Int Tax Public Finance (2008) 15: 527–545
DOI 10.1007/s10797-008-9076-x

On taxing capital income with income shifting

Vidar Christiansen · Matti Tuomala

Published online: 16 May 2008
© Springer Science+Business Media, LLC 2008

Abstract We examine a linear capital income tax and a nonlinear labor income tax in a two-type model where individuals live for two periods. We assume that taxes are paid only in the second period in which the agents receive both labor and capital income and may shift income from labor to capital. The two types of individuals may differ with respect to wage rate and initial resource endowments. In the absence of income shifting, endowment variation motivates a capital income tax which would not exist where there is pure wage rate variation. In the latter circumstance, income shifting would indeed establish a case for a capital income tax while adding variation in resource endowments would ambiguously affect the case. The asymmetric information case for a capital income tax must be traded off against distortionary effects not only on savings, but also on labor as an agent may earn labor income which is reported and taxed as capital income.

Keywords Capital taxation · Income shifting

JEL Classification H21 · H24

1 Introduction

One of the key issues in the optimal tax literature is how to differentiate taxes on labor income and capital income, respectively. A number of arguments have been

V. Christiansen
University of Oslo, Oslo, Norway

M. Tuomala (✉)
University of Tampere, Tampere, Finland
e-mail: matti.tuomala@kolumbus.fi

put forward in favor of a low tax or no tax at all on capital.[1] As is well known, due to Atkinson and Stiglitz (1976), under a mild separability assumption, labor income taxation does not need be supplemented by other taxes. The Atkinson–Stiglitz result has been subject to considerable scrutiny in the literature, and special attention has been devoted to circumstances in which it is violated. Naito (1999) has shown that it does not hold with nonlinear technology. Cremer et al. (2001) show that when individuals differ in several unobservable characteristics differential commodity taxation is desirable even with separability. Boadway et al. (2000) and Cremer et al. (2003) in turn show that where skill and inherited wealth are not observable a capital income tax might become desirable even with separable preferences. A major focus of these two papers is to study capital income taxation as an instrument to indirectly tax inherited wealth, particularly when it is not perfectly correlated with skills. Saez (2002) argues that the Atkinson–Stiglitz result of commodity taxes holds when individuals have identical tastes. In the context of present and future consumption as two kinds of consumer goods, it is argued that individuals with higher earnings save relatively more, which suggests that high skill individuals are more likely to have higher discount factors. Pirttilä and Tuomala (2001), on the other hand, have shown that capital income taxation may be desirable even when preferences are separable across time where future relative wages are sensitive to current savings via their effect on capital income. In the context of an OLG model where the concern is intergenerational redistribution, Blackorby and Brett (2000) show that capital income taxes can implement parts of the Pareto frontier that would otherwise be unattainable.

A typical assumption in the theoretical literature is that the tax authority can easily distinguish the respective types of income. In practice, however, this is far from easy as has been realized and discussed at length in connection with the dual income tax in the Nordic countries.[2] Norway witnessed extensive circumvention of the income splitting model introduced by the 1992 tax reform. Rather than having a substantial part of their income taxed as labor income, many entrepreneurs found ways to have it taxed much more leniently at the rate applied to capital income.[3] This problem was a major motivation for appointing a tax reform committee (the Skauge committee) that

[1] The result that no capital tax is optimal arises in a model with infinitely-lived individuals, see Judd (1985) and Chamley (1986). There have been views on capital income taxation that deviate from the Chamley–Judd prescription. For example, Leif Johansen writes in his book *Public Economics* (Johansen 1965) "In some countries income derived from wealth has been regarded as a surer and more permanent form of income than earned income, and it has therefore been considered that a definite amount of an income of this kind provided a higher tax ability than a corresponding amount of other income, within a particular period." (p. 197) Jim Mirrlees (2000) in turn writes "some of variations in the return on capital are the result of the application of the skill and effort: but most is surely the result of risky outcomes. To that extent, there might be advantage in a high tax on the returns, offset by a subsidy on the capital; for that would provide people with insurance against investment risks. When one then takes account of the redistributive element of taxation, there is case for taxing wealth... I suggest there is a case for a rather progressive tax on income from capital after all, with perhaps some small offset related to capital value." (p. 8).

[2] Pirttilä and Selin (2006) provides empirical evidence on income shifting in Finland.

[3] In particular, many active owners of corporations have escaped the split model. One way to do this was to invite more passive owners into the company to bring the ownership share of active owners below 66%. Between 1992 and 2000, the percentage of corporations subject to income splitting fell from 55 to 32. By avoiding mandatory income splitting, the owners were free to work for a very low official salary, while reaping large dividends.

delivered its report in 2003 and led to the introduction of a new shareholder income tax with a marginal tax rate close to that of the labor income tax in order to remove the motivation for artificially channeling the income earned through a corporation into dividends or capital gains.

In this paper, we will assume that labor income can be camouflaged as capital income, but only at a cost so that if a tax relief can be achieved by converting capital income into taxable labor income, this will be done to the extent that the marginal tax saving exceeds the marginal cost of transforming the former type of income to the latter. The issue of fiscal manipulation in the form of income shifting has also received much attention in the context of the US tax reform 1986. Gordon and Slemrod (1998) have argued that a large part of the response observable in the tax return was due to income shifting between the corporate sector and the individual sector.

Our approach differs from previous ones in a couple of notable respects. While a capital income tax is sometimes assumed to be the only available instrument for certain types of redistribution (e.g., between those who inherit and those who do not as in Boadway et al. 2000), in principle any redistribution might in our models be achieved by means of a nonlinear tax on earnings. The question is whether imposing a tax on capital income may still be desirable. Income shifting is sometimes related to the choice between being an entrepreneur or a wage earner as in Gordon and MacKie-Mason (1995) where the role of the corporate income tax is to prevent a distortion of this choice under income shifting. We do not model this choice explicitly, but one may conceive of income reported as capital income as being earned as business income. A more important aspect of our model is that with differential taxation of labor and capital, income shifting allows the remuneration of marginal labor supply to be taxed at the capital tax rate. As in Gordon and MacKie-Mason, the reporting does not involve any tax evasion, but is conditional on making use of perfectly legal ways to organize one's economic activities.

Fuest and Huber (2001) discussed optimal taxes of labor and capital in a model with income shifting, but in other respects, their model is different from ours as we shall detail below after presenting the main structure of our model.

We will use the simplest possible model capturing the essentials of the problem by making a number of assumptions that allow us to use a two-type, asymmetric information model of nonlinear income taxation bearing strong resemblance to the model of Stern (1982) and Stiglitz (1982). The two types of persons, labeled L and H, are endowed with productivities (skill levels) that are reflected by their respective wage rates w^{L} and $w^{\mathrm{H}} > w^{\mathrm{L}}$. The two-type model has been used to analyze labor and capital taxation in overlapping generation models. It is typically assumed that each generation lives for two periods, but is work active, earns income, and pays taxes on labor income only in the first period. This is a useful simplification, implying that one can focus on the labor income tax of an agent in a single period, but it means that the model fails to capture the problems that arise when an agent is due to pay labor and capital taxes in the same period. To include the latter concern, we shall consider a model in which the agent works and pays labor taxes in the second period when he receives the return to capital accumulated in the first period. As further simplifications, we assume that wage rates are constant over time, and that there is a fixed rate of return to savings, which may be justified by assuming that we consider a small open economy facing a world capital market (the same assumption as in Salanié 2003). To establish a benchmark close to previous models, we shall start out by considering the case in which labor income and capital income can indeed be perfectly distinguished.

