THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2008/08

Optimal Nonlinear Income Taxation with Learning-by-Doing

By

Alan Krause

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

Optimal Nonlinear Income Taxation with Learning-by-Doing

Alan Krause*

7 May 2008

Abstract

This paper examines a two-period model of optimal nonlinear income taxation with learning-by-doing, in which second-period wages are an increasing function of first-period labour supply. We consider the cases when the government can and cannot commit to its second-period tax policy. In both cases, the canonical Mirrlees/Stiglitz results regarding optimal marginal tax rates no longer apply. In particular, if the government cannot commit and skill-type information is revealed, it is optimal to distort the high-skill consumer's labour supply downwards through a positive marginal tax rate to relax the incentive-compatibility constraint. Alternatively, if the government cannot commit and skill-type information is concealed, it is optimal to distort the high-skill consumer's labour supply upwards to relax the incentive-compatibility constraint, but due to some other factors at work the high-skill consumer's marginal tax rate cannot be signed. Our analysis therefore identifies a setting in which a positive marginal tax rate on the highest-skill individual can be justified, despite its depressing effect on labour supply and wages.

Keywords: Income taxation, learning-by-doing, commitment.

JEL classifications: H21, H24.

---

*Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, U.K. Email: ak519@york.ac.uk. I thank seminar participants at the University of Exeter for helpful comments. Any errors are my responsibility.

1 Introduction

In recent years, a literature known as the 'new dynamic public finance' has emerged that extends the static Mirrlees [1971] model of optimal nonlinear income taxation to a dynamic setting. For the most part, this literature assumes that random productivity shocks determine future wages, and that the government can commit to its future tax policy.[1] In this paper, we assume that wages are determined by 'learning-by-doing', i.e., an individual who works longer in the present becomes more productive through work experience, and therefore enjoys higher wages in the future. Our interest in learning-by-doing stems from the observation that, while the role of education in raising wages has received a great deal of attention in terms of its implications for redistributive taxation, as far as we know the similar role of learning-by-doing has received no attention.[2] Given that work experience is arguably at least as important as formal education in raising productivity in many occupations, the implications of learning-by-doing for redistributive taxation are potentially important. We consider the case when the government can commit to its future tax policy, but we also consider the case when the government cannot commit. We think the no-commitment case is particularly relevant, since the second-best nature of the Mirrlees framework stems from the assumption that an individual's skill type is private information. But taxation in earlier periods may result in this information being revealed to the government, which would enable the government to implement first-best taxation in latter periods. As a result, some individuals may be reluctant to reveal their skill type in earlier periods, in order to avoid being subjected to first-best taxation in latter periods.

We work with the two-type version of the Mirrlees model with a single consumer of each type, but extend it to a two-period setting. It is well known that in the static two-type model, a government with redistributive goals will impose a positive marginal tax

[1] See Golosov, et al. [2006] for a review of the 'new dynamic public finance' literature. This literature has been developed by macroeconomists who recognise that the representative-agent (Ramsey) approach to optimal taxation omits some important features that are relevant for determining optimal taxes. Recent contributions to this literature include Kocherlakota [2005], Albanesi and Sleet [2006], and Werning [2007].

[2] Learning-by-doing has featured in growth models with taxation, but the focus is on how the government can set taxes to smooth the business cycle. For example, see Martin and Rogers [2000].

rate on the low-skill type and a zero marginal tax rate on the high-skill type.[3] The rationale for the positive marginal tax rate on the low-skill type is to distort her labour supply downwards to relax the high-skill type's incentive-compatibility constraint. One might expect that learning-by-doing simply gives the government an additional motive for marginal distortions, e.g., distorting the low-skill type's first-period labour supply upwards may facilitate redistribution by increasing her second-period wage. Or distorting both types first-period labour supply upwards may increase social welfare via higher second-period wages. However, the only motive the government has to implement marginal tax rate distortions remains that to relax the high-skill type's incentive-compatibility constraint. This is because the consumers rationally consider the effect on their second-period wage when deciding their first-period labour supply. Thus the government has no reason to distort individual behaviour to correct any sort of dynamic inconsistency.[4] Nevertheless, we show that the static optimal marginal tax rate results no longer apply, even when the government can commit to its second-period tax policy.

When the government can commit, it may be optimal for the low-skill type to face a negative marginal tax rate in the first period in order to relax the high-skill type's incentive-compatibility constraint. This result also applies when the government cannot commit. Moreover, when the government cannot commit, the standard 'no-distortion-at-the-top' result no longer holds. If the consumers are separated in period 1, thus giving the government enough information to implement first-best taxation in period 2, it is optimal for the high-skill type to face a positive marginal tax rate in period 1. This is because the government wants to distort the high-skill type's first-period labour supply downwards to relax the incentive-compatibility constraint. If the consumers are pooled in period 1, thus constraining the government to use second-best taxation in period 2, the government wants to distort the high-skill type's first-period labour supply upwards

[3] There is a large literature that works with the two-type version of the Mirrlees [1971] model introduced by Stiglitz [1982]. This is due to its simplicity, but also because theory alone sheds little light on the pattern of optimal marginal tax rates over the intermediate skills range. An exception is Boadway and Jacquet [2006], who show that some features of the entire optimal income tax schedule can be characterised theoretically if the government's objective is a maxi-min social welfare function.

[4] In general, if consumers exhibit dynamically inconsistent preferences, then a clear-cut case can be made for corrective taxation. For example, see O'Donoghue and Rabin [2006] for a time-inconsistency argument in favour of taxes on unhealthy foods.

to relax the incentive-compatibility constraint. This suggests that the high-skill type should face a negative marginal tax rate in period 1, but due to some other factors at work the high-skill type's first-period optimal marginal tax rate cannot be signed.

Since the second period of our model is the last period, the optimal tax problem in the second period is identical to that in a static model, and therefore the static results apply. The only exception is when first-best taxation is possible in the second period, in which case both consumers naturally face zero marginal tax rates. Our focus therefore is on optimal taxation in the first period, since it is the first period that captures the essential challenge of dynamic taxation. That is, when choosing its present tax policy, the government must also consider how this affects its taxation possibilities in the future.

The present paper is related to recent work by Berliant and Ledyard [2005], Apps and Rees [2006], and Brett and Weymark [2008c]. These papers also employ two-period nonlinear income tax models in which the government cannot commit, although learning-by-doing does not feature in their models.[5] Instead, they assume that wages are fixed and constant through time. In Berliant and Ledyard [2005] there is a continuum of types, and their focus is on deriving conditions under which the consumers are separated in the first period. They contrast this possibility with the infinite-horizon model of Roberts [1984], in which the consumers are never separated. In Apps and Rees [2006] and Brett and Weymark [2008c] there are only two types, but in Apps and Rees [2006] there is a continuum of consumers of each type while in Brett and Weymark [2008c] there is a single consumer of each type. This makes partial pooling possible in Apps and Rees [2006], while in Brett and Weymark [2008c] the consumers are either separated or pooled.[6] Our model is therefore most closely related to that of Brett and Weymark [2008c]. They show that the static optimal marginal tax rate results remain intact if the government can commit. The static results also remain intact in the first period if the government

[5] Gaube [2007] also considers a two-period model of nonlinear income taxation without learning-by-doing, but assumes that the government can commit. His focus is on showing that, if the government cannot control consumption in each period due to 'hidden' savings, then the government cannot implement the optimal long-term tax contract with a pair of short-term tax contracts. This problem does not arise in our model, however, since we assume there are no savings.

