CCES Discussion Paper Series
Center for Research on Contemporary Economic Systems

Graduate School of Economics
Hitotsubashi University

Optimal human capital policies under the endogenous choice of educational types
(Revised)

Takuya Obara
(Hitotsubashi University)

Optimal Human Capital Policies under the Endogenous Choice of Educational Types[*]

Takuya Obara[†]

June 12, 2018

Abstract

This study examines optimal human capital policies under non-linear labor and capital income taxes in the presence of consumption value of education in a two-period setting. We show that when individuals can choose educational types differing by the relative importance of consumption value and production value, education subsidies for low-type individuals should not equal an efficient level that offsets distortions induced by non-linear taxes on labor and capital income. Our findings imply that education policy does not restore efficiency, or the Diamond–Mirrlees production efficiency theorem fails. Moreover, capital income taxation is optimal, which means that the Atkinson–Stiglitz theorem breaks down.

---

1. Introduction

Although investment in human capital plays an important role in enriching lives, income taxation affects investment in human capital.[1] Labor income taxes prevent individuals from investing in human capital by capturing part of the return to human capital, and capital income taxes distort the choice between physical and human capital.[2] To alleviate tax distortions and foster human capital accumulation, OECD countries heavily subsidize higher education. From the efficiency concern that the government's intervention should not distort individual's decision-making, the optimal design of education policies under labor and capital income taxation is a research issue of interest for many economists.

A common assumption in previous literature on optimal education policies is that investment in human capital results in only a production value. Put differently, these studies have considered that the time invested in education contributes only to labor productivity, which leads to higher wages. However, there is growing empirical evidence for the existence of consumption value (Schaafsma (1976), Lazear (1977), Kodde and Ritzen (1984), Gullason (1989), Heckman et al. (1999), Carneiro et al. (2003), Arcidiacono (2004), and Alstadsæter (2011)). For example, education yields joy and satisfaction in learning new things, meeting new people, and participating in lectures and campus activities. Moreover, higher education generates opportunities for obtaining higher social status and finding interesting jobs. Therefore, the motivation underlining the educational choices of individuals stems from not only production value but also consumption value.[3] In addition, the importance of consumption value or production value differs between individuals. Alstadsæter (2011) shows that teachers' colleges in Norway are an educational type with a higher consumption value and a lower production value than business schools. Walker and Zhu (2003) report a negative wage return to an art degree in the UK, while there is a substantial positive wage return to an engineering degree. This implies that art graduates are willing to forgo future wages to enjoy the consumption value in education. These findings suggest that these returns from education vary across educational types and individuals choose an educational type depending on their own preferences.

The present study introduces the consumption value of education into the model and allows individuals to choose an educational type differing in the ratio between consumption value and production value. If educational choices consist of both production value and consumption value, the optimal design of non-linear education subsidies crucially depends on the observability of the two types of education.[4] When the government can observe both

---

[1]For example, Abramitzky and Lavy (2014) estimate the responsiveness of investment in education to changes in redistributive taxation, and provide experimental evidence that it affects education decisions.

[2]Trostel (1993) shows that, if all costs of education are effectively tax deductible, then labor income taxation is neutral with respect to human capital investment. However, many educational costs including tuition fees cannot be deducted from income tax, which implies that labor income taxation affects human capital investment (e.g., Bovenberg and Jacobs (2005)).

[3]According to Trostel et al. (2002), estimated rates of return to education in Norway and Sweden are about one-half of a worldwide average rate of return to education even with control for the differences in the level of the human capital investment across countries. This implies that students in these countries accept lower returns to productive human capital investment and choose to invest, instead, in more consumption value of education. Thus, overconsumption in human capital could be important in the Scandinavian countries.

[4]Our paper allows the government to employ non-linear education subsidies. The justification of non-linear

production and consumption values, education subsidies can be made contingent on the type of education. In this case, while the marginal subsidy on consumption value is zero from the Atkinson and Stiglitz (1976) theorem, the marginal subsidy on production value equals the marginal labor income tax rate from the Diamond and Mirrlees (1971) production efficiency theorem. However, if the government cannot verify the production and consumption values, it can only implement education subsidies on total human capital investment.[5] In this case, subsidies on education not only alleviate distortions in productive human capital but also lead to overconsumption of human capital. This means that an educational policy neglecting the consumption value of education extremely distorts the composition of consumption. Consequently, the optimal policy faces a trade-off between reducing distortions on productive human capital and avoiding overconsumption of consumptive human capital. The objective of this study is to theoretically clarify the optimal structure of non-linear education subsidies under non-linear taxes on labor and capital income when the type of human capital investment is not verifiable.

Our framework consists of a dynamic setting without uncertainty in which there are two types of individuals who differ only in exogenous ability, that is, a modified version of the two-type model developed by Stern (1982) and Stiglitz (1982).[6] These individuals live for two periods. In the first period, they consume, invest in education with a consumption value and a production value, and transfer resources through savings. The former value directly affects individuals' utility and the latter value raises the effective labor supply. In the second period, individuals work and then consume by spending their earnings and assets. Their earnings are a function of ability, labor supply, and education with the production value. We assume that the government can observe labor and capital income and education for each type, but cannot distinguish two types of value in education. This measurement problem does not allow the government to subsidize only the contribution to human capital. Therefore, the government can employ three sorts of non-linear tax schemes: non-linear labor and capital income taxes and non-linear education subsidies.

The first contribution of the study is to show that optimal education policies attaining

---

education subsidies stems from the fact that education is not easily retradable compared to commodities, as stated by Guo and Krause (2013).

[5]It implies that it is difficult for the government to observe how individuals allocate their time into consumptive and productive education while the years spent in formal education is verifiable. On the one hand, the assumption, that education subsidies cannot differ across productive and consumptive human capital, seems reasonable. This is because, to implement education subsidies contingent on the return from education, the government is required to grasp consumption value differing not only across different degree subjects but also within degree subjects since consumption value may differ depending on individual decisions (see Malchow-Møller et al. (2011)). On the other hand, the assumption, that total human capital investment is verifiable, is justified since foregone earnings are tax deductible and enrollment in higher education is subsidized (see Jacobs and Bovenberg (2010)).

[6]There is a growing body of literature analyzing optimal tax policies with human capital investment in a dynamic setting with certainty, for example, Nielsen and Sørensen (1997), Jacobs (2005), Bohacek and Kapička (2008), Jacobs and Bovenberg (2010), Schindler (2011), Kapička (2015), Jacobs and Yang (2016). By contrast, Eaton and Rosen (1980), Hamilton (1987), Anderberg and Andersson (2003), Da Costa and Maestri (2007), Anderberg (2009), Grochulski and Piskorski (2010), Jacobs et al. (2012), Kapička and Neira (2015), Schindler and Yang (2015), Findeisen and Sachs (2016), and Stantcheva (2017) investigate optimal tax policies in the presence of a stochastic risk factor on endogenous human capital formation in a dynamic setting.

an efficient level with respect to education choice should be modified under endogenous choice of educational type. Therefore, optimal education policies should not be set at a level to achieve efficiency concerns, which means that the production efficiency theorem of Diamond and Mirrlees (1971) breaks down. The second contribution of this study is to show that an individual's behavior reflecting a choice of educational types can justify taxation on capital income even if the utility function is separable between consumption and labor supply. This result presents the case in which the theorem of Atkinson and Stiglitz (1976) fails. These findings crucially depend on preference heterogeneity in educational types, which are endogenously generated. Allowing for choice of educational types, low-type individuals prefer production value to consumption value more than high-type individuals. Following the logic of Saez (2002), the additional information is useful to relax the binding incentive constraint, and therefore, the standard result is modified. As usual, high-type individuals face zero marginal tax rate on labor and capital income and education, that is, the result with no-distortion at the top remains. The present study highlights the importance of recognizing individuals' choice of educational types when implementing education policies.

Since the seminal contribution of the information-based approach to tax policy emanated from Mirrlees (1971), many economists have analyzed how education policies should be designed under non-linear labor income taxes when individuals have private information. Our study is closely related to Bovenberg and Jacobs (2005), who introduce education choices as one of the individual's behaviors into the framework of Mirrlees (1971) and show that the role of education subsidies is to eliminate the distortion on educational efforts induced by labor income taxes.[7] The findings suggest that education subsidies restore efficiency in education choices, that is, the Diamond–Mirrlees production efficiency theorem is valid. Moreover, the findings demonstrate that the result continues even in the presence of non-pecuniary benefit in education as long as the utility function is separable between work effort and non-pecuniary benefit in education. Our study differs in two ways from the framework of Bovenberg and Jacobs (2005). First, we extend their model as a two-period setting to explore the desirability of capital income taxes. Second, the authors assume that the choice of consumption value in education is exogenous, as in Alstadsæter (2003). The present study assumes that the choice of consumption value in education is endogenous.[8] Under the setting, we address the desirability of capital income taxation in addition to the distortion on learning, which differs in the statement of Bovenberg and Jacobs (2005).

In this study, the government can employ not only non-linear labor income taxes and education subsidies but also non-linear capital income taxes. It is well known that Ordover and

---

[7]Ulph (1977), Hare and Ulph (1979), and Krause (2006) are previous works exploring optimal tax systems with both income taxes and education expenditure. However, since these studies focus on publicly provided education, individuals do not decide on educational effort. Tuomala (1986) examines how educational choices should be reflected in optimal income taxation by allowing individuals to choose their educational choices, but education subsidies are not introduced in his model. By contrast, the present study analyzes education policies under non-linear income taxes when individuals can decide not only the level of labor supply but also educational effort, in line with Bovenberg and Jacobs (2005).

[8]Malchow-Møller et al. (2011) examine linear progressive taxes on labor income and tuition fees under endogenous choice of educational types when the tax rate on capital income is given exogenously. However, we investigate the optimal design of income taxes and education policies in the context of non-linear taxation. Furthermore, we allow the government to optimize capital income taxation and show that capital income taxation is not superfluous.

4

Phelps (1979) examine optimal non-linear taxation on income and savings in an overlapping-generations economy in the case of unobservable earnings ability, and state that if preferences are weakly separable between private goods and leisure, then taxes on savings are redundant. This is consistent with the Atkinson and Stiglitz (1976) theorem. Compared to the result, Jacobs and Bovenberg (2010) show that capital income taxation is useful to alleviate the tax distortion caused by labor income taxes instead of education subsidies when part of educational investment is non-verifiable, even under the weak separability condition. However, the authors also conclude that capital income taxes drop to zero as soon as all educational investments are verifiable. The present study demonstrates that even if all educational investments were verifiable, capital income taxation would not become redundant. This is because consumption value in education will endogenously trigger type-specific discount factors and differences in consumption preferences without any need for assuming multidimensional heterogeneity. The findings are closely related to Saez (2002) explaining the desirability of capital income taxes based on heterogeneous tastes for goods between high- and low-income earners. However, our model does not require multidimensional heterogeneity to show the desirability of capital income taxes in contrast with Saez (2002) and other related studies (e.g., Boadway et al. (2000), Cremer et al. (2001), Diamond and Spinnewijn (2011)).

The remainder of this paper is organized as follows. Section 2 describes the basic framework of the model. Section 3 characterizes and investigates optimal tax policies. Section 4 provides an additional analysis and section 5 concludes the paper.

2. The model

We consider a partial-equilibrium two-period model without uncertainty. The economy consists of two types of individuals who live for two periods, $t = 1, 2$, high-ability and low-ability, indexed by $i = H, L$. The population size is normalized to one. The proportion of high-ability individuals is π_H and the proportion of low-ability individuals is π_L. The amount of educational investment for type-i individuals is denoted by q_i, whose price is normalized to one.[9] For example, q_i can be interpreted as years spent in formal education. Let educational investments q_i consist of consumption value and production value. h_i is the share of q_i with consumption value. Correspondingly, $1 - h_i$ is the share of q_i with production value. We assume that h_i is an endogenous variable over $[0, 1]$, that is, individuals can choose any combination of consumption value and production value. Therefore, $x_i \equiv h_i q_i$ is consumption value, which directly affects utility, and $e_i \equiv (1 - h_i) q_i$ is production value, which augments effective labor supply. An individual's preference for type i is defined over consumption in the first period c_i^1, consumption in the second period c_i^2, consumption value in education x_i, and work effort ℓ_i. We assume separability between consumption in the first and second periods, c_i^1 and c_i^2, and work effort ℓ_i, and between work effort ℓ_i and consumption value x_i.[10]

---

[9]In our model, costs of education indicate direct costs (tuition fees) and the opportunity costs of education (foregone earnings when learning).

[10]The former assumption is made not only for the sake of simplicity but also to keep in line with the Atkinson-Stiglitz theorem. Also, the latter one is related to the Diamond-Mirrlees production efficiency theorem (see footnote 20). This specification thus allows us to clearly show the robustness of these results in the presence of the endogenous choice of educational types.

Then, type i's preference is expressed by

$$U(c_i^1, c_i^2, x_i, \ell_i) = u(c_i^1, c_i^2, x_i) - v(\ell_i) \tag{1}$$

Following conventional assumptions, we assume that $u(\cdot)$ is twice differentiable, strictly concave, and strictly increasing while $v(\cdot)$ is twice differentiable, strictly convex, and strictly increasing.

The accumulation of human capital is given by $g_i = a_i \phi(e_i)$, where a_i is the exogenous ability to benefit from educational investment and $\phi(\cdot)$ is the production function for human capital, where $\phi(\cdot)$ is twice differentiable, strictly increasing, and strictly concave, that is, $\phi'(\cdot) > 0$ and $\phi''(\cdot) < 0$. We suppose $a_H > a_L$, that is, high-ability individuals can learn more effectively from the same amount of educational investment. The elasticity of the production function for type i is defined as $\eta_i \equiv \frac{e_i}{g_i} \frac{\partial g_i}{\partial e_i}$. In the setting, we obtain $\eta_i = \frac{e_i \phi'(e_i)}{\phi(e_i)}$, which is constant with respect to ability and labor supply.[11]

We denote labor income of type i by $Y_i \equiv g_i \ell_i = a_i \phi(e_i) \ell_i$.[12] The government can observe labor income, capital income, and educational investment for each type, and thus, it can levy non-linear labor income taxes $T(Y_i)$, capital income taxes $\Phi(rs_i)$, and education subsidies $S(q_i)$ for type-i individuals, where r is the interest rate and s_i is the savings of type-i individuals.[13][14] However, the government cannot distinguish consumption value and production value of human capital investment.