The rest of the paper is organized as follows. In Sect. 2, we establish the basic structure of our model by setting up a simple benchmark model in which we address taxation of savings in a two-type model where individuals live two periods. In order to fully retain the simplicity of the conventional model, we make the assumption that the young pay no taxes. In Sects. 3.1 and 3.2, we provide the formulas for linear taxation of savings without and with income shifting. Section 4 concludes.

2 A benchmark model

The classic optimal income tax model, Mirrlees (1971), treats differences in observed income as being due to unobserved differences in ability. We will in general assume that individuals do not differ only in ability, but also in initial endowments, denoted by e. In our model, the initial endowment may be interpreted as representing various factors affecting capital income. Beyond representing a tangible asset, say in terms of inherited wealth or exogenous labor income, it may liberally be interpreted as representing entrepreneurial skill, family background, social and business networks, and other circumstances that are conducive to earning capital income. It is also quite plausible to assume that in reality both ability and endowment are unobservable. This may be more plausible for intangible assets, but in practice, there are a number of nontransparent ways in which even tangible assets can be transferred from one generation to the next.[4] First best taxation is not feasible in this economy because we cannot distinguish ex ante between the two types. Thus, only anonymous tax systems are feasible.

There is a total of four types of individuals. To avoid difficulties with multidimensional optimal tax problems variation is restricted to two types by assuming that there are only two fixed w, e bundles ($w^{\mathrm{H}}, e^{\mathrm{H}}$ and $w^{\mathrm{L}}, e^{\mathrm{L}}$)[5] where we either assume that $e^{\mathrm{H}} > e^{\mathrm{L}}$ or $e^{\mathrm{H}} = e^{\mathrm{L}}$. This framework allows us to consider as special cases either variation in initial resource endowments or in skill, but also the case where one type is more richly endowed with both initial resources and skill. The latter case may be justified as an approximation to reality as there is evidence of a strong positive correlation between the two characteristics.[6] Multidimensional variation would pose serious problems of tractability that are likely to require numerical computations.

Each agent supplies h units of labor in the second period. The labor market is perfectly competitive so that an individual's effective labor supply equals his or her gross income, $z = wh$. To simplify the exposition and notation without loss of generality, we consider the case with an equal number of individuals of each type. The

[4] This is a key point in Cremer et al. (2003).

[5] This is a less restrictive way to avoid multidimensional variation than the pure endowment variation adopted in Fuest and Huber (2001).

[6] See Pekkarinen et al. (1985).

government wishes to design a tax system that may redistribute income between individuals. There is asymmetric information in the sense that the tax authority is informed neither about individual skill levels and labor supply nor endowments. To introduce returns to capital and the possible taxation thereof, it is useful to consider a two-period model wherein an individual starts out with the endowment e. The economy lasts two periods. Individuals are free to divide their first period (when young) endowment between consumption, denoted by c and savings, s. Each unit of savings yields a consumer $1+r$ additional units of consumption in the second period. Denote after-tax income by B. Consumption in each period is given by

$$c_1^i = e^i - s^i, \quad i = \mathrm{L}, \mathrm{H}, \tag{1}$$

$$c_2^i = B_2^i + (1+r)s^i, \quad i = \mathrm{L}, \mathrm{H}. \tag{2}$$

Labor is supplied (elastically) only in the second period and all taxes are imposed in that same period. (Exogenous labor income may be permitted in the first period.) The analysis may be generalized to (elastic) labor supply in both periods, but only by adding considerable analytical complexity. The individuals have identical and additively separable preferences over first and second period consumption and labor supply, represented by the utility function

$$U^i = u\left(c_1^i\right) + \psi\left(c_2^i\right) - v\left(h^i\right), \quad i = \mathrm{L}, \mathrm{H}. \tag{3}$$

Unless otherwise stated, the functions u and ψ are increasing, strictly concave and twice differentiable. The function v is increasing, strictly convex, and twice continuously differentiable. We also assume that all goods are normal.

In this setting, where taxes on both earnings and savings income are available, we examine whether or not the return to savings ought to be taxed.

A main difference between our model and that of Fuest and Huber (2001, henceforth referred to as FH) is that the latter paper is concerned with endowments as the sole source of inequality as agents are assumed to face a uniform wage rate. We shall partly consider a model allowing both different endowments and nonuniform wage rates (reflecting skill diversity), and we shall partly focus on the opposite case of FH where inequality is solely due to differences in skill. Another major difference is that FH consider an atemporal (or single period) model without savings while we model inter-temporal behavior where savings incentives are central. Some further differences will be noted below. In particular, somewhat different variants of income taxes are considered.

3 Nonlinear labor income tax and linear taxation of savings

3.1 No income shifting

Sticking to our benchmark model, we shall address a tax system similar to the Nordic dual income taxation where capital income is taxed at a fixed rate (proportional tax) and a nonlinear tax is levied on labor income. That means that each type of tax is conditioned only on one type of income. This tax system is similar to that of Boadway et al. (2000), but is different from the taxes examined by FH who postulate a nonlinear tax function with both capital income and labor income as arguments. (In addition, FH consider a source tax on capital and extends the analysis to fiscal competition.) While both regimes are of interest, we have wanted to examine what is basically the Nordic type of tax system which is based on a mixture of principles and practical considerations that have motivated this particular dual design.[7]

It is well known from the tax theory with one-dimensional population that an important role for linear taxes may be to alleviate the self-selection constraint imposed under asymmetric information. This may be a role also where individuals differ both in skills and initial endowments. Adopting standard procedures, we can characterize Pareto efficient second best taxes. This is done by maximizing the utility of type L

$$U^{\mathrm{L}} = u\left(e^{\mathrm{L}} - s^{\mathrm{L}}\right) + \psi\left(B^{\mathrm{L}} + (1+\bar{r})s^{\mathrm{L}}\right) - v\left(h^{\mathrm{L}}\right)$$

for a fixed utility assigned to type H, $U^{\mathrm{H}} = \overline{U}^{\mathrm{H}}$, and subject to the preset revenue constraint

$$\sum\left(z^i - B^i + trs^i\right) = R, \quad i = \mathrm{L}, \mathrm{H}$$

and the self-selection constraint

$$\overline{U}^{\mathrm{H}} \geq u\left(e^{\mathrm{H}} - \hat{s}\right) + \psi\left(B^{\mathrm{L}} + (1+\bar{r})\hat{s}\right) - v\left(\frac{w^{\mathrm{L}}}{w^{\mathrm{H}}} h^{\mathrm{L}}\right)$$

where t is the tax rate for capital income, $r(1-t) = \bar{r}$, R denotes the required tax revenue, and $\overline{U}^{\mathrm{H}}$ is the pre-assigned utility level of type H, and where "hat" is used to indicate type H as a mimicker who would choose the bundle intended for type L. The maximization amounts to choosing bundles of gross and net incomes, z^{H}, B^{H} and z^{L}, B^{L}, while it is implicit that each type chooses the corresponding utility-maximizing level of savings conditional on the values of e, B, and $\bar{r}$.