[6] The focus of Brett and Weymark [2008c] is on the desirability of nonlinear savings taxation; there are no savings in the Apps and Rees [2006] model.

cannot commit and there is separation. However, if there is pooling, the high-skill type faces a positive marginal tax rate and the low-skill type faces a negative marginal tax rate in the first period.[7] Our analysis shows that these results no longer hold—and are often reversed—when wages are determined by learning-by-doing.

The remainder of the paper is organised as follows. Section 2 presents the key features of the model that we consider. Section 3 examines optimal income taxation with commitment, while Section 4 examines optimal income taxation without commitment. Section 5 contains some closing remarks, while proofs are relegated to an appendix.

2 The Economy

There are two consumers in the economy, who both live for two periods. Consumption by consumer i ($i = 1, 2$) in period t ($t = 1, 2$) is denoted by c_i^t, and labour supply by consumer i in period t is denoted by l_i^t. Consumer i's wage is equal to w_i^1 in period 1 and $w_i^2 = w_i^2(l_i^1)$ in period 2, where it is assumed that $\partial w_i^2(\cdot)/\partial l_i^1 > 0$ which captures the notion of learning-by-doing. That is, an increase in consumer i's labour supply in period 1 raises her productivity and hence wage rate in period 2. We assume throughout that $w_2^1 > w_1^1$ and $w_2^2(l_2^1) > w_1^2(l_1^1)$ for all levels of l_2^1 and l_1^1, so consumer 1 is the low-skill worker and consumer 2 is the high-skill worker. Consumer i's pre-tax income in period t is denoted by $y_i^t = w_i^t l_i^t$.

The consumers have identical preferences over consumption and labour in each period, which are represented by the additively separable utility function $u(c_i^t) - v(l_i^t)$, where $u(\cdot)$ is increasing and strictly concave and $v(\cdot)$ is increasing and strictly convex. To obtain an expression for the marginal tax rate faced by consumer i in period t, suppose the consumers faced smooth nonlinear income tax functions $T^1(y_i^1)$ and $T^2(y_i^2)$ in periods 1 and 2, respectively.[8] Then consumer i's behaviour can be described by the

[7] It should be noted that these distortions are not implemented to relax an incentive-compatibility constraint. When there is pooling, the government offers a single tax treatment which in effect is chosen based on an average of the high-skill and low-skill wage rates. This results in the high-skill type's labour supply being distorted downwards, and the low-skill type's labour supply being distorted upwards, to earn the same level of pre-tax income.

[8] It is well known that in models with a finite number of consumers, the optimal income tax schedule

following programme:

$$\max_{c_i^1,\, l_i^1,\, c_i^2,\, l_i^2} \{u(c_i^1) - v(l_i^1) + u(c_i^2) - v(l_i^2) \mid c_i^1 \leq y_i^1 - T^1(y_i^1) \ \wedge\ c_i^2 \leq y_i^2 - T^2(y_i^2)\} \quad (2.1)$$

where for simplicity it is assumed that there are no savings and future utility is not discounted. Thus a consumer's lifetime utility is the simple sum of her utility in each period, and consumption in each period cannot exceed post-tax income in that period. It is shown in the appendix that the solution to programme (2.1) yields the following expressions for the marginal tax rates:

$$MTR_i^1 := \frac{\partial T^1(\cdot)}{\partial y_i^1} = 1 - \frac{v'(l_i^1)}{u'(c_i^1)w_i^1} + \frac{v'(l_i^2)l_i^2}{u'(c_i^1)w_i^1 w_i^2}\frac{\partial w_i^2(\cdot)}{\partial l_i^1} \quad (2.2)$$

$$MTR_i^2 := \frac{\partial T^2(\cdot)}{\partial y_i^2} = 1 - \frac{v'(l_i^2)}{u'(c_i^2)w_i^2} \quad (2.3)$$

where MTR_i^t denotes the marginal tax rate faced by consumer i in period t. Equation (2.3) shows that the marginal tax rate in period 2 is equal to one minus the marginal rate of substitution of pre-tax income for consumption in period 2, which is the same result as obtained in static models.[9] This follows simply from the fact that period 2 is the last period of our model. Similarly, equation (2.2) shows that the marginal tax rate in period 1 is equal to one minus the marginal rate of substitution of first-period pre-tax income for consumption. However, the marginal rate of substitution in (2.2) is complicated by the fact that a marginal increase in y_i^1, which necessitates a marginal increase in l_i^1, results in a *ceteris paribus* increase in utility in period 2 via higher wages. It is the last term in (2.2) that captures this effect.

may not be differentiable. Thus we follow the standard practice of deriving expressions for 'implicit' marginal tax rates in terms of derivatives of the utility function.

[9] See, e.g., Stiglitz [1982], Weymark [1987], and Brett and Weymark [2008a, 2008b].

3 Optimal Income Taxation with Commitment

If in period 1 the government can commit to its tax policy in period 2, the government cannot exploit any information that may be revealed in period 1 to redesign its second-period tax system. In this case, the government can be described as choosing a 'tax contract' $\langle c_1^1, y_1^1, c_1^2, y_1^2 \rangle$ for the low-skill consumer and $\langle c_2^1, y_2^1, c_2^2, y_2^2 \rangle$ for the high-skill consumer to maximise:[10]

$$u(c_1^1) - v\left(\frac{y_1^1}{w_1^1}\right) + u(c_1^2) - v\left(\frac{y_1^2}{w_1^2}\right) + u(c_2^1) - v\left(\frac{y_2^1}{w_2^1}\right) + u(c_2^2) - v\left(\frac{y_2^2}{w_2^2}\right) \tag{3.1}$$

subject to:

$$y_1^1 - c_1^1 + y_2^1 - c_2^1 \geq 0 \tag{3.2}$$

$$y_1^2 - c_1^2 + y_2^2 - c_2^2 \geq 0 \tag{3.3}$$

$$u(c_1^1) - v\left(\frac{y_1^1}{w_1^1}\right) + u(c_1^2) - v\left(\frac{y_1^2}{w_1^2}\right) \geq u(c_2^1) - v\left(\frac{y_2^1}{w_1^1}\right) + u(c_2^2) - v\left(\frac{y_2^2}{\widehat{w}_1^2}\right) \tag{3.4}$$

$$u(c_2^1) - v\left(\frac{y_2^1}{w_2^1}\right) + u(c_2^2) - v\left(\frac{y_2^2}{w_2^2}\right) \geq u(c_1^1) - v\left(\frac{y_1^1}{w_2^1}\right) + u(c_1^2) - v\left(\frac{y_1^2}{\widehat{w}_2^2}\right) \tag{3.5}$$

where $w_1^2 = w_1^2\left(\frac{y_1^1}{w_1^1}\right)$, $w_2^2 = w_2^2\left(\frac{y_2^1}{w_2^1}\right)$, $\widehat{w}_1^2 = w_1^2\left(\frac{y_2^1}{w_1^1}\right)$, and $\widehat{w}_2^2 = w_2^2\left(\frac{y_1^1}{w_2^1}\right)$. The objective function (3.1) is a utilitarian social welfare function, where the utility functions have been written in terms of the government's choice variables c_i^t and y_i^t. Equations (3.2) and (3.3) are budget constraints requiring that total tax revenues be non-negative in each period.[11] Equations (3.4) and (3.5) are incentive-compatibility constraints for consumer 1 and consumer 2, respectively. Following the standard practice, we assume that the government knows there is one low-skill and one high-skill consumer in the economy, but each individual's skill type is private information. The government must therefore satisfy the incentive-compatibility constraints to induce the consumers to choose their intended tax contracts, rather than 'mimicking' the other consumer by choosing the

[10] A tax contract consists of pre-tax income and post-tax income (which is equal to consumption) in each period. The difference between pre-tax income and consumption is total taxes paid (or transfers received). While we do not observe such a tax system in practice, the 'Revelation Principle' implies that any tax system (or any mechanism) can be replicated by an incentive-compatible direct mechanism.

