisin, A., Rampini, A. A., 2006. Markets as beneficial constraints on government. Journal of Public Economics 90, 601–629.

Boadway, R., 2012. From Optimal Tax Theory to Tax Policy: Retrospective and Prospective Views. MIT Press, Cambridge, MA.

Boadway, R., Marceau, N., Marchand, M., 1996. Investment in education and the time inconsistency of redistributive tax policy. Economica 63, 171–189.

Brito, D. L., Hamilton, J. A., Slutsky, S. M., Stiglitz, J. E., 1990. Pareto efficient tax structures. Oxford Economic Papers 42, 61–77.

Browning, M., Burbidge, J., 1990. Consumption and income taxation. Oxford Economic Papers 42, 281–292.

Dillén, M., Lundholm, M., 1996. Dynamic income taxation, redistribution, and the ratchet effect. Journal of Public Economics 59, 69–93.

Farhi, E., Sleet, C., Werning, I., Yeltekin, S., 2012. Non-linear capital taxation without commitment. Review of Economic Studies 79, 1469–1493.

Freixas, X., Guesnerie, R., Tirole, J., 1985. Planning under incomplete information and the ratchet effect. Review of Economic Studies 52, 173–191.

Gaube, T., 2010. Taxation of annual income as a commitment device, unpublished manuscript, University of Osnabrück.

Golosov, M., Iovino, L., 2015. Social insurance, information revelation, and lack of commitment, unpublished manuscript, Princeton University and Bocconi University.

Golosov, M., Kocherlakota, N., Tsyvinski, A., 2003. Optimal indirect and capital taxation. Review of Economic Studies 70, 569–587.

Golosov, M., Tsyvinski, A., 2015. Policy implications of dynamic public finance. Annual Review of Economics 7, 147–171.

Golosov, M., Tsyvinski, A., Werning, I., 2007. New dynamic public finance: A user's guide. In: Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), NBER Macroeconomics Annual 2006. Vol. 21. MIT Press, Cambridge, MA, pp. 317–363.

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

Hammond, P. J., 1979. Straightforward individual incentive compatibility in large economies. Review of Economic Studies 46, 263–282.

Kocherlakota, N. R., 2005. Zero expected wealth taxes: A Mirrlees approach to dynamic optimal taxation. Econometrica 73, 1587–1621.

Kocherlakota, N. R., 2006. Advances in dynamic optimal taxation. In: Blundell, R., Newey, W. K., Persson, T. (Eds.), Advances In Economics and Econometrics. Theory and Applications, Ninth World Congress. Vol. I. Cambridge University Press, Cambridge, pp. 269–297.

Kocherlakota, N. R., 2010. The New Dynamic Public Finance. Princeton University

Press, Princeton, NJ.
Krause, A., 2009. Optimal nonlinear income taxation with learning-by-doing. Journal of Public Economics 93, 1098–1110.
Mirrlees, J. A., 1971. An exploration in the theory of optimum income taxation. Review of Economic Studies 38, 175–208.
Ordover, J. A., Phelps, E. S., 1979. The concept of optimal taxation in the overlapping–generations model of capital and wealth. Journal of Public Economics 12, 1–26.
Pirttilä, J., Tuomala, M., 2001. On optimal non-linear taxation and public good provision in an overlapping generations economy. Journal of Public Economics 79, 485–501.
Roberts, K., 1984. The theoretical limits to redistribution. Review of Economic Studies 51, 177–195.
Scheuer, F., Wolitzky, A., 2016. Capital taxation under political constraints. American Economic Review 106, forthcoming.
Stiglitz, J. E., 1982. Self-selection and Pareto efficient taxation. Journal of Public Economics 17, 213–240.
Sundaram, R. K., 1996. A First Course in Optimization Theory. Cambridge University Press, Cambridge.
Tuomala, M., 2016. Optimal Redistributive Taxation. Oxford University Press, Oxford.
Weitzman, M. L., 1980. The ratchet principle and performance incentives. Bell Journal of Economics 11, 302–308.
Weymark, J. A., 1986. Bunching properties of optimal nonlinear income taxes. Social Choice and Welfare 3, 213–232.

