

Optimal Nonlinear Taxation of Income and Education Expenditures

Jang-Ting Guo*

University of California, Riverside

Alan Krause†

University of York

4 April 2011

Abstract

Despite using a variety of models and assumptions, the existing literature has overwhelmingly concluded that education policy should be regressive. In this paper, we examine a two-period model in which the government may impose nonlinear taxes on both labour income and education expenditures. Individuals undertake education in the first period to increase their second-period wages. Our main result is that optimal education policy in our model is progressive. Specifically, if the government can commit, it is optimal for high-skill individuals to face a zero marginal tax rate on their education expenditures, while that for low-skill individuals is negative. If the government cannot commit, the optimal marginal tax rate on education expenditures by high-skill individuals is positive, while that for low-skill individuals remains negative. Optimal education policy is therefore more progressive when the government cannot commit.

Keywords: education policy; dynamic nonlinear taxation; commitment.

JEL Classifications: H21, H24.

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

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

1 Introduction

A fundamental question in public economics is how government policy might improve upon the allocation achieved by the free market, which tends to be efficient but unfair. To a large extent, this question has been addressed using the Mirrlees [1971] and Stiglitz [1982] models of nonlinear income taxation. In the Mirrlees/Stiglitz model, inequality arises from differences in individuals' skill levels, which results in differences in their income-earning abilities. However, the government cannot implement (the first-best) personalised lump-sum taxation based on skill levels as the Second Welfare Theorem would recommend, owing to the assumption that each individual's skill level is private information. Instead, the government is constrained to use (the second-best) incentive-compatible nonlinear income taxation, in which optimal taxation involves some efficiency loss due to the necessity of inducing marginal distortions to achieve redistribution.

The nonlinear income taxation literature typically assumes that skill levels are fixed. In reality though, skill levels can change over time, in particular by individuals investing in education. This observation immediately suggests that education policy may play an important role alongside income taxation policy, and a large literature has arisen exploring their relationship (reviewed below). Despite using different models and making a variety of assumptions, the literature has overwhelmingly concluded that education should be subsidised and/or education policy should be *regressive*.¹ That is, policy should seek to encourage higher-skill individuals to acquire education relatively more than lower-skill individuals.

In this paper, we revisit optimal income taxation and education policy, but with two key features that together distinguish our analysis from the existing literature: (i) the government may impose nonlinear taxation on both labour income and education expenditures,² and (ii) our model is dynamic, and we focus on the possibility that the government cannot commit to its future tax policy. Specifically, in the two-period model

¹A uniform education subsidy is often regressive in its effect, since higher-skill individuals tend to acquire more education than lower-skill individuals. Also, public provision of education can be equivalent to subsidising private investments in education.

²For convenience, we will continue to refer to the *taxation* of income and education, where it is understood that a negative tax is a subsidy.

that we employ,³ the government may take advantage of skill-type information revealed in period 1 to implement lump-sum taxes in period 2. We allow the government to impose nonlinear taxes on education expenditures because the usual argument made against nonlinear taxation of commodities, based on the possibility of retrade opportunities, is weak for the case of education as it is not easily tradable. Therefore, restricting attention to linear taxation of education expenditures would seem difficult to justify. We consider a multi-period model because the education process is intrinsically dynamic, i.e., current investments in education lead to higher skills and wages in the future. Dynamic models of taxation raise the issue of whether the government can commit to its future tax policy. We consider optimal taxation with and without commitment, because an argument can be made for each case.⁴ That said, we think the no-commitment case is more relevant, because the second-best nature of the Mirrlees/Stiglitz model stems entirely from the assumption that the government cannot observe each individual's skill type. This is no longer true in period 2 of our model, however, if individuals revealed their types in period 1. To ignore the no-commitment case would be to ignore an essential feature of the Mirrlees/Stiglitz approach to optimal taxation—that no ad hoc constraints be placed on the nature of the optimal tax function, and that the tax instruments available to the government be constrained only by the information structure. Therefore, analysis of the no-commitment case is the main focus and contribution of our paper.

Our main result is that optimal education policy in our model is *progressive*. When the government can commit, low-skill individuals have their education levels distorted upwards via a negative marginal tax rate, while high-skill individuals face a zero marginal tax rate on their education expenditures. Distorting the low-skill type's education upwards facilitates redistribution by increasing their second-period wage, but may also undermine redistribution by creating an incentive for high-skill individuals to mimic, i.e., pretend to be low skill. However, since low-skill individuals work longer than do mimick-

³Since the optimal nonlinear tax problem without commitment is most easily solved by backwards induction, we use a finite-horizon model.

⁴One could argue in favour of the commitment assumption on the basis that real-world tax systems are not frequently redesigned. Gaube [2007], for example, makes this argument. On the other hand, the commitment assumption has been criticised as being unrealistic, since the present government cannot easily impose binding constraints on the tax policies of future governments.

ing high-skill individuals, the benefits of distorting the low-skill type's education upwards outweigh the costs. This argument also helps justify a negative marginal tax rate on education expenditures by low-skill individuals when the government cannot commit, but importantly there is now an additional force working in its favour. Under lump-sum taxation in period 2, an individual's utility is *decreasing* in their wage rate. Thus distorting the low-skill type's education upwards decreases their second-period utility, and correspondingly increases that of high-skill individuals. This facilitates redistributive income taxation by making mimicking less attractive. Analogously, it is optimal to impose a positive marginal tax rate on education expenditures by high-skill individuals when the government cannot commit, since this increases their second-period utility and decreases that of low-skill individuals. This again makes redistributive income taxation more effective by making mimicking less attractive. Optimal education policy in our model is therefore progressive, and it is more progressive when the government cannot commit. A progressive education policy is optimal in our model because it mitigates the information constraint faced by the government and makes nonlinear income taxation more efficient. Finally, we quantitatively verify the progressive nature of the optimal education policy with a numerical example.

The remainder of the paper is organised as follows. Section 2 reviews the related literature, while Section 3 outlines the analytical framework that we consider. Section 4 examines optimal taxation with commitment, while Section 5 examines optimal taxation without commitment. Section 6 presents a numerical example to further illustrate the progressive nature of optimal education policy in our model. Section 7 concludes, while proofs and some other mathematical details are relegated to an appendix.

2 Related Literature

The literature on the relationship between education policy and optimal income taxation is quite extensive. An early study by Ulph [1977] examines how the government should distribute a given amount of education resources amongst a population of individuals who differ in their skill levels, when the government can also use redistributive

income taxation. He concludes that education policy should be regressive, in that higher-skill individuals should receive more education than lower-skill individuals. Hare and Ulph [1979] examine education policy when education may be provided publicly or privately. They find that public provision is desirable when the government cannot implement lump-sum taxes, but is constrained instead to use (distortionary) income taxation. Grout [1983] also examines public versus private provision of education. He concludes that previous arguments made against a free market in education are overstated once imperfect information on part of the government is taken into account.

Tuomala [1986] and Brett and Weymark [2003] investigate how individuals' educational choices affect the design of an optimal nonlinear income tax system. They find that some standard results, such as the optimality of zero marginal tax rates at the endpoints of the tax schedule, may no longer hold. Boadway, et al. [1996] argue that education should be subsidised to offset underinvestment in education by individuals, which arises if the government cannot commit to a future redistributive tax system. That is, education subsidies can alleviate the "hold-up" problem associated with investments in education. Andersson and Konrad [2003] show that the argument for education subsidies based on the hold-up problem is weakened if individuals are internationally mobile. Krause [2009a] also examines a model in which individuals can be internationally mobile, but concludes that optimal education policy is ambiguous in that it can be optimal to distort education levels upwards, downwards, or not at all.