[11] As with the consumers, for simplicity we do not permit the government to save.

other consumer's tax contract. We focus on what Stiglitz [1982] calls the 'normal' case and what Guesnerie [1995] calls 'redistributive equilibria', in the sense that we assume the high-skill consumer's incentive-compatibility constraint (3.5) binds at an optimum while the low-skill consumer's incentive-compatibility constraint (3.4) is slack.[12] Most of the literature has focused on this case, the rationale being that the government uses its taxation powers to redistribute from high-skill to low-skill consumers, which creates an incentive for high-skill consumers to mimic low-skill consumers, but not vice versa.

It is shown in the appendix that the solution to programme (3.1) – (3.5) yields:

Proposition 1 *Optimal income taxation with learning-by-doing and when the government can commit to its second-period tax policy is characterised by*: MTR_1^1 *is ambiguous*, $MTR_2^1 = 0$, $MTR_1^2 > 0$, *and* $MTR_2^2 = 0$.

The pattern of marginal tax rate distortions in the second period—namely, that consumer 1 faces a positive marginal tax rate and consumer 2 faces a zero marginal tax rate—is the same as that in a static model. It is straightforward to show that the consumers are separated in the first period,[13] so the government actually has enough information to implement first-best (lump-sum) taxes in the second period. However, because the government has committed to not exploit any information revealed in period 1, it is obligated to use second-best (incentive-compatible) taxation in period 2. Hence the usual pattern of marginal tax rate distortions is obtained.

Consumer 2 also faces a zero marginal tax rate in period 1, but consumer 1's marginal tax rate cannot be signed. In particular, it is now possible that the government will want to distort consumer 1's first-period labour supply *upwards* via a negative marginal tax rate to relax consumer 2's incentive-compatibility constraint. There are two forces at work here. On the one hand, consumer 1 works longer than consumer 2 when both choose to earn y_1^1 (since $w_1^1 < w_2^1$). Therefore, consumer 1 suffers a greater disutility from labour supply in period 1 than does a mimicking consumer 2. This gives the

[12] We will continue to assume that the low-skill consumer's incentive-compatibility constraint never binds; it is therefore omitted throughout the remainder of the paper.

[13] Using equations (A.6) and (A.8) in the appendix, it can be shown that $c_2^1 > c_1^1$. Hence the consumers make different choices in the first period, which allows the government to identify the low-skill consumer and the high-skill consumer.

government the usual motive to distort consumer 1's labour supply downwards via a positive marginal tax rate to deter mimicking. But on the other hand, learning-by-doing implies that second-period wages are increasing in first-period labour supply. Consumer 1 may therefore obtain a greater increase in her second-period utility from her first-period labour supply than does a mimicking consumer 2. Accordingly, it is possible that the *lifetime* marginal disutility that consumer 1 incurs from additional first-period labour is less than that incurred by a mimicking consumer 2, even though consumer 2 when mimicking works less than consumer 1. If this is the case, the government can relax the incentive-compatibility constraint by distorting consumer 1's first-period labour supply upwards through a negative marginal tax rate. It is also possible that the lifetime marginal disutilities that consumer 1 and a mimicking consumer 2 incur from additional first-period labour are the same. In this case, consumer 1 will face a zero marginal tax rate because distortions to her first-period labour supply will not relax the incentive-compatibility constraint.

4 Optimal Income Taxation without Commitment

If in the first period the government cannot commit to its second-period tax policy, there are two possibilities: (i) the consumers are separated in the first period, giving the government enough information to implement first-best taxation in the second period. (ii) the consumers are pooled in the first period, meaning no information is revealed and the government must use second-best taxation in the second period. Since the consumers—in particular the high-skill consumer—know that if they reveal their types in period 1 they will be subjected to first-best taxation in period 2, the high-skill consumer must be offered a relatively attractive tax contract in period 1 to compensate for the unfavourable tax treatment she will receive in period 2. From a social welfare point of view, the lack of redistribution in period 1 required to obtain type information may be too costly. Instead, the government may be better off offering the same tax contract to both consumers in period 1 so that no type information is revealed, even though it is then constrained to use second-best taxation in period 2. As Brett and Weymark [2008c] note, deciding whether the government is better off with a tax system that separates or pools in the first period requires a comparison of the maximised values of the social welfare function in each case. In general, such comparisons depend upon the exact form of the utility function and the distribution of wages. We therefore examine both possibilities.

4.1 Separation in Period 1 and First-Best Taxation in Period 2

If the consumers are separated in the first period, the government's behaviour in the second period can be described as follows. Choose $\langle c_1^2, y_1^2 \rangle$ and $\langle c_2^2, y_2^2 \rangle$ to maximise:

$$u(c_1^2) - v\left(\frac{y_1^2}{w_1^2}\right) + u(c_2^2) - v\left(\frac{y_2^2}{w_2^2}\right) \tag{4.1}$$

subject to:

$$y_1^2 - c_1^2 + y_2^2 - c_2^2 \geq 0 \tag{4.2}$$

where (4.1) is social welfare in period 2, and (4.2) is the second-period budget constraint. Since the government can identify the consumers, consumer 1 must accept $\langle c_1^2, y_1^2 \rangle$ and consumer 2 must accept $\langle c_2^2, y_2^2 \rangle$. That is, the government is not constrained by incentive-compatibility constraints.

The solution to the above programme yields the functions $c_1^2(y_1^1, w_1^1, y_2^1, w_2^1)$, $y_1^2(y_1^1, w_1^1, y_2^1, w_2^1)$, $c_2^2(y_1^1, w_1^1, y_2^1, w_2^1)$ and $y_2^2(y_1^1, w_1^1, y_2^1, w_2^1)$. Substituting these into (4.1) yields the value function $W^2(y_1^1, w_1^1, y_2^1, w_2^1)$.

Both consumers and the government know that, if there is separation in period 1, the government will solve programme (4.1) – (4.2) in period 2. Therefore, the government in period 1 can be described as choosing $\langle c_1^1, y_1^1 \rangle$ and $\langle c_2^1, y_2^1 \rangle$ to maximise:

$$u(c_1^1) - v\left(\frac{y_1^1}{w_1^1}\right) + u(c_2^1) - v\left(\frac{y_2^1}{w_2^1}\right) + W^2(y_1^1, w_1^1, y_2^1, w_2^1) \tag{4.3}$$

subject to:

$$y_1^1 - c_1^1 + y_2^1 - c_2^1 \geq 0 \tag{4.4}$$

$$u(c_2^1) - v\left(\frac{y_2^1}{w_2^1}\right) + u(c_2^2(\cdot)) - v\left(\frac{y_2^2(\cdot)}{w_2^2}\right) \geq u(c_1^1) - v\left(\frac{y_1^1}{w_2^1}\right) + u(\widehat{c}_2^2(\cdot)) - v\left(\frac{\widehat{y}_2^2(\cdot)}{\widehat{w}_2^2}\right) \tag{4.5}$$