Appendix B. Supplementary File

Derivation of (A.41). We implicitly differentiate (A.19)–(A.23) with respect to the endogenous variables $(x_1^2, x_2^2, y_1^2, y_2^2, \lambda)$ and the parameter s_2 to obtain

$$A \begin{bmatrix} dx_1^2 \\ dx_2^2 \\ dy_1^2 \\ dy_2^2 \\ d\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ -(1-\alpha)(1+r)v''(c_2^2) \\ 0 \\ 0 \\ 0 \end{bmatrix} ds_2. \tag{B.1}$$

Because A is invertible, we can now use Cramer's Rule to compute the derivatives of the allocation functions with respect to s_2. For x_1^2, we have

$$|A|\frac{\partial x_1^2}{\partial s_2} = \begin{vmatrix} 0 & 0 & 0 & 0 & -\alpha \\ -(1-\alpha)(1+r)v''(c_2^2) & (1-\alpha)v''(c_2^2) & 0 & 0 & -(1-\alpha) \\ 0 & 0 & -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ 0 & -(1-\alpha) & \alpha & (1-\alpha) & 0 \end{vmatrix}. \tag{B.2}$$

Computing the determinant by expanding along the first column,

$$|A|\frac{\partial x_1^2}{\partial s_2} = (1-\alpha)(1+r)v''(c_2^2)(1-\alpha) \begin{vmatrix} 0 & 0 & -\alpha \\ -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \end{vmatrix} \tag{B.3}$$

or, equivalently,

$$|A|\frac{\partial x_1^2}{\partial s_2} = -\alpha^2(1-\alpha)^3(1+r)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2}. \tag{B.4}$$

For x_2^2, we have

$$|A|\frac{\partial x_2^2}{\partial s_2} = \begin{vmatrix} \alpha v''(c_1^2) & 0 & 0 & 0 & -\alpha \\ 0 & -(1-\alpha)(1+r)v''(c_2^2) & 0 & 0 & -(1-\alpha) \\ 0 & 0 & -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ -\alpha & 0 & \alpha & (1-\alpha) & 0 \end{vmatrix}. \tag{B.5}$$

Computing the determinant by expanding along the second column,

$$|A|\frac{\partial x_2^2}{\partial s_2} = -(1-\alpha)(1+r)v''(c_2^2)\left\{\alpha v''(c_1^2)\begin{vmatrix} -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ \alpha & (1-\alpha) & 0 \end{vmatrix} + \alpha\begin{vmatrix} 0 & 0 & -\alpha \\ -\frac{\alpha h''(l_1^2)}{(w_1)^2} & 0 & \alpha \\ 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \end{vmatrix}\right\} \tag{B.6}$$

or, equivalently,

$$|A|\frac{\partial x_2^2}{\partial s_2} = -\alpha^2(1-\alpha)^2(1+r)v''(c_2^2) \times \left[(1-\alpha)v''(c_1^2)\frac{h''(l_1^2)}{(w_1)^2} + \alpha v''(c_1^2)\frac{h''(l_2^2)}{(w_2)^2} - \alpha\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2}\right]. \tag{B.7}$$

For y_1^2, we have

$$|A|\frac{\partial y_1^2}{\partial s_2} = \begin{vmatrix} \alpha v''(c_1^2) & 0 & 0 & 0 & -\alpha \\ 0 & (1-\alpha)v''(c_2^2) & -(1-\alpha)(1+r)v''(c_2^2) & 0 & -(1-\alpha) \\ 0 & 0 & 0 & 0 & \alpha \\ 0 & 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} & (1-\alpha) \\ -\alpha & -(1-\alpha) & 0 & (1-\alpha) & 0 \end{vmatrix}. \tag{B.8}$$