Fleurbaey, et al. [2002] examine education policy in a model in which each individual's cost of acquiring education is private information. They show that an increase in the government's aversion to inequality leads to more inequality in educational achievements, even though education policy *per se* is progressive. De Fraja [2002] examines a model in which households are distinguished by their incomes and the abilities of their children. Households therefore differ in both their ability to pay for education and their ability to benefit from education. In this setting, De Fraja [2002] shows that the optimal education policy is regressive.

Bovenberg and Jacobs [2005] argue that education subsidies and redistributive taxation are "Siamese twins", in that education subsidies offset the distortions created by

redistributive income taxation. Krause [2006] examines redistributive taxation and education policy in a model in which education is skill specific, with low-skill individuals benefiting from vocational training and high-skill individuals benefiting from professional education. He shows that it is optimal to distort professional education upwards or vocational training downwards relative to their first-best levels. The paper most closely related to ours is that by Maldonado [2008]. Maldonado examines a static model (as does most of the literature) and therefore cannot consider the no-commitment case, but like us he allows the government to impose nonlinear taxes on labour income and education expenditures. He shows that high-skill individuals should face a zero marginal tax rate on their education expenditures, while the sign of the optimal marginal tax rate applicable to education expenditures by low-skill individuals is generally ambiguous. Blumkin and Sadka [2008] show, contrary to most of the literature, that a case can be made for taxing education. Starting with an optimal linear or nonlinear income tax system, they show that introducing a small tax on education is desirable provided the correlation between innate skill and learning ability is high, and the social welfare function is sufficiently egalitarian.

As our model is dynamic, our paper is also related to the “new dynamic public finance” literature which extends the static Mirrlees [1971] nonlinear income tax model to a dynamic setting. Surveys of this literature are provided by Golosov, et al. [2006, 2010], while a textbook treatment is provided by Kocherlakota [2010]. For the most part, this literature has assumed that the government can commit to its future tax policy. Models of dynamic nonlinear income taxation without commitment that are similar to the one we employ have been used by Apps and Rees [2006], Brett and Weymark [2008], Krause [2009b], and Guo and Krause [2010], among others. However, none of these papers examine education policy.

3 Analytical Framework

There is a unit measure of individuals in the economy who live for two periods, with a proportion $\phi \in (0, 1)$ being high-skill workers and the remaining $(1 - \phi)$ being low-

skill workers. First-period wages are fixed and are denoted by w_i^1 , where $i \in \{H, L\}$ represents the high-skill and low-skill types, respectively. Second-period wages depend on education and are denoted by $w_i^2 = w_i^1 + \alpha e_i$, where $\alpha > 0$ and e_i is the amount of education acquired by a type i individual in period 1. The assumption that second-period wages are a linear function of education is made for simplicity, but also because there does not seem to be any *a priori* reason to assume that $\partial w_i^2(\cdot)/\partial e_i$ differs by type if individuals undertake education appropriate to their needs. For example, it seems reasonable to expect that low-skill workers will undertake additional training in, say, the use of computers, which would increase their skills considerably but be of little benefit to high-skill workers (who would already have these skills). Similarly, high-skill workers would benefit from additional training in, say, managerial decision making, but such training would be inaccessible to low-skill workers.⁵ Accordingly, there does not seem to be any particular reason to assume that $\partial w_i^2(\cdot)/\partial e_i$ is greater for one type of individual than the other, thus we postulate that they are the same.⁶ It is assumed that the government can observe the level of an individual's education expenditures, but not its composition. Therefore, taxation cannot be contingent upon the type of education an individual undertakes. This assumption is analogous to that regarding income in the Mirrlees/Stiglitz model, in which the government can observe an individual's income level, but not its composition. It is also assumed that $w_H^1 > w_L^1$ and $w_H^2 > w_L^2$ for all relevant levels of education, which simplifies the analysis considerably by ruling out the possibility of skill reversals.

In period 1, individuals spend their non-leisure time working and acquiring education to increase their second-period wage. Both types of individual have the same preferences, which are represented by the utility function $u(c_i^1) - v(l_i^1) - z(e_i)$, where c_i^1 and l_i^1 represent type i 's first-period consumption and labour supply, respectively. The function $u(\cdot)$ is increasing and strictly concave, while $v(\cdot)$ and $z(\cdot)$ are increasing and strictly

⁵It should be noted that our formulation of the relationship between education and wages is not inconsistent with the common assumption that higher-skill individuals are more able to benefit from education than lower-skill individuals. As our example suggests, it is possible that only higher-skill individuals may benefit from additional training in managerial decision making.

⁶Nevertheless, in the concluding section we discuss the implications of relaxing this assumption.

convex.⁷ Since period 2 is the last period, no one acquires education in period 2. Second-period utility is therefore $u(c_i^2) - v(l_i^2)$, where c_i^2 and l_i^2 represent type i 's second-period consumption and labour, respectively. Finally, type i 's pre-tax income in period t is equal to $y_i^t = w_i^t l_i^t$, and to isolate the dynamic effects of education and the (possible) revelation and use of skill-type information, we assume there are no savings.⁸

It is assumed that the government can impose nonlinear taxation on both income and education expenditures. In order to derive expressions for the implicit marginal tax rates applicable to income and education, suppose the individuals faced smooth nonlinear tax functions $T^1(y_i^1, e_i)$ in period 1 and $T^2(y_i^2)$ in period 2. Each individual would then solve the following programme:

$$\max_{c_i^1, l_i^1, e_i, c_i^2, l_i^2} u(c_i^1) - v(l_i^1) - z(e_i) + \delta [u(c_i^2) - v(l_i^2)] \quad (3.1)$$

subject to:

$$c_i^1 + e_i \leq y_i^1 - T^1(y_i^1, e_i) \quad (3.2)$$

$$c_i^2 \leq y_i^2 - T^2(y_i^2) \quad (3.3)$$

where (3.1) is lifetime utility, (3.2) and (3.3) are first- and second-period budget constraints, respectively, and $\delta \in (0, 1]$ is the discount factor.

It is shown in the Appendix that the solution to programme (3.1) – (3.3) yields the following expressions for the marginal tax rates:

$$MTRI_i^t := \frac{\partial T^t(\cdot)}{\partial y_i^t} = 1 - \frac{v'(l_i^t)}{u'(c_i^t)w_i^t} \quad (3.4)$$

$$MTR E_i := \frac{\partial T^1(\cdot)}{\partial e_i} = \frac{\alpha \delta l_i^2 v'(l_i^2)}{u'(c_i^1)w_i^2} - \frac{z'(e_i)}{u'(c_i^1)} - 1 \quad (3.5)$$

where $MTRI_i^t$ is the marginal income tax rate faced by type i individuals in period t ,

⁷By assuming that $z(\cdot)$ is increasing and strictly convex, we are implicitly assuming that there is disutility in acquiring education in much the same way as there is disutility in working. It would make no difference to our main conclusions if instead utility was assumed to be increasing in education acquired, say due to the existence of non-pecuniary benefits from education.