where $\widehat{c}_2^2(\cdot) = \widehat{c}_2^2(y_1^1, w_1^1, w_2^1)$ and $\widehat{y}_2^2(\cdot) = \widehat{y}_2^2(y_1^1, w_1^1, w_2^1)$ represent the tax contract that consumer 2 would receive in the second period if she mimicked in the first period. The government chooses $\langle c_1^1, y_1^1 \rangle$ and $\langle c_2^1, y_2^1 \rangle$ while taking into account how its choice will affect social welfare in period 2. Its first-period objective function (4.3) therefore includes the second-period value function $W^2(\cdot)$. Equation (4.4) is the first-period budget constraint, while (4.5) is consumer 2's incentive-compatibility constraint. In order for consumer 2 to be willing to reveal her type in period 1, the utility she obtains from $\langle c_2^1, y_2^1 \rangle$ in period 1 plus the utility she obtains from the first-best tax contract $\langle c_2^2(\cdot), y_2^2(\cdot) \rangle$ that she must accept in period 2 has to be greater than or equal to the utility she could obtain from $\langle c_1^1, y_1^1 \rangle$ in period 1 plus the utility from the second-best tax contract $\langle \widehat{c}_2^2(\cdot), \widehat{y}_2^2(\cdot) \rangle$ she would choose in period 2. That is, if consumer 2 chooses $\langle c_1^1, y_1^1 \rangle$ in period 1, the consumers are pooled and no type information is revealed. Therefore, conditional on y_1^1, w_1^1, and w_2^1, the government offers consumer 2 an incentive-compatible tax contract $\langle \widehat{c}_2^2(\cdot), \widehat{y}_2^2(\cdot) \rangle$ in period 2.

It is shown in the appendix that the solutions to programmes (4.1) – (4.2) and (4.3) – (4.5) together imply:

Proposition 2 *Optimal income taxation with learning-by-doing, when the government cannot commit to its second-period tax policy, and when the consumers are separated in the first period is characterised by:* MTR_1^1 *is ambiguous,* $MTR_2^1 > 0$, $MTR_1^2 = 0$, *and* $MTR_2^2 = 0$.

The zero marginal tax rate faced by both consumers in period 2 follows simply from the first-best nature of taxation in that period. What is more interesting is the pattern of marginal tax rate distortions in period 1. In particular, the high-skill consumer necessarily faces a positive marginal tax rate. The reason is as follows. The first-best allocation in period 2 involves both consumers receiving the same level of consumption, but consumer 2 works longer than consumer 1.[14] Therefore, consumer 2 obtains a lower level of utility than consumer 1 in the second period. Indeed, it can be shown that for the many-consumer case, first-best taxation has utility decreasing in wages, since

[14]Using equations (A.21) and (A.23) in the appendix we obtain $u'(c_1^2) = u'(c_2^2)$ which implies that $c_1^2 = c_2^2$. Using (A.22) and (A.24) we obtain $v'(l_1^2)/v'(l_2^2) = w_1^2/w_2^2$ which implies that $l_1^2 < l_2^2$.

all consumers receive the same level of consumption but labour supply is increasing in skill type. By distorting consumer 2's labour supply downwards in the first period, the government is decreasing her second-period wage, but actually increasing her second-period utility. Consumer 2's consumption in the second period falls, but her labour supply falls by more, resulting in a net increase in utility. This makes consumer 2 more willing to reveal her type in period 1, i.e., the incentive-compatibility constraint is relaxed.

The sign of the marginal tax rate faced by consumer 1 in period 1 is ambiguous. This is for the same reasons as to why it is ambiguous when the government can commit (see Section 3), but now there are some additional complications. On the one hand, distorting consumer 1's first-period labour supply upwards raises her second-period wage. This reduces the extent of redistribution undertaken using first-best taxation in period 2, which makes consumer 2 better off and thereby relaxes the incentive-compatibility constraint. On the other hand, if consumer 2 were to mimic consumer 1 in the first period, any upward distortion to consumer 1's labour supply would also involve an increase in a mimicking consumer 2's labour supply. As it is not clear how an increase in both consumers' second-period wages affects consumer 2's welfare in period 2, the effect on the incentive-compatibility constraint is also unclear. Thus these additional factors further serve to make consumer 1's first-period marginal tax rate ambiguous.

4.2 Pooling in Period 1 and Second-Best Taxation in Period 2

If the consumers are pooled in the first period, the government's behaviour in the second period can be described as follows. Choose $\langle c_1^2, y_1^2 \rangle$ and $\langle c_2^2, y_2^2 \rangle$ to maximise:

$$u(c_1^2) - v\left(\frac{y_1^2}{w_1^2}\right) + u(c_2^2) - v\left(\frac{y_2^2}{w_2^2}\right) \tag{4.6}$$

subject to:

$$y_1^2 - c_1^2 + y_2^2 - c_2^2 \geq 0 \tag{4.7}$$

$$u(c_2^2) - v\left(\frac{y_2^2}{w_2^2}\right) \geq u(c_1^2) - v\left(\frac{y_1^2}{w_2^2}\right) \tag{4.8}$$

where (4.6) is social welfare in period 2, (4.7) is the second-period budget constraint, and

(4.8) is consumer 2's incentive-compatibility constraint. The solution to the above programme yields the functions $c_1^2(y^1, w_1^1, w_2^1)$, $y_1^2(y^1, w_1^1, w_2^1)$, $c_2^2(y^1, w_1^1, w_2^1)$ and $y_2^2(y^1, w_1^1, w_2^1)$, where y^1 denotes the pre-tax income earned by both consumers in the first period. Substituting these functions into (4.6) yields the value function $W^2(y^1, w_1^1, w_2^1)$.

Both consumers and the government know that, if there is pooling in period 1, the government will solve programme (4.6) – (4.8) in period 2. Therefore, the government in period 1 can be described as choosing $\langle c^1, y^1 \rangle$ to maximise:

$$u(c^1) - v\left(\frac{y^1}{w_1^1}\right) + u(c^1) - v\left(\frac{y^1}{w_2^1}\right) + W^2(y^1, w_1^1, w_2^1) \tag{4.9}$$

subject to:

$$y^1 - c^1 \geq 0 \tag{4.10}$$

where $\langle c^1, y^1 \rangle$ is the tax contract offered to both consumers in period 1. When choosing $\langle c^1, y^1 \rangle$, the government considers how its choice will affect social welfare in period 2. Its first-period objective function (4.9) therefore includes the second-period value function $W^2(\cdot)$. Equation (4.10) is the first-period budget constraint. As both consumers are offered the single choice of $\langle c^1, y^1 \rangle$, the government does not face an incentive-compatibility constraint in the first period.[15]

It is shown in the appendix that the solutions to programmes (4.6) – (4.8) and (4.9) – (4.10) together imply:

Proposition 3 *Optimal income taxation with learning-by-doing, when the government cannot commit to its second-period tax policy, and when the consumers are pooled in the first period is characterised by*: MTR_1^1 *is ambiguous*, MTR_2^1 *is ambiguous*, $MTR_1^2 > 0$, *and* $MTR_2^2 = 0$. *Moreover*, $MTR_1^1 + MTR_2^1 < 0$ *which implies that* $MTR_1^1 < 0$ *and/or* $MTR_2^1 < 0$.