The self-selection constraint requires that an individual weakly prefers the bundle, over the two time periods, intended for him or her to the bundle designated for the other individual. We consider the more interesting case, where only the incentive compatibility or self-selection constraint of the high-skilled type binds. This amounts to the region where redistribution takes place from high-skilled to low-skilled. The natural limit of redistribution is that if taken too far, such redistribution might induce the high-skilled type to pretend to be the low-skilled type. Such mimicking implies that the high-skilled would choose a labor supply $w^{\mathrm{L}}h^{\mathrm{L}}/w^{\mathrm{H}}$ to make his income equal to that of the low-skilled, and he would choose savings $\hat{s}$ to maximize intertemporal utility given the second period income he would earn as a mimicker. Imposing (6) precludes mimicking. Multipliers α, μ, and λ are assigned to the constraints,

[7] One relevant concern has been the desire to have equal tax rates on corporate profits and (other) capital income (interest, dividends, etc.), and with a constant corporate tax rate, there will a fixed tax rate on capital income in general.

and the Lagrange function of the optimization problem is

$$\begin{aligned}\Lambda &= u\big(e^{\mathrm{L}} - s^{\mathrm{L}}\big) + \psi\big(B^{\mathrm{L}} + (1+\bar{r})s^{\mathrm{L}}\big) - v\big(h^{\mathrm{L}}\big) + \alpha\big(u^{\mathrm{H}}\big(e^{\mathrm{H}} - s^{\mathrm{H}}\big)\\ &\quad + \psi\big(B^{\mathrm{H}} + (1+\bar{r})s^{\mathrm{H}}\big) - v\big(h^{\mathrm{H}}\big) - \overline{U}^{\mathrm{H}}\big)\\ &\quad + \mu\Big[\sum\big(w^i h^i - B^i + trs^i - R\big)\Big] + \lambda\Big[\overline{U}^{\mathrm{H}} - u\big(e^{\mathrm{H}} - \hat{s}\big) - \psi\big(B^{\mathrm{L}} + (1+\bar{r})\hat{s}\big)\\ &\quad + v\bigg(\frac{w^{\mathrm{L}}h^{\mathrm{L}}}{w^{\mathrm{H}}}\bigg)\Big]. \end{aligned} \tag{4}$$

Let subscripts denote partial derivatives. The first order conditions for interior solution with respect to B^i, $i = \mathrm{L}, \mathrm{H}$, are

$$\frac{\partial \Lambda}{\partial B^{\mathrm{L}}} = \psi_B^{\mathrm{L}} - \lambda\hat{\psi}_B - \mu + \mu trs_B^{\mathrm{L}} = 0, \tag{5}$$

$$\frac{\partial \Lambda}{\partial B^{\mathrm{H}}} = \alpha\psi_B^{\mathrm{H}} - \mu + \mu trs_B^{\mathrm{H}} = 0. \tag{6}$$

A "hat" is assigned to a function where the suppressed arguments are those of the mimicker.

Using Roy's theorem and the Slutsky decomposition, the first order condition with respect to t can be written as follows:

$$\begin{aligned}\frac{\partial \Lambda}{\partial t} &= -rs^{\mathrm{L}}\psi_B^{\mathrm{L}} - \alpha rs^{\mathrm{H}}\psi_B^{\mathrm{H}} + \lambda r\hat{\psi}_B\hat{s} + \mu\big(rs^{\mathrm{L}} + rs^{\mathrm{H}}\big) + \mu trs_B^{\mathrm{L}}\big(-rs^{\mathrm{L}}\big)\\ &\quad + \mu trs_B^{\mathrm{H}}\big(-rs^{\mathrm{H}}\big) + \mu tr\frac{\partial s^{c\mathrm{L}}}{\partial t} + \mu tr\frac{\partial s^{c\mathrm{H}}}{\partial t} = 0,\end{aligned} \tag{7}$$

where superscript c indicates compensated effects. Multiplying (5) by $-rs^{\mathrm{L}}$ and (6) by $-rs^{\mathrm{H}}$ and substituting in (7), we have

$$\lambda\hat{\psi}_B\big(\hat{s} - s^{\mathrm{L}}\big) + \mu t\bigg(\frac{\partial s^{c\mathrm{L}}}{\partial t} + \frac{\partial s^{c\mathrm{H}}}{\partial t}\bigg) = 0 \tag{8}$$

or

$$t = \frac{\frac{\lambda}{\mu}\hat{\psi}_B(\hat{s} - s^{\mathrm{L}})}{-(\frac{\partial s^{c\mathrm{L}}}{\partial t} + \frac{\partial s^{c\mathrm{H}}}{\partial t})}, \tag{9}$$

where $e^{\mathrm{L}} = e^{\mathrm{H}}$, the mimicker and the genuine low-skilled type have the same initial endowment and the same second period income (being a property of mimicking), and they will have the same level of savings. A larger initial endowment will induce larger first period savings. We can state the following proposition.

Proposition 1

(i) *If $e^{\mathrm{L}} = e^{\mathrm{H}}$ and if preferences of individuals are additively separable, $\hat{s} = s^{\mathrm{L}}$, and there is no taxation of capital income at the optimum.*

(ii) *If* $e^{\mathrm{H}} > e^{\mathrm{L}}$, *the mimickers have a higher savings level, and there is a case for taxing capital income at the optimum.*

We shall leave further interpretation to the next section.

3.2 Income shifting

Above, labor income was treated as observable. In reality, the government cannot directly observe an individual's true labor income and capital income, but individuals have to report their labor income and capital income for tax purposes. Now, the government faces an information problem not only because information on skill and wealth is private, but even the true labor income is unknown since individuals have the possibility to shift labor income to capital income. Let z be the amount of income reported as labor income in the second period whilst Δ is the labor income which is converted to capital income. Hence, the actual labor income is $z + \Delta$, and the labor supply is $h = \frac{z+\Delta}{w}$. Income shifting involves costs which could be modeled in different ways. As the crucial role for these costs is to determine an optimum extent of income shifting, while their exact nature is less important, we choose the simplest possible way by assuming that shifting an amount of income Δ inflicts a loss of net income $k(\Delta)$ on the taxpayer,[8] where $k'(\Delta) > 0$ and $k''(\Delta) > 0$. As above, each type of agent will choose savings and income shifting to maximize his utility.

In practice, income shifting from labor to capital is an important issue as the marginal tax on labor income exceeds the tax rate on capital income in countries with a dual income tax, but in general, whether an agent wants to shift income from labor to capital depends on relative marginal tax rates. However, we shall simply rule out the possibility of reversing the income shifting by shifting income from capital to labor (i.e., setting $\Delta < 0$) as it is an issue of minor practical interest. Thus, we simply impose the restriction that $\Delta > 0$. As we shall see and discuss further below, there may be theoretical cases where this constraint will be binding.

As a first step toward exploring optimal taxation in these circumstances, consider the behavior of an arbitrary agent supposed to maximize

$$U = u(e - s) + \psi\big(B + (1 + \bar{r})s + (1 - t)\Delta - k(\Delta)\big) - v\left(\frac{z + \Delta}{w}\right) \tag{10}$$

w.r.t. s and Δ, which yields the first order conditions

$$U_s = -u'(e - s) + \psi'\big(B + (1 + \bar{r})s + (1 - t)\Delta - k(\Delta)\big)(1 + \bar{r}) = 0, \tag{11}$$

$$\begin{aligned} U_\Delta = {} & \psi'\big(B + (1 + \bar{r})s + (1 - t)\Delta - k\big(\Delta^{\mathrm{L}}\big)\big)\big(1 - t - k'\big) \\ & - v'\left(\frac{z + \Delta}{w}\right)\frac{1}{w} \le 0 \end{aligned} \tag{12}$$

[8] Alternatively, one might have assumed that the cost would reduce taxable income, and hence the tax liability, or that it would imply use of own labor.

implicitly defining s and Δ as functions of z, B and t. At this point, we have found it handy to let a prime denote derivatives. The inequality applies where income shifting reversal would be desirable and $\Delta > 0$ is binding.