⁸Brett and Weymark [2008] examine the optimal nonlinear taxation of savings in a dynamic Mirrlees/Stiglitz model without commitment.

and $MTR E_i$ is the marginal tax rate applicable to education expenditures by type i individuals in period 1. Equation (3.4) is the same as the expression obtained in static models, that the marginal income tax rate is equal to one minus the marginal rate of substitution of pre-tax income for consumption (see, e.g., Stiglitz [1982]). Equation (3.5) is the expression for the marginal tax rate on education. The first term on the right-hand side of (3.5) represents the marginal benefit of an additional unit of education, obtained via higher second-period wages, divided by the marginal utility of first-period consumption. The second term on the right-hand side of (3.5) represents the marginal cost of an additional unit of education, which reflects the effort required to obtain education, divided by the marginal utility of first-period consumption.

4 Optimal Nonlinear Taxation with Commitment

While analysis of the no-commitment case is the main focus and contribution of our paper, we begin with the commitment case to provide a benchmark. If the government can commit to its second-period tax policy, it cannot use skill-type information revealed in period 1 to implement lump-sum taxes in period 2, even though in period 2 it would be optimal to do so. Therefore, when the government can commit, it can be described as choosing lifetime tax treatments $\langle c_L^1, y_L^1, e_L, c_L^2, y_L^2 \rangle$ and $\langle c_H^1, y_H^1, e_H, c_H^2, y_H^2 \rangle$ for the low-skill and high-skill types, respectively, to maximise:⁹

$$(1 - \phi) \left[u(c_L^1) - v \left(\frac{y_L^1}{w_L^1} \right) - z(e_L) \right] + \phi \left[u(c_H^1) - v \left(\frac{y_H^1}{w_H^1} \right) - z(e_H) \right] \\ + \delta(1 - \phi) \left[u(c_L^2) - v \left(\frac{y_L^2}{w_L^2(e_L)} \right) \right] + \delta\phi \left[u(c_H^2) - v \left(\frac{y_H^2}{w_H^2(e_H)} \right) \right] \quad (4.1)$$

subject to:

$$(1 - \phi) [y_L^1 - c_L^1 - e_L] + \phi [y_H^1 - c_H^1 - e_H] \geq 0 \quad (4.2)$$

$$(1 - \phi) [y_L^2 - c_L^2] + \phi [y_H^2 - c_H^2] \geq 0 \quad (4.3)$$

⁹While we do not observe such tax treatments in practice, the Revelation Principle implies that one can model the government as choosing each type's allocation directly subject to budget and incentive-compatibility constraints, rather than as specifying nonlinear tax schedules.

$$\begin{aligned}
& u(c_H^1) - v\left(\frac{y_H^1}{w_H^1}\right) - z(e_H) + \delta \left[u(c_H^2) - v\left(\frac{y_H^2}{w_H^2(e_H)}\right) \right] \geq \\
& u(c_L^1) - v\left(\frac{y_L^1}{w_H^1}\right) - z(e_L) + \delta \left[u(c_L^2) - v\left(\frac{y_L^2}{\widehat{w}_H^2(e_L)}\right) \right]
\end{aligned} \tag{4.4}$$

where the objective function (4.1) is a utilitarian social welfare function, (4.2) and (4.3) are the government's first- and second-period budget constraints,¹⁰ and (4.4) is the high-skill type's incentive-compatibility constraint. Following the standard practice, we assume that the government knows the distribution of skill types, but each individual's type is private information. The government must therefore satisfy incentive-compatibility constraints to ensure that each individual chooses their intended tax treatment, rather than mimicking the other type by choosing the other type's tax treatment. However, we omit the low-skill type's incentive-compatibility constraint, as we focus on what Stiglitz [1982] calls the "normal" case and what Guesnerie [1995] calls "redistributive equilibria". That is, with the postulated parameter values under consideration, we assume that the government will be using its taxation powers to redistribute from high-skill to low-skill individuals, which creates an incentive for high-skill individuals to mimic low-skill individuals, but not vice versa. Accordingly, the high-skill type's incentive-compatibility constraint will bind at an optimum, whereas the low-skill type's incentive-compatibility constraint will be slack. Finally, we use $\widehat{w}_H^2 = w_H^1 + \alpha e_L$ to denote a high-skill individual's second-period wage when mimicking.

It is shown in the Appendix that the solution to programme (4.1) – (4.4) yields:

Proposition 1 *Optimal nonlinear taxation of income and education expenditures when the government can commit to its second-period tax policy has: $MTRI_L^1 > 0$, $MTRI_H^1 = 0$, $MTRI_L^2 > 0$, $MTRI_H^2 = 0$, $MTR E_L < 0$, and $MTR E_H = 0$.*

The findings that it is optimal for low-skill individuals to face positive marginal income tax rates in both periods, and for high-skill individuals to face zero marginal income tax rates in both periods, are nothing but the standard "no-distortion-at-the-top" and "downward-distortion-at-the-bottom" results that typify static nonlinear income tax models. Likewise, the finding that the optimal marginal tax rate applicable to education

¹⁰As with the individuals, we assume that the government cannot save or borrow.

expenditures by high-skill individuals is zero is another example of the no-distortion-at-the-top result. However, it is optimal for low-skill individuals to face a negative marginal tax rate on their education expenditures. It follows that the optimal education policy is *progressive*. Since low-skill individuals have a lower wage than high-skill individuals, the former must work longer to earn any given level of pre-tax income. This means that low-skill individuals work longer than does a mimicking high-skill individual. Therefore, low-skill individuals benefit more from an increase in their second-period wage via more education than does a mimicking high-skill individual. The government can therefore make low-skill individuals better-off by distorting their education levels upwards via a negative marginal tax rate, without unduly enticing high-skill individuals to mimic.

Under the assumption that the government can commit, our results are similar to those obtained in a static model by Maldonado [2008], and can be interpreted as applications of the Atkinson and Stiglitz [1976], Mirrlees [1976], and Christiansen [1984], among others, arguments for supplementing income taxation with commodity taxation. The usefulness of commodity taxation in this literature follows from its ability to reduce the distortions associated with redistributive income taxation, and the same logic applies when the commodity happens to be education. Since low-skill individuals work longer and benefit more from education than does a mimicking high-skill individual, education and labour are in some sense complements. Following the arguments in the literature for subsidising commodities that are complements with labour, it is therefore optimal to apply a negative marginal tax rate to the low-skill type's education expenditures.

5 Optimal Nonlinear Taxation without Commitment

If the government cannot commit to its second-period tax policy, it will take advantage of skill-type information revealed in period 1 to implement the first-best personalised lump-sum taxes in period 2. In this case, the government's behaviour in period 2 can be described as follows. Choose tax treatments $\langle c_L^2, y_L^2 \rangle$ and $\langle c_H^2, y_H^2 \rangle$ for the low-skill and

high-skill types, respectively, to maximise:

$$(1 - \phi) \left[u(c_L^2) - v \left(\frac{y_L^2}{w_L^2(e_L)} \right) \right] + \phi \left[u(c_H^2) - v \left(\frac{y_H^2}{w_H^2(e_H)} \right) \right] \quad (5.1)$$

subject to:

$$(1 - \phi) [y_L^2 - c_L^2] + \phi [y_H^2 - c_H^2] \geq 0 \quad (5.2)$$

where (5.1) is second-period social welfare, and (5.2) is the government's second-period budget constraint. Since in period 2 the government knows each individual's type, low-skill individuals must accept $\langle c_L^2, y_L^2 \rangle$ and high-skill individuals must accept $\langle c_H^2, y_H^2 \rangle$, i.e., the tax treatments need not satisfy incentive-compatibility constraints. The solution to programme (5.1) – (5.2) yields the functions $c_L^2(\phi, w_L^2(e_L), w_H^2(e_H))$, $y_L^2(\cdot)$, $c_H^2(\cdot)$ and $y_H^2(\cdot)$. Substituting these functions into (5.1) yields the value function $W^2(\cdot)$.