When there is pooling in the first period, the second-period optimal tax problem is identical to that in a static model. Hence the usual pattern of marginal tax rate distortions is obtained in period 2. To understand why the marginal tax rates faced by

[15] Although the government does face the second-period incentive-compatibility constraint (4.8) indirectly through the value function $W^2(\cdot)$.

both consumers in period 1 are ambiguous, suppose learning-by-doing was absent from the model. Then without taxation, consumer 2 would choose to earn a higher income than consumer 1 (as both consumers have the same preferences, but $w_2^1 > w_1^1$). When both consumers are subjected to the same tax treatment in period 1, the government in effect chooses y^1 based on an average of w_1^1 and w_2^1. This results in consumer 1's labour supply being distorted upwards to earn y^1 and consumer 2's labour supply being distorted downwards to earn y^1. Therefore, without learning-by-doing, consumer 1 would face a negative marginal tax rate and consumer 2 would face a positive marginal tax rate. However, with learning-by-doing, it is not necessarily the case that, in the absence of taxation, consumer 2 would choose to earn a higher income than consumer 1 in the first period. This is because the lifetime marginal disutility that consumer 1 incurs from first-period labour may be less than that incurred by consumer 2. (The reasoning is similar to that for the case when the government can commit, as discussed in Section 3.) Therefore, with learning-by-doing MTR_1^1 and MTR_2^1 cannot be signed as it is not clear whether each consumer's labour supply is being distorted upwards or downwards to earn y^1. Furthermore, any marginal increase in y^1 will increase consumer 2's first-period labour supply, which increases her second-period wage and thereby relaxes the incentive-compatibility constraint. An increase in w_2^2 relaxes the incentive-compatibility constraint because, in the second period, consumer 2 works longer when revealing herself than when mimicking. Therefore, there is a higher utility payoff under the former from a wage increase, which reduces the incentive to mimic. This gives the government a motive to distort consumer 2's first-period labour supply upwards to relax the incentive-compatibility constraint, which is the opposite of the case when there is separation. But since the consumers are pooled in period 1, any increase in y^1 used to increase consumer 2's labour supply will also increase consumer 1's labour supply. Thus it can be determined only that, in aggregate, first-period labour will be distorted upwards, i.e., the sum of the first-period marginal tax rates must be negative.

5 Concluding Comments

The 'new dynamic public finance' literature that extends Mirrlees [1971] to a dynamic setting has assumed that random productivity shocks determine future wages, and that the government can commit to its future tax policy. The assumption that the government can commit is a strong one, since the present government cannot commit future governments. For example, Auerbach [2006] cites a recent proposal made to resolve the U.S. Social Security system's imbalance, which includes a tax increase to be made in 2045! As Auerbach notes, such a proposal cannot be taken seriously.

Recent contributions by Berliant and Ledyard [2005], Apps and Rees [2006], and Brett and Weymark [2008c] have dropped the commitment assumption, but they assume that wages are fixed. By contrast, we have assumed that learning-by-doing determines future wages, and that the government may not be able to commit. Given that the sole source of heterogeneity in the Mirrlees framework is wage differentials, understanding how optimal marginal tax rates respond to changes in wages seems particularly relevant. It has long been known that endogenous wages in static models make it optimal for the high-skill type to face a negative marginal tax rate (see Stiglitz [1982]). Recently, Simula [2007] and Brett and Weymark [2008b] have derived a number of comparative static results for exogenous changes in wages. However, the 'no-distortion-at-the-top' result remains intact. Our analysis shows that in a dynamic model with wages determined by learning-by-doing, the 'no-distortion-at-the-top' result no longer applies, and a positive marginal tax rate on the high-skill type can be justified despite its depressing effect on labour supply and wages.

6 Appendix

Derivation of Equations (2.2) and (2.3)

The Lagrangian corresponding to programme (2.1) can be written as:

$$L = u(c_i^1) - v(l_i^1) + u(c_i^2) - v(l_i^2)$$

$$+ \alpha^1 \left[w_i^1 l_i^1 - T^1(w_i^1 l_i^1) - c_i^1\right] + \alpha^2 \left[w_i^2(l_i^1) l_i^2 - T^2(w_i^2(l_i^1) l_i^2) - c_i^2\right] \tag{A.1}$$

where α^1 and α^2 are Lagrange multipliers. The relevant first-order conditions can be written as:

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

$$-v'(l_i^1) + \alpha^1 w_i^1 \left[1 - \frac{\partial T^1(\cdot)}{\partial y_i^1}\right] + \alpha^2 l_i^2 \frac{\partial w_i^2(\cdot)}{\partial l_i^1} \left[1 - \frac{\partial T^2(\cdot)}{\partial y_i^2}\right] = 0 \tag{A.3}$$

$$u'(c_i^2) - \alpha^2 = 0 \tag{A.4}$$

$$-v'(l_i^2) + \alpha^2 w_i^2 \left[1 - \frac{\partial T^2(\cdot)}{\partial y_i^2}\right] = 0 \tag{A.5}$$

Straightforward manipulation of (A.4) and (A.5) yields equation (2.3). After substituting (2.3) into (A.3) and using (A.2) and (A.4) to eliminate the Lagrange multipliers, equation (A.3) can be manipulated to yield equation (2.2). ■

Proof of Proposition 1

The relevant first-order conditions corresponding to programme (3.1) – (3.5) are:

$$(1-\theta)u'(c_1^1) - \lambda^1 = 0 \tag{A.6}$$

$$-v'\left(\frac{y_1^1}{w_1^1}\right)\frac{1}{w_1^1} + v'\left(\frac{y_1^2}{w_1^2}\right)\frac{y_1^2}{w_1^2 w_1^2 w_1^1}\frac{\partial w_1^2(\cdot)}{\partial l_1^1} + \lambda^1 + \theta v'\left(\frac{y_1^1}{w_2^1}\right)\frac{1}{w_2^1} - \theta v'\left(\frac{y_1^2}{\widehat{w}_2^2}\right)\frac{y_1^2}{\widehat{w}_2^2 \widehat{w}_2^2 w_2^1}\frac{\partial w_2^2(\cdot)}{\partial \widehat{l}_2^1} = 0 \tag{A.7}$$

$$(1+\theta)u'(c_2^1) - \lambda^1 = 0 \tag{A.8}$$

$$-(1+\theta)v'\left(\frac{y_2^1}{w_2^1}\right)\frac{1}{w_2^1} + (1+\theta)v'\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2 w_2^2 w_2^1}\frac{\partial w_2^2(\cdot)}{\partial l_2^1} + \lambda^1 = 0 \tag{A.9}$$

$$(1-\theta)u'(c_1^2) - \lambda^2 = 0 \tag{A.10}$$

$$-v'\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2} + \lambda^2 + \theta v'\left(\frac{y_1^2}{\widehat{w}_2^2}\right)\frac{1}{\widehat{w}_2^2} = 0 \tag{A.11}$$

$$(1+\theta)u'(c_2^2) - \lambda^2 = 0 \tag{A.12}$$

$$-(1+\theta)v'\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2} + \lambda^2 = 0 \tag{A.13}$$

where λ^1 is the multiplier on the first-period budget constraint (3.2), λ^2 is the multiplier on the second-period budget constraint (3.3), θ is the multiplier on consumer 2's incentive-compatibility constraint (3.5), and $\widehat{l}_2^1 = y_1^1/w_2^1$. Equations (A.8) and (A.12) imply that $\lambda^1 > 0$ and $\lambda^2 > 0$, and therefore both budget constraints are binding. By assumption (3.5) is binding, and therefore $\theta > 0$.