To characterize Pareto efficient second best taxes, we now maximize the utility of type L

$$u\big(e^{\mathrm{L}} - s^{\mathrm{L}}\big) + \psi\big(B^{\mathrm{L}} + (1+\bar{r})s^{\mathrm{L}} + (1-t)\Delta^{\mathrm{L}} - k\big(\Delta^{\mathrm{L}}\big)\big) - v\left(\frac{z^{\mathrm{L}} + \Delta^{\mathrm{L}}}{w^{\mathrm{L}}}\right)$$

for a fixed utility $\overline{U}^{\mathrm{H}}$ assigned to type H and subject to the revenue constraint

$$\sum\big(z^i - B^i + t\big(rs^i + \Delta^i\big)\big) = R, \quad i = 1, 2$$

and the self-selection constraint

$$\overline{U}^{\mathrm{H}} \geq u\big(e^{\mathrm{H}} - \hat{r}\big) + \psi\big(B^{\mathrm{L}} + \big(1 + r(1-t)\big)\hat{s} + (1-t)\hat{\Delta} - k\big(\hat{\Delta}\big)\big) - v\left(\frac{z^{\mathrm{L}} + \hat{\Delta}}{w^{\mathrm{H}}}\right).$$

The following multipliers are assigned to the constraints: α to the preassigned utility of type H, μ to the revenue constraint, and λ to the self-selection constraint, and the corresponding Lagrangian is

$$\begin{aligned}
\Lambda = {} & u\big(e^{\mathrm{L}} - s^{\mathrm{L}}\big) + \psi\big(B^{\mathrm{L}} + (1+\bar{r})s^{\mathrm{L}} + (1-t)\Delta^{\mathrm{L}} - k\big(\Delta^{\mathrm{L}}\big)\big) - v\left(\frac{z^{\mathrm{L}} + \Delta^{\mathrm{L}}}{w^{\mathrm{L}}}\right) \\
& + \alpha\Bigg[u\big(e^{\mathrm{H}} - s^{\mathrm{H}}\big) + \psi\big(B^{\mathrm{H}} + (1+\bar{r})s^{\mathrm{H}} + (1-t)\Delta^{\mathrm{H}} - k\big(\Delta^{\mathrm{H}}\big)\big) \\
& - v\left(\frac{z^{\mathrm{H}} + \Delta^{\mathrm{H}}}{w^{\mathrm{H}}}\right) - \overline{U}^{\mathrm{H}}\Bigg] \\
& - \lambda\Bigg[u\big(e^{\mathrm{H}} - \hat{s}\big) + \psi\big(B^{\mathrm{L}} + (1+\bar{r})\hat{s} + (1-t)\hat{\Delta} - k\big(\hat{\Delta}\big)\big) \\
& - v\left(\frac{z^{\mathrm{L}} + \hat{\Delta}}{w^{\mathrm{H}}}\right) - \overline{U}^{\mathrm{H}}\Bigg] \\
& + \mu\big[z^{\mathrm{L}} - B^{\mathrm{L}} + z^{\mathrm{H}} - B^{\mathrm{H}} + trs^{\mathrm{L}} + trs^{\mathrm{H}} + t\Delta^{\mathrm{L}} + t\Delta^{\mathrm{H}} - R\big].
\end{aligned} \tag{13}$$

The associated first order conditions for an interior solution with respect to z^i and B^i, $i = \mathrm{L}, \mathrm{H}$, are

$$\begin{aligned}
\frac{\partial\Lambda}{\partial z^{\mathrm{L}}} = {} & -v'\left(\frac{z^{\mathrm{L}} + \Delta^{\mathrm{L}}}{w^{\mathrm{L}}}\right)\frac{1}{w^{\mathrm{L}}} + \lambda v'\left(\frac{z^{\mathrm{L}} + \hat{\Delta}}{w^{\mathrm{H}}}\right)\frac{1}{w^{\mathrm{H}}} + \mu + \mu trs_z^{\mathrm{L}} \\
& + \mu t\Delta_z^{\mathrm{L}} = 0,
\end{aligned} \tag{14}$$

$$\frac{\partial\Lambda}{\partial z^{\mathrm{H}}} = -\alpha v'\left(\frac{z^{\mathrm{H}} + \Delta^{\mathrm{H}}}{w^{\mathrm{H}}}\right)\frac{1}{w^{\mathrm{H}}} + \mu + \mu trs_z^{\mathrm{H}} + \mu t\Delta_z^{\mathrm{H}} = 0, \tag{15}$$

$$\frac{\partial \Lambda}{\partial B^{\mathrm{L}}} = \psi_B^{\mathrm{L}} - \lambda \hat{\psi}_B - \mu + \mu t\left(r s_B^{\mathrm{L}} + \Delta_B^{\mathrm{L}}\right) = 0, \tag{16}$$

$$\frac{\partial \Lambda}{\partial B^{\mathrm{H}}} = \alpha \psi_B^{\mathrm{H}} - \mu + \mu t\left(r s_B^{\mathrm{H}} + \Delta_B^{\mathrm{H}}\right) = 0. \tag{17}$$

Using Roy's theorem and the Slutsky decomposition, the first order condition with respect to t can be written as follows:

$$\begin{aligned}\frac{\partial \Lambda}{\partial t} = {} & \left(-r s^{\mathrm{L}} - \Delta^{\mathrm{L}}\right)\psi_B^{\mathrm{L}} + \left(-r s^{\mathrm{H}} - \Delta^{\mathrm{H}}\right)\alpha \psi_B^{\mathrm{H}} + \lambda \hat{\psi}_B\left(r\hat{s} + \hat{\Delta}\right) \\ & \times \mu\left(r s^{\mathrm{L}} + \Delta^{\mathrm{L}} + r s^{\mathrm{H}} + \Delta^{\mathrm{H}}\right) \\ & + \mu t r s_B^{\mathrm{L}}\left(-r s^{\mathrm{L}} - \Delta^{\mathrm{L}}\right) + \mu t r s_B^{\mathrm{H}}\left(-r s^{\mathrm{H}} - \Delta^{\mathrm{H}}\right) + \mu t \Delta_B^{\mathrm{H}}\left(-r s^{\mathrm{L}} - \Delta^{\mathrm{L}}\right) \\ & + \mu t \Delta_B^{\mathrm{H}}\left(-r s^{\mathrm{H}} - \Delta^{\mathrm{H}}\right) \\ & + \mu t r s_t^{c\mathrm{L}} + \mu t r s_t^{c\mathrm{H}} + \mu t \Delta_t^{c\mathrm{L}} + \mu t \Delta_t^{c\mathrm{H}} = 0. \end{aligned} \tag{18}$$

As above, superscript c indicates compensated effects. Multiplying (16) by $(-r s^{\mathrm{L}} - \Delta^{\mathrm{L}})$ and (17) by $(-r s^{\mathrm{H}} - \Delta^{\mathrm{H}})$ and substituting in (18), we have

$$\lambda \hat{\psi}_B\left(r\hat{s} - r s^{\mathrm{L}} + \hat{\Delta} - \Delta^{\mathrm{L}}\right) + \mu t\left(r\frac{\partial s^{c\mathrm{L}}}{\partial t} + r\frac{\partial s^{c\mathrm{H}}}{\partial t} + \frac{\partial \Delta^{c\mathrm{L}}}{\partial t} + \frac{\partial \Delta^{c\mathrm{H}}}{\partial t}\right) = 0. \tag{19}$$

Solving for t, we obtain the formula for the optimal capital tax.

$$t = \frac{\frac{\lambda}{\mu}\hat{\psi}_B(r\hat{s} - r s^{\mathrm{L}} + \hat{\Delta} - \Delta^{\mathrm{L}})}{-[r\frac{\partial s^{c\mathrm{L}}}{\partial t} + r\frac{\partial s^{c\mathrm{H}}}{\partial t}] - [\frac{\partial \Delta^{c\mathrm{L}}}{\partial t} + \frac{\partial \Delta^{c\mathrm{H}}}{\partial t}]}. \tag{20}$$

The formula for the optimal capital income tax shows that three effects should be taken into account. Where the mimicker would report a larger capital income $(rs + \Delta)$ than the genuinely low-skilled type, the mimicker is hit harder when the return to capital (rs) and the concealed labor income (Δ) are taxed more strongly, and the self-selection constraint is alleviated. This effect is captured by the numerator. On the other hand, the tax causes a distortion of the intertemporal consumption trade-off, represented by the former term in the denominator. Finally, as it is possible to increase labor supply and have it taxed as capital income due to income shifting, the capital tax also distorts the labor supply. This effect is captured by the latter term in the denominator. The alleviation of the self-selection constraint should be traded off against the distortions. Below, we will review in further detail the contents of each term in the formula for t, starting with the numerator.