All individuals know that if they reveal their type in period 1, the government will solve programme (5.1) – (5.2) in period 2. The government's behaviour in period 1 can therefore be described as follows. Choose tax treatments $\langle c_L^1, y_L^1, e_L \rangle$ and $\langle c_H^1, y_H^1, e_H \rangle$ for the low-skill and high-skill types, respectively, to maximise:

$$(1 - \phi) \left[u(c_L^1) - v \left(\frac{y_L^1}{w_L^1} \right) - z(e_L) \right] + \phi \left[u(c_H^1) - v \left(\frac{y_H^1}{w_H^1} \right) - z(e_H) \right] + \delta W^2(\cdot) \quad (5.3)$$

subject to:

$$(1 - \phi) [y_L^1 - c_L^1 - e_L] + \phi [y_H^1 - c_H^1 - e_H] \geq 0 \quad (5.4)$$

$$\begin{aligned} u(c_H^1) - v \left(\frac{y_H^1}{w_H^1} \right) - z(e_H) + \delta \left[u(c_H^2(\cdot)) - v \left(\frac{y_H^2(\cdot)}{w_H^2(e_H)} \right) \right] \geq \\ u(c_L^1) - v \left(\frac{y_L^1}{w_H^1} \right) - z(e_L) + \delta \left[u(c_L^2(\cdot)) - v \left(\frac{y_L^2(\cdot)}{\widehat{w}_H^2(e_L)} \right) \right] \end{aligned} \quad (5.5)$$

where (5.3) is first-period social welfare plus the discounted value of second-period social welfare, (5.4) is the government's first-period budget constraint, and (5.5) is the high-skill type's incentive-compatibility constraint. Since the levels of education undertaken in period 1 affect the level of social welfare attainable in period 2, the government takes this into account when designing the first-period tax system. Therefore, the objective

function (5.3) includes the second-period value function. As in the commitment case, we assume that the redistributive goals of the government imply that the high-skill type's incentive-compatibility constraint binds at an optimum, whereas the low-skill type's incentive-compatibility constraint is slack.¹¹ In order for high-skill individuals to be willing to reveal their type, the utility they obtain from choosing $\langle c_H^1, y_H^1, e_H \rangle$ in period 1 and thus revealing their type, plus the utility they obtain from the first-best tax treatment $\langle c_H^2(\cdot), y_H^2(\cdot) \rangle$ they will face in period 2, must be greater than or equal to the utility they could obtain by pretending to be low skill. In particular, if a high-skill individual pretends to be low skill by choosing $\langle c_L^1, y_L^1, e_L \rangle$ in period 1, they will be treated as such by the government in period 2, i.e., they will receive the low-skill type's tax treatment $\langle c_L^2(\cdot), y_L^2(\cdot) \rangle$. Finally, note that the second-period tax treatments depend in part upon the levels of education undertaken in period 1.

It is shown in the Appendix that the solutions to programmes (5.1) – (5.2) and (5.3) – (5.5) together imply:

Proposition 2 *Optimal nonlinear taxation of income and education expenditures when the government cannot commit to its second-period tax policy has: $MTRI_L^1 > 0$, $MTRI_H^1 = 0$, $MTRI_L^2 = 0$, $MTRI_H^2 = 0$, $MTR E_L < 0$, and $MTR E_H > 0$.*

As in the commitment case, in period 1 the standard pattern of optimal marginal income tax rates applies. Both types face zero marginal income tax rates in period 2 simply because first-best taxation is used in that period. Low-skill individuals continue to face a negative marginal tax rate on their education expenditures, but that for high-skill individuals is now positive. First-best taxation in period 2 involves both types receiving the same level of consumption, but high-skill individuals are required to work longer.¹² Therefore, given that both types have the same preferences, high-skill individuals obtain less utility than low-skill individuals in period 2. Indeed, individual utility is

¹¹Indeed, since high-skill individuals know that they will face first-best taxation in the second period if they reveal their type in the first period, their incentive to mimic low-skill individuals is stronger than when the government can commit. This creates the possibility that it may be optimal in period 1 to pool some or all of the high-skill individuals with the low-skill individuals. However, we focus on the more economically interesting case in which high-skill and low-skill individuals receive different tax treatments.

¹²This has led some to describe first-best income taxation as Marxist in nature, because it takes from each individual according to their ability and gives to each individual according to their need.

generally decreasing in the wage rate under first-best taxation, since everyone receives the same level of consumption but higher-wage individuals must work longer. By imposing a positive marginal tax rate on education expenditures by high-skill individuals, the government is distorting their education and second-period wage downwards, but in fact raising their second-period utility. Concurrently, this reduces the gap between low-skill and high-skill wages, and reduces the low-skill type's second-period utility. The increase in the high-skill type's second-period utility and decrease in the low-skill type's second-period utility make mimicking by high-skill individuals less attractive. The end result of imposing a positive marginal tax rate on the high-skill type's education expenditures is an increase in social welfare, obtained by relaxing the incentive-compatibility constraint.

Low-skill individuals face a negative marginal tax rate on their education expenditures for the same reason as when the government can commit, but now there is an additional factor at work. Imposing a negative marginal tax rate on education expenditures by low-skill individuals distorts their education and second-period wage upwards. By an analogous argument made above for the positive marginal tax rate on the high-skill type's education expenditures, a negative marginal tax rate on education expenditures by low-skill individuals decreases their utility under first-best taxation in period 2, and increases that for high-skill individuals. This again makes mimicking less attractive, and increases social welfare by relaxing the incentive-compatibility constraint. Overall, the optimal education policy is more progressive than when the government can commit.

6 A Numerical Example

To further illustrate the progressive nature of optimal education policy in our model, we present a numerical example. To this end, the utility function is assumed to take the following form:

$$\ln(c_i^1) - \frac{1}{1+\gamma}(l_i^1)^{1+\gamma} - \frac{1}{1+\gamma}(e_i)^{1+\gamma} + \delta \left[\ln(c_i^2) - \frac{1}{1+\gamma}(l_i^2)^{1+\gamma} \right] \quad (6.1)$$

where $\gamma > 0$ is a preference parameter. For our example we set $\gamma = 2$, as this implies a labour supply elasticity of 0.5 (see Appendix) which is in line with empirical estimates. For simplicity, we assume that the disutility involved in acquiring education is the same as the disutility from working.

The remaining parameters of our model are chosen on the following basis. The OECD [2010] reports that on average across OECD countries, approximately one-quarter of all adults have attained tertiary level education. We therefore assume that 25% of individuals are high-skill workers, i.e., we set $\phi = 0.25$. We assume an annual discount rate of 5%, which is consistent with common practice. Since most individuals work for around 40 years of their lives, we take each period to be 20 years in length. An annual discount rate of 5% then corresponds to a 20-year discount factor of $\delta = 0.38$. Fang [2006] and Goldin and Katz [2007] estimate that the college wage premium, i.e., the average difference in the wages of university graduates over high-school graduates, is approximately 60%. We therefore normalise the low-skill type's first-period wage to equal 10, and correspondingly set the high-skill type's first-period wage equal to 16. The parameter which determines the productivity of education, α , is arbitrarily set equal to 5, which ensures an interior solution and rules out skill reversals.