In the first period, individuals with a common level of initial assets s_0 consume and invest in education. The first-period budget constraint is given by

$$c_i^1 + s_i + q_i - S(q_i) = s_0 \tag{2}$$

In the second period, the individuals consume, work, and consume their assets or repay their debts. The second-period budget constraint is given by

$$c_i^2 = Y_i - T(Y_i) + (1+r)s_i - \Phi(rs_i) \tag{3}$$

---

[11]Maldonado (2008) examines education policies under the assumption of complementarity between ability and educational investment. In this setting, the elasticity of the production function can vary with ability. Jacobs and Bovenberg (2011) generalize the model of Maldonado (2008) by allowing for the elasticity of the production function to depend on not only ability but also labor supply.

[12]The interpretation is that Y_i is the product of the wage rate, which is normalized to one, and effective labor supply $g_i \ell_i$. Alternatively, if we consider $\phi(e_i)\ell_i$ as the effective labor supply, we can interpret ability a_i as the wage rate per effective labor supply.

[13]Pirttilä and Tuomala (2001) show that the production efficiency theorem breaks down and capital income taxes are required under endogenous factor prices determined in general equilibrium. In addition, Jacobs (2013) presents the implication of optimal education policies in the presence of the general equilibrium effect, and shows that non-linear education policies play a redistributive role, which leads to the production efficiency theorem breaking down. In the model, we assume no general-equilibrium effects of input prices to clarify our contribution, and therefore, wage rates and interest rates are exogenous.

[14]For nonlinear capital taxation, the amount in a savings account must be observed at an individual level. Therefore, savings by family members other than the official holder of the account must be excluded (e.g., Blomquist and Micheletto (2008)). Although the government's ability to implement nonlinear capital income taxation seems to be limited in the real world, many countries impose a progressive tax on capital income (see Saez (2013) for more details).

2.1 Individual's behavior

The government must take account of optimal individual behavior when implementing non-linear labor and capital income taxes and education subsidies. Type-i individuals choose c_i^1, c_i^2, s_i, q_i, h_i, and Y_i to maximize their utility subject to the individual's budget constraint (equations (2) and (3)). This is formally defined as

$$
\max_{c_i^1, c_i^2, s_i, q_i, h_i, Y_i} \quad u(c_i^1, c_i^2, x_i) - v\left(\frac{Y_i}{a_i \phi(e_i)}\right)
$$
$$
s.t. \quad c_i^1 + s_i + q_i - S(q_i) = s_0
$$
$$
c_i^2 = Y_i - T(Y_i) + (1+r)s_i - \Phi(rs_i)
$$
(4)

According to Stiglitz (1982), it is convenient to express the utility function in terms of the variables the government is able to observe, in line with the self-selection problem. Since the government cannot observe h_i, it can control h_i only indirectly by assigning observable variables (c_i^1, c_i^2, q_i, and Y_i). It is equivalent to considering the individual's problem that determines the level of h_i for a given allocation. Given the individual's behavior for h_i, the government must rule out individuals choosing an allocation intended for another type to characterize the second-best planning solution using the revelation principle (see equation (16)). Thus, we rewrite the above problem as follows:

$$
\max_{c_i^1, c_i^2, s_i, q_i, Y_i} \quad u(c_i^1, c_i^2, x_i^*) - v\left(\frac{Y_i}{a_i \phi(e_i^*)}\right)
$$
$$
s.t. \quad c_i^1 + s_i + q_i - S(q_i) = s_0
$$
$$
c_i^2 = Y_i - T(Y_i) + (1+r)s_i - \Phi(rs_i)
$$
(5)

where $x_i^* \equiv h_i^* q_i$ is the choice for consumption value in education associated with the case in which each type will be truth-telling and $e_i^* \equiv (1 - h_i^*)q_i$ is the choice for production value in education associated with the same case. Also, h_i^* is used as a shorthand for $h_i(c_i^1, c_i^2, q_i, Y_i; a_i)$ which is given by

$$
h_i(c_i^1, c_i^2, q_i, Y_i; a_i) = \arg\max_{h_i} u(c_i^1, c_i^2, h_i q_i) - v\left(\frac{Y_i}{a_i \phi((1 - h_i)q_i)}\right)
$$
(6)

This means that the government can control h_i only indirectly through nonlinear tax/subsidy schedules by assigning observable variables directly.[15] Note that h_i^* satisfies the first-order condition for equation (6), that is,

$$
v_\ell(\ell_i^*)\frac{\ell_i^* \phi'(e_i^*)}{\phi(e_i^*)} = u_x(c_i^1, c_i^2, x_i^*)
$$
(7)

---

[15]For example, Findeisen and Sachs (2018) describe an unobservable variable as a function of observable variables due to the self-selection problem in line with Stiglitz (1982) when the government can only levy linear labor income taxes under nonlinear education subsidies.

where $\ell_i^* \equiv \frac{Y_i}{a_i\phi(e_i^*)}$ is the choice for labor supply associated with the case in which each type will be truth-telling, $u_x(c_i^1, c_i^2, x_i) \equiv \frac{\partial u(c_i^1, c_i^2, x_i)}{\partial x_i}$ the marginal utility of consumption value in education for type i, and $v_\ell(\ell_i) \equiv \frac{\partial v(\ell_i)}{\partial \ell_i}$ the marginal disutility of labor for type i. This condition indicates that the marginal utility of consumption value in education should equal that of production value in education, which implies that subsidies on human capital not only reduce distortions in productive human capital but also lead to overconsumption of human capital. Thus, the government needs to take account of this problem to alleviate distortions in the composition of consumption.

Using equations (7), the optimization problem given by equation (5) yields the first-order conditions (Appendix A):

$$MRS_{c^1q}^i \equiv \frac{u_x(c_i^1, c_i^2, x_i^*)}{u_c^1(c_i^1, c_i^2, x_i^*)} = \frac{v_\ell(\ell_i^*)}{u_c^1(c_i^1, c_i^2, x_i^*)}\frac{\ell_i^*\phi'(e_i^*)}{\phi(e_i^*)} = 1 - S'(q_i) \tag{8}$$

$$MRS_{c^2\ell}^i \equiv \frac{v_\ell(\ell_i^*)}{a_i\phi(e_i^*)u_c^2(c_i^1, c_i^2, x_i^*)} = 1 - T'(Y_i) \tag{9}$$

$$MRS_{c^1c^2}^i \equiv \frac{u_c^1(c_i^1, c_i^2, x_i^*)}{u_c^2(c_i^1, c_i^2, x_i^*)} = 1 + r - r\Phi'(rs_i) \tag{10}$$

where $u_c^1(c_i^1, c_i^2, x_i) \equiv \frac{\partial u(c_i^1, c_i^2, x_i)}{\partial c_i^1}$ denotes marginal utility of consumption in the first period, $u_c^2(c_i^1, c^2, x_i) \equiv \frac{\partial u(c_i^1, c^2, x_i)}{\partial c_i^2}$ that in the second period, $S'(q_i) \equiv \frac{\partial S(q_i)}{\partial q_i}$ the marginal subsidy rate for education, $T'(Y_i) \equiv \frac{\partial T(Y_i)}{\partial Y_i}$ the marginal labor income tax rate, and $\Phi'(rs_i) \equiv \frac{\partial \Phi(rs_i)}{\partial rs_i}$ the marginal capital income tax rate. Combining equations (8), (9), and (10) yields

$$MRT_{q\ell}^i \equiv \frac{Y_i\phi'(e_i^*)}{\phi(e_i^*)} = \frac{1 - S'(q_i)}{1 - T'(Y_i)}(1 + r - r\Phi'(rs_i)) \tag{11}$$

To measure the extent to which the tax (subsidy) instruments decrease (increase) the marginal returns to learning, we denote the total net tax wedge on learning for type i by

$$\begin{aligned}\Delta_i &\equiv T'(\cdot)\frac{Y_i\phi'(e_i^*)}{\phi(e_i^*)} - r\Phi'(\cdot) - S'(\cdot)(1 + r - r\Phi'(\cdot)) \\ &= \frac{T'(\cdot)}{1 - T'(\cdot)}R(1 - S'(\cdot)) - r\Phi'(\cdot) - S'(\cdot)R\end{aligned} \tag{12}$$

where $R \equiv 1 + r - r\Phi'(\cdot)$ is the discount factor.[16] The equality is derived using equation (11). From equation (12), while labor income taxes distort decision-making in terms of

---

[16]Compared to the model of Bovenberg and Jacobs (2005) in the presence of non-pecuniary benefit, Δ_i includes the distortion on consumption benefit from education caused by labor income taxes. In their model, optimal education subsidies fall compared to labor income tax in order to restore production efficiency, because non-pecuniary benefit can escape the distortion caused by labor income tax. On the other hand, if individuals can choose any combination of non-pecuniary benefit and pecuniary benefit, non-pecuniary benefit cannot escape from the distortion, since the distortion on the choice for production value in education caused by labor income tax indirectly affects the choice for consumption value in education, as presented in equation (7).

education, capital income taxes and education subsidies alleviate the distortion caused by labor income taxes. In particular, capital income taxes act as a subsidy for education by raising the present value of the marginal benefit of education. If $\Delta_i = 0$, the intertemporal marginal rate of transformation between education and labor supply for type i $MRT_{q\ell}^i$ equals $1 + r$, that is, equation (11) coincides with the first-order condition without any tax policy. As in the previous literature, if we consider education and labor supply as two inputs in the household production problem, in the situation of $\Delta_i = 0$, the result of Diamond and Mirrlees (1971) applies, that is, the government should ensure efficiency in the production side of the economy. Jacobs and Bovenberg (2010) show that if all educational investments are verifiable, that is, can be subsidized, then education subsidies eliminate the entire distortion on education due to labor income taxes without levying capital income taxes, which implies that the production efficiency theorem of Diamond and Mirrlees (1971) in addition to the theorem of Atkinson and Stiglitz (1976) are desirable. Our concern is whether both these theorems are robust, even if individuals possess endogenous choice of educational type.

2.2   The government

The objective of the government is to maximize the sum of utility for type i expressed by observable variables, which is given by

$$W = \sum_i \pi_i \{u(c_i^1, c_i^2, x_i^*) - v(\frac{Y_i}{a_i \phi(e_i^*)})\} \tag{13}$$

The government levies a non-linear tax on labor income and capital income to subsidize human capital investment. The budget constraint of the government takes the following form:

$$\sum_{i=H,L} \pi_i \left[-S(q_i) + \frac{1}{1+r}(T(Y_i) + \Phi(rs_i)) \right] = 0 \tag{14}$$

Using the budget constraint that individuals face, equation (14) can be rewritten as

$$\sum_{i=H,L} \pi_i \left[s_0 - c_i^1 - q_i + \frac{1}{1+r}(Y_i - c_i^2) \right] = 0 \tag{15}$$

The informational assumptions are in line with the optimal taxation literature analyzing the second-best allocation: the government cannot directly observe labor supply and ability. Additionally, h_i is not observable to the government in our model, which is indirectly controlled according to equation (6). By the revelation principle, the government must design the allocation to induce individuals to reveal their true types by choosing observable variables directly. We focus on the case in which the incentive constraint preventing high-ability individuals from mimicking low-ability individuals is binding. Therefore, the incentive constraint is

$$u(c_H^1, c_H^2, x_H^*) - v(\frac{Y_H}{a_H \phi(e_H^*)}) \geq u(c_L^1, c_L^2, \hat{x}^*) - v(\frac{Y_L}{a_H \phi(\hat{e}^*)}) \tag{16}$$

9

where $\hat{x}^* \equiv \hat{h}^* q_L$ is the choice for the consumption value associated with the case where high-ability individuals (mimickers) choose the allocation of low-ability individuals (the person being mimicked) and $\hat{e}^* \equiv (1 - \hat{h}^*) q_L$ is the choice for the production value associated with the same case. Given c_L^1, c_L^2, q_L, and Y_L, $\hat{h}^*$ is formally defined as

$$\hat{h}^* \equiv h_H(c_L^1, c_L^2, q_L, Y_L; a_H) = \arg\max_{h_H} u(c_L^1, c_L^2, h_H q_L) - v\left(\frac{Y_L}{a_H \phi((1 - h_H) q_L)}\right) \tag{17}$$

Note that $\hat{h}^*$ satisfies the first-order condition for equation (17), that is,

$$v_\ell(\hat{\ell}^*) \frac{\hat{\ell}^* \phi'(\hat{e}^*)}{\phi(\hat{e}^*)} = u_x(c_L^1, c_L^2, \hat{x}^*) \tag{18}$$

Let $\hat{\ell}^* \equiv \frac{Y_L}{a_H \phi(\hat{e}^*)}$ be the labor supply of the mimicker.

In summary, the government maximizes the social welfare function (13) subject to the government's budget constraint (15) and the incentive constraint (16) by choosing the allocation with respect to consumption in the first and second periods, educational investment, and labor income for each type. The corresponding Lagrangian is

$$\max_{\{c_i^1, c_i^2, q_i, Y_i\}_i} \mathcal{L} = \sum_i \pi_i \left\{ u(c_i^1, c_i^2, x_i^*) - v\left(\frac{Y_i}{a_i \phi(e_i^*)}\right) \right\}$$

$$+ \lambda \left[\sum_{i=H,L} \pi_i \left\{ s_0 - c_i^1 - q_i + \frac{1}{1+r}(Y_i - c_i^2) \right\} \right] \tag{19}$$

$$+ \gamma \left[u(c_H^1, c_H^2, x_H^*) - v\left(\frac{Y_H}{a_H \phi(e_H^*)}\right) - u(c_L^1, c_L^2, \hat{x}^*) + v\left(\frac{Y_L}{a_H \phi(\hat{e}^*)}\right) \right]$$

Let λ be the Lagrange multiplier of the government's budget constraint and γ the Lagrange multiplier of the incentive constraint.