The mimicker can deviate from the true low-skilled agent by earning a higher wage rate (reflecting superior skill) and/or by possessing a larger endowment. First, consider a pure wage rate (ability) difference so that $w^{\mathrm{L}} < w^{\mathrm{H}}$ and $e^{\mathrm{L}} = e^{\mathrm{H}}$. It is easily recognized that if neither the mimicker nor the true low-skilled agent shifts income ($\hat{\Delta} = \Delta^{\mathrm{L}} = 0$) both will have the same savings ($\hat{s} = s^{\mathrm{L}}$) and there would be no capital income tax as $t = 0$ according to (19). However, if $\hat{\Delta} > 0$ and $\Delta^{\mathrm{L}} > 0$, things will be different. We shall now proceed on that assumption and come back to this question below. By comparative statics, we can show that *cet. par.* the ability difference implies that the mimicker will report more capital income as follows from the comparative statics leading to (A.16) in the Appendix. The productivity advantage will induce the mimicker to work more than type L and to report the extra income as capital income such that $\hat{\Delta} - \Delta^{\mathrm{L}} > 0$. The income effect generated by the extra labor supply implies a weaker incentive to save, but the net effect is that $r\hat{s} - rs^{\mathrm{L}} + \hat{\Delta} - \Delta^{\mathrm{L}} > 0$.

We should note that this result differs from that of the benchmark model without income shifting and that of FH where equal endowments would imply identical savings and equal taxable capital income of the mimicker and the true low-skilled type. Thus, income shifting creates a difference in the capital income and establishes a case for a capital income tax where none otherwise existed. Mimicking no longer implies equal labor income. It only implies equal reported labor income while the mimicker earns additional labor income to be reported as capital income.

Where endowments differ, $e^{\mathrm{L}} < e^{\mathrm{H}}$, the partial effect of the mimicker having a larger endowment is to stimulate savings which in turn discourages labor supply and transformation of labor income into capital income in the second period. These ambiguous effects (displayed in (A.22)) imply that we cannot know for sure that a larger endowment will generate a larger reported capital income and establish a case for a capital income tax.

A special case occurs where the utility function belongs to the quasi-linear class $u(c_1) + c_2 - v(\frac{z+\Delta}{w})$ and $s = e - c_1(\bar{r})$. Then the mimicker and the true low-skilled type will have the same savings level if they only differ w.r.t. to the wage rate, and $r\hat{s} - rs^{\mathrm{L}} + \hat{\Delta} - \Delta^{\mathrm{L}}$ reduce to $\hat{\Delta} - \Delta^{\mathrm{L}}$. Where $e^{\mathrm{L}} - e^{\mathrm{H}}$ (and $w^{\mathrm{L}} \leq w^{\mathrm{H}}$), we always have $\hat{s} > s^{\mathrm{L}}$ and $r\hat{s} - rs^{\mathrm{L}} + \hat{\Delta} - \Delta^{\mathrm{L}} > 0$.

Now turn to the denominator. By comparative statics, we can show that for an unspecified agent $r\frac{\partial s^c}{\partial t} + \frac{\partial \Delta^c}{\partial t} < 0$ as follows from (A.12) in the Appendix. (We may note that where $\Delta = 0$, $\partial \Delta^c/\partial t = 0$.) The reported capital income declines in response to an increase in the tax rate on capital income and the denominator is negative.

Note that in this model income shifting does not mean converting part of a fixed labor income into capital income. Rather, it means that there is an opportunity for choosing working hours while having the marginal earnings taxed as capital income, but at a cost since it is costly to transform labor income into capital income. An important role for income shifting in this model is to give high-income people an opportunity for mimicking low-income people without lowering their labor supply as much as they would otherwise have to do in order to earn the same income as the low-skilled people. The reason is that while officially their labor income is reduced to the level of the low-skilled people, they do in fact work and earn more, but what they earn over and above the earnings of the low-skilled person is taxed as capital income.

The change in labor income reported as capital income reflects a change in labor supply which means that interactions between savings and labor supply are crucial determinants behind the total effect on reported capital income. To make these interactions more transparent, consider the special case where the utility function belongs to the quasi-separable class:

$$U = c_1 + \psi(c_2) - v\left(\frac{z+\Delta}{w}\right). \tag{21}$$

Then making use of the budget constraint $s = \frac{c_2}{1+\bar{r}} - \frac{B}{1+\bar{r}} - \frac{(1-t)\Delta}{1+\bar{r}} - \frac{k(\Delta)}{1+\bar{r}}$,

$$s_t = \frac{1}{1+\bar{r}} c_{2t} - \frac{1-t-k'}{1+\bar{r}} \Delta_t^c, \tag{22}$$

where the former term on the right-hand side is negative reflecting that a larger capital income tax discourages savings and second period consumption. The latter term is positive as a larger t imposes a larger tax on labor supply where labor income is reported as capital income. As less labor involves lower income in the second period, the partial effect is to encourage savings in the first period.

Proposition 2 *The optimal tax is characterized by* (20). *If* $e^{\mathrm{L}} = e^{\mathrm{H}}$, *and there is income shifting*, *taxing capital income is part of the optimal tax policy.*

In the absence of income shifting, there would be no tax on capital income. This would correspond to the two-period model with labor supplied only in period one and preferences where consumption in the two periods is weakly separable from labor (leisure) as discussed in Salanié (2003, Chap. 6.3). Considering the expression for the tax rate t in (20) above, we note that with income shifting, the numerator is positive and the denominator is negative, and due to the minus sign t is positive. Hence, there is a case for taxing capital due to the income shifting.

It is interesting to note that there is a case for a positive capital income tax even though it is conceivable that no income shifting is observed. Conversely, observing no (or little) income shifting would not be a valid argument for stating that since income shifting is not perceived to be a problem no capital income tax is needed. The reason is that the case for the tax rests on the presumption that the high-skilled type would shift income if he were to mimic the low-skilled. This is not to argue that no actual income shifting is plausible. A feature of our model that may appear surprising or even weird is that the high-skilled person will not shift any income. The intuition for this result is that we get the well-known zero marginal tax at the top, and then there is obviously no inducement to shift income to a base with a lower marginal tax. We would like to play down the significance of this specific result which may easily be exaggerated because of the simplifications that we have made by considering only two types of agents and a quasi-linear utility function. From the standard Mirrlees type of optimum tax theory with a continuum of individuals, we know that normally it is only at the very top of the distribution that the marginal tax is zero. Indeed the marginal tax may be quite large very close to the top (as discussed in Tuomala 1984). Then the result would apply only to a tiny fraction of the population and its role would appear more modest that in the two type model where all high-skilled individuals are at the very type since by assumption there is only one type of high-skilled agents. Moreover, it is the case that with two periods (unlike the conventional Mirrlees model) the zero marginal tax at the top will not necessarily obtain if we abandon the assumption of quasi-linearity. For a further discussion in the absence of income shifting, see Gaube (2006).