Computing the determinant by expanding along the third column,

$$|A|\frac{\partial y_1^2}{\partial s_2} = (1-\alpha)(1+r)v''(c_2^2)\alpha\begin{vmatrix} \alpha v''(c_1^2) & 0 & 0 \\ 0 & 0 & -\frac{(1-\alpha)h''(l_2^2)}{(w_2)^2} \\ -\alpha & -(1-\alpha) & (1-\alpha) \end{vmatrix} \tag{B.9}$$

or, equivalently,

$$|A|\frac{\partial y_1^2}{\partial s_2} = -\alpha^2(1-\alpha)^3(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_2^2)}{(w_2)^2}. \tag{B.10}$$

Finally, for y_2^2, we have

$$|A|\frac{\partial y_2^2}{\partial s_2} = \begin{vmatrix} \alpha v''(c_1^2) & 0 & 0 & 0 & -\alpha \\ 0 & (1-\alpha)v''(c_2^2) & 0 & -(1-\alpha)(1+r)v''(c_2^2) & -(1-\alpha) \\ 0 & 0 & -\frac{\alpha h''(l_1^2)}{(w_2)^1} & 0 & \alpha \\ 0 & 0 & 0 & 0 & (1-\alpha) \\ -\alpha & -(1-\alpha) & \alpha & 0 & 0 \end{vmatrix}. \tag{B.11}$$

Computing the determinant by expanding along the fourth column,

$$|A|\frac{\partial y_2^2}{\partial s_2} = (1-\alpha)(1+r)v''(c_2^2)(1-\alpha)\begin{vmatrix} \alpha v''(c_1^2) & 0 & 0 \\ 0 & 0 & -\frac{\alpha)h''(l_1^2)}{(w_1)^2} \\ -\alpha & -(1-\alpha) & \alpha \end{vmatrix} \tag{B.12}$$

or, equivalently,

$$|A|\frac{\partial y_2^2}{\partial s_2} = -\alpha^2(1-\alpha)^3(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2}. \tag{B.13}$$

Combining (B.7)–(B.13) yields

$$\begin{aligned} |A|\left[\frac{\partial x_1^2}{\partial s_2} - \frac{\partial x_2^2}{\partial s_2} - \frac{\partial y_1^2}{\partial s_2} + \frac{\partial y_2^2}{\partial s_2}\right] &= \alpha^2(1-\alpha)^3(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2} \\ + \alpha^3(1-\alpha)^2(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_2^2)}{(w_2)^2} &- \alpha^2(1-\alpha)^3(1+r)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2} \\ - \alpha^3(1-\alpha)^2(1+r)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2}\frac{h''(l_2^2)}{(w_2)^2} &+ \alpha^2(1-\alpha)^3(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_2^2)}{(w_2)^2} \\ - \alpha^2(1-\alpha)^3(1+r)v''(c_1^2)v''(c_2^2)\frac{h''(l_1^2)}{(w_1)^2}. \end{aligned} \tag{B.14}$$

The first and sixth terms on the right-hand side of (B.14) cancel. Factoring common elements, grouping the second term with the fifth, and the third with the fourth yields (A.41).

The analog of (A.52) *for* s_2. An argument similar to the one used to establish (A.52) may be used to show that (A.48) is equivalent to

$$\begin{aligned} &\frac{u'(c_2^1)}{v'(c_2^2)} + \left(\frac{\psi}{1-\alpha+\psi}\right)\left[\frac{\partial x_1^2}{\partial s_2} - \frac{\partial x_2^2}{\partial s_2} - \frac{\partial y_1^2}{\partial s_2} + \frac{\partial y_2^2}{\partial s_2}\right] \\ &\qquad + \left(\frac{\psi}{1-\alpha+\psi}\right)\frac{1}{v'(c_2^2)}\left[\frac{1}{w_1}h'\left(\frac{y_1^2}{w_1}\right) - \frac{1}{w_2}h'\left(\frac{y_1^2}{w_2}\right)\right]\frac{\partial y_1^2}{\partial s_2} = (1+r). \end{aligned} \tag{B.15}$$

By part (iv) of Lemma 1, the second term on the left-hand side of (B.15) is positive. In the discussion following (A.52), we have shown that the term in square brackets in the third term is positive. It then follows from (B.10) that the third term is negative. Because the second and third terms have opposite signs, the sign of the savings distortion for the high skilled is indeterminate.