ELSEVIER

Journal of Public Economics 71 (1999) 1–26

JOURNAL OF PUBLIC ECONOMICS

Optimal taxation and spending in general competitive growth models

Kenneth L. Judd[a,b,*]

[a]*Hoover Institution, Stanford, CA, 94305, USA*
[b]*National Bureau of Economic Research, City, USA*

Received 30 June 1995; accepted 21 April 1998

Abstract

We find that the optimal long-run tax on capital income is zero even if the capital stock does not converge to a steady state nor to a steady state growth rate. The optimal tax on human capital is also zero if human capital is not a final good, but the long-run wage tax is not generally zero. We argue that ''consumption'' tax proposals, such as the Flat Tax, are not consumption taxes, and are biased against human capital.

Keywords: Optimal taxation; Human capital; Growth; Consumption taxation

JEL classification: H21

1. Introduction

This paper explores the proposition that the optimal long-run tax on capital income is zero. In contrast with earlier studies, we distinguish between human and physical capital, we include public goods, we allow general, nonstationary production functions, and we do not assume convergence of dynamic equilibrium to any kind of steady state. The general result we find is that the optimal tax on physical capital is zero *on average* except for an initial period. Furthermore, this zero-average-tax result holds also for human capital if it has no final consumption

*Tel.: +1-650-723-5866; fax: +1-650-723-1687; e-mail: judd@hoover.stanford.edu

0047-2727/99/$ – see front matter
PII: S0047-2727(98)00054-1

value, but that labor income will generally be taxed in the long-run. This can be accomplished by taxing all labor income but subsidizing human capital inputs. We argue that all these results follow from optimal commodity taxation theory.

There have been many analyses which argue for a zero long-run capital income tax rate. Early arguments, such as Atkinson and Sandmo (1980); Auerbach (1979), and Diamond (1973), relied heavily on separability assumptions and identical agents in each cohort. Judd (1985) found a zero optimal long-run capital income tax rate for steady states of competitive dynamic general equilibrium with heterogeneous infinitely-lived agents and nonseparable preferences. Others have explored taxation issues in models of unbounded growth. Eaton (1981) showed that capital income taxation reduces the steady state growth rate and Hamilton (1987) demonstrated that asymmetric treatment of different forms of capital has a high welfare cost. Jones et al. (1997) and Bull (1993) argued that the zero-tax result also holds for all factor income, labor as well as both human and physical capital, in models generalizing the Eaton model. Jones et al. (1997) also argue that revenue constraints may lead to taxation of capital income in the long-run. We show that these results arise from special and unrealistic assumptions they make. Also, this paper proves its results without making any convergence assumptions; instead we include public goods to the analysis and find that assuming nondegenerate expenditures on public goods serves as a substitute for assuming convergence.

One general problem with this literature is the lack of economic intuition. The analyses in Judd (1985); Bull (1993), and Jones et al. (1997) are presented in formal ways not clearly connected with basic principles of optimal taxation; in particular, they strongly use the assumption of convergence to a steady state. Other analyses have focussed on special long-run dynamic features. For example, Auerbach (1979) conjectured that the zero optimal capital income tax result arises from the infinite long-run elasticity of savings in representative agent models with separable utility. Judd (1985) showed that this long-run elasticity property is not relevant since the same long-run zero tax results even if the long-run saving elasticity is finite and differs across individuals; however, no alternative intuition was offered.

In this paper, we ignore simple dynamic features such as the steady-state behavior or long-run elasticities, and instead put the zero long-run tax results on more economically appealing foundations. To do this, we look to the commodity tax literature. Two results from that literature apply here; first, the optimality of uniform taxation with separable and sufficiently symmetric utility, and, second, the prohibition against intermediate good taxation derived in Diamond and Mirrlees, 1971. Our methods generalize previous work and tie the results to the commodity tax literature, a change which helps us understand why we often find that the average tax rate on capital income is zero in the optimal policy.