Above, we have focused on the capital income tax. To characterize the optimal labor income tax and overall tax policy, we shall highlight a number of marginal tax rates which are also crucial for the existence of income shifting brushed aside above. First, we note that there is a need to distinguish two tax rates on labor income, namely the marginal tax rate on reported labor income and the marginal tax rate on labor income reported as capital income. The latter is the more transparent one. As capital income is taxed at a fixed rate equal to t, this is simply the marginal tax rate on labor income disguised as capital income.

In our model, behavior and optimality conditions are formulated in terms of marginal rates of substitution rather than explicit tax rates, but as has become conventional in the literature, we may interpret the marginal rate of substitution between gross and net income as one minus the marginal income tax $\frac{v'(h)}{w\psi'(c_2)} = 1 - T'(z)$, which would be equivalent to the characterization of the labor supply of an unspecified type of agent facing a labor income tax function $T(z)$. Where there is also a tax on capital income, the total tax burden is the sum of the two taxes. Denoting by $\tau(z)$ the total or effective tax as defined in Edwards et al. (1994):

$$\tau(z) = T(z) + trs + t\Delta.$$

The corresponding effective marginal tax $\tau'(z)$ is discussed in the Appendix. From (A.37) and (A.39), the high-skilled type faces a zero marginal effective tax rate

$$\tau'\left(z^{\mathrm{H}}\right) = 0, \tag{23}$$

while the effective marginal tax rate for the low-skilled is positive:

$$\tau'\left(z^{\mathrm{L}}\right) = \frac{\lambda\hat{\psi}_B}{\mu}\left(\frac{v'(h^{\mathrm{L}})}{\psi_B^{\mathrm{L}}}\frac{1}{w^{\mathrm{L}}} - \frac{\hat{v}'}{\hat{\psi}_B}\frac{1}{w^{\mathrm{H}}}\right) > 0, \tag{24}$$

where the sign follows from imposing the standard agent monotonicity (or single crossing property) that through any given point in B, z space the mimicker's indifference curve is flatter than that of the true low-skilled type. Our marginal tax results reproduce in a two-period setting the effective marginal tax results shown by Edwards et al. (1994) for the single period setting.

It follows from the individual's optimal choice of Δ:

$$U_\Delta = \psi'\left(B + (1+\bar{r})s + (1-t)\Delta - k\left(\Delta^{\mathrm{L}}\right)\right)\left(1 - t - k'\right) - v'\left(\frac{z+\Delta}{w}\right)\frac{1}{w} = 0$$

that

$$t = 1 - \frac{v'}{\psi'}\frac{1}{w} - k'.$$

Defining the marginal income tax by $T' = 1 - \frac{v'}{\psi'}\frac{1}{w}$ as above, the following equation must hold:

$$T' = t + k'. \tag{25}$$

The interpretation is that where an agent shifts income the marginal income tax on reported labor income must equal the tax rate on labor income reported as capital income plus the marginal cost of income-shifting. With a lower marginal income tax on reported labor income, there would be no incentive to shift income. We can rewrite (25) as $k'(\Delta) = T' - t$. A tax difference motivates an unproductive and wasteful use of resources as taxpayers incur a cost in pursuit of a tax saving by shifting income. Indeed, we can take this expression as a measure of the extent of income shifting since the left-hand side is monotonically increasing in Δ.

Consider now whether the respective agents will actually shift income such that $\Delta^i > 0$ (for $i = \mathrm{L}$ or H or both). We focus on the case where the agents are distinguished by ability ($w^{\mathrm{L}} < w^{\mathrm{H}}$ and $e^{\mathrm{L}} = e^{\mathrm{H}}$); supposed to start with that $t = 0$. Then $\tau'(z^{\mathrm{H}}) = T'(z^{\mathrm{H}}) = t = 0$ and $\tau'(z^{\mathrm{L}}) = T'(z^{\mathrm{L}}) > t = 0$. Assuming that $k'(0)$ is sufficiently small, the low-skilled agent has an incentive to shift income to capital (setting $\Delta^{\mathrm{L}} > 0$). As the mimicker being more productive has an even stronger incentive to shift income, we have $\hat{\Delta} > \Delta^{\mathrm{L}} > 0$ as formally confirmed by (A.15). Then income shifting will persist for sufficiently small values of t, and we have established the premise underlying Proposition 2 that income shifting takes place.

However, there is no implication that all agents will actually shift income at the optimum such that $\Delta^{\mathrm{L}} > 0$ and $\Delta^{\mathrm{H}} > 0$. A necessary condition for this to happen is that $T' > 0$, otherwise (20) is violated. From (A.33), we have $\tau' = T' + trs_z + t\Delta_z + (1 - T')(trs_B + t\Delta_B)$. From (A.6), (A.31), (A.25), and (A.26) we know that $trs_z + t\Delta_z < 0$ and $trs_B + t\Delta_B < 0$. It follows that $T' > 0$ where $\tau' \geq 0$ and $t > 0$. We also note that the case for $T' > 0$ is stronger where $\tau' > 0$. We see that the necessary condition for income shifting that $T' > 0$ is fulfilled for both agents where $t > 0$. Finally, it follows that $T' > 0$ and $t > 0$ may be compatible with income shifting for either type, but whether they will indeed shift income is an unsettled issue. The above results may seem surprising as we are used to having a zero marginal tax rate at the top (as is the case in FH with no source-based tax) but with a linear tax on capital income neither t nor T' need to be zero at the top.

It is important to note that the crucial condition for using a capital income tax to relax the self-selection constraint is that the mimicker would like to shift income to a larger extent than the genuinely low-skilled. It is immaterial whether the high-skilled agent also shifts income. It is not even crucial that the low-skilled type shifts income as our result would also go through with $\hat{\Delta} > \Delta^{\mathrm{L}} = 0$. However, raising t to a level where income shifting both by the mimicker and the true low-skilled type is quenched off would not be efficient as it would then fail to relax the self-selection constraint.

4 Concluding remarks

We have addressed nonlinear taxation of labor income and linear taxation of capital income in a two period model with agents who differ in their earnings capacity. As is well known, a capital income tax may be redundant in such a setting given certain separability assumptions on preferences. Beyond imposing those separability assumptions, we assume that the agents may organize their economic activity in perfectly legal but costly ways that allow (part of) their labor earnings to be taxed as capital income, which is a favorable option granted that the capital tax rate is lower. Even if an agent makes use of this opportunity, the government can always tax him harder by increasing the income tax on labor. A tax on capital income is strictly required neither for raising revenue nor redistributing income. We show that there may still be a role for a capital income tax as taxation takes place under the usual asymmetric information constraint causing tax distortions, and under income shifting the capital income tax may alleviate the effects of the information constraint.

We show our result in a simple two-period model with taxation only in a single period and where high-skilled and low-skilled individuals may have different non-observable resource endowments beyond their different earnings capacity. An important implication is that such endowments may impact on tax policy even without introducing a new distributional dimension (as between those who inherit or not in Boadway et al. 2000).

Unlike Fuest and Huber (2001), we have considered an intertemporal model and also allowed for differences in wage rates. An important finding is that with pure wage rate variation there is no case for a capital income tax unless there is income shifting. With variation in initial endowments, there is a case for capital taxation even without income shifting in our model with savings, but income shifting will rather weaken this case as the larger savers will have less incentive to earn additional labor income to be reported as capital income.

The asymmetric information argument for a capital income tax must be traded off against its distortionary effect not only on savings, as in the conventional model of capital taxation, but also on labor as due to income shifting, the capital tax now also becomes a marginal tax on labor. The interaction between tax shifting and labor supply is a key element. Inducing a larger reported labor income weakens the incentive to earn labor income to be reported as capital income as do a low marginal tax on labor income and a large tax rate on capital income. Our model highlights a central and complex interaction between taxes on income from labor and capital, respectively. The crucial marginal tax rate is the *effective* marginal tax rate capturing both the marginal tax on labor income tax and the marginal tax due to its income effect on savings, and hence on capital income. The latter effect will to a large extent influence the discrepancy between the tax rate on (reported) labor income and the tax rate on (reported) capital income at the optimum and thereby govern the inducement to shift income.