TABLE 1

Numerical Example

| | | | | | | |
|--------------------------------|------|----------|-------|---------|-------------|-------------|
| ϕ | 0.25 | δ | 0.38 | w_L^1 | 10.00 | |
| γ | 2.00 | α | 5.00 | w_H^1 | 16.00 | |
| | | | e_L | e_H | MTR_{E_L} | MTR_{E_H} |
| First-best | | | 0.222 | 0.290 | 0.000 | 0.000 |
| Second-best with commitment | | | 0.247 | 0.231 | -0.117 | 0.000 |
| Second-best without commitment | | | 0.261 | 0.205 | -0.324 | 0.928 |

The results of our numerical example are presented in Table 1. It can be seen that the low-skill type's education is distorted upwards, and the high-skill type's education is

distorted downwards, relative to their first-best levels, and these distortions are greater when the government cannot commit. The intuition for this outcome, as discussed earlier, is that in the no-commitment case the government has additional reasons to implement a progressive education policy.

7 Final Remarks

This paper has analysed a stylised two-period variant of the Mirrlees/Stiglitz model in which the government may impose nonlinear taxes on both labour income and education expenditures. Given that our model is dynamic, we have considered the possibility that the government may not be able to commit to its future tax policy. However, in both the commitment and no-commitment cases, optimal taxation of education expenditures is *progressive*, and it is more progressive when the government cannot commit.

For simplicity, we have assumed that second-period wages are a linear function of education, $w_i^2 = w_i^1 + \alpha e_i$, because for the reasons discussed earlier there does not seem to be any *a priori* reason to assume that $\partial w_i^2(\cdot)/\partial e_i$ is greater for one type of individual than the other. This assumption is in the same spirit as the standard assumption that both types have the same preferences. That said, one may wonder how our results change if a more general wage function, $w_i^2 = w_i^2(e_i)$, is assumed. As it turns out, assuming a general wage function has no effect on our results regarding the optimal marginal tax rates applicable to the high-skill type's education expenditures. That is, they still face a zero marginal tax rate when the government can commit, and a positive marginal tax rate when the government cannot commit. Our results that the low-skill type should face a negative marginal tax rate on their education expenditures when the government can and cannot commit are also unaffected provided that $\partial w_L^2(\cdot)/\partial e_L \geq \partial \hat{w}_H^2(\cdot)/\partial e_L$, but otherwise they cannot be signed. This is because if $\partial w_L^2(\cdot)/\partial e_L < \partial \hat{w}_H^2(\cdot)/\partial e_L$, our argument that low-skill individuals benefit more from additional education than does a mimicking high-skill individual may no longer be valid.

We have also assumed that $w_H^2 > w_L^2$ for all relevant levels of education, which rules out the possibility of skill reversals. Allowing low-skill individuals to become high skill

through investments in education would significantly complicate the analysis, because it would no longer be clear as to which type's incentive-compatibility constraint binds. Nevertheless, skill reversals are a real-world possibility, and therefore present an interesting avenue for future research.

8 Appendix

Derivation of Equations (3.4) and (3.5)

The relevant first-order conditions corresponding to programme (3.1) – (3.3) can be written as:

$$u'(c_i^1) - \lambda_i^1 = 0 \quad (\text{A.1})$$

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

$$-z'(e_i) - \lambda_i^1 \left[1 + \frac{\partial T^1(\cdot)}{\partial e_i} \right] + \lambda_i^2 \alpha l_i^2 \left[1 - \frac{\partial T^2(\cdot)}{\partial y_i^2} \right] = 0 \quad (\text{A.3})$$

$$\delta u'(c_i^2) - \lambda_i^2 = 0 \quad (\text{A.4})$$

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

where $\lambda_i^1 \geq 0$ and $\lambda_i^2 \geq 0$ are the multipliers on individual i 's budget constraints (3.2) and (3.3), respectively. Straightforward manipulation of equations (A.1) and (A.2), and of (A.4) and (A.5), yields equation (3.4). Likewise, substituting (A.5) into (A.3), and combining the result with (A.1), yields equation (3.5). ■

Proof of Proposition 1

The relevant first-order conditions corresponding to programme (4.1) – (4.4) can be written as:

$$(1 - \phi - \theta_H)u'(c_L^1) - \lambda^1(1 - \phi) = 0 \quad (\text{A.6})$$

$$-(1 - \phi)v' \left(\frac{y_L^1}{w_L^1} \right) \frac{1}{w_L^1} + \lambda^1(1 - \phi) + \theta_H v' \left(\frac{y_L^1}{w_H^1} \right) \frac{1}{w_H^1} = 0 \quad (\text{A.7})$$

$$-(1 - \phi - \theta_H)z'(e_L) - \lambda^1(1 - \phi) + \frac{\delta(1 - \phi)\alpha y_L^2}{w_L^2(e_L)w_L^2(e_L)} v' \left(\frac{y_L^2}{w_L^2(e_L)} \right) - \frac{\delta\theta_H \alpha y_L^2}{\hat{w}_H^2(e_L)\hat{w}_H^2(e_L)} v' \left(\frac{y_L^2}{\hat{w}_H^2(e_L)} \right) = 0 \quad (\text{A.8})$$

$$(\phi + \theta_H)u'(c_H^1) - \lambda^1\phi = 0 \quad (\text{A.9})$$

$$-(\phi + \theta_H)v' \left(\frac{y_H^1}{w_H^1} \right) \frac{1}{w_H^1} + \lambda^1\phi = 0 \quad (\text{A.10})$$

$$-(\phi + \theta_H)z'(e_H) - \lambda^1\phi + \frac{\delta(\phi + \theta_H)\alpha y_H^2}{w_H^2(e_H)w_H^2(e_H)}v' \left(\frac{y_H^2}{w_H^2(e_H)} \right) = 0 \quad (\text{A.11})$$

$$\delta(1 - \phi - \theta_H)u'(c_L^2) - \lambda^2(1 - \phi) = 0 \quad (\text{A.12})$$

$$-\delta(1 - \phi)v' \left(\frac{y_L^2}{w_L^2(e_L)} \right) \frac{1}{w_L^2(e_L)} + \lambda^2(1 - \phi) + \delta\theta_Hv' \left(\frac{y_L^2}{\widehat{w}_H^2(e_L)} \right) \frac{1}{\widehat{w}_H^2(e_L)} = 0 \quad (\text{A.13})$$

$$\delta(\phi + \theta_H)u'(c_H^2) - \lambda^2\phi = 0 \quad (\text{A.14})$$

$$-\delta(\phi + \theta_H)v' \left(\frac{y_H^2}{w_H^2(e_H)} \right) \frac{1}{w_H^2(e_H)} + \lambda^2\phi = 0 \quad (\text{A.15})$$

where $\lambda^1 \geq 0$ and $\lambda^2 \geq 0$ are the multipliers on the first- and second-period budget constraints (4.2) and (4.3), respectively, while $\theta_H \geq 0$ is the multiplier on the high-skill type's incentive-compatibility constraint (4.4).