3.   Optimal tax policy

From the first-order conditions with respect to equation (19), we characterize the optimal marginal subsidy rate on education and the optimal marginal labor income and capital income tax rate for each type (Appendix B):

$$S'(q_H) = 0 \tag{20}$$

$$S'(q_L) = \frac{\gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L} \left[\frac{u_x(c_L^1, c_L^2, x_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)} - \frac{u_x(c_L^1, c_L^2, \hat{x}^*)}{u_c^1(c_L^1, c_L^2, \hat{x}^*)} \right]$$

$$\equiv \frac{\gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L} \left[MRS_{c^1 q}^L - \hat{MRS}_{c^1 q} \right] \tag{21}$$

$$T'(Y_H) = 0 \tag{22}$$

$$\frac{T'(Y_L)}{1+r} = \frac{\gamma u_c^2(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L} \left[\frac{v_\ell(\ell_L^*)}{u_c^2(c_L^1, c_L^2, x_L^*)} \frac{1}{a_L \phi(e_L^*)} - \frac{v_\ell(\hat{\ell}^*)}{u_c^2(c_L^1, c_L^2, \hat{x}^*)} \frac{1}{a_H \phi(\hat{e}^*)} \right]$$
$$\equiv \frac{\gamma u_c^2(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L} \left[MRS_{c^2\ell}^L - M\hat{R}S_{c^2\ell} \right] \tag{23}$$

$$\Phi'(rs_H) = 0 \tag{24}$$

$$\frac{\Phi'(rs_L)}{1+r} = \frac{\gamma u_c^2(c_L^1, c_L^2, \hat{x}^*)}{r \lambda \pi_L} \left[\frac{u_c^1(c_L^1, c_L^2, x_L^*)}{u_c^2(c_L^1, c_L^2, x_L^*)} - \frac{u_c^1(c_L^1, c_L^2, \hat{x}^*)}{u_c^2(c_L^1, c_L^2, \hat{x}^*)} \right]$$
$$\equiv \frac{\gamma u_c^2(c_L^1, c_L^2, \hat{x}^*)}{r \lambda \pi_L} \left[MRS_{c^1 c^2}^L - M\hat{R}S_{c^1 c^2} \right] \tag{25}$$

From these results, we clarify our concerns about (i) the sign of $\Phi'(\cdot)$, that is, the justification of capital income taxes, and (ii) the sign of Δ_i, that is, the direction of the overall distortion on education induced by the three sorts of tax instruments. Before investigating these concerns, we present the following lemma.

Lemma 1. *The production value of education for low-type individuals is greater than that of mimicker, that is, $e_L^* > \hat{e}^*$. On the other hand, the consumption value of education for low-type individuals is lower than mimicker's one, that is $x_L^* < \hat{x}^*$.*

This proof is shown in Appendix C. This result stems from the fact that the mimicker reduces its labor supply to mimic low-type individuals. As shown in Appendix C, consumptive human capital investments are a substitutionary relationship for labor supply. This means that the reduction in labor supply increases consumptive human capital investments and decreases productive human capital investments. Intuitively, since the mimicker can earn a level of labor income for low-type individuals with less work than low-type individuals themselves perform, this leads to underinvestment in productive human capital of the mimicker.

3.1   Optimal capital income taxation

First, we investigate the desirability of capital income taxes. Whereas the marginal capital income tax rate is zero for high-type individuals from equation (24), the situation is different for low-type individuals. Note that deviating from the Atkinson–Stiglitz theorem crucially relies on the sign of the bracket on the right-hand side of equation (25), which is determined by the complementarity of consumption in each period with consumption value in education. For example, education can affect the level of consumption such as books, computers, and tobacco. Let us consider the situation in which consumption in the second period is more complementary to the consumption value than consumption in the first period. In this case, the bracket is positive, which means that $\Phi'(\cdot)$ is positive. Therefore, the Atkinson–Stiglitz theorem breaks down, as capital income taxation is not redundant. The intuition is as follows: since the mimicker values consumption in the second period more than the mimicked because of $MRS_{c^1 c^2}^L > M\hat{R}S_{c^1 c^2}$, imposing capital income taxes hurts the mimicker more than the mimicked, and thus, the government relaxes the incentive constraint for high-type individuals.[17] This finding is related to Banks and Diamond (2010) and Diamond and

---

[17] This is analogous to the result in the context of commodity taxation. Mirrlees (1976) establishes conditions for optimal mixed taxation consisting of labor income and commodity taxation, and shows the desir-

Spinnewijn (2011), who argue that capital income should be taxed if individuals with higher ability have lower discount rates, which implies that they tend to save more. This condition is observationally equivalent to stating that consumptive human capital is more complementary to future consumption than to current consumption.[18]

According to lemma 1, the justification of capital income taxes stems from heterogeneous preferences for educational types between high- and low-income earners, which are endogenously generated. The heterogeneous preferences create type-specific discount factors, and thus, allow the intertemporal marginal rate of substitution to vary between the mimicker and the mimicked. This means that the government can obtain additional information to relax the incentive constraint. In contrast to Saez (2002), the findings of our study show the desirability of capital income taxes without requiring multidimensional heterogeneity. The following proposition summarizes the main results of this section.

Proposition 1. *When the consumption value in education is more complementary to (substitutionary for) consumption in the second period than in the first, the marginal capital income tax rate is positive (negative) for low-type individuals and zero for high-type individuals.*

This result is closely related to Jacobs and Bovenberg (2010). The authors show that the capital tax is always positive to cancel out the tax distortion caused by non-verifiable educational investment. In contrast, this paper justifies capital income taxation even under verifiable educational investment since consumption value in education endogenously creates a difference in preferences for intertemporal consumption and then capital income taxation relaxes the incentive constraint by using the additional information. Furthermore, we show that a negative tax on capital income is desirable in order to relax the incentive constraint when the mimicker prefers first-period consumption more than the mimicked.

However, the Atkinson–Stiglitz theorem can be restored if we assume that preferences are weakly separable in the sense of the following functional form: $u(c_L^1, c_L^2, x) = u(f(c_L^1, c_L^2), x)$. In that case, the sign of the bracket is zero, owing to no impact of heterogeneous tastes on the intertemporal marginal rate of substitution. Therefore, capital income taxes are no longer required.

Finally, we provide economic implications for applied tax policies. First, if consumptive human capital exists, taxes/subsidies on savings can achieve more income redistribution than labor income taxes alone. Thus, an optimal tax policy is determined by a fundamental economic trade-off between additional income redistribution and saving distortions. Second, empirical studies show that savings are positively correlated with education, taken as a proxy for earnings ability, conditional on labor income (e.g., Lawrance (1991) and Dynan et al. (2004)). This means that high-type individuals have a higher taste for savings. Therefore, it is plausible to think that taxing savings is desirable to redistribute income.

---

ability of commodity taxation on goods that are relatively more preferred by high income earners.

[18]The pattern of complementarity between savings and consumptive human capital that determines the sign of the marginal capital income tax rate does not depend on whether consumptive human capital emerges in the first period or the second period. This is because the utility function is expressed by equation (1) regardless of the timing of consumption value, which means that the form of equation (25) does not depend on the timing.

3.2 Production inefficiency

The next item of interest is examination of the sign of Δ_i, that is, the direction of the overall distortion on education caused by the three kinds of tax instruments used in this study. To observe this, we combine equations (21), (23), and (25), which yields (Appendix D):

$$
\begin{aligned}
\frac{\Delta_L}{1+r} &= \frac{\gamma u_x(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L} \left[\frac{Y_L \phi'(\hat{e}^*)}{\phi(\hat{e}^*)} - \frac{Y_L \phi'(e_L^*)}{\phi(e_L^*)} \right] \\
&\equiv \frac{\gamma u_x(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L} \left[\hat{MRT}_{q\ell} - MRT_{q\ell}^L \right]
\end{aligned}
\tag{26}
$$

In contrast to Bovenberg and Jacobs (2005), when individuals have education choice between consumption value and production value, the novel term appears even if the utility function is separable between work effort and consumption value.[19][20] The deviation from household production efficiency crucially depends on the sign of the bracket on the right-hand side of equation (26), which creates an informational advantage for the government. From lemma 1, the intertemporal marginal rate of transformation between education and labor supply for the mimicker is greater than that for low-type individuals, that is, we obtain $\hat{MRT}_{q\ell} > MRT_{q\ell}^L$.[21] Therefore, $\Delta_L > 0$ is optimal, which means that education subsidies for low-type individuals should not completely offset the distortions of labor income taxes that are alleviated (augmented) by positive (negative) capital income taxes. The intuition is that distortions in learning for low-type individuals harm the mimicker more than the mimicked, and thereby allow the government to relax the binding incentive constraint, since the mimicker prefers education to labor supply measured by the present value relative to low-type individuals. It then follows that the Diamond–Mirrlees production efficiency theorem breaks down. Also, the main finding has a strong connection with Jacobs and Bovenberg (2011). The authors check the robustness of full efficiency in human capital formation by generalizing the earnings function, and show that it is justified as long as education is weakly separable from labor and ability. This is because education policies do not affect the incentive constraint under the special form of the earnings function. However, as shown in equation (26), when the consumption value of education is introduced and individuals can choose the ratio between consumption value and production value, the production efficiency theorem fails for low-type

---

[19]As mentioned in footnote 16, labor income taxes distort decision-making in terms of educational choice between consumption and production value. Consequently, consumption benefit in education cannot escape from the distortion induced by labor income taxes. Equation (26) implies that education subsidies should offset the distortion on the benefit. Thus, the first term in the bracket of equation (43) in the model of Bovenberg and Jacobs (2005) disappears.

[20]Under non-separability between work effort and consumption value, education subsidies increase or decrease to mitigate the distortion stemming from redistributive taxes, depending on the complementarity between them (see Bovenberg and Jacobs (2005)).

[21]The first derivative of $\frac{\phi'(e)}{\phi(e)}$ with respect to e is

$$
\frac{\partial \frac{\phi'(e)}{\phi(e)}}{\partial e} = \frac{\phi''(e)}{\phi(e)} - \left(\frac{\phi'(e)}{\phi(e)} \right)^2
$$

From the concavity of $\phi(\cdot)$, it is negative. Thus, we can obtain $\hat{MRT}_{q\ell} > MRT_{q\ell}^L$ from lemma 1.

individuals even if the special form of the earnings function is satisfied. From equation (20), for high-type individuals, education subsidies are redundant, since learning is not distorted by labor and capital income taxes from equations (22) and (24), which implies $\Delta_H = 0$. In summary, we propose as follows.

Proposition 2. *Low-ability types face a downward distortion on learning, that is, Δ_L is positive; high-ability types face no distortion on learning, that is, $\Delta_H = 0$.*

Intuitively, this finding can be explained as follows. When the government cannot employ a separate subsidy on productive and consumptive human capital investment, subsidies on total human capital investment not only reduce distortions in productive human capital but also cause overconsumption of consumptive human capital. This is because it is not desirable to subsidize consumptive human capital investment if the government can observe the consumption and production values of human capital investment for each type, as shown in Appendix E. This is consistent with the Atkinson and Stiglitz (1976) theorem. As a result, the optimal education policy faces a trade-off between mitigating distortions on productive human capital and avoiding overconsumption of consumptive human capital. Therefore, it is not desirable to subsidize human capital until full efficiency in human capital is achieved.

The main result also stems from heterogeneous tastes in educational types from lemma 1. As the bracket of equation (26) shows, the sign of the bracket is positive as long as heterogeneity occurs. Therefore, if individuals differ in exogenous skill ability and the incentive constraint is binding, then production inefficiency is always desirable in our model.

3.3 Optimal labor income taxation and education subsidies

Finally, we check the sign of $S'(q_i)$ and $T'(Y_i)$. For high-type individuals, both marginal subsidy and tax rate are zero from equations (20) and (22). In other words, no distortion for the top result holds. On the other hand, the sign of $S'(q_L)$ and $T'(Y_L)$ for low-type individuals depends on the complementarity or substitutability between consumption and consumption value. First, we take the marginal subsidy rate on education for low-type individuals in equation (21). Using lemma 1, if consumption value is complementary to or has no relationship with consumption in the first period, that is, $u_{c^1 x} \geq 0$, then the sign of $S(q_L)$ is positive which means that education is subsidized.[22] This is because the single-crossing property holds, in other words, the marginal rate of substitution between consumption and consumption value in education for the mimicker is lower than that for low-type individuals. The intuition of positive $S'(q_L)$ is that the mimicker prefers consumption to consumption value over education, and therefore, distorting educational efforts for low-type individuals upward can relax the incentive constraint. On the other hand, if consumption value in education is substitutionary for consumption, that is, $u_{c^1 x} < 0$, then $S'(q_L)$ cannot be signed,

---

[22]The first derivative of the marginal rate of substitution between consumption in the first period and consumption value is

$$\frac{\partial \frac{u_x(c^1, c^2, x)}{u_c^1(c^1, c^2, x)}}{\partial x} = \frac{u_{xx}}{u_c^1} - \frac{u_x}{u_c^1} u_{c^1 x}$$

If $u_{c^1 x} \geq 0$, it is a decreasing function with respect to consumption value. Thus, the sign of the bracket on the right-hand side in equation (21) is positive, which means $S'(q_L) > 0$.

since the single-crossing property does not hold. Second, we take the marginal tax rate on labor income for low-type individuals in equation (23). Following the standard result in a two-class economy suggested by Stiglitz (1982), the marginal labor income tax rate is positive for low-type individuals if a single-crossing condition holds. However, as for the discussion on marginal subsidy rates, it is ambiguous whether the single-crossing condition holds in our model. To observe this, we rewrite the marginal income tax rate for low-type individuals by substituting equations (7) and (18) into (23), as follows:

$$
\begin{aligned}
\frac{T'(Y_L)}{1+r} &= \frac{\gamma u_c^2(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L} \left[\frac{u_x(c_L^1, c_L^2, x_L^*)}{u_c^2(c_L^1, c_L^2, x_L^*)} \frac{\phi(e_L^*)}{Y_L \phi'(e_L^*)} - \frac{u_x(c_L^1, c_L^2, \hat{x}^*)}{u_c^2(c_L^1, c_L^2, \hat{x}^*)} \frac{\phi(\hat{e}^*)}{Y_L \phi'(\hat{e}^*)} \right] \\
&\equiv \frac{\gamma u_c(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L} \left[\frac{MRS_{c^2q}^L}{MRT_{q\ell}} - \frac{\hat{MRS}_{c^2q}}{\hat{MRT}_{q\ell}} \right]
\end{aligned}
\tag{27}
$$

From the fact $\hat{MRT}_{q\ell} > MRT_{q\ell}$, the sign of $T'(Y_L)$ crucially depends on the sign of u_{c^2x}. If consumption value in education is complementary to or has no relationship with consumption in the second period, that is, $u_{c^2x} \geq 0$, $\hat{MRS}_{c^2q}$ is less than MRS_{c^2q} which leads to the conclusion that the single-crossing property holds. This case means that the marginal labor income tax rate is positive at the bottom. However, if consumption value in education is substitutionary for consumption, that is, $u_{c^2x} < 0$, then $T'(Y_L)$ cannot be signed, since the single-crossing condition does not hold. As a result, the marginal income tax rate is not necessarily positive for low-type individuals. Thus, we summarize the sign of the marginal subsidy rate on education and the marginal tax rate on labor income as follows.