Acknowledgements We are grateful to the Editor, Richard Cornes, and an anonymous referee for their useful comments. We also thank Peter Birch Sørensen for his valuable comments on an early version of the paper at the Nordic Workshop on Tax Policy and Public Economic, Bergen, 2006, Tomas Sjögren (our discussant at the 2007 IIPF Congress in Warwick), and seminar participants at Umeå University for many helpful comments.

Appendix

Below, we present the comparative statics of individual behavior starting out from the first order conditions of an interior optimum

$$U_s = -u'(e-s) + \psi'\big(B + (1+\bar{r})s + (1-t)\Delta - k(\Delta)\big)(1+\bar{r}) = 0, \quad \text{(A.1)}$$

$$U_\Delta = \psi'\big(B + (1+\bar{r})s + (1-t)\Delta - k\big(\Delta^{\mathrm{L}}\big)\big)(1-t-k') - v'\left(\frac{z+\Delta}{w}\right)\frac{1}{w} = 0. \quad \text{(A.2)}$$

Second order conditions

$$U_{ss} = u''(e-s) + \psi''(1+\bar{r})^2 < 0, \tag{A.3}$$

$$U_{\Delta\Delta} = \psi''\big(1-t-k'\big)^2 - \psi'k'' - v''\frac{1}{w^2} < 0, \tag{A.4}$$

$$U_{\Delta s} = \psi''\big(1-t-k'\big)(1+\bar{r}) < 0, \tag{A.5}$$

$$D = \det\begin{pmatrix} U_{ss} & U_{s\Delta} \\ U_{\Delta s} & U_{\Delta\Delta} \end{pmatrix} > 0. \tag{A.6}$$

We can derive the compensated tax effects on s and Δ

$$U_{ss}s_t^c + U_{s\Delta}\Delta_t^c = \psi' r, \tag{A.7}$$

$$U_{\Delta s}s_t^c + U_{\Delta\Delta}\Delta_t^c = \psi', \tag{A.8}$$

$$s_t^c = \frac{1}{D}\big(\psi' r U_{\Delta\Delta} - \psi' U_{s\Delta}\big), \tag{A.9}$$

$$\begin{aligned} rU_{\Delta\Delta} - U_{s\Delta} = {} & r\psi''\big(1-t-k'\big)^2 - r\psi'k'' - rv''\frac{1}{w^2} \\ & - \psi''\big(1-t-k'\big)(1+\bar{r}), \end{aligned} \tag{A.10}$$

$$\begin{aligned} \Delta_t^c &= \frac{1}{D}\big(\psi' U_{ss} - \psi' U_{\Delta s} r\big) \\ &= \frac{\psi'}{D}\Big(u'' + \psi''(1+\bar{r})^2 - r\psi''\big(1-t-k'\big)(1+\bar{r})\Big), \end{aligned} \tag{A.11}$$

$$\begin{aligned} rs_t^c + \Delta_t^c &= \frac{\psi'}{D}\big(r^2 U_{\Delta\Delta} - rU_{s\Delta} + U_{ss} - rU_{s\Delta}\big) \\ &= \frac{\psi'}{D}\Big(r^2\psi''\big(1-t-k'\big)^2 - r^2\psi'k'' - r^2 v''\frac{1}{w^2} \\ &\quad - 2r\psi''\big(1-t-k'\big)(1+\bar{r}) + u'' + \psi''(1+\bar{r})^2\Big) \\ &= \frac{\psi'}{D}\left(\psi''\Big(r\big(1-t-k'\big) - (1+\bar{r})\Big)^2 \right. \\ &\quad \left. - r^2\psi'k'' - r^2 v''\frac{1}{w^2} + u''\right) < 0. \end{aligned} \tag{A.12}$$

Wage effects:

$$\begin{aligned} & U_{ss}s_w + U_{s\Delta}\Delta_w = 0, \\ & U_{\Delta s}s_w + U_{\Delta\Delta}\Delta_w = -\frac{z+\Delta}{w^3}v'' - v'\frac{1}{w^2} = N < 0, \end{aligned} \tag{A.13}$$

$$s_w = -\frac{1}{D}NU_{s\Delta} < 0, \tag{A.14}$$

$$\Delta_w = \frac{1}{D}NU_{ss} > 0, \tag{A.15}$$

$$rs_w + \Delta_w = \frac{N}{D}(U_{ss} - rU_{s\Delta}) > 0, \tag{A.16}$$

since

$$\begin{aligned} U_{ss} - rU_{s\Delta} &= u'' + \psi''(1+\bar{r})^2 - r\psi''(1-t)(1+\bar{r}) + r\psi''k'(1+\bar{r}) \\ &= u'' + \psi''(1+\bar{r}) + r\psi''k'(1+\bar{r}) < 0. \end{aligned} \tag{A.17}$$

Endowment effects:

$$U_{ss}s_e + U_{s\Delta}\Delta_e = u'', \tag{A.18}$$

$$U_{\Delta s}s_e + U_{\Delta\Delta}\Delta_e = 0, \tag{A.19}$$

$$s_e = \frac{1}{D}u''U_{\Delta\Delta} > 0, \tag{A.20}$$

$$\Delta_e = -\frac{1}{D}u''U_{\Delta s} < 0, \tag{A.21}$$

$$\begin{aligned} \Delta_e + rs_e &= \frac{u''}{D}\left(-\psi''\left(1-t-k'\right)(1+\bar{r}) + r\psi''\left(1-t-k'\right)^2 - r\psi'k'' - rv''\frac{1}{w^2}\right) \\ &= \frac{u''}{D}\left(\psi''\left(1-t-k'\right)\left(r\left(1-t-k'\right) - (1+\bar{r})\right) - r\psi'k'' - rv''\frac{1}{w^2}\right) \\ &= \frac{u''}{D}\left(\psi''\left(1-t-k'\right)\left(\bar{r} - rk' - 1 - \bar{r}\right) - r\psi'k'' - rv''\frac{1}{w^2}\right) \\ &= \frac{u''}{D}\left(\psi''\left(1-t-k'\right)\left(-1-rk'\right) - r\psi'k'' - rv''\frac{1}{w^2}\right). \end{aligned} \tag{A.22}$$

We note that there are conflicting effects and the overall sign is indeterminate.

Effects of B:

$$U_{ss}s_B + U_{s\Delta}\Delta_B = -\psi''(1+\bar{r}), \tag{A.23}$$

$$U_{\Delta s}s_B + U_{\Delta\Delta}\Delta_B = -\psi''\left(1-t-k'\right), \tag{A.24}$$

$$\begin{aligned} Ds_B = &-\psi''(1+\bar{r})\left(\psi''\left(1-t-k'\right)^2 - \psi'k'' - v''\frac{1}{w^2}\right) \\ &+ \psi''\left(1-t-k'\right)\psi''\left(1-t-k'\right)(1+\bar{r}) \\ = &\ \psi''(1+\bar{r})\left(\psi'k'' + v''\frac{1}{w^2}\right) < 0, \end{aligned} \tag{A.25}$$

$$\begin{aligned} D\Delta_B &= \left(u'' + \psi''(1+\bar{r})^2\right)\left(-\psi''\right)\left(1-t-k'\right) + \psi''\left(1-t-k'\right)(1+\bar{r})\psi''(1+\bar{r}) \\ &= -u''\psi''\left(1-t-k'\right) < 0. \end{aligned} \tag{A.26}$$

Effects of z:

$$U_{ss}s_z + U_{s\Delta}\Delta_z = 0, \tag{A.27}$$

$$U_{\Delta s}S_z + U_{\Delta\Delta}\Delta_z = v''\frac{1}{w^2}, \tag{A.28}$$

$$Ds_z = -v''\frac{1}{w^2}U_{s\Delta} > 0, \tag{A.29}$$

$$D\Delta_z = U_{ss}v''\frac{1}{w^2} < 0, \tag{A.30}$$

$$D(rs_z + \Delta_z) = v''\frac{1}{w^2}(U_{ss} - rU_{s\Delta}) < 0, \tag{A.31}$$

where $U_{ss} - rU_{s\Delta}$ is given in (A.17).