The issues examined here are central to current policy debates in the U.S. A key feature of consumption tax proposals, such as those described in Bradford (1986);

Hall and Rabushka (1995); McClure and Zodrow (1996), and Weidenbaum (1997) is the zero effective tax rate on income from new physical capital investments; however, none propose a true consumption tax. The difference arises because they would not give human capital investments the same treatment they advocate for physical capital investment. These ''consumption tax'' proposals essentially advocate reducing the tax burden of physical capital but increasing the tax burden of some human capital investments, and do so without offering any explanation for this policy preference over a true consumption tax. The analysis below argues for symmetric treatment of physical and human capital, a type of ''level playing field'' argument which follows the Diamond–Mirrlees case against intermediate good taxation.

2. Basic intuition from optimal commodity tax theory

The optimal factor taxation literature has generally derived its results through dynamic optimization methods in ways which do not illustrate the underlying economic logic. In contrast, the results of the optimal commodity taxation literature are stated in more intuitive ways. In this section we review two basic optimal commodity taxation ideas—the inverse elasticity rule and the nontaxation of intermediate goods—and show how these ideas can be used to understand the optimal factor taxation results.

We first consider a simple problem wherein the optimal commodity tax is uniform. Suppose that we have an additively separable, isoelastic utility function over n commodities,

$$U = \sum_{i=0}^{n} u(c_i)$$

Also assume that good 0 is the numeraire and is untaxed, p_i is the consumer price of good i, q_i is the producer price of good i, and the government's problem is to tax the n commodities so as to maximize U subject to a revenue constraint. The inverse elasticity rule says that the optimal tax rates on goods $i \neq 0$, $(p_i - q_i)/q_i$, are equal across all commodities[1]. Such a tax structure would produce a uniform distortion between the marginal rate of substitution and the marginal rate of transformation between each good c_i and the numeraire. More precisely, for all $i \neq 0$,

$$MRS_{i,0} \equiv \frac{u'(c_i)}{u'(c_0)} = p_i = (1+\tau)q_i \equiv (1+\tau)MRT_{i,0} \tag{1}$$

[1]See Atkinson and Stiglitz (1972) for a discussion of the inverse elasticity rule, and its general validity in the case of separable utility.

Note that this implies that there is no distortion in the choice between goods $i, j \neq 0$ since (1) implies $MRS_{i,j} = MRT_{i,j}$. This result also generalizes to the case where we have n untaxed ''leisures'', l_i, $i = 0, 1, ..., n$, and the utility function is a weighted sum, as in

$$U = \sum_{i=0}^{n} w_i^c u(c_i) + \sum_{i=0}^{n} w_i^l v(l_i) \tag{2}$$

Suppose next that the indices i refer to the same commodity at different dates, and that the weights w_i^c and w_i^l in (2) equal the discount factor for utility in period i relative to period 0. From general equilibrium theory we know that the Arrow–Debreu model applies to this dynamic context; similarly, so does the ''static'' commodity tax literature[2]. The resulting optimal uniform tax policy creates uniform MRS/MRT distortions between the untaxed good and every other good.

Income taxation implies a pattern of distortions across consumption and leisure at various dates. For example, if we have an asset at time 0 which we liquidated at time t to finance consumption, then a tax on the investment income essentially taxes consumption at time t. However, income taxation cannot perfectly implement commodity taxation. The difference lies in the initial time periods. For example, if the interest rate is 10% per period, then a 100% tax on interest income results in the cost of period 1 consumption equalling one unit of period 0 consumption and implies an effective commodity tax rate of 11% on period 1 consumption. Any higher interest tax income rate is avoidable by just holding wealth in the form of zero interest rate assets, such as money[3]. Therefore, this 11% commodity tax rate between periods 0 and 0 is the highest possible commodity tax which can be implemented by an income tax system, even though a higher commodity tax rate may be desirable. The upper bounds on implementable commodity tax rates indicates that there will be an initial period of 100% interest tax rates; after this initial period, the optimal plan would presumably return to the desirable uniform pattern of distortion between these goods and period 0 consumption.

While this is all stated in commodity tax terms, the principle can be readily translated into income taxation. We can sharply illustrate the points in a simple continuous time model, similar to that we use below. Suppose $U = \int_0^\infty e^{-\rho t}(u(c) + v(l))\, dt$, r is the marginal product of capital, and τ is the interest tax rate. In that case, the social cost of one unit of consumption at time t in units of the time 0 good is e^{-rt} and the after-tax price is $e^{-r(1-\tau)t}$, implying a MRS/MRT ratio of $e^{r\tau t}$. This is displayed in Fig. 1, which shows the demand for the time t good with the time 0 good being the numeraire. This income tax is equivalent to a commodity

[2]This relation between optimal commodity tax theory and intertemporal models was recognized in Atkinson and Stiglitz.