Using (A.6) and (A.7), we obtain:

$$(1 - \phi - \theta_H)u'(c_L^1) = (1 - \phi)v' \left(\frac{y_L^1}{w_L^1} \right) \frac{1}{w_L^1} - \theta_Hv' \left(\frac{y_L^1}{w_H^1} \right) \frac{1}{w_H^1} \quad (\text{A.16})$$

Since $w_H^1 > w_L^1$ and $v(\cdot)$ is strictly convex:

$$(1 - \phi)v' \left(\frac{y_L^1}{w_L^1} \right) \frac{1}{w_L^1} - \theta_Hv' \left(\frac{y_L^1}{w_H^1} \right) \frac{1}{w_H^1} > (1 - \phi)v' \left(\frac{y_L^1}{w_L^1} \right) \frac{1}{w_L^1} - \theta_Hv' \left(\frac{y_L^1}{w_L^1} \right) \frac{1}{w_L^1} \quad (\text{A.17})$$

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

$$(1 - \phi - \theta_H)u'(c_L^1) > (1 - \phi - \theta_H)v' \left(\frac{y_L^1}{w_L^1} \right) \frac{1}{w_L^1} \quad (\text{A.18})$$

Equations (A.6) and (A.9) imply that $1 - \phi - \theta_H > 0$. Hence, (A.18) can be rearranged to yield:

$$1 > \frac{v'(l_L^1)}{u'(c_L^1)w_L^1} \quad (\text{A.19})$$

which using (3.4) establishes that $MTRI_L^1 > 0$. Analogous manipulations of equations (A.12) and (A.13) establish that $MTRI_L^2 > 0$.

Dividing (A.10) by (A.9) and rearranging yields:

$$\frac{v'(l_H^1)}{u'(c_H^1)w_H^1} = 1 \quad (\text{A.20})$$

which using (3.4) establishes that $MTRI_H^1 = 0$. Analogous manipulations of equations (A.14) and (A.15) establish that $MTRI_H^2 = 0$.

Equations (A.6) and (A.8) can be combined to yield:

$$\begin{aligned} (1 - \phi - \theta_H) \frac{\alpha \delta l_L^2 v'(l_L^2)}{w_L^2(e_L)} - (1 - \phi - \theta_H) z'(e_L) - (1 - \phi - \theta_H) u'(c_L^1) \\ = \delta \theta_H \alpha y_L^2 \left[\frac{v' \left(\frac{y_L^2}{\widehat{w}_H^2(e_L)} \right)}{\widehat{w}_H^2(e_L) \widehat{w}_H^2(e_L)} - \frac{v' \left(\frac{y_L^2}{w_L^2(e_L)} \right)}{w_L^2(e_L) w_L^2(e_L)} \right] \end{aligned} \quad (\text{A.21})$$

Dividing both sides of (A.21) by $(1 - \phi - \theta_H) u'(c_L^1)$ and using (3.5), we obtain:

$$MTRE_L = \frac{\delta \theta_H \alpha y_L^2}{(1 - \phi - \theta_H) u'(c_L^1)} \left[\frac{v' \left(\frac{y_L^2}{\widehat{w}_H^2(e_L)} \right)}{\widehat{w}_H^2(e_L) \widehat{w}_H^2(e_L)} - \frac{v' \left(\frac{y_L^2}{w_L^2(e_L)} \right)}{w_L^2(e_L) w_L^2(e_L)} \right] \quad (\text{A.22})$$

Since $\widehat{w}_H^2(e_L) > w_L^2(e_L)$ and $v(\cdot)$ is strictly convex, the term in square brackets in (A.22) is negative. Hence $MTRE_L < 0$.

Equations (A.9) and (A.11) can be combined to yield:

$$(\phi + \theta_H) \frac{\alpha \delta l_H^2 v'(l_H^2)}{w_H^2(e_H)} - (\phi + \theta_H) z'(e_H) - (\phi + \theta_H) u'(c_H^1) = 0 \quad (\text{A.23})$$

Dividing (A.23) by $(\phi + \theta_H) u'(c_H^1)$ and using (3.5) establishes that $MTRE_H = 0$. ■

Proof of Proposition 2

The first-order conditions corresponding to programme (5.1) – (5.2) are:

$$(1 - \phi) u'(c_L^2) - \lambda^2 (1 - \phi) = 0 \quad (\text{A.24})$$

$$-(1 - \phi) v' \left(\frac{y_L^2}{w_L^2(e_L)} \right) \frac{1}{w_L^2(e_L)} + \lambda^2 (1 - \phi) = 0 \quad (\text{A.25})$$

$$\phi u'(c_H^2) - \lambda^2 \phi = 0 \quad (\text{A.26})$$

$$-\phi v' \left(\frac{y_H^2}{w_H^2(e_H)} \right) \frac{1}{w_H^2(e_H)} + \lambda^2 \phi = 0 \quad (\text{A.27})$$

$$(1 - \phi) [y_L^2 - c_L^2] + \phi [y_H^2 - c_H^2] = 0 \quad (\text{A.28})$$

where $\lambda^2 \geq 0$ is the multiplier on the second-period budget constraint (5.2).

Dividing (A.25) by (A.24) and rearranging yields:

$$\frac{v'(l_L^2)}{u'(c_L^2)w_L^2(e_L)} = 1 \quad (\text{A.29})$$

while dividing (A.27) by (A.26) and rearranging yields:

$$\frac{v'(l_H^2)}{u'(c_H^2)w_H^2(e_H)} = 1 \quad (\text{A.30})$$

which using (3.4) establishes that $MTRI_L^2 = 0$ and $MTRI_H^2 = 0$.

The relevant first-order conditions corresponding to programme (5.3) – (5.5) can be written as:

$$(1 - \phi - \theta_H)u'(c_L^1) - \lambda^1(1 - \phi) = 0 \quad (\text{A.31})$$

$$-(1 - \phi)v' \left(\frac{y_L^1}{w_L^1} \right) \frac{1}{w_L^1} + \lambda^1(1 - \phi) + \theta_H v' \left(\frac{y_H^1}{w_H^1} \right) \frac{1}{w_H^1} = 0 \quad (\text{A.32})$$

$$\begin{aligned} & -(1 - \phi - \theta_H)z'(e_L) - \lambda^1(1 - \phi) + \delta \frac{\partial W^2(\cdot)}{\partial e_L} - \frac{\delta \theta_H \alpha y_L^2}{\widehat{w}_H^2(e_L)\widehat{w}_H^2(e_L)} v' \left(\frac{y_L^2}{\widehat{w}_H^2(e_L)} \right) \\ & + \delta \theta_H \left[u'(c_H^2) \frac{\partial c_H^2(\cdot)}{\partial e_L} - \frac{v' \left(\frac{y_H^2}{w_H^2(e_H)} \right)}{w_H^2(e_H)} \frac{\partial y_H^2(\cdot)}{\partial e_L} - u'(c_L^2) \frac{\partial c_L^2(\cdot)}{\partial e_L} + \frac{v' \left(\frac{y_L^2}{\widehat{w}_H^2(e_L)} \right)}{\widehat{w}_H^2(e_L)} \frac{\partial y_L^2(\cdot)}{\partial e_L} \right] = 0 \end{aligned} \quad (\text{A.33})$$

$$(\phi + \theta_H)u'(c_H^1) - \lambda^1 \phi = 0 \quad (\text{A.34})$$

$$-(\phi + \theta_H)v' \left(\frac{y_H^1}{w_H^1} \right) \frac{1}{w_H^1} + \lambda^1 \phi = 0 \quad (\text{A.35})$$

$$\begin{aligned} & -(\phi + \theta_H)z'(e_H) - \lambda^1 \phi + \delta \frac{\partial W^2(\cdot)}{\partial e_H} + \frac{\delta \theta_H \alpha y_H^2}{w_H^2(e_H)w_H^2(e_H)} v' \left(\frac{y_H^2}{w_H^2(e_H)} \right) \\ & + \delta \theta_H \left[u'(c_H^2) \frac{\partial c_H^2(\cdot)}{\partial e_H} - \frac{v' \left(\frac{y_H^2}{w_H^2(e_H)} \right)}{w_H^2(e_H)} \frac{\partial y_H^2(\cdot)}{\partial e_H} - u'(c_L^2) \frac{\partial c_L^2(\cdot)}{\partial e_H} + \frac{v' \left(\frac{y_L^2}{\widehat{w}_H^2(e_L)} \right)}{\widehat{w}_H^2(e_L)} \frac{\partial y_L^2(\cdot)}{\partial e_H} \right] = 0 \end{aligned} \quad (\text{A.36})$$

where $\lambda^1 \geq 0$ is the multiplier on the first-period budget constraint (5.4), and $\theta_H \geq 0$ is the multiplier on the high-skill type's incentive-compatibility constraint (5.5).