Effective tax rates:

$$\tau(z) = T(z) + trs + t\Delta, \tag{A.32}$$

$$\tau' = T' + trs_z + t\Delta_z + \big(1 - T'\big)(trs_B + t\Delta_B), \tag{A.33}$$

$$\frac{v'}{\psi_B}\frac{1}{w} = 1 - T' \quad \text{and} \quad T' = 1 - \frac{v'}{\psi_B}\frac{1}{w}, \tag{A.34}$$

$$\tau' = 1 - \frac{v'}{\psi_B}\frac{1}{w} + trs_z + t\Delta_z + \frac{v'}{\psi_B}\frac{1}{w}(trs_B + t\Delta_B), \tag{A.35}$$

$$\tau'\psi_B = \psi_B - \frac{v'}{w} + \psi_B(trs_z + t\Delta_z) + \frac{v'}{w}(trs_B + t\Delta_B). \tag{A.36}$$

To find $\tau'(z^{\mathrm{H}})$ and $\tau'(z^{\mathrm{L}})$ we insert the respective expressions for ψ_B^{H}, $\frac{v'}{w^{\mathrm{H}}}$, and ψ_B^{L}, $\frac{v'}{w^{\mathrm{L}}}$ from (17), (15) and (16), (14) in the equation above and find that

$$\tau'\big(z^{\mathrm{H}}\big) = 0 \tag{A.37}$$

and

$$\tau'\big(z^{\mathrm{L}}\big)\psi_B^{\mathrm{L}} = \lambda\hat{\psi}_B\big(1 + trs_z^{\mathrm{L}} + t\Delta_z^{\mathrm{L}}\big) - \lambda\hat{v}'\frac{1}{w^{\mathrm{H}}}\big(1 - trs_B^{\mathrm{L}} - t\Delta_B^{\mathrm{L}}\big). \tag{A.38}$$

Then invoking (14) and (16),

$$\begin{aligned}\tau'\big(z^{\mathrm{L}}\big)\psi_B^{\mathrm{L}} &= \lambda\hat{\psi}_B\left(\frac{v'}{w^{\mathrm{L}}} - \lambda\hat{v}'\frac{1}{w^{\mathrm{H}}}\right)\frac{1}{\mu} - \lambda\hat{V}'\frac{1}{w^{\mathrm{H}}}\big(\psi_B^{\mathrm{L}} - \lambda\hat{\psi}_B\big)\frac{1}{\mu} \\ &= \frac{\lambda\hat{\psi}_B\psi_B^{\mathrm{L}}}{\mu}\left(\frac{v_{\mathrm{L}}'}{\psi_B^{\mathrm{L}}}\frac{1}{w^{\mathrm{L}}} - \frac{\hat{v}'}{\hat{\psi}_B}\frac{1}{w^{\mathrm{H}}}\right) > 0,\end{aligned} \tag{A.39}$$

where the sign follows from agent monotonicity. In brief, $\tau'(z^{\mathrm{H}}) = 0$ and $\tau'(z^{\mathrm{L}}) > 0$.

References

Atkinson, A. B., & Stiglitz, J. E. (1976). The design of tax structure: direct versus indirect taxation. *Journal of Public Economics*, *6*, 55–75. doi:10.1016/0047-2727(76)90041-4.

Blackorby, C., & Brett, C. (2000). *Capital taxation in a simple finite horizon OLG model* (Discussion Paper 00-11). Department of Economics, University of British Columbia.

Boadway, R., Marchand, M., & Pestieau, P. (2000). Redistribution with unobservable bequests: A case for taxing capital income. *Scandinavian Journal of Economics*, *102*, 253–267. doi:10.1111/1467-9442.00198.

Chamley, C. (1986). Optimal taxation of capital income in general equilibrium model with infinite lives. *Econometrica*, *54*, 607–622. doi:10.2307/1911310.

Christiansen, V. (1985). The choice of excise taxes when savings and labor decisions are distorted. *Journal of Public Economics*, *28*, 95–110. doi:10.1016/0047-2727(85)90022-2.

Cremer, H., Pestieau, P., & Rochet, J.-C. (2001). Direct versus indirect taxation: The design of the tax structure revisited. *International Economic Review*, *42*(3), 781–799.

Cremer, H., Pestieau, P., & Rochet, J.-C. (2003). Capital income taxation when inherited wealth is not observable. *Journal of Public Economics*, *87*, 2475–2490. doi:10.1016/S0047-2727(02)00050-6.

Edwards, J., Keen, M., & Tuomala, M. (1994). Income tax, commodity taxes and public good provision. *FinanzArchiv*, *51*, 472–487.

Fuest, C., & Huber, B. (2001). Labor and capital taxation, fiscal competition and the distribution of wealth. *Journal of Public Economics*, *79*, 71–91. doi:10.1016/S0047-2727(00)00096-7.

Gaube, T. (2006). Optimum taxation of each year's income. *Journal of Public Economic Theory*, *9*, 127–150. doi:10.1111/j.1467-9779.2007.00301.x.

Gordon, R. H., & MacKie-Mason, J. K. (1995). Why is there corporate taxation in a small open economy? The role of transfer pricing and income shifting. In M. Feldstein, J. R. Hines & R. G. Hubbard (Eds.), *The effects of taxation on multinational corporations*. Chicago: University of Chicago Press.

Gordon, R., & Slemrod, J. (1998). *Are "real" responses to taxes simply income shifting between corporate and personal tax bases*? (NBER Working Paper 6576).

Johansen, L. (1965). *Public economics*. Amsterdam: North-Holland.

Judd, K. (1985). Redistributive taxation in a simple perfect foresight model. *Journal of Public Economics*, *28*, 59–83. doi:10.1016/0047-2727(85)90020-9.

Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, *38*, 175–208. doi:10.2307/2296779.

Mirrlees, J. (2000). What taxes should there be? In *IDEI conference*, University of Toulouse.

Naito, H. (1999). Re-examination of uniform commodity taxes under a non-linear income tax system and its implications for production efficiency. *Journal of Public Economics*, *17*, 181–211.

Pekkarinen, J., Takala, K., & Tuomala, M. (1985). *Distribution and composition of wealth in Finland 1981* (Research Reports no. 49). Department of Economics, University of Helsinki.

Pirttilä, J., & Selin, H. (2006). *Income shifting and the elasticity of taxable income: Evidence from the Finnish dual tax reform* (CESifo Working Papers no. 1875).

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. doi:10.1016/S0047-2727(99)00105-X.

Saez, E. (2002). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics*, *83*, 217–230. doi:10.1016/S0047-2727(00)00159-6.

Salanié, B. (2003). *The economics of taxation*. Cambridge: MIT Press.

Stern, N. H. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, *17*, 181–211. doi:10.1016/0047-2727(82)90019-6.

Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics*, *17*, 213–240. doi:10.1016/0047-2727(82)90020-2.

Tuomala, M. (1984). On the optimal income taxation: some further numerical results. *Journal of Public Economics*, *23*, 351–366. doi:10.1016/0047-2727(84)90035-5.