Dividing (A.13) by (A.12) and rearranging yields:

$$\frac{v'(l_2^2)}{u'(c_2^2)w_2^2} = 1 \tag{A.14}$$

which using (2.3) establishes that $MTR_2^2 = 0$. Similarly, dividing (A.9) by (A.8) and rearranging yields:

$$\frac{v'(l_2^1)}{u'(c_2^1)w_2^1} - \frac{v'(l_2^2)l_2^2}{u'(c_2^1)w_2^1 w_2^2}\frac{\partial w_2^2(\cdot)}{\partial l_2^1} = 1 \tag{A.15}$$

which using (2.2) establishes that $MTR_2^1 = 0$.

Using (A.10) and (A.11) we obtain:

$$(1-\theta)u'(c_1^2) = v'\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2} - \theta v'\left(\frac{y_1^2}{\widehat{w}_2^2}\right)\frac{1}{\widehat{w}_2^2} \tag{A.16}$$

Because $\widehat{w}_2^2 > w_1^2$ and $v(\cdot)$ is strictly convex:

$$v'\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2} - \theta v'\left(\frac{y_1^2}{\widehat{w}_2^2}\right)\frac{1}{\widehat{w}_2^2} > v'\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2} - \theta v'\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2} \tag{A.17}$$

Therefore, (A.16) and (A.17) imply that:

$$(1-\theta)u'(c_1^2) > (1-\theta)v'\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2} \tag{A.18}$$

Using (A.6) it follows that $\theta \in (0,1)$. Hence (A.18) can be rearranged to yield:

$$1 > \frac{v'(l_1^2)}{u'(c_1^2)w_1^2} \tag{A.19}$$

which using (2.3) establishes that $MTR_1^2 > 0$.

To show that MTR_1^1 is ambiguous, use (A.6) and (A.7) to obtain:

$$(1-\theta)u'(c_1^1) = v'\left(\frac{y_1^1}{w_1^1}\right)\frac{1}{w_1^1} - v'\left(\frac{y_1^2}{w_1^2}\right)\frac{y_1^2}{w_1^2 w_1^2 w_1^1}\frac{\partial w_1^2(\cdot)}{\partial l_1^1} - \theta\left[v'\left(\frac{y_1^1}{w_2^1}\right)\frac{1}{w_2^1} - v'\left(\frac{y_1^2}{\widehat{w}_2^2}\right)\frac{y_1^2}{\widehat{w}_2^2\widehat{w}_2^2 w_2^1}\frac{\partial w_2^2(\cdot)}{\partial \widehat{l}_2^1}\right] \tag{A.20}$$

The first two terms on the right-hand side of (A.20) together represent consumer 1's marginal disutility of first-period labour (MDL_1^1). It is equal to the direct disutility incurred from working longer in period 1 minus the utility obtained in period 2 via an increase in the second-period wage. Likewise, the two terms in square brackets on the right-hand side of (A.20) together represent consumer 2's marginal disutility of first-period labour ($\widehat{MDL}_2^1$), but for the case when consumer 2 mimics consumer 1. Since $MDL_1^1 - \widehat{MDL}_2^1$ cannot be signed, MTR_1^1 is ambiguous. Specifically, by undertaking manipulations of (A.20) analogous to those of (A.16) used to sign MTR_1^2, it can be shown that $MTR_1^1 > 0$ if and only if $MDL_1^1 > \widehat{MDL}_2^1$, $MTR_1^1 = 0$ if and only if $MDL_1^1 = \widehat{MDL}_2^1$, and $MTR_1^1 < 0$ if and only if $MDL_1^1 < \widehat{MDL}_2^1$. ■

Proof of Proposition 2

The first-order conditions corresponding to programme (4.1) – (4.2) are:

$$u'(c_1^2) - \lambda^2 = 0 \tag{A.21}$$

$$-v'\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2} + \lambda^2 = 0 \tag{A.22}$$

$$u'(c_2^2) - \lambda^2 = 0 \tag{A.23}$$

$$-v'\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2} + \lambda^2 = 0 \tag{A.24}$$

$$y_1^2 - c_1^2 + y_2^2 - c_2^2 = 0 \tag{A.25}$$

where λ^2 is the multiplier on the second-period budget constraint (4.2). Dividing (A.22) by (A.21) and rearranging yields:

$$\frac{v'(l_1^2)}{u'(c_1^2)w_1^2} = 1 \tag{A.26}$$

while dividing (A.24) by (A.23) and rearranging yields:

$$\frac{v'(l_2^2)}{u'(c_2^2)w_2^2} = 1 \tag{A.27}$$

which using (2.3) establish that $MTR_1^2 = 0$ and $MTR_2^2 = 0$.

The relevant first-order conditions corresponding to programme (4.3) – (4.5) can be written as:

$$(1-\theta)u'(c_1^1) - \lambda^1 = 0 \tag{A.28}$$

$$-v'\left(\frac{y_1^1}{w_1^1}\right)\frac{1}{w_1^1} + \frac{\partial W^2(\cdot)}{\partial y_1^1} + \lambda^1 + \theta v'\left(\frac{y_1^1}{w_2^1}\right)\frac{1}{w_2^1} - \theta v'\left(\frac{\widehat{y}_2^2}{\widehat{w}_2^2}\right)\frac{\widehat{y}_2^2}{\widehat{w}_2^2\widehat{w}_2^2 w_2^1}\frac{\partial \widehat{w}_2^2(\cdot)}{\partial \widehat{l}_2^1}$$
$$+\,\theta\left[u'(c_2^2)\frac{\partial c_2^2(\cdot)}{\partial y_1^1} - v'\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2}\frac{\partial y_2^2(\cdot)}{\partial y_1^1}\right] - \theta\left[u'(\widehat{c}_2^2)\frac{\partial \widehat{c}_2^2(\cdot)}{\partial y_1^1} - v'\left(\frac{\widehat{y}_2^2}{\widehat{w}_2^2}\right)\frac{1}{\widehat{w}_2^2}\frac{\partial \widehat{y}_2^2(\cdot)}{\partial y_1^1}\right] = 0 \tag{A.29}$$

$$(1+\theta)u'(c_2^1) - \lambda^1 = 0 \tag{A.30}$$

$$-v'\left(\frac{y_2^1}{w_2^1}\right)\frac{1}{w_2^1} + \frac{\partial W^2(\cdot)}{\partial y_2^1} + \lambda^1 - \theta v'\left(\frac{y_2^1}{w_2^1}\right)\frac{1}{w_2^1} + \theta v'\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2 w_2^2 w_2^1}\frac{\partial w_2^2(\cdot)}{\partial l_2^1}$$
$$+\,\theta\left[u'(c_2^2)\frac{\partial c_2^2(\cdot)}{\partial y_2^1} - v'\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2}\frac{\partial y_2^2(\cdot)}{\partial y_2^1}\right] = 0 \tag{A.31}$$

where λ^1 is the multiplier on the first-period budget constraint (4.4), and θ is the multiplier on the incentive-compatibility constraint (4.5). To derive expressions for $\partial W^2(\cdot)/\partial y_1^1$ and $\partial W^2(\cdot)/\partial y_2^1$, note that the Lagrangian corresponding to programme (4.1) – (4.2) can be written as:

$$L = u(c_1^2) - v\left(\frac{y_1^2}{w_1^2}\right) + u(c_2^2) - v\left(\frac{y_2^2}{w_2^2}\right) + \lambda^2\left[y_1^2 - c_1^2 + y_2^2 - c_2^2\right] \tag{A.32}$$