Proposition 3. *For low-type individuals, when consumption value in education is complementary to or has no relationship with consumption in the first and second periods, both the marginal subsidy rate on education and the marginal tax rate on labor income are positive, and otherwise, the sign of either or both of them is ambiguous. For high-type individuals, both marginal education subsidy and labor income tax rate are zero.*

As a result, a sufficient condition to hold each single-crossing property satisfies a complementary relationship between consumption value in education and consumption in the first and second periods, that is, $u_{c^1x} \geq 0$ and $u_{c^2x} \geq 0$. For example, the sub-utility function is the constant elasticity of substitution (CES) form: $u(c_i^1, c_i^2, x_i) = (\alpha(c_i^1)^{-\rho} + \beta(c_i^2)^{-\rho} + \delta x_i^{-\rho})^{-\frac{1}{\rho}}$ with $\alpha + \beta + \delta = 1$ and $\rho \geq -1$. In this case, the marginal subsidy rate on education and labor income tax rate for low-type individuals are always positive, because $u_{c^1x} \geq 0$ and $u_{c^2x} \geq 0$ regardless of the level of ρ.[23]

4. Non-pecuniary costs of education

So far, we have assumed that education yields utility, that is, the marginal utility of consumption value from education u_x is positive. In contrast, some empirical evidences suggest

---

[23]The cross-derivative of the sub-utility function is

$$
u_{c^1x} = \alpha\delta(1+\rho)(\alpha(c_i^1)^{-\rho} + \beta(c_i^2)^{-\rho} + \delta x_i^{-\rho})^{-\frac{1}{\rho}-2}(c_i^1 x_i)^{-\rho-1}
$$

Therefore, it is non-negative for any $\rho \geq -1$. Using a similar method, we obtain $u_{c^2x} \geq 0$.

that education costs utility, that is, u_x is negative. For example, Heckman et al. (2006) refer to the important role of psychic costs to explain why the college attendance rate is so low when monetary returns are so high. This means that people invest in human capital based on not only monetary returns and pecuniary costs but also non-pecuniary costs. Also, Palacios-Huerta (2003, 2006) suggest that the human capital premium cannot be explained by riskiness alone, implying that psychic costs are one of the sources that explains the size of the risk-adjusted premium. Therefore, we analyze the case in which education costs utility.

In this case, note that every person can avoid utility costs by investing only in productive human capital. Indeed, the first-order condition of equation (6) with respect to h_i is

$$\frac{\partial U_i}{\partial h_i} = u_x(c_i^1, c_i^2, x_i) - v_\ell(\ell_i)\frac{\ell_i \phi'(e_i)}{\phi(e_i)} < 0 \tag{28}$$

where $U_i \equiv u(c_i^1, c_i^2, h_i q_i) - v(\frac{Y_i}{a_i \phi((1-h_i)q_i)})$ denotes individual's utility for type i. Therefore, the corner solution is optimal, that is, h_i^* is zero. Similarly, the result applies to the mimicker from the first-order condition of equation (17), that is, $\hat{h}^*$ is zero. This means that both the mimicker and the mimicked invest only in the productive value of education to avoid utility costs. This corresponds to the standard model of Bovenberg and Jacobs (2005). Therefore, assuming that $u(c_i^1, c_i^2, 0)$ is strictly increasing with respect to c_i^1 and c_i^2, the subsidy rate equals the marginal labor income tax rate, and capital income taxation is redundant. Moreover, the marginal labor income tax rate is zero at the top and positive at the bottom, which means that the marginal education subsidy rate is zero at the top and positive at the bottom. These results stem from the fact that preference heterogeneity does not occur.[24]

Proposition 4. *If education costs utility, that is, u_x is negative, both the Diamond-Mirrlees production efficiency theorem and the Atkinson-Stiglitz theorem are valid. Also, both the marginal subsidy rate on education and the marginal tax rate on labor income are zero for high-type individuals and positive for low-type individuals.*

5. Concluding Remarks

This study examines optimal human capital policies under non-linear labor and capital income taxes when education has consumption value and production value, and individuals can choose an educational type. The former value generates a direct utility gain and the latter value promotes effective labor supply. Since the government can observe labor income, capital income, and educational investment, but is unable to distinguish the two types of returns from education, it can implement non-linear labor and capital income taxes and subsidies for

---

[24]As shown above, if h_i is endogenously determined, investment in productive human capital does not generate utility costs. In contrast, Bovenberg and Jacobs (2005) examine optimal education subsidies under non-linear labor income taxes when productive human capital requires utility costs in addition to investment of resources. Basically, their setting corresponds to our model under the case where h_i is exogenously given and identical between individuals. They show that the subsidy rate is larger than the marginal tax rate to remedy low investment due to utility costs (see equation (42) in the model of Bovenberg and Jacobs (2005)). However, the role of the additional subsidies is to ensure efficiency in human capital accumulation, not improve equity. Thus, such a policy does not overturn the production efficiency theorem (see Appendix F).

education. To the best of our knowledge, no previous study characterizes optimal education policies under non-linear income tax instruments when individuals have endogenous choice of educational types.

Under endogenous choice of educational types, the optimal tax policies are modified in this study. First, we show that capital income taxation can be necessary for low-type individuals, even when individuals differ in a single dimension; this result is in contrast to several studies that have highlighted how the theorem of Atkinson and Stiglitz (1976) breaks down when individuals differ along more than one dimension. Second, the direction of the overall distortion on learning induced by the three sorts of tax instruments shifts down, which means that the production efficiency theorem of Diamond and Mirrlees (1971) fails. The two novel findings stem from the preference heterogeneity in education between the mimicker and the mimicked, which is endogenously generated. The additional information is useful to relax the incentive constraint and thus, the deviation from the tax policy attaining efficiency is optimal. On the other hand, the result of no distortion at the top remains. The present study highlights the importance of recognizing individuals' choice of educational types when implementing education policies.

Appendix A

The Lagrangian for the optimization problem for type i (equation (5)) is formulated by

$$
\begin{aligned}
\max_{c_i^1, c_i^2, q_i, Y_i, s_i} L = u(c_i^1, c_i^2, x_i^*) - v\left(\frac{Y_i}{a_i \phi(e_i^*)}\right) \\
+ \lambda_1 \left[s_0 - c_i^1 - s_i - q_i + S(q_i) \right] \\
+ \lambda_2 \left[Y_i - T(Y_i) + (1+r)s_i - \Phi(rs_i) - c_i^2 \right]
\end{aligned}
\tag{A.1}
$$

where λ_1 denotes the Lagrangian multiplier of the individual's budget constraint in the first-period and λ_2 that in the second-period. The first-order conditions with respect to c_i^1, c_i^2, q_i, Y_i, and s_i are

$$
\frac{\partial L}{\partial c_i^1} = u_c^1(c_i^1, c_i^2, x_i^*) + \frac{\partial h_i^*}{\partial c_i^1} q_i \left[u_x(c_i^1, c_i^2, x_i^*) - v_\ell(\ell_i^*) \frac{\ell_i^* \phi'(e_i^*)}{\phi(e_i^*)} \right] - \lambda_1 = 0
\tag{A.2}
$$

$$
\frac{\partial L}{\partial c_i^2} = u_c^2(c_i^1, c_i^2, x_i^*) + \frac{\partial h_i^*}{\partial c_i^2} q_i \left[u_x(c_i^1, c_i^2, x_i^*) - v_\ell(\ell_i^*) \frac{\ell_i^* \phi'(e_i^*)}{\phi(e_i^*)} \right] - \lambda_2 = 0
\tag{A.3}
$$

$$
\frac{\partial L}{\partial q_i} = v_\ell(\ell_i^*) \frac{\ell_i^* \phi'(e_i^*)}{\phi(e_i^*)} + \left[h_i^* + \frac{\partial h_i^*}{\partial q_i} q_i \right] \left[u_x(c_i^1, c_i^2, x_i^*) - v_\ell(\ell_i^*) \frac{\ell_i^* \phi'(e_i^*)}{\phi(e_i^*)} \right] - \lambda_1(1 - S'(q_i)) = 0
\tag{A.4}
$$

$$
\frac{\partial L}{\partial Y_i} = -\frac{v_\ell(\ell_i^*)}{a_i \phi(e_i^*)} + \frac{\partial h_i^*}{\partial Y_i} q_i \left[u_x(c_i^1, c_i^2, x_i^*) - v_\ell(\ell_i^*) \frac{\ell_i^* \phi'(e_i^*)}{\phi(e_i^*)} \right] + \lambda_2(1 - T'(Y_i)) = 0
\tag{A.5}
$$

$$
\frac{\partial L}{\partial s_i} = -\lambda_1 + \lambda_2(1 + r - r\Phi'(rs_i)) = 0
\tag{A.6}
$$

Using equation (7), equations (A.2)-(A.5) can be rewritten as

$$\frac{\partial L}{\partial c_i^1} = u_c^1(c_i^1, c_i^2, x_i^*) - \lambda_1 = 0 \tag{A.7}$$

$$\frac{\partial L}{\partial c_i^2} = u_c^2(c_i^1, c_i^2, x_i^*) - \lambda_2 = 0 \tag{A.8}$$

$$\frac{\partial L}{\partial q_i} = v_\ell(\ell_i^*)\frac{\ell_i^* \phi'(e_i^*)}{\phi(e_i^*)} - \lambda_1(1 - S'(q_i)) = 0 \tag{A.9}$$

$$\frac{\partial L}{\partial Y_i} = -\frac{v_\ell(\ell_i^*)}{a_i \phi(e_i^*)} + \lambda_2(1 - T'(Y_i)) = 0 \tag{A.10}$$

Thus, we can obtain equation (8) from equation (7), (A.7), and (A.9), equation (9) from (A.8) and (A.10), and equation (10) from (A.6), (A.7), and (A.8). □

Appendix B

The first-order conditions associated with c_i^1, c_i^2, q_i, and Y_i, $i = H, L$, of equation (19) are as follows:

$$\frac{\partial \mathcal{L}}{\partial c_L^1} = \pi_L u_c^1(c_L^1, c_L^2, x_L^*) + \pi_L \frac{\partial h_L^*}{\partial c_L^1} q_L \left[u_x(c_L^1, c_L^2, x_L^*) - v_\ell(\ell_L^*)\frac{\ell_L^* \phi'(e_L^*)}{\phi(e_L^*)} \right]$$
$$- \gamma u_c^1(c_L^1, c_L^2, \hat{x}^*) - \gamma \frac{\partial \hat{h}^*}{\partial c_L^1} q_L \left[u_x(c_L^1, c_L^2, \hat{x}^*) - v_\ell(\hat{\ell}^*)\frac{\hat{\ell}^* \phi'(\hat{e}^*)}{\phi(\hat{e}^*)} \right] - \lambda \pi_L = 0 \tag{B.1}$$

$$\frac{\partial \mathcal{L}}{\partial c_H^1} = \pi_H u_c^1(c_H^1, c_H^2, x_H^*) + \pi_H \frac{\partial h_H^*}{\partial c_H^1} q_H \left[u_x(c_H^1, c_H^2, x_H^*) - v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)} \right]$$
$$+ \gamma u_c^1(c_H^1, c_H^2, x_H^*) + \gamma \frac{\partial h_H^*}{\partial c_H^1} q_H \left[u_x(c_H^1, c_H^2, x_H^*) - v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)} \right] - \lambda \pi_H = 0 \tag{B.2}$$

$$\frac{\partial \mathcal{L}}{\partial c_L^2} = \pi_L u_c^2(c_L^1, c_L^2, x_L^*) + \pi_L \frac{\partial h_L^*}{\partial c_L^2} q_L \left[u_x(c_L^1, c_L^2, x_L^*) - v_\ell(\ell_L^*)\frac{\ell_L^* \phi'(e_L^*)}{\phi(e_L^*)} \right]$$
$$- \gamma u_c^2(c_L^1, c_L^2, \hat{x}^*) - \gamma \frac{\partial \hat{h}^*}{\partial c_L^2} q_L \left[u_x(c_L^1, c_L^2, \hat{x}^*) - v_\ell(\hat{\ell}^*)\frac{\hat{\ell}^* \phi'(\hat{e}^*)}{\phi(\hat{e}^*)} \right] - \frac{1}{1+r}\lambda \pi_L = 0 \tag{B.3}$$

$$\frac{\partial \mathcal{L}}{\partial c_H^2} = \pi_H u_c^2(c_H^1, c_H^2, x_H^*) + \pi_H \frac{\partial h_H^*}{\partial c_H^2} q_H \left[u_x(c_H^1, c_H^2, x_H^*) - v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)} \right]$$
$$+ \gamma u_c^2(c_H^1, c_H^2, x_H^*) + \gamma \frac{\partial h_H^*}{\partial c_H^2} q_H \left[u_x(c_H^1, c_H^2, x_H^*) - v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)} \right] - \frac{1}{1+r}\lambda \pi_H = 0 \tag{B.4}$$