[3]This discussion implicitly assumes no inflation. The presence of inflation complicates the discussion, but the essential features remain.

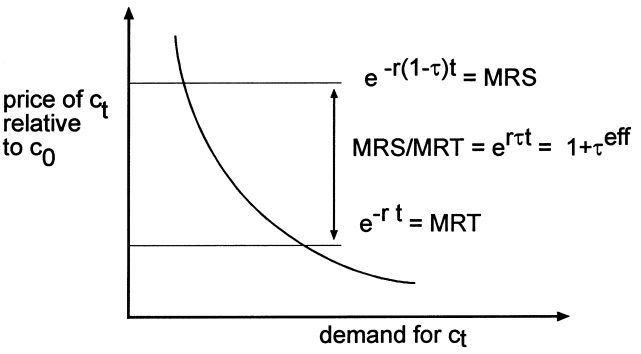

Fig. 1. Commodity tax equivalent of interest taxation.

tax on time t consumption equal to $e^{r\tau t}-1$ per unit of the time t good, as displayed in Fig. 1. Since we are assuming a time-homogeneous (in fact, time additive) utility function, the demand curve in Fig. 1 is the same for goods at all times. Therefore, a constant positive interest tax is equivalent to a commodity tax on the time t good which grows exponentially in t. This so strongly violates the intuition for uniform taxation that it is clear that a constant income tax rate on interest is not optimal.

If we instead assumed that the elasticity of demand for consumption fell over time in just the right way, then a constant interest rate tax would be optimal; this would require the demand curve in Fig. 1 for the time t good to become less elastic as t increases. However, that approach to justifying capital income taxation has not been implemented[4]. This example shows how a so-called ''constant'' tax on interest income is equivalent to a nonconstant, explosive commodity tax rate. The happens only for the interest tax case, and does not apply to pure wage taxation. Suppose that w_t is the before-tax wage at time t in terms of the time t consumption good. Let τ be a constant wage tax, and assume there is no interest tax. Then leisure demand at time t is governed by the optimality condition

$$e^{-\rho t}v'(l_t) = e^{-\rho t}(1-\tau)w_t u'(c_t) = (1-\tau)w_t e^{-rt}u'(c_0)$$

implying the time-independent, hence uniform, distortion $MRS/MRT=(1-\tau)^{-1}$ between l_t and c_0.

The second key principal we invoke is the Diamond–Mirrlees argument against taxation of intermediate goods. This is relevant here since capital goods, physical

[4]Here is a point at which the OLG literature and our results may diverge. Perhaps it is the case that the elasticity of demand for consumption falls as one ages, but that all generations have the same life-cycle tastes. Then there would be no trend in the aggregate elasticity of demand for consumption, but interest taxation would be desirable. I know of no empirical evidence on age-specific intertemporal elasticities of substitution in consumption which could justify observed rates of capital income taxation.

and human, are intermediate goods. In fact, income taxation is equivalent to sales taxation of intermediate goods. This can be seen by noting, for example, that a 100% sales tax on capital equipment is equivalent to a 50% tax on the income flow from capital equipment. Since intermediate good taxation will generally put an economy on the interior of its production possibilities frontier, capital income taxation is likely to produce similar factor distortions, particularly if there are many capital goods. Therefore, an optimal tax structure would tax only final goods.

These two principles together have strong implications for optimal tax policy in both the static and dynamic contexts. In the next sections we will make all this more precise, exploring how far we can apply these ideas to dynamic general equilibrium models.

3. A simple aggregate model

We begin by presenting the main results in a simple aggregate model. We will examine models more general than the usual ones. In particular, we will allow social increasing returns to scale. We will then examine generalizations involving other types of capital.