By the Envelope Theorem:

$$\frac{\partial W^2(\cdot)}{\partial y_1^1} = \frac{\partial L(\cdot)}{\partial y_1^1} = v'\left(\frac{y_1^2}{w_1^2}\right)\frac{y_1^2}{w_1^2 w_1^2 w_1^1}\frac{\partial w_1^2(\cdot)}{\partial l_1^1} \tag{A.33}$$

$$\frac{\partial W^2(\cdot)}{\partial y_2^1} = \frac{\partial L(\cdot)}{\partial y_2^1} = v'\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2 w_2^2 w_2^1}\frac{\partial w_2^2(\cdot)}{\partial l_2^1} \tag{A.34}$$

Substituting (A.34) into (A.31) and combining the result with (A.30) yields:

$$(1+\theta)v'\left(\frac{y_2^1}{w_2^1}\right)\frac{1}{w_2^1}-(1+\theta)v'\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2w_2^2w_2^1}\frac{\partial w_2^2(\cdot)}{\partial l_2^1}$$

$$=(1+\theta)u'(c_2^1)+\theta\left[u'(c_2^2)\frac{\partial c_2^2(\cdot)}{\partial y_2^1}-v'\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2}\frac{\partial y_2^2(\cdot)}{\partial y_2^1}\right] \tag{A.35}$$

Dividing both sides of (A.35) by $(1+\theta)u'(c_2^1)$ and rearranging yields:

$$1-\frac{v'(l_2^1)}{u'(c_2^1)w_2^1}+\frac{v'(l_2^2)l_2^2}{u'(c_2^1)w_2^1w_2^2}\frac{\partial w_2^2(\cdot)}{\partial l_2^1}=\frac{-\theta}{(1+\theta)u'(c_2^1)}\left[u'(c_2^2)\frac{\partial c_2^2(\cdot)}{\partial y_2^1}-v'\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2}\frac{\partial y_2^2(\cdot)}{\partial y_2^1}\right] \tag{A.36}$$

Using (2.2), (A.30), (A.23) and (A.24), equation (A.36) can be simplified to:

$$MTR_2^1=\frac{-\theta u'(c_2^2)}{\lambda^1}\left[\frac{\partial c_2^2(\cdot)}{\partial y_2^1}-\frac{\partial y_2^2(\cdot)}{\partial y_2^1}\right] \tag{A.37}$$

We now show that $\frac{\partial c_2^2(\cdot)}{\partial y_2^1}-\frac{\partial y_2^2(\cdot)}{\partial y_2^1}<0$, which establishes that $MTR_2^1>0$. Application of the Implicit Function Theorem to (A.21) – (A.25) yields:

$$\frac{\partial c_2^2(\cdot)}{\partial y_2^1}=\frac{-u''(c_1^2)v''\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2w_1^2}\left[v''\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2}+v'\left(\frac{y_2^2}{w_2^2}\right)\right]\frac{1}{w_2^1w_2^2w_2^2}\frac{\partial w_2^2(\cdot)}{\partial l_2^1}}{\mid A\mid}>0 \tag{A.38}$$

$$\frac{\partial y_2^2(\cdot)}{\partial y_2^1}=\frac{\left[v''\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2}+v'\left(\frac{y_2^2}{w_2^2}\right)\right]\frac{1}{w_2^1w_2^2w_2^2}\frac{\partial w_2^2(\cdot)}{\partial l_2^1}\left[u''(c_1^2)u''(c_2^2)-v''\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2w_1^2}\left[u''(c_1^2)+u''(c_2^2)\right]\right]}{\mid A\mid}>0 \tag{A.39}$$

where A is the Hessian associated with (A.21) – (A.25):

$$A=\begin{bmatrix} u''(c_1^2) & 0 & 0 & 0 & -1\\ 0 & -v''\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2w_1^2} & 0 & 0 & 1\\ 0 & 0 & u''(c_2^2) & 0 & -1\\ 0 & 0 & 0 & -v''\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2w_2^2} & 1\\ -1 & 1 & -1 & 1 & 0 \end{bmatrix} \tag{A.40}$$

and the determinant of A is given by:

$$\mid A \mid = u''(c_1^2)u''(c_2^2)\left[v''\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2 w_1^2} + v''\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2 w_2^2}\right]$$

$$- v''\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2 w_1^2} v''\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2 w_2^2}\left[u''(c_1^2) + u''(c_2^2)\right] > 0 \qquad \text{(A.41)}$$

Therefore, using (A.38) and (A.39):

$$\frac{\partial c_2^2(\cdot)}{\partial y_2^1} - \frac{\partial y_2^2(\cdot)}{\partial y_2^1} = \frac{-\left[v''\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2} + v'\left(\frac{y_2^2}{w_2^2}\right)\right]\frac{u''(c_2^2)}{w_2^1 w_2^2 w_2^2}\frac{\partial w_2^2(\cdot)}{\partial l_2^1}\left[u''(c_1^2) - v''\left(\frac{y_1^2}{w_1^2}\right)\frac{1}{w_1^2 w_1^2}\right]}{\mid A \mid} < 0 \qquad \text{(A.42)}$$

To show that MTR_1^1 is ambiguous, use (A.28), (A.29) and (A.33) to obtain:

$$(1-\theta)u'(c_1^1) = v'\left(\frac{y_1^1}{w_1^1}\right)\frac{1}{w_1^1} - v'\left(\frac{y_1^2}{w_1^2}\right)\frac{y_1^2}{w_1^2 w_1^2 w_1^1}\frac{\partial w_1^2(\cdot)}{\partial l_1^1} - \theta\left[v'\left(\frac{y_1^1}{w_2^1}\right)\frac{1}{w_2^1} - v'\left(\frac{\widehat{y}_2^2}{\widehat{w}_2^2}\right)\frac{\widehat{y}_2^2}{\widehat{w}_2^2\widehat{w}_2^2 w_2^1}\frac{\partial \widehat{w}_2^2(\cdot)}{\partial \widehat{l}_2^1}\right]$$

$$- \theta\left[u'(c_2^2)\frac{\partial c_2^2(\cdot)}{\partial y_1^1} - v'\left(\frac{y_2^2}{w_2^2}\right)\frac{1}{w_2^2}\frac{\partial y_2^2(\cdot)}{\partial y_1^1}\right] + \theta\left[u'(\widehat{c}_2^2)\frac{\partial \widehat{c}_2^2(\cdot)}{\partial y_1^1} - v'\left(\frac{\widehat{y}_2^2}{\widehat{w}_2^2}\right)\frac{1}{\widehat{w}_2^2}\frac{\partial \widehat{y}_2^2(\cdot)}{\partial y_1^1}\right] \qquad \text{(A.43)}$$

Using (2.2), (A.23) and (A.24), equation (A.43) can be manipulated to produce:

$$MTR_1^1 = \theta\left[1 - \frac{1}{u'(c_1^1)}\left[v'\left(\frac{y_1^1}{w_2^1}\right)\frac{1}{w_2^1} - v'\left(\frac{\widehat{y}_2^2}{\widehat{w}_2^2}\right)\frac{\widehat{y}_2^2}{\widehat{w}_2^2\widehat{w}_2^2 w_2^1}\frac{\partial \widehat{w}_2^2(\cdot)}{\partial \widehat{l}_2^1}\right]\right]$$

$$- \frac{\theta u'(c_2^2)}{u'(c_1^1)}\left[\frac{\partial c_2^2(\cdot)}{\partial y_1^1} - \frac{\partial y_2^2(\cdot)}{\partial y_1^1}\right] + \frac{\theta}{u'(c_1^1)}\left[u'(\widehat{c}_2^2)\frac{\partial \widehat{c}_2^2(\cdot)}{\partial y_1^1} - v'\left(\frac{\widehat{y}_2^2}{\widehat{w}_2^2}\right)\frac{1}{\widehat{w}_2^2}\frac{\partial \widehat{y}_2^2(\cdot)}{\partial y_1^1}\right] \qquad \text{(A.44)}$$