18

$$\frac{\partial \mathcal{L}}{\partial q_L} = \pi_L v_\ell(\ell_L^*)\frac{\ell_L^* \phi'(e_L^*)}{\phi(e_L^*)} + \pi_L\left[h_L^* + \frac{\partial h_L^*}{\partial q_L}q_L\right]\left[u_x(c_L^1, c_L^2, x_L^*) - v_\ell(\ell_L^*)\frac{\ell_L^* \phi'(e_L^*)}{\phi(e_L^*)}\right]$$
$$- \gamma v_\ell(\hat{\ell}^*)\frac{\hat{\ell}^* \phi'(\hat{e}^*)}{\phi(\hat{e}^*)} - \gamma\left[\hat{h}^* + \frac{\partial \hat{h}^*}{\partial q_L}q_L\right]\left[u_x(c_L^1, c_L^2, \hat{x}^*) - v_\ell(\hat{\ell}^*)\frac{\hat{\ell}^* \phi'(\hat{e}^*)}{\phi(\hat{e}^*)}\right] - \lambda\pi_L = 0 \tag{B.5}$$

$$\frac{\partial \mathcal{L}}{\partial q_H} = \pi_H v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)} + \pi_H\left[h_H^* + \frac{\partial h_H^*}{\partial q_H}q_H\right]\left[u_x(c_H^1, c_H^2, x_H^*) - v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)}\right]$$
$$+ \gamma v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)} + \gamma\left[h_H^* + \frac{\partial h_H^*}{\partial q_H}q_H\right]\left[u_x(c_H^1, c_H^2, x_H^*) - v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)}\right] - \lambda\pi_H = 0 \tag{B.6}$$

$$\frac{\partial \mathcal{L}}{\partial Y_L} = -\pi_L v_\ell(\ell_L^*)\frac{1}{a_L\phi(e_L^*)} + \pi_L\frac{\partial h_L^*}{\partial Y_L}q_L\left[u_x(c_L^1, c_L^2, x_L^*) - v_\ell(\ell_L^*)\frac{\ell_L^* \phi'(e_L^*)}{\phi(e_L^*)}\right]$$
$$+ \gamma v_\ell(\hat{\ell}^*)\frac{1}{a_H\phi(\hat{e}^*)} - \gamma\frac{\partial \hat{h}^*}{\partial Y_L}q_L\left[u_x(c_L^1, c_L^2, \hat{x}^*) - v_\ell(\hat{\ell}^*)\frac{\hat{\ell}^* \phi'(\hat{e}^*)}{\phi(\hat{e}^*)}\right] + \frac{1}{1+r}\lambda\pi_L = 0 \tag{B.7}$$

$$\frac{\partial \mathcal{L}}{\partial Y_H} = -\pi_H v_\ell(\ell_H^*)\frac{1}{a_H\phi(e_H^*)} + \pi_H\frac{\partial h_H^*}{\partial Y_H}q_H\left[u_x(c_H^1, c_H^2, x_H^*) - v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)}\right]$$
$$- \gamma v_\ell(\ell_H^*)\frac{1}{a_H\phi(e_H^*)} + \gamma\frac{\partial h_H^*}{\partial Y_H}q_H\left[u_x(c_H^1, c_H^2, x_H^*) - v_\ell(\ell_H^*)\frac{\ell_H^* \phi'(e_H^*)}{\phi(e_H^*)}\right] + \frac{1}{1+r}\lambda\pi_H = 0 \tag{B.8}$$

Using equations (7) and (18), (B.1)-(B.8) can be rewritten as

$$\frac{\partial \mathcal{L}}{\partial c_L^1} = \pi_L u_c^1(c_L^1, c_L^2, x_L^*) - \gamma u_c^1(c_L^1, c_L^2, \hat{x}^*) - \lambda\pi_L = 0 \tag{B.9}$$

$$\frac{\partial \mathcal{L}}{\partial c_H^1} = \pi_H u_c^1(c_H^1, c_H^2, x_H^*) + \gamma u_c^1(c_H^1, c_H^2, x_H^*) - \lambda\pi_H = 0 \tag{B.10}$$

$$\frac{\partial \mathcal{L}}{\partial c_L^2} = \pi_L u_c^2(c_L^1, c_L^2, x_L^*) - \gamma u_c^2(c_L^1, c_L^2, \hat{x}^*) - \frac{1}{1+r}\lambda\pi_L = 0 \tag{B.11}$$

$$\frac{\partial \mathcal{L}}{\partial c_H^2} = \pi_H u_c^2(c_H^1, c_H^2, x_H^*) + \gamma u_c^2(c_H^1, c_H^2, x_H^*) - \frac{1}{1+r}\lambda\pi_H = 0 \tag{B.12}$$

$$\frac{\partial \mathcal{L}}{\partial q_L} = \pi_L u_x(c_L^1, c_L^2, x_L^*) - \gamma u_x(c_L^1, c_L^2, \hat{x}^*) - \lambda\pi_L = 0 \tag{B.13}$$

$$\frac{\partial \mathcal{L}}{\partial q_H} = \pi_H u_x(c_H^1, c_H^2, x_H^*) + \gamma u_x(c_H^1, c_H^2, x_H^*) - \lambda\pi_H = 0 \tag{B.14}$$

$$\frac{\partial \mathcal{L}}{\partial Y_L} = -\pi_L v_\ell(\ell_L^*)\frac{1}{a_L\phi(e_L^*)} + \gamma v_\ell(\hat{\ell}^*)\frac{1}{a_H\phi(\hat{e}^*)} + \frac{1}{1+r}\lambda\pi_L = 0 \tag{B.15}$$

$$\frac{\partial \mathcal{L}}{\partial Y_H} = -\pi_H v_\ell(\ell_H^*)\frac{1}{a_H\phi(e_H^*)} - \gamma v_\ell(\ell_H^*)\frac{1}{a_H\phi(e_H^*)} + \frac{1}{1+r}\lambda\pi_H = 0 \tag{B.16}$$

First, we derive the marginal subsidy rate on education at the optimum. Combining (B.9)

19

with (B.13) yields

$$\pi_L\{u_x(c_L^1, c_L^2, x_L^*) - u_c^1(c_L^1, c_L^2, x_L^*)\} = \gamma\{u_x(c_L^1, c_L^2, \hat{x}^*) - u_c^1(c_L^1, c_L^2, \hat{x}^*)\} \tag{B.17}$$

(B.17) is rewritten as

$$\pi_L u_c^1(c_L^1, c_L^2, x_L^*)\left[\frac{u_x(c_L^1, c_L^2, x_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)} - 1\right] = \gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)\left[\frac{u_x(c_L^1, c_L^2, \hat{x}^*)}{u_c^1(c_L^1, c_L^2, \hat{x}^*)} - 1\right] \tag{B.18}$$

Using equation (8), the bracket in the left hand side $\left(\frac{u_x(c_L^1, c_L^2, x_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)} - 1\right)$ is equal to $-S'(q_L)$ and the bracket in the right hand side $\left(\frac{u_x(c_L^1, c_L^2, \hat{x}^*)}{u_c^1(c_L^1, c_L^2, \hat{x}^*)} - 1\right)$ is equal to $\frac{u_x(c_L^1, c_L^2, \hat{x}^*)}{u_c^1(c_L^1, c_L^2, \hat{x}^*)} - \frac{u_x(c_L^1, c_L^2, x_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)} - S'(q_L)$.
Substituting these relationships into (B.18) yields

$$-\pi_L u_c^1(c_L^1, c_L^2, x_L^*)S'(q_L) = \gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)\left[\frac{u_x(c_L^1, c_L^2, \hat{x}^*)}{u_c^1(c_L^1, c_L^2, \hat{x}^*)} - \frac{u_x(c_L^1, c_L^2, x_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)} - S'(q_L)\right] \tag{B.19}$$

Rearranging (B.19) yields

$$\{\pi_L u_c^1(c_L^1, c_L^2, x_L^*) - \gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)\}S'(q_L) = \gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)\left[\frac{u_x(c_L^1, c_L^2, x_L^*)}{u_c(c_L^1, c_L^2, x_L^*)} - \frac{u_x(c_L^1, c_L^2, \hat{x}^*)}{u_c(c_L^1, c_L^2, \hat{x}^*)}\right] \tag{B.20}$$

Substituting (B.9) into the term in the brackets of the left-hand side, we obtain equation (21). Similarly, combining (B.10) with (B.14) yields

$$\pi_H\{u_x(c_H^1, c_H^2, x_H^*) - u_c^1(c_H^1, c_H^2, x_H^*)\} = -\gamma\{u_x(c_H^1, c_H^2, x_H^*) - u_c^1(c_H^1, c_H^2, x_H^*)\} \tag{B.21}$$

Using equation (8), $u_x(c_H^1, c_H^2, x_H^*) - u_c^1(c_H^1, c_H^2, x_H^*) = -u_c^1(c_H^1, c_H^2, x_H^*)S'(q_H)$. Thus, (B.21) can be rewritten as follows:

$$(\pi_H + \gamma)u_c^1(c_H^1, c_H^2, x_H^*)S'(q_H) = 0 \tag{B.22}$$

From (B.10), $\pi_H + \gamma$ is positive, which implies that $S'(q_H)$ is zero.

Second, we turn to the derivation of the optimal marginal labor income tax rate. Combining (B.11) with (B.15) yields

$$\pi_L\left[u_c^2(c_L^1, c_L^2, x_L^*) - v_\ell(\ell_L^*)\frac{1}{a_L\phi(e_L^*)}\right] = \gamma\left[u_c^2(c_L^1, c_L^2, \hat{x}^*) - v_\ell(\hat{\ell}^*)\frac{1}{a_H\phi(\hat{e}^*)}\right] \tag{B.23}$$

(B.23) is rewritten as

$$\pi_L u_c^2(c_L^1, c_L^2, x_L^*)\left[1 - \frac{v_\ell(\ell_L^*)}{u_c^2(c_L^1, c_L^2, x_L^*)}\frac{1}{a_L\phi(e_L^*)}\right] = \gamma u_c^2(c_L^1, c_L^2, \hat{x}^*)\left[1 - \frac{v_\ell(\hat{\ell}^*)}{u_c^2(c_L^1, c_L^2, \hat{x}^*)}\frac{1}{a_H\phi(\hat{e}^*)}\right] \tag{B.24}$$

Using equation (9), the bracket in the left hand side $\left(1 - \frac{v_\ell(\ell_L^*)}{u_c^2(c_L^1, c_L^2, x_L^*)}\frac{1}{a_L\phi(e_L^*)}\right)$ is equal to $T'(Y_L)$ and the bracket in the right hand side $\left(1 - \frac{v_\ell(\hat{\ell}^*)}{u_c^2(c_L^1, c_L^2, \hat{x}^*)}\frac{1}{a_H\phi(\hat{e}^*)}\right)$ is equal to $T'(Y_L) +$

$\frac{v_\ell(\ell_L^*)}{u_c^2(c_L^1,c_L^2,x_L^*)}\frac{1}{a_L\phi(e_L^*)} - \frac{v_\ell(\hat{\ell}^*)}{u_c^2(c_L^1,c_L^2,\hat{x}^*)}\frac{1}{a_H\phi(\hat{e}^*)}$. Substituting these relationships into (B.24) yields

$$\pi_L u_c^2(c_L^1,c_L^2,x_L^*)T'(Y_L) = \gamma u_c^2(c_L^1,c_L^2,\hat{x}^*)\left[T'(Y_L) + \frac{v_\ell(\ell_L^*)}{u_c^2(c_L^1,c_L^2,x_L^*)}\frac{1}{a_L\phi(e_L^*)} - \frac{v_\ell(\hat{\ell}^*)}{u_c^2(c_L^1,c_L^2,\hat{x}^*)}\frac{1}{a_H\phi(\hat{e}^*)}\right]$$
(B.25)

Rearranging (B.25) and then substituting equation (9) yields

$$\{\pi_L u_c^2(c_L^1,c_L^2,x_L^*) - \gamma u_c^2(c_L^1,c_L^2,\hat{x}^*)\}T'(Y_L)$$
$$= \gamma u_c^2(c_L^1,c_L^2,\hat{x}^*)\left[\frac{v_\ell(\ell_L^*)}{u_c^2(c_L^1,c_L^2,x_L^*)}\frac{1}{a_L\phi(e_L^*)} - \frac{v_\ell(\hat{\ell}^*)}{u_c^2(c_L^1,c_L^2,\hat{x}^*)}\frac{1}{a_H\phi(\hat{e}^*)}\right]$$
(B.26)

Substituting (B.11) into the term in the brackets of the left-hand side, we obtain equation (23). Similarly, combining (B.12) with (B.16) yields

$$\pi_H\left[u_c^2(c_H^1,c_H^2,x_H^*) - v_\ell(\ell_H^*)\frac{1}{a_H\phi(e_H^*)}\right] = -\gamma\left[u_c^2(c_H^1,c_H^2,x_H^*) - v_\ell(\ell_H^*)\frac{1}{a_H\phi(e_H^*)}\right]$$
(B.27)

Using equation (9), $u_c^2(c_H^1,c_H^2,x_H^*) - v_\ell(\ell_H^*)\frac{1}{a_H\phi(e_H^*)} = u_c^2(c_H^1,c_H^2,x_H^*)T'(Y_H)$. Thus, (B.27) can be rewritten as follows:

$$(\pi_H + \gamma)u_c^2(c_H^1,c_H^2,x_H^*)T'(Y_H) = 0$$
(B.28)

From (B.12), $\pi_H + \gamma$ is positive, which implies that $T'(Y_H)$ is zero.