Using (A.31) and (A.32) it can be shown that $MTRI_L^1 > 0$, and using (A.34) and (A.35) it can be shown that $MTRI_H^1 = 0$. The proofs of these results are analogous to those for the case when the government can commit, and are therefore omitted.

To derive expressions for $\partial W^2(\cdot)/\partial e_L$ and $\partial W^2(\cdot)/\partial e_H$, note that the Lagrangian corresponding to programme (5.1) – (5.2) can be written as:

$$\mathcal{L} = (1-\phi) \left[u(c_L^2) - v \left(\frac{y_L^2}{w_L^2(e_L)} \right) \right] + \phi \left[u(c_H^2) - v \left(\frac{y_H^2}{w_H^2(e_H)} \right) \right] + \lambda^2 [(1-\phi) [y_L^2 - c_L^2] + \phi [y_H^2 - c_H^2]] \quad (\text{A.37})$$

By the Envelope Theorem:

$$\frac{\partial W^2(\cdot)}{\partial e_L} = \frac{\partial \mathcal{L}(\cdot)}{\partial e_L} = \frac{\alpha(1-\phi)y_L^2}{w_L^2(e_L)w_L^2(e_L)} v' \left(\frac{y_L^2}{w_L^2(e_L)} \right) \quad (\text{A.38})$$

$$\frac{\partial W^2(\cdot)}{\partial e_H} = \frac{\partial \mathcal{L}(\cdot)}{\partial e_H} = \frac{\alpha\phi y_H^2}{w_H^2(e_H)w_H^2(e_H)} v' \left(\frac{y_H^2}{w_H^2(e_H)} \right) \quad (\text{A.39})$$

Substituting (A.38) into (A.33) and combining the result with (A.31) yields:

$$\begin{aligned} & (1-\phi-\theta_H) \frac{\alpha \delta l_L^2 v'(l_L^2)}{w_L^2(e_L)} - (1-\phi-\theta_H) z'(e_L) - (1-\phi-\theta_H) u'(c_L^1) \\ &= \delta \theta_H \alpha y_L^2 \left[\frac{v' \left(\frac{y_L^2}{\hat{w}_H^2(e_L)} \right)}{\hat{w}_H^2(e_L) \hat{w}_H^2(e_L)} - \frac{v' \left(\frac{y_L^2}{w_L^2(e_L)} \right)}{w_L^2(e_L) w_L^2(e_L)} \right] + \delta \theta_H u'(c_L^2) \left[\frac{\partial c_L^2(\cdot)}{\partial e_L} - \frac{\partial c_H^2(\cdot)}{\partial e_L} \right] \\ & \quad + \delta \theta_H \left[\frac{v' \left(\frac{y_H^2}{w_H^2(e_H)} \right)}{w_H^2(e_H)} \frac{\partial y_H^2(\cdot)}{\partial e_L} - \frac{v' \left(\frac{y_L^2}{\hat{w}_H^2(e_L)} \right)}{\hat{w}_H^2(e_L)} \frac{\partial y_L^2(\cdot)}{\partial e_L} \right] \quad (\text{A.40}) \end{aligned}$$

where use has been made of $u'(c_L^2) = u'(c_H^2)$, which follows from (A.24) and (A.26).

Dividing both sides of (A.40) by $(1-\phi-\theta_H)u'(c_L^1)$ and using (3.5), we obtain:

$$MTRE_L = \frac{\delta \theta_H \alpha y_L^2}{(1-\phi-\theta_H)u'(c_L^1)} \left[\frac{v' \left(\frac{y_L^2}{\hat{w}_H^2(e_L)} \right)}{\hat{w}_H^2(e_L) \hat{w}_H^2(e_L)} - \frac{v' \left(\frac{y_L^2}{w_L^2(e_L)} \right)}{w_L^2(e_L) w_L^2(e_L)} \right]$$

$$\begin{aligned}
& + \frac{\delta\theta_H u'(c_L^2)}{(1-\phi-\theta_H)u'(c_L^1)} \left[\frac{\partial c_L^2(\cdot)}{\partial e_L} - \frac{\partial c_H^2(\cdot)}{\partial e_L} \right] \\
& + \frac{\delta\theta_H}{(1-\phi-\theta_H)u'(c_L^1)} \left[\frac{v'\left(\frac{y_H^2}{w_H^2(e_H)}\right)}{w_H^2(e_H)} \frac{\partial y_H^2(\cdot)}{\partial e_L} - \frac{v'\left(\frac{y_L^2}{\widehat{w}_H^2(e_L)}\right)}{\widehat{w}_H^2(e_L)} \frac{\partial y_L^2(\cdot)}{\partial e_L} \right] \tag{A.41}
\end{aligned}$$

It follows from (A.31) and (A.34) that $1 - \phi - \theta_H > 0$. And since $\widehat{w}_H^2(e_L) > w_L^2(e_L)$ and $v(\cdot)$ is strictly convex, the first term in (A.41) is negative. By applying the Implicit Function Theorem and Cramer's Rule to (A.24) – (A.28), it can be shown that:

$$\frac{\partial c_L^2(\cdot)}{\partial e_L} = \frac{\partial c_H^2(\cdot)}{\partial e_L} > 0 \quad \frac{\partial y_H^2(\cdot)}{\partial e_L} < 0 \quad \frac{\partial y_L^2(\cdot)}{\partial e_L} > 0 \tag{A.42}$$

Therefore, the second term in (A.41) equals zero, while the third term is negative. This establishes that $MTRE_L < 0$.