The first term on the right-hand side of (A.44) is ambiguous for the same reasons that MTR_1^1 is ambiguous when the government can commit, i.e., $MDL_1^1 - \widehat{MDL_2}^1$ cannot be signed. Using techniques similar to those used to sign $\frac{\partial c_2^2(\cdot)}{\partial y_2^1} - \frac{\partial y_2^2(\cdot)}{\partial y_2^1}$, it can be shown that $\frac{\partial c_2^2(\cdot)}{\partial y_1^1} - \frac{\partial y_2^2(\cdot)}{\partial y_1^1} > 0$ and therefore the second term on the right-hand side of (A.44) is negative. In principle, one could also use these techniques to sign the last term on the right-hand side of (A.44), but this would require determining the comparative statics of a second-best nonlinear income tax problem, which are generally too complex to yield tractable results.[16] Nevertheless, even if the last term in (A.44) could be signed, MTR_1^1 would remain ambiguous. ■

Proof of Proposition 3

When there is pooling in period 1, the second-period optimal tax problem is identical to that in a static model. We therefore omit the proof of the results that $MTR_1^2 > 0$ and $MTR_2^2 = 0$.

The relevant first-order conditions corresponding to programme (4.9) – (4.10) can be written as:

$$2u'(c^1) - \lambda^1 = 0 \tag{A.45}$$

$$-v'\left(\frac{y^1}{w_1^1}\right)\frac{1}{w_1^1} - v'\left(\frac{y^1}{w_2^1}\right)\frac{1}{w_2^1} + \frac{\partial W^2(\cdot)}{\partial y^1} + \lambda^1 = 0 \tag{A.46}$$

where λ^1 is the multiplier on the first-period budget constraint (4.10). To derive an expression for $\partial W^2(\cdot)/\partial y^1$, note that the Lagrangian corresponding to programme (4.6) – (4.8) can be written as:

$$\begin{aligned} L = {} & u(c_1^2) - v\left(\frac{y_1^2}{w_1^2}\right) + u(c_2^2) - v\left(\frac{y_2^2}{w_2^2}\right) \\ & + \lambda^2\left[y_1^2 - c_1^2 + y_2^2 - c_2^2\right] + \theta\left[u(c_2^2) - v\left(\frac{y_2^2}{w_2^2}\right) - u(c_1^2) + v\left(\frac{y_1^2}{w_2^2}\right)\right] \end{aligned} \tag{A.47}$$

where λ^2 is the multiplier on the second-period budget constraint (4.7), and θ is the multiplier on the incentive-compatibility constraint (4.8). By the Envelope Theorem:

$$\begin{aligned} \frac{\partial W^2(\cdot)}{\partial y^1} = \frac{\partial L(\cdot)}{\partial y^1} = {} & v'\left(\frac{y_1^2}{w_1^2}\right)\frac{y_1^2}{w_1^2 w_1^2 w_1^1}\frac{\partial w_1^2(\cdot)}{\partial l_1^1} + v'\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2 w_2^2 w_2^1}\frac{\partial w_2^2(\cdot)}{\partial l_2^1} \\ & + \theta v'\left(\frac{y_2^2}{w_2^2}\right)\frac{y_2^2}{w_2^2 w_2^2 w_2^1}\frac{\partial w_2^2(\cdot)}{\partial l_2^1} - \theta v'\left(\frac{y_1^2}{w_2^2}\right)\frac{y_1^2}{w_2^2 w_2^2 w_2^1}\frac{\partial w_2^2(\cdot)}{\partial l_2^1} \end{aligned} \tag{A.48}$$

[16] Thus far, it has only been possible to derive the comparative static properties of optimal nonlinear income taxes for the case when preferences are quasi-linear. See Weymark [1987] and Brett and Weymark [2008a, 2008b] for the case when preferences are quasi-linear in labour, and Simula [2007] for the case when preferences are quasi-linear in consumption. It is not possible to impose quasi-linearity in our model, since it renders either the first-best or second-best taxation problems indeterminate.

Using equations (A.45), (A.46) and (A.48) we obtain:

$$MRS_1^1 + MRS_2^1 = 2 + \frac{\theta}{u'(c^1)w_2^2 w_2^2 w_2^1} \frac{\partial w_2^2(\cdot)}{\partial l_2^1} \left[v'\left(\frac{y_2^2}{w_2^2}\right) y_2^2 - v'\left(\frac{y_1^2}{w_2^2}\right) y_1^2 \right] \tag{A.49}$$

where:

$$MRS_i^1 := \frac{v'(l_i^1)}{u'(c^1)w_i^1} - \frac{v'(l_i^2)l_i^2}{u'(c^1)w_i^1 w_i^2} \frac{\partial w_i^2(\cdot)}{\partial l_i^1} \tag{A.50}$$

denotes consumer i's marginal rate of substitution of y^1 for c^1. Using (2.2) we know that $MTR_i^1 = 1 - MRS_i^1$. Moreover, since $v(\cdot)$ is strictly convex and $y_2^2 > y_1^2$,[17] the last term in (A.49) is positive, which implies that $MRS_1^1 + MRS_2^1 > 2$. This establishes that $MTR_1^1 + MTR_2^1 < 0$. However, the signs of MTR_1^1 and MTR_2^1 are ambiguous, since (A.49) provides no information as to whether each consumer's marginal rate of substitution is greater than, equal to, or less than one. ■

[17] Since the second-period optimal tax problem is identical to that in a static model, $y_2^2 > y_1^2$ follows from the well-known result that it is optimal for the high-skill type to earn a higher pre-tax income than the low-skill type in static models.

References

[1] Albanesi, S. and C. Sleet [2006], "Dynamic Optimal Taxation with Private Information", *Review of Economic Studies*, 73, 1-30.

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

[3] Auerbach, A [2006], "Budget Windows, Sunsets, and Fiscal Control", *Journal of Public Economics*, 90, 87-100.

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

[5] Boadway, R. and L. Jacquet [2006], "Optimal Marginal and Average Income Taxation under Maxi-min", *Journal of Economic Theory*, forthcoming.

[6] Brett, C. and J. Weymark [2008a], "Public Good Provision and the Comparative Statics of Optimal Nonlinear Income Taxation", *International Economic Review*, 49, 255-290.

[7] Brett, C. and J. Weymark [2008b], "The Impact of Changing Skill Levels on Optimal Nonlinear Income Taxes", *Journal of Public Economics*, 92, 1765-1771.

[8] Brett, C. and J. Weymark [2008c], "Optimal Nonlinear Taxation of Income and Savings without Commitment", *Vanderbilt University Working Paper*, 08-W05.

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

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

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

[12] Kocherlakota, N [2005], "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation", *Econometrica*, 73, 1587-1621.

[13] Martin, P. and C. Rogers [2000], "Optimal Stabilization Policy in the Presence of Learning by Doing", *Journal of Public Economic Theory*, 2, 213-241.

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

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

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

[17] Simula, L [2007], "Optimality Conditions and Comparative Static Properties of Nonlinear Income Taxes Revisited", *Paris School of Economics*, WP2007-15.

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

[19] Werning, I [2007], "Optimal Fiscal Policy with Redistribution", *Quarterly Journal of Economics*, 122, 925-967.

[20] Weymark, J [1987], "Comparative Static Properties of Optimal Nonlinear Income Taxes", *Econometrica*, 55, 1165-1185.