Finally, we derive the optimal marginal capital income tax rate. Combining (B.9) with (B.11) yields

$$\pi_L\{u_c^1(c_L^1,c_L^2,x_L^*) - (1+r)u_c^2(c_L^1,c_L^2,x_L^*)\} = \gamma\{u_c^1(c_L^1,c_L^2,\hat{x}^*) - (1+r)u_c^2(c_L^1,c_L^2,\hat{x}^*)\}$$
(B.29)

(B.29) is rewritten as

$$\pi_L u_c^2(c_L^1,c_L^2,x_L^*)\left[\frac{u_c^1(c_L^1,c_L^2,x_L^*)}{u_c^2(c_L^1,c_L^2,x_L^*)} - (1+r)\right] = \gamma u_c^2(c_L^1,c_L^2,\hat{x}^*)\left[\frac{u_c^1(c_L^1,c_L^2,\hat{x}^*)}{u_c^2(c_L^1,c_L^2,\hat{x}^*)} - (1+r)\right]$$
(B.30)

Using equation (10), the bracket in the left hand side ($\frac{u_c^1(c_L^1,c_L^2,x_L^*)}{u_c^2(c_L^1,c_L^2,x_L^*)} - (1+r)$) is equal to $-r\Phi'(rs_L)$ and the bracket in the right hand side ($\frac{u_c^1(c_L^1,c_L^2,\hat{x}^*)}{u_c^2(c_L^1,c_L^2,\hat{x}^*)} - (1+r)$) is equal to $\frac{u_c^1(c_L^1,c_L^2,\hat{x}^*)}{u_c^2(c_L^1,c_L^2,\hat{x}^*)} - \frac{u_c^1(c_L^1,c_L^2,x_L^*)}{u_c^2(c_L^1,c_L^2,x_L^*)} - r\Phi'(rs_L)$. Substituting these relationships into (B.30) yields

$$-\pi_L u_c^2(c_L^1,c_L^2,x_L^*)r\Phi'(rs_L) = \gamma u_c^2(c_L^1,c_L^2,\hat{x}^*)\left[\frac{u_c^1(c_L^1,c_L^2,\hat{x}^*)}{u_c^2(c_L^1,c_L^2,\hat{x}^*)} - \frac{u_c^1(c_L^1,c_L^2,x_L^*)}{u_c^2(c_L^1,c_L^2,x_L^*)} - r\Phi'(rs_L)\right]$$
(B.31)

Rearranging (B.31) and substituting equation (10) yields

$$\{\pi_L u_c^2(c_L^1,c_L^2,x_L^*) - \gamma u_c^2(c_L^1,c_L^2,\hat{x}^*)\}r\Phi'(rs_L) = \gamma u_c^2(c_L^1,c_L^2,\hat{x}^*)\left[\frac{u_c^1(c_L^1,c_L^2,x_L^*)}{u_c^2(c_L^1,c_L^2,x_L^*)} - \frac{u_c^1(c_L^1,c_L^2,\hat{x}^*)}{u_c^2(c_L^1,c_L^2,\hat{x}^*)}\right]$$
(B.32)

Substituting (B.11) into the term in the brackets of the left-hand side, we obtain equation

(25). Similarly, combining (B.10) with (B.12) yields

$$\pi_H\{u_c^1(c_H^1, c_H^2, x_H^*) - (1+r)u_c^2(c_H^1, c_H^2, x_H^*)\} = -\gamma\{u_c^1(c_H^1, c_H^2, x_H^*) - (1+r)u_c^2(c_H^1, c_H^2, x_H^*)\}$$
(B.33)

Using equation (10), $u_c^1(c_H^1, c_H^2, x_H^*) - (1+r)u_c^2(c_H^1, c_H^2, x_H^*) = -u_c^2(c_H^1, c_H^2, x_H^*)r\Phi'(rs_H)$. Thus, (B.33) can be rewritten as follows:

$$(\pi_H + \gamma)u_c^2(c_H^1, c_H^2, x_H^*)r\Phi'(rs_H) = 0$$
(B.34)

From (B.12), $\pi_H + \gamma$ is positive, which implies that $\Phi'(rs_H)$ is zero.

$\square$

Appendix C

Given Y_L, c_L^1, c_L^2, s_L, and q_L, the first-order condition associated with equation (6) for low-type individuals is given by

$$\frac{\partial U_L}{\partial h_L} = u_x(c_L^1, c_L^2, x_L)q_L - v_\ell\left(\frac{Y_L}{a_L\phi(e_L)}\right)\frac{Y_L}{a_L\phi(e_L)}\frac{\phi'(e_L)}{\phi(e_L)}q_L = 0$$
(C.1)

where $U_L \equiv u(c_L^1, c_L^2, h_L q_L) - v\left(\frac{Y_L}{a_L\phi((1-h_L)q_L)}\right)$ denotes low-type individuals' utility. Moreover, the second-order condition with respect to h_L is as follows:

$$\frac{1}{q_L^2}\frac{\partial^2 U_L}{\partial h_L^2} = u_{xx}(c_L^1, c_L^2, x_L) - v_{\ell\ell}\left(\frac{Y_L}{a_L\phi(e_L)}\right)\left(\frac{Y_L}{a_L\phi(e_L)}\frac{\phi'(e_L)}{\phi(e_L)}\right)^2$$
$$- 2v_\ell\left(\frac{Y_L}{a_L\phi(e_L)}\right)\frac{Y_L}{a_L\phi(e_L)}\left(\frac{\phi'(e_L)}{\phi(e_L)}\right)^2 + v_\ell\left(\frac{Y_L}{a_L\phi(e_L)}\right)\frac{Y_L}{a_L\phi(e_L)}\frac{\phi''(e_L)}{\phi(e_L)} < 0$$
(C.2)

Therefore, h_L^* is a locally maximized solution, because the second-order condition is satisfied from the assumption of $u(\cdot)$, $v(\cdot)$, and $\phi(\cdot)$ on the curvature. Now, we present the comparative statics of an individual's behavior due to the change of a_L. From equation (C.1), we derive it as follows:

$$\frac{\partial h_L^*}{\partial a_L}\frac{\partial^2 U_L}{\partial h_L^2}\bigg|_* = -v_{\ell\ell}\left(\frac{Y_L}{a_L\phi(e_L^*)}\right)\left(\frac{Y_L}{a_L\phi(e_L^*)}\right)^2\frac{\phi'(e_L^*)}{\phi(e_L^*)}\frac{1}{a_L}q_L$$
$$- v_\ell\left(\frac{Y_L}{a_L\phi(e_L^*)}\right)\frac{Y_L}{(a_L)^2\phi(e_L^*)}\frac{\phi'(e_L^*)}{\phi(e_L^*)}q_L$$
(C.3)

where $\frac{\partial^2 U_L}{\partial h_L^2}\big|_*$ denotes $\frac{\partial^2 U_L}{\partial h_L^2}$ evaluated at $h_L = h_L^*$. Since the sign of the sum of the two terms on the right-hand side is negative, the sign of $\frac{\partial h_L^*}{\partial a_L}$ is positive. Here, note that the mimicker faces the same allocations as expressed by equation (17). Thus, we can conclude that $\hat{h}^* > h_L^*$ because $a_H > a_L$, which means that $x_L^* < \hat{x}^*$ and $e_L^* > \hat{e}^*$. The result is related to the fact that the mimicker reduces his/her labor supply to mimic low-type individuals. Substituting

h_L^* only into $\frac{Y_L}{a_L\phi(e_L)}$ in equation (C.1) yields:

$$u_x(c_L^1, c_L^2, h_L q_L) = v_\ell(\ell_L^*)\frac{\ell_L^* \phi'((1-h_L)q_L)}{\phi((1-h_L)q_L)} \tag{C.4}$$

From equation (C.4), we derive h_L as a function of ℓ_L^*, denoted by $\bar{h}_L(c_L^1, c_L^2, q_L, \ell_L^*)$. Obviously, we have $h_L^* = \bar{h}_L(c_L^1, c_L^2, q_L, \ell_L^*)$. Differentiating it with respect to a_L yields $\frac{\partial h_L^*}{\partial a_L} = \frac{\partial \bar{h}_L(\cdot)}{\partial \ell_L}\frac{\partial \ell_L^*}{\partial a_L}$. Substituting $\bar{h}_L(\cdot)$ into equation (C.4) and then differentiating it with respect to ℓ_L^* yields:

$$\frac{\partial \bar{h}_L(\cdot)}{\partial \ell_L^*}\Theta = v_{\ell\ell}(\ell_L^*)\frac{\ell_L^* \phi'(e_L^*)}{\phi(e_L^*)}q_L + v_\ell(\ell_L^*)\frac{\phi'(e_L^*)}{\phi(e_L^*)}q_L \tag{C.5}$$

where $\Theta \equiv \left[u_{xx}(c_L^1, c_L^2, x_L^*) - v_\ell(\ell_L^*)\ell_L^*\left(\frac{\phi'(e_L^*)}{\phi(e_L^*)}\right)^2 + v_\ell(\ell_L^*)\frac{\ell_L^* \phi''(e_L^*)}{\phi(e_L^*)}\right](q_L)^2 < 0$. Note that we use the relationship $x_L^* = h_L^* q_L = \bar{h}_L(\cdot)q_L$ and $e_L^* = (1-h_L^*)q_L = (1-\bar{h}_L(\cdot))q_L$ in equation (C.5). From equation (C.5), $\frac{\partial \bar{h}_L(\cdot)}{\partial \ell_L}$ is negative. This means that consumptive human capital investments are a substitutionary relationship for labor supply. As a result, $\frac{\partial \ell_L^*}{\partial a_L}$ is negative. Thus, the reduction in labor supply to mimic induces the mimicker to increase consumptive human capital investments and decrease productive human capital investments. Consequently, the mimicker prefers consumptive human capital more than low-type individuals.

$\square$

Appendix D

Using equations (7) and (18), equation (21) can be rewritten as follows:

$$S'(q_L) = \frac{\gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L}\left[\frac{v_\ell(\ell_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)}\frac{\ell_L^* \phi'(e_L^*)}{\phi(e_L^*)} - \frac{v_\ell(\hat{\ell}^*)}{u_c^1(c_L^1, c_L^2, \hat{x}^*)}\frac{\hat{\ell}^* \phi'(\hat{e}^*)}{\phi(\hat{e}^*)}\right] \tag{D.1}$$

By the definition of ℓ_L^* and $\hat{\ell}^*$,

$$S'(q_L) = \frac{\gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L}\frac{v_\ell(\ell_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)}\frac{Y_L}{a_L\phi(e_L^*)}\frac{\phi'(e_L^*)}{\phi(e_L^*)} - \frac{\gamma}{\lambda \pi_L}\frac{v_\ell(\hat{\ell}^*)}{a_H\phi(\hat{e}^*)}\frac{Y_L\phi'(\hat{e}^*)}{\phi(\hat{e}^*)} \tag{D.2}$$

Here, we rearrange equation (23) as follows:

$$\frac{\gamma u_c^1(c_L^1, c_L^2, \hat{x}^*)}{\lambda \pi_L}\frac{v_\ell(\ell_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)}\frac{1}{a_L\phi(e_L^*)} = \Gamma\left[\frac{T'(Y_L)}{1+r} + \frac{\gamma}{\lambda \pi_L}\frac{v_\ell(\hat{\ell}^*)}{a_H\phi(\hat{e}^*)}\right] \tag{D.3}$$

where, $\Gamma \equiv \frac{u_c^2(c_L^1, c_L^2, x_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)}\frac{u_c^1(c_L^1, c_L^2, \hat{x}^*)}{u_c^2(c_L^1, c_L^2, \hat{x}^*)}$
Substituting (D.3) into the first term of (D.2),

$$S'(q_L) = -\frac{\gamma}{\lambda \pi_L}\frac{v_\ell(\hat{\ell}^*)}{a_H\phi(\hat{e}^*)}\frac{Y_L\phi'(\hat{e}^*)}{\phi(\hat{e}^*)} + \frac{Y_L\phi'(e_L^*)}{\phi(e_L^*)}\Gamma\left[\frac{T'(Y_L)}{1+r} + \frac{\gamma}{\lambda \pi_L}\frac{v_\ell(\hat{\ell}^*)}{a_H\phi(\hat{e}^*)}\right] \tag{D.4}$$

23

On the other hand, we rearrange equation (25) as follows:

$$\Gamma = 1 - \frac{r\lambda\pi_L}{\gamma u_c^2(c_L^1, c_L^2, \hat{x}^*)} \frac{u_c^2(c_L^1, c_L^2, x_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)} \frac{\Phi'(rs_L)}{1+r} \tag{D.5}$$

Substituting (D.5) into (D.4) yields

$$
\begin{aligned}
S'(q_L) = &-\frac{\gamma}{\lambda\pi_L} \frac{v_\ell(\hat{\ell}^*)}{a_H\phi(\hat{e}^*)} \left[\frac{Y_L\phi'(\hat{e}^*)}{\phi(\hat{e}^*)} - \frac{Y_L\phi'(e_L^*)}{\phi(e_L^*)} \right] + \frac{Y_L\phi'(e_L^*)}{\phi(e_L^*)} \frac{T'(Y_L)}{1+r} \\
&- \frac{Y_L\phi'(e_L^*)}{\phi(e_L^*)} \left[\frac{T'(Y_L)}{1+r} + \frac{\gamma}{\lambda\pi_L} \frac{v_\ell(\hat{\ell}^*)}{a_H\phi(\hat{e}^*)} \right] \frac{\lambda\pi_L}{\gamma u_c^2(c_L^1, c_L^2, \hat{x}^*)} \frac{u_c^2(c_L^1, c_L^2, x_L^*)}{u_c^1(c_L^1, c_L^2, x_L^*)} \frac{r\Phi'(rs_L)}{1+r}
\end{aligned}
\tag{D.6}
$$

Using equations (7), (8), and (23), the last term of equation (D.6) reduces to $(1 - S'(q_L))\frac{r\Phi'(\cdot)}{1+r}$. Therefore, (D.6) can be rewritten as follows:

$$-S'(q_L) + \frac{Y_L\phi'(e_L^*)}{\phi(e_L^*)} \frac{T'(Y_L)}{1+r} - (1 - S'(q_L))\frac{r\Phi'(\cdot)}{1+r} = \frac{\gamma}{\lambda\pi_L} \frac{v_\ell(\hat{\ell}^*)}{a_H\phi(\hat{e}^*)} \left[\frac{Y_L\phi'(\hat{e}^*)}{\phi(\hat{e}^*)} - \frac{Y_L\phi'(e_L^*)}{\phi(e_L^*)} \right] \tag{D.7}$$

Using equation (11) and the definition of Δ, the left-hand side equals $\frac{\Delta}{1+r}$. Finally, applying equation (18) to the right-hand side, we obtain equation (26). $\square$

Appendix E

In this Appendix, we now show that both the Atkinson and Stiglitz (1976) theorem and the Diamond and Mirrlees (1971) production efficiency theorem are valid when the government can observe the share of educational investment with consumption value, h_i, that is, the government can control h_i directly in contrast with the preceding analysis. This means that we do not need to consider the individual's problem that determines the level of h_i under an assigned allocation. In the setting, education subsidies are redefined as $S(x_i, e_i)$ depending on each component separately since the government can observe both x_i and e_i. Now, we consider that individuals choose $x_i = h_i q_i$ and $e_i = (1 - h_i)q_i$ as control variables instead of h_i and q_i, and thus, the individual's optimization problem is formulated by

$$
\begin{aligned}
\max_{c_i^1, c_i^2, s_i, x_i, e_i, Y_i} \quad & u(c_i^1, c_i^2, x_i) - v\left(\frac{Y_i}{a_i\phi(e_i)}\right) \\
\text{s.t.} \quad & c_i^1 + s_i + x_i + e_i - S(x_i, e_i) = s_0 \\
& c_i^2 = Y_i - T(Y_i) + (1+r)s_i - \Phi(rs_i)
\end{aligned}
\tag{E.1}
$$