Substituting (A.39) into (A.36) and combining the result with (A.34) yields:

$$\begin{aligned}
& (\phi + \theta_H) \frac{\alpha \delta l_H^2 v'(l_H^2)}{w_H^2(e_H)} - (\phi + \theta_H) z'(e_H) - (\phi + \theta_H) u'(c_H^1) \\
& = \delta\theta_H u'(c_H^2) \left[\frac{\partial c_L^2(\cdot)}{\partial e_H} - \frac{\partial c_H^2(\cdot)}{\partial e_H} \right] + \delta\theta_H \left[\frac{v'\left(\frac{y_H^2}{w_H^2(e_H)}\right)}{w_H^2(e_H)} \frac{\partial y_H^2(\cdot)}{\partial e_H} - \frac{v'\left(\frac{y_L^2}{\widehat{w}_H^2(e_L)}\right)}{\widehat{w}_H^2(e_L)} \frac{\partial y_L^2(\cdot)}{\partial e_H} \right] \tag{A.43}
\end{aligned}$$

where use has been made of the fact that $u'(c_L^2) = u'(c_H^2)$. Dividing both sides of (A.43) by $(\phi + \theta_H)u'(c_H^1)$ and using (3.5), we obtain:

$$\begin{aligned}
MTRE_H & = \frac{\delta\theta_H u'(c_H^2)}{(\phi + \theta_H)u'(c_H^1)} \left[\frac{\partial c_L^2(\cdot)}{\partial e_H} - \frac{\partial c_H^2(\cdot)}{\partial e_H} \right] \\
& + \frac{\delta\theta_H}{(\phi + \theta_H)u'(c_H^1)} \left[\frac{v'\left(\frac{y_H^2}{w_H^2(e_H)}\right)}{w_H^2(e_H)} \frac{\partial y_H^2(\cdot)}{\partial e_H} - \frac{v'\left(\frac{y_L^2}{\widehat{w}_H^2(e_L)}\right)}{\widehat{w}_H^2(e_L)} \frac{\partial y_L^2(\cdot)}{\partial e_H} \right] \tag{A.44}
\end{aligned}$$

By applying the Implicit Function Theorem and Cramer's Rule to (A.24) – (A.28), it can be shown that:

$$\frac{\partial c_L^2(\cdot)}{\partial e_H} = \frac{\partial c_H^2(\cdot)}{\partial e_H} > 0 \quad \frac{\partial y_H^2(\cdot)}{\partial e_H} > 0 \quad \frac{\partial y_L^2(\cdot)}{\partial e_H} < 0 \tag{A.45}$$

Therefore, the first term in (A.44) equals zero, while the second term is positive. This establishes that $MTRE_H > 0$. ■

Derivation of the Labour Supply Elasticity

In the absence of taxation, each individual would solve the following programme:

$$\max_{c_i^1, l_i^1, e_i, c_i^2, l_i^2} \ln(c_i^1) - \frac{1}{1+\gamma}(l_i^1)^{1+\gamma} - \frac{1}{1+\gamma}(e_i)^{1+\gamma} + \delta \left[\ln(c_i^2) - \frac{1}{1+\gamma}(l_i^2)^{1+\gamma} \right] \quad (\text{A.46})$$

subject to:

$$c_i^1 + e_i \leq w_i^1 l_i^1 \quad (\text{A.47})$$

$$c_i^2 \leq w_i^2 l_i^2 = (w_i^1 + \alpha e_i) l_i^2 \quad (\text{A.48})$$

The first-order conditions corresponding to programme (A.46) – (A.48) can be manipulated to yield:

$$(c_i^t)(l_i^t)^\gamma = w_i^t \quad \text{or} \quad \ln(c_i^t) + \gamma \ln(l_i^t) = \ln(w_i^t) \quad (\text{A.49})$$

which implies a labour supply elasticity of $1/\gamma$. ■

References

- [1] Andersson, F. and K. Konrad [2003], “Human Capital Investment and Globalization in Extortionary States”, *Journal of Public Economics*, 87, 1539-1555.
- [2] Apps, P. and R. Rees [2006], “Repeated Optimal Nonlinear Income Taxation”, *mimeo*.
- [3] Atkinson, A. and J. Stiglitz [1976], “The Design of Tax Structure: Direct Versus Indirect Taxation”, *Journal of Public Economics*, 6, 55-75.
- [4] Blumkin, T. and E. Sadka [2008], “A Case for Taxing Education”, *International Tax and Public Finance*, 15, 145-163.
- [5] Boadway, R., N. Marceau and M. Marchand [1996], “Investment in Education and the Time Inconsistency of Redistributive Tax Policy”, *Economica*, 63, 171-189.
- [6] Bovenberg, A. and B. Jacobs [2005], “Redistribution and Education Subsidies are Siamese Twins”, *Journal of Public Economics*, 89, 2005-2035.
- [7] Brett, C. and J. Weymark [2003], “Financing Education Using Optimal Redistributive Taxation”, *Journal of Public Economics*, 87, 2549-2569.
- [8] Brett, C. and J. Weymark [2008], “Optimal Nonlinear Taxation of Income and Savings without Commitment”, *Vanderbilt University Working Paper*, No. 08-W05.
- [9] Christiansen, V [1984], “Which Commodity Taxes Should Supplement the Income Tax?”, *Journal of Public Economics*, 24, 195-220.
- [10] De Fraja, G [2002], “The Design of Optimal Education Policies”, *Review of Economic Studies*, 69, 437-466.
- [11] Fang, H [2006], “Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices”, *International Economic Review*, 47, 1151-1185.
- [12] Fleurbaey, M., R. Gary-Bobo and D. Maguain [2002], “Education, Distributive Justice, and Adverse Selection”, *Journal of Public Economics*, 84, 113-150.
- [13] Gaube, T [2007], “Optimum Taxation of Each Year’s Income”, *Journal of Public Economic Theory*, 9, 127-150.
- [14] Goldin, C. and L. Katz [2007], “The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890-2005”, *NBER Working Paper*, No. 12984.
- [15] Golosov, M., M. Troshkin and A. Tsyvinski [2010], “Optimal Taxation: Merging Micro and Macro Approaches”, *Journal of Money, Credit and Banking*, forthcoming.

- [16] Golosov, M., A. Tsyvinski and I. Werning [2006], “New Dynamic Public Finance: A User’s Guide”, *NBER Macroeconomics Annual*, 2006, 317-387.
- [17] Grout, P [1983], “Imperfect Information, Markets and Public Provision of Education”, *Journal of Public Economics*, 22, 113-121.
- [18] Guesnerie, R [1995], *A Contribution to the Pure Theory of Taxation*, Cambridge University Press.
- [19] Guo, J-T. and A. Krause [2010], “Optimal Nonlinear Income Taxation with Habit Formation”, *Journal of Public Economic Theory*, forthcoming.
- [20] Hare, P. and D. Ulph [1979], “On Education and Distribution”, *Journal of Political Economy*, 87, S193-S212.
- [21] Kocherlakota, N [2010], *The New Dynamic Public Finance*, Princeton University Press.
- [22] Krause, A [2006], “Redistributive Taxation and Public Education”, *Journal of Public Economic Theory*, 8, 807-819.
- [23] Krause, A [2009a], “Education and Taxation Policies in the Presence of Counter-vailing Incentives”, *Economica*, 76, 387-399.
- [24] Krause, A [2009b], “Optimal Nonlinear Income Taxation with Learning-by-Doing”, *Journal of Public Economics*, 93, 1098-1110.
- [25] Maldonado, D [2008], “Education Policies and Optimal Taxation”, *International Tax and Public Finance*, 15, 131-143.
- [26] Mirrlees, J [1971], “An Exploration in the Theory of Optimum Income Taxation”, *Review of Economic Studies*, 38, 175-208.
- [27] Mirrlees, J [1976], “Optimal Tax Theory: A Synthesis”, *Journal of Public Economics*, 6, 327-358.
- [28] OECD [2010], *Highlights from Education at a Glance*, OECD publications 2010.
- [29] Stiglitz, J [1982], “Self-selection and Pareto Efficient Taxation”, *Journal of Public Economics*, 17, 213-240.
- [30] Tuomala, M [1986], “On the Optimal Income Taxation and Educational Decisions”, *Journal of Public Economics*, 30, 183-198.
- [31] Ulph, D [1977], “On the Optimal Distribution of Income and Educational Expenditure”, *Journal of Public Economics*, 8, 341-356.