This problem yields the first-order conditions:

$$MRS_{c^1x}^i \equiv \frac{u_x(c_i^1, c_i^2, x_i)}{u_c^1(c_i^1, c_i^2, x_i)} = 1 - \frac{\partial S(x_i, e_i)}{\partial x_i} \equiv 1 - S_{x_i}(x_i, e_i) \tag{E.2}$$

$$MRS_{c^1e}^i \equiv \frac{v_\ell(\ell_i)}{u_c^1(c_i^1, c_i^2, x_i)} \frac{\ell_i\phi'(e_i)}{\phi(e_i)} = 1 - \frac{\partial S(x_i, e_i)}{\partial e_i} \equiv 1 - S_{e_i}(x_i, e_i) \tag{E.3}$$

$$MRS^i_{c^2\ell} \equiv \frac{v_\ell(\ell_i)}{a_i\phi(e_i)u_c^2(c_i^1, c_i^2, x_i)} = 1 - T'(Y_i) \tag{E.4}$$

$$MRS^i_{c^1c^2} \equiv \frac{u_c^1(c_i^1, c_i^2, x_i)}{u_c^2(c_i^1, c_i^2, x_i)} = 1 + r - r\Phi'(rs_i) \tag{E.5}$$

Combining (E.3), (E.4), and (E.5) yields

$$MRT^i_{e\ell} \equiv \frac{Y_i\phi'(e_i)}{\phi(e_i)} = \frac{1 - S_{e_i}(x_i, e_i)}{1 - T'(Y_i)}(1 + r - r\Phi'(rs_i)) \tag{E.6}$$

From the same manner of equation (12), the total net tax wedge on learning for type i is

$$\Delta_i = \frac{T'(\cdot)}{1 - T'(\cdot)}R(1 - S_{e_i}(\cdot)) - r\Phi'(\cdot) - S_{e_i}(\cdot)R \tag{E.7}$$

We now turn to the analysis of the government optimization problem. The objective of the government, the public budget constraint, and the incentive constraint are given by

$$W = \sum_i \pi^i\{u(c_i^1, c_i^2, x_i) - v(\frac{Y_i}{a_i\phi(e_i)})\} \tag{E.8}$$

$$\sum_{i=H,L} \pi_i\left[s_0 - c_i^1 - x_i - e_i + \frac{1}{1+r}(Y_i - c_i^2)\right] = 0 \tag{E.9}$$

$$u(c_H^1, c_H^2, x_H) - v(\frac{Y_H}{a_H\phi(e_H)}) \geq u(c_L^1, c_L^2, x_L) - v(\frac{Y_L}{a_H\phi(e_L)}) \tag{E.10}$$

Therefore, the government maximizes the social welfare function (E.8) subject to the government's budget constraint (E.9) and the incentive constraint (E.10) by selecting the allocations with respect to c_i^1, c_i^2, x_i, e_i, and Y_i for type i. The corresponding Lagrangian is

$$\max_{\{c_i^1, c_i^2, x_i, e_i, Y_i\}_i} \tilde{\mathcal{L}} = \sum_i \pi^i\{u(c_i^1, c_i^2, x_i) - v(\frac{Y_i}{a_i\phi(e_i)})\}$$
$$+ \lambda\left[\sum_{i=H,L} \pi_i\{s_0 - c_i^1 - x_i - e_i + \frac{1}{1+r}(Y_i - c_i^2)\}\right] \tag{E.11}$$
$$+ \gamma\left[u(c_H^1, c_H^2, x_H) - v(\frac{Y_H}{a_H\phi(e_H)}) - u(c_L^1, c_L^2, x_L) + v(\frac{Y_L}{a_H\phi(e_L)})\right]$$

The first-order conditions with respect to c_i^1, c_i^2, x_i, e_i, and Y_i, $i = H, L$, are given by:

$$\frac{\partial\tilde{\mathcal{L}}}{\partial c_L^1} = \pi_L u_c^1(c_L^1, c_L^2, x_L) - \gamma u_c^1(c_L^1, c_L^2, x_L) - \lambda\pi_L = 0 \tag{E.12}$$

$$\frac{\partial\tilde{\mathcal{L}}}{\partial c_H^1} = \pi_H u_c^1(c_H^1, c_H^2, x_H) + \gamma u_c^1(c_H^1, c_H^2, x_H) - \lambda\pi_H = 0 \tag{E.13}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial c_L^2} = \pi_L u_c^2(c_L^1, c_L^2, x_L) - \gamma u_c^2(c_L^1, c_L^2, x_L) - \frac{1}{1+r}\lambda \pi_L = 0 \tag{E.14}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial c_H^2} = \pi_H u_c^2(c_H^1, c_H^2, x_H) + \gamma u_c^2(c_H^1, c_H^2, x_H) - \frac{1}{1+r}\lambda \pi_H = 0 \tag{E.15}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial x_L} = \pi_L u_x(c_L^1, c_L^2, x_L) - \gamma u_x(c_L^1, c_L^2, x_L) - \lambda \pi_L = 0 \tag{E.16}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial x_H} = \pi_H u_x(c_H^1, c_H^2, x_H) + \gamma u_x(c_H^1, c_H^2, x_H) - \lambda \pi_H = 0 \tag{E.17}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial e_L} = \pi_L v_\ell(\ell_L)\frac{\ell_L \phi'(e_L)}{\phi(e_L)} - \gamma v_\ell(\hat{\ell})\frac{\hat{\ell} \phi'(e_L)}{\phi(e_L)} - \lambda \pi_L = 0 \tag{E.18}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial e_H} = \pi_H v_\ell(\ell_H)\frac{\ell_H \phi'(e_H)}{\phi(e_H)} - \gamma v_\ell(\ell_H)\frac{\ell_H \phi'(e_H)}{\phi(e_H)} - \lambda \pi_H = 0 \tag{E.19}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial Y_L} = -\pi_L v_\ell(\ell_L)\frac{1}{a_L \phi(e_L)} + \gamma v_\ell(\hat{\ell})\frac{1}{a_H \phi(e_L)} + \frac{1}{1+r}\lambda \pi_L = 0 \tag{E.20}$$

$$\frac{\partial \tilde{\mathcal{L}}}{\partial Y_H} = -\pi_H v_\ell(\ell_H)\frac{1}{a_H \phi(e_H)} - \gamma v_\ell(\ell_H)\frac{1}{a_H \phi(e_H)} + \frac{1}{1+r}\lambda \pi_H = 0 \tag{E.21}$$

where $\hat{\ell} = \frac{Y_L}{a_H \phi(e_L)}$. Using the first-order conditions (E.12)-(E.21) and equations (E.2)-(E.5), we characterize optimal conditions for each type with respect to the marginal tax rate on educational investment with consumption value, the marginal subsidy rate for educational investment with production value, and the marginal labor and capital income tax rate as follows.

$$S_{x_H}(x_H, e_H) = S_{x_L}(x_L, e_L) = 0 \tag{E.22}$$

$$\Phi'(rs_H) = \Phi'(rs_L) = 0 \tag{E.23}$$

$$T'(Y_H) = 0 \tag{E.24}$$

$$\frac{T'(Y_L)}{1+r} = \frac{\gamma u_c^2(c_L^1, c_L^2, x_L)}{\lambda \pi_L}\left[\frac{v_\ell(\ell_L)}{u_c^2(c_L^1, c_L^2, x_L)}\frac{1}{a_L \phi(e_L)} - \frac{v_\ell(\hat{\ell})}{u_c^2(c_L^1, c_L^2, x_L)}\frac{1}{a_H \phi(e_L)}\right] > 0 \tag{E.25}$$

$$S_{e_H}(x_H, e_H) = 0 \tag{E.26}$$

$$S_{e_L}(x_L, e_L) = \frac{\gamma u_c^1(c_L^1, c_L^2, x_L)}{\lambda \pi_L}\left[\frac{v_\ell(\ell_L)}{u_c^1(c_L^1, c_L^2, x_L)}\frac{\ell_L \phi'(e_L)}{\phi(e_L)} - \frac{v_\ell(\hat{\ell})}{u_c^1(c_L^1, c_L^2, x_L)}\frac{\hat{\ell} \phi'(e_L)}{\phi(e_L)}\right] > 0 \tag{E.27}$$

We omit the derivations of equations (E.22)-(E.27) since we can obtain them by the similar way in Appendix B. Equations (E.22) and (E.23) mean that the differential tax on educational investment with consumption value and capital income is superfluous and correspond to the canonical result of Atkinson and Stiglitz (1976), which states that nonlinear income taxes are only needed if the utility function is weakly separable between consumption and labor supply. Equations (E.24) and (E.25) are consistent with the result of Stiglitz (1982) analyzing optimal labor income taxation in the economy consisting of two types of individuals. The marginal income tax rate is zero at the top and positive at the

bottom. Equations (E.26) and (E.27) suggest that the marginal subsidy for educational investment with production value is zero for high-type and positive for low-type. Here, we rewrite equation (E.25) as $\frac{T'(Y_L)}{1+r} = \frac{\gamma}{\lambda \pi_L}\{v_\ell(\ell_L)\frac{1}{a_L\phi(e_L)} - v_\ell(\hat{\ell})\frac{1}{a_H\phi(e_L)}\}$ and equation (E.27) as $S_{e_L}(x_L, e_L) = \frac{\gamma}{\lambda \pi_L}\frac{Y_L\phi'(e_L)}{\phi(e_L)}\{v_\ell(\ell_L)\frac{1}{a_L\phi(e_L)} - v_\ell(\hat{\ell})\frac{1}{a_H\phi(e_L)}\}$. Combining these equations yields $S_{e_L}(x_L, e_L) = \frac{Y_L\phi'(e_L)}{\phi(e_L)}\frac{T'(Y_L)}{1+r}$. Moreover, substituting equation (E.6) into it and then using equation (E.23) yields $S_{e_L}(x_L, e_L) = \frac{1-S_{e_L}(x_L,e_L)}{1-T'(Y_L)}T'(Y_L)$. Finally, we can obtain the following relationship:

$$S_{e_L}(x_L, e_L) = T'(Y_L) \tag{E.28}$$

Therefore, the marginal subsidy rate for productive human capital investment should equal the marginal labor income tax rate across the two types. This result implies that the government ought to improve the distortion of individual's behavior with respect to human capital investment due to labor income taxation by subsidizing educational investment. From the above results, we have $\Delta_H = \Delta_L = 0$, which means that the Diamond-Mirrlees production efficiency theorem holds. This is consistent with Bovenberg and Jacobs (2005). Hence, all standard results in the traditional literature appear. $\square$

Appendix F

In the Appendix, we consider that h_i is exogenously given and identical between individuals, that is, $h_H = h_L \equiv h$ to replicate the result of Bovenberg and Jacobs (2005) showing that the production efficiency theorem is robust even if education subsidies are greater than marginal labor income tax rates in the presence of non-pecuniary costs. In the setting, individual optimization problem is

$$
\begin{aligned}
\max_{c_i^1, c_i^2, s_i, q_i, Y_i} \quad & u(c_i^1, c_i^2, x_i) - v\left(\frac{Y_i}{a_i\phi(e_i)}\right) \\
s.t. \quad & c_i^1 + s_i + q_i - S(q_i) = s_0 \\
& c_i^2 = Y_i - T(Y_i) + (1+r)s_i - \Phi(rs_i)
\end{aligned} \tag{F.1}
$$

This problem yields the first-order conditions:

$$MRS_{c^1 q}^i \equiv \frac{u_x(c_i^1, c_i^2, x_i)}{u_c^1(c_i^1, c_i^2, x_i)}h + \frac{v_\ell(\ell_i)}{u_c^1(c_i^1, c_i^2, x_i)}\frac{\ell_i\phi'(e_i)}{\phi(e_i)}(1-h) = 1 - S'(q_i) \tag{F.2}$$

$$MRS_{c^2 \ell}^i \equiv \frac{v_\ell(\ell_i)}{a_i\phi(e_i)u_c^2(c_i^1, c_i^2, x_i)} = 1 - T'(Y_i) \tag{F.3}$$

$$MRS_{c^1 c^2}^i \equiv \frac{u_c^1(c_i^1, c_i^2, x_i)}{u_c^2(c_i^1, c_i^2, x_i)} = 1 + r - r\Phi'(rs_i) \tag{F.4}$$

Combining (F.2), (F.3), and (F.4) yields

$$MRT_{q\ell}^i \equiv \frac{Y_i\phi'(e_i)}{\phi(e_i)}(1-h) = \frac{1-S'(q_i)}{1-T'(Y_i)}(1+r-r\Phi'(rs_i))(1-\Omega_i) \tag{F.5}$$

where $\Omega_i \equiv \frac{\frac{u_x(c_i^1, c_i^2, x_i)}{u_c^1(c_i^1, c_i^2, x_i)} h}{1 - S'(q_i)} < 0$ denotes the ratio between utility costs and material costs. The total net tax wedge on learning for type i is

$$\Delta_i = \frac{T'(\cdot)}{1 - T'(\cdot)} R(1 - S'(q_i))(1 - \Omega_i) - r\Phi'(\cdot) - S'(q_i)R + r\Phi'(\cdot)\Omega_i(1 - S'(q_i)) \qquad \text{(F.6)}$$

Therefore, if $\Delta_i = 0$, $MRT_{q\ell}^i$ corresponds to the condition without any tax policy, that is, $MRT_{q\ell}^i = (1+r)(1 - \frac{u_x(c_i^1, c_i^2, x_i)}{u_c^1(c_i^1, c_i^2, x_i)} h)$. In particular, when $\Phi'(rs_i)$ is zero, equation (F.6) reduces to

$$\frac{\Delta_i}{1 + r} = \frac{T'(\cdot)}{1 - T'(\cdot)}(1 - S'(q_i))(1 - \Omega_i) - S'(q_i) \qquad \text{(F.7)}$$

Except for zero tax/subsidy rate, equation (F.7) means that $S'(q_i)$ is larger than $T'(Y_i)$ to ensure efficiency in human capital formation. Indeed, when $S'(q_i) = \frac{1 - \Omega_i}{1 - \Omega_i T'(Y_i)} T'(Y_i) > T'(Y_i)$, Δ_i is zero. Of course, the condition to ensure full efficiency in human capital formation is $S'(q_i) = T'(Y_i)$ if $h_H = h_L = 0$.

Here, we examine the government optimization problem. The problem is formulated by the same manner, and the corresponding Lagrangian is

$$\max_{\{c_i^1, c_i^2, q_i, Y_i\}_i} \bar{\mathcal{L}} = \sum_i \pi^i \{u(c_i^1, c_i^2, x_i) - v(\frac{Y_i}{a_i\phi(e_i)})\}$$
$$+ \lambda \left[\sum_{i=H,L} \pi_i \{s_0 - c_i^1 - q_i + \frac{1}{1+r}(Y_i - c_i^2)\} \right] \qquad \text{(F.8)}$$
$$+ \gamma \left[u(c_H^1, c_H^2, x_H) - v(\frac{Y_H}{a_H\phi(e_H)}) - u(c_L^1, c_L^2, x_L) + v(\frac{Y_L}{a_H\phi(e_L)}) \right]$$

Hereafter, we focus on the characterization of the optimal policy for low-type individuals. The first-order conditions with respect to c_L^1, c_L^2, q_L, and Y_L are given by:

$$\frac{\partial \bar{\mathcal{L}}}{\partial c_L^1} = \pi_L u_c^1(c_L^1, c_L^2, x_L) - \gamma u_c^1(c_L^1, c_L^2, x_L) - \lambda \pi_L = 0 \qquad \text{(F.9)}$$

$$\frac{\partial \bar{\mathcal{L}}}{\partial c_L^2} = \pi_L u_c^2(c_L^1, c_L^2, x_L) - \gamma u_c^2(c_L^1, c_L^2, x_L) - \frac{1}{1+r}\lambda \pi_L = 0 \qquad \text{(F.10)}$$

$$\frac{\partial \bar{\mathcal{L}}}{\partial q_L} = \pi_L \left[u_x(c_L^1, c_L^2, x_L)h + v_\ell(\ell_L)\frac{\ell_L \phi'(e_L)}{\phi(e_L)}(1-h) \right]$$
$$- \gamma \left[u_x(c_L^1, c_L^2, x_L)h + v_\ell(\hat{\ell})\frac{\hat{\ell}\phi'(e_L)}{\phi(e_L)}(1-h) \right] - \lambda \pi_L = 0 \qquad \text{(F.11)}$$

$$\frac{\partial \bar{\mathcal{L}}}{\partial Y_L} = -\pi_L v_\ell(\ell_L)\frac{1}{a_L \phi(e_L)} + \gamma v_\ell(\hat{\ell})\frac{1}{a_H \phi(e_L)} + \frac{1}{1+r}\lambda \pi_L = 0 \qquad \text{(F.12)}$$

From equations (F.4), (F.9), and (F.10), the marginal capital income tax rate is zero for low-type individuals, that is, $\Phi'(rs_L) = 0$. Also, from equations (F.2), (F.9), and (F.11), the

marginal education subsidy rate is characterized by

$$S'(q_L) = \frac{\gamma u_c^1(c_L^1, c_L^2, x_L)}{\lambda \pi^L} \left[\frac{v_\ell(\ell_L)}{u_c^1(c_L^1, c_L^2, x_L)} \frac{\ell_L \phi'(e_L)}{\phi(e_L)} - \frac{v_\ell(\hat{\ell})}{u_c^1(c_L^1, c_L^2, x_L)} \frac{\hat{\ell} \phi'(e_L)}{\phi(e_L)} \right] (1 - h) \quad \text{(F.13)}$$

Moreover, from equations (F.3), (F.10), and (F.12), the marginal labor income rate is given by

$$\frac{T'(Y_L)}{1+r} = \frac{\gamma u_c^2(c_L^1, c_L^2, x_L)}{\lambda \pi_L} \left[\frac{v_\ell(\ell_L)}{u_c^2(c_L^1, c_L^2, x_L)} \frac{1}{a_L \phi(e_L)} - \frac{v_\ell(\hat{\ell})}{u_c^2(c_L^1, c_L^2, x_L)} \frac{1}{a_H \phi(e_L)} \right] \quad \text{(F.14)}$$

These results are obtained by the same method in Appendix B. Now, we rewrite equation (E.13) as $S'(q_L) = \frac{\gamma}{\lambda \pi_L} \frac{Y_L \phi'(e_L)}{\phi(e_L)}(1-h)\{v_\ell(\ell_L)\frac{1}{a_L \phi(e_L)} - v_\ell(\hat{\ell})\frac{1}{a_H \phi(e_L)}\}$ and equation (E.14) as $\frac{T'(Y_L)}{1+r} = \frac{\gamma}{\lambda \pi_L}\{v_\ell(\ell_L)\frac{1}{a_L \phi(e_L)} - v_\ell(\hat{\ell})\frac{1}{a_H \phi(e_L)}\}$. Combining these equations yields $S'(q_L) = \frac{Y_L \phi'(e_L)}{\phi(e_L)}(1-h)\frac{T'(Y_L)}{1+r}$. Moreover, substituting equation (F.5) into it and then using the fact that $\Phi'(rs_L)$ is zero yields $S'(q_L) = \frac{1-S'(q_i)}{1-T'(Y_i)}(1-\Omega_i)T'(Y_L)$. Rearranging this yields $S'(q_i) = \frac{1-\Omega_i}{1-\Omega_i T'(Y_i)}T'(Y_i)$. Thus, although education subsidies are larger than marginal labor income taxes, production efficiency theorem is robust. $\square$

References

Abramitzky, R., Lavy, V., 2014. How responsive is investment in schooling to changes in redistributive policies and in returns? Econometrica 82, 1241–1272.

Alstadsæter, A., 2003. Does the tax system encourage too much education? FinanzArchiv 59, 27–48.

Alstadsæter, A., 2011. Measuring the consumption value of higher education. CESifo Economics Studies 57, 458–479.

Anderberg, D., 2009. Optimal policy and the risk properties of human capital reconsidered. Journal of Public Economics 93, 1017–1026.

Anderberg, D., Andersson, F., 2003. Investments in human capital, wage uncertainty, and public policy. Journal of Public Economics 87, 1521–1537.

Arcidiacono, P., 2004. Ability sorting and the returns to college major. Journal of Econometrics 121, 343–375.

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

Banks, J., Diamond, P. A., 2010. The Base for Direct Taxation, Chap.6 in *Dimensions of Tax Design: The Mirrlees Review*, ed. by the Institution for Fiscal Studies, 548–648. Oxford, UK: Oxford University Press.

Blomquist, S., Micheletto, L., 2008. Age-related optimal income taxation. Scandinavian Journal of Economics 110, 45–71.

Boadway, R., Marchand, M., Pestieau, P., 2000. Redistribution with unobservable bequests: a case for taxing capital income. Scandinavian Journal of Economics 102, 253–267.

Bohacek, R., Kapička, M., 2008. Optimal human capital policies. Journal of Monetary Economics 55, 1–16.

Bovenberg, A. L., Jacobs, B., 2005. Redistribution and education subsidies are Siamese twins. Journal of Public Economics 89, 2005–2035.

Carneiro, P., Hansen, K. T., Heckman, J. J., 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty of college choice. International Economic Review 44, 361–422.

Cremer, H., Pestieau, P., Rochet, J.-C., 2001. Direct versus indirect taxation: the design of the tax structure revisited. International Economic Review 42, 781–800.

Da Costa, C. E., Maestri, L. J., 2007. The risk properties of human capital and the design of government policies. European Economic Review 51, 695–713.

Diamond, P., Mirrlees, J., 1971. Optimal taxation and public production I: production efficiency. American Economic Review 61, 8–27.

Diamond, P. A., Spinnewijn, J., 2011. Capital income taxes with heterogeneous discount rates. American Economic Journal: Economic Policy 3, 52–76.

Dynan, K. E., Skinner, J., Zeldes, S. P., 2004. Do the rich save more? Journal of Political Economy 112, 397–444.

Eaton, J., Rosen, H. S., 1980. Taxation, human capital, and uncertainty. American Economic Review 70, 705–715.

Findeisen, S., Sachs, D., 2016. Education and optimal dynamic taxation: the role of income-contingent student loans. Journal of Public Economics 138, 1–21.

Findeisen, S., Sachs, D., 2018. Education policies and taxation without commitment. Scandinavian Journal of Economics (forthcoming).

Grochulski, B., Piskorski, T., 2010. Risky human capital and deferred capital income taxation. Journal of Economic Theory 145, 908–943.

Gullason, E. T., 1989. The consumption value of schooling: an empirical estimate of one aspect. Journal of Human Resources 24, 287–298.

Guo, J.-T., Krause, A., 2013. Optimal nonlinear taxation of income and education expenditure. Oxford Economic Papers 65, 74–95.

Hamilton, J. H., 1987. Optimal wage and income taxation with wage uncertainty. Journal of Economic Theory 28, 373–388.

Hare, P. G., Ulph, D. T., 1979. On education and distribution. Journal of Political Economy 87, 193–212.

Heckman, J. J., Lochner, L., Taber, C., 1999. General-equilibrium cost-benefit analysis of education and tax policies, in G. Ranis and L.K. Raut, eds. Trade, Growth, and Development, Elsevier, Amsterdam, pp.291-349.

Heckman, J. J., Lochner, L. J., Todd, P. E., 2006. Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, Chap.7 in *Handbook of the Economics of Education*, ed. 307–458. Amsterdam: North-Holland.

Jacobs, B., 2005. Optimal income taxation with endogenous human capital. Journal of Public Economic Theory 7, 295–315.

Jacobs, B., 2013. Optimal redistributive tax and education policies in general equilibrium. International Tax and Public Finance 20, 312–337.

Jacobs, B., Bovenberg, A. L., 2010. Human capital and optimal positive taxation of capital income. International Tax and Public Finance 17, 451–478.

Jacobs, B., Bovenberg, A. L., 2011. Optimal taxation of human capital and the earnings function. Journal of Public Economic Theory 13, 957–971.

Jacobs, B., Schindler, D., Yang, H., 2012. Optimal taxation of risky human capital. Scandinavian Journal of Economics 114, 908–931.

Jacobs, B., Yang, H., 2016. Second-best income taxation and education policy with endogenous human capital and borrowing constraints. International Tax and Public Finance 20, 312–337.

Kapička, M., 2015. Optimal Mirrleesean taxation in a Ben-Porath economy. American Economic Journal: Macroeconomics 7, 219–248.

Kapička, M., Neira, J., 2015. Optimal taxation with risky human capital, CERGE-EI Working Paper Series No.553.

Kodde, D. A., Ritzen, J. M. M., 1984. Integrating consumption and investment motives in a neoclassical model of demand for education. KyKlos 37, 598–605.

Krause, A., 2006. Redistributive taxation and public education. Journal of Public Economic Theory 8, 807–819.

Lawrance, E. C., 1991. Poverty and the rate of time preference: evidence from panel data. Journal of Political Economy 99, 54–77.

Lazear, E., 1977. Education: consumption or production? Journal of Political Economy 85, 569–598.

Malchow-Møller, N., Nielsen, S. B., Skaksen, J. R., 2011. Taxes, tuition fees, and education for pleasure. Journal of Public Economic Theory 13, 189–215.

Maldonado, D., 2008. Education policies and optimal taxation. International Tax and Public Finance 15, 131–143.

Mirrlees, J., 1971. An exploration in the theory of optimum income taxation. Review of Economic Studies 38, 175–208.

Mirrlees, J., 1976. Optimum tax theory: a synthesis. Journal of Public Economics 6, 327–358.

Nielsen, S. B., Sørensen, P. B., 1997. On the optimality of the Nordic system of dual income taxation. Journal of Public Economics 63, 311–329.

Ordover, J. A., Phelps, E. S., 1979. The concept of optimal taxation in the overlapping-generations model of capital and wealth. Journal of Public Economics 12, 1–26.

Palacios-Huerta, I., 2003. An empirical analysis of the risk properties of human capital returns. American Economic Review 93, 948–964.

Palacios-Huerta, I., 2006. The human capital premium puzzle, unpublished manuscript.

Pirttilä, J., Tuomala, M., 2001. On optimal non-linear taxation and public good provision in an overlapping generations economy. Journal of Public Economics 79, 485–501.

Saez, E., 2002. The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. Journal of Public Economics 83, 217–230.

Saez, E., 2013. Optimal progressive capital income taxes in the infinite horizon model. Journal of Public Economics 97, 61–74.

Schaafsma, J., 1976. The consumption and investment aspects of the demand for education. Journal of Human Resources 11, 233–242.

Schindler, D., 2011. Tuition fees and the dual income tax: the optimality of the Nordic income tax system reconsidered. German Economic Review 12, 59–84.

Schindler, D., Yang, H., 2015. Catalysts for social insurance: education subsidies versus physical capital taxation. International Tax and Public Finance 22, 274–310.

Stantcheva, S., 2017. Optimal taxation and human capital policies over the life cycle. Journal of Political Economy 125, 1931–1990.

Stern, N., 1982. Optimum taxation with errors in administration. Journal of Public Economics 17, 181–211.

Stiglitz, J. E., 1982. Self-selection and Pareto efficient taxation. Journal of Public Economics 17, 213–240.

Trostel, P., 1993. The effect of taxation on human capital. Journal of Political Economy 101, 327–350.

Trostel, P., Walker, I., Woolley, P., 2002. Estimates of the economic return to schooling for 28 countries. Labour Economics 9, 1–16.

Tuomala, M., 1986. On the optimal income taxation and educational decisions. Journal of Public Economics 30, 183–198.

Ulph, D. T., 1977. On the optimal distribution of income and educational expenditure. Journal of Public Economics 8, 341–356.

Walker, L., Zhu, Y., 2003. Education, earnings and productivity: recent UK evidence. Labour Market Trends 111, 145–152.