3.1. The Representative Agent's Problem

We assume that the representative agent's utility function depends on consumption, c, labor supply, n, and a vector of government expenditures, g, and that the utility function is globally concave in the individual decisions, but not necessarily in g. We first explore the individual's dynamic problem. If we let A denote an individual's assets, $\bar{r}$ the after-tax return on A, and $\bar{w}$ the after-tax wage rate, then the individual's maximization problem is

$$\begin{aligned} &\max_{c,n} \quad \int_0^\infty \mathrm{e}^{-\rho t} u(c, n, g)\, \mathrm{d}t \\ &\text{s.t.} \qquad \dot{A} = \bar{r}A + \bar{w}n - c \end{aligned} \tag{3}$$

His current value Hamiltonian is $u(c, n, g) + \lambda(\bar{r}A + \bar{w}n - c)$ where λ is the current value shadow price of A. The costate equation for λ is

$$\dot{\lambda} = \lambda(\rho - \bar{r}) \tag{4}$$

The first-order conditions for the choices of c and n are[5]

$$0 = u_c - \lambda, \; u_n = -\lambda\bar{w} \tag{5}$$

[5]We drop arguments of u and other functions to prevent notational clutter when those arguments are clear from context.

3.2. Representative Agent versus Alternative Models

At this point we should point out the reasons why we choose an infinite-lived, representative agent model over an overlapping generations (OLG) model, the approach taken, for example, in Atkinson and Sandmo (1980), or a model with agent heterogeneity. In the OLG literature, some results revolve around the relative social weight put on successive generations versus the period-to-period discounting of an individual. Here, there will be no such conflicts and the results are driven purely by efficiency considerations.

Furthermore, the typical OLG model assumes agents live for only two periods. This specification makes it difficult to match empirical data concerning the relevant elasticities to the two-period OLG model given the high amount of intertemporal aggregation present in the latter. Also, intertemporal nonseparability plays a critical role in the typical OLG analysis. As Judd (1985) emphasizes, the zero-tax result is far less sensitive to assumptions concerning tastes in the representative agent case than in the OLG case. Finally, most OLG analysis examines the steady state nature of optimal policy, whereas we will not make steady state assumptions and will derive more robust dynamic implications. This is easiest to do in the representative agent framework.

It is unclear how much heterogeneous tastes and productivity would alter the results. Judd (1985) shows that the zero taxation of capital is optimal even in some models with substantial agent heterogeneity and redistributive motives. In general, this paper abstracts from redistributive issues, both inter- and intragenerational.

The purpose of this paper is to pursue the logic of the zero-tax result for capital and examine its relevance for human capital. This exercise can be accomplished most cleanly in our simple infinitely-lived, representative agent model with intertemporally separable utility. In particular, the commodity tax literature, such as Diamond and Mirrlees, presumes an Arrow–Debreu general equilibrium framework, whereas the Arrow–Debreu analysis does not apply to OLG models. Since we want to investigate the intuitions of commodity tax theory in a dynamic framework, it is most natural to do in the representative agent framework. Both the OLG and representative agent extremes are flawed approximations of reality. I conjecture that the infinite-life model is a better approximation of reality than the two-period OLG model, but that conjecture remains to be confirmed. Further work is needed to see how robust these results are to alternative demographic specifications, empirically reasonable intertemporal utility functions, and distributional concerns.

3.3. Social Problem

The optimal tax problem for our representative agent model is to maximize the utility of the representative agent subject to the dynamic revenue constraint and

keeping the agent on his demand and supply curves. This is summarized in the optimal control problem

$$\begin{aligned}
&\max_{c,n,g,\bar{w},\bar{r}} && \int_0^\infty \mathrm{e}^{-\rho t} u(c, n, g)\,\mathrm{d}t \\
&\text{s.t.} && \dot{k} = f(k, n, g, t) - c - g \\
& && \dot{\lambda} = \lambda(\rho - \bar{r}) \\
& && \dot{B} = \bar{r}B - (f(k, n, g, t) - r\bar{k} - w\bar{n} - g) \\
& && \lim_{t\to\infty} |B| < \infty \\
& && 0 = u_c - \lambda,\ 0 = u_n + \lambda\bar{w},\ \bar{r} \geq 0
\end{aligned} \tag{6}$$

where we include time t in the production function $f(k, n, g, t)$ so as to allow exogenous growth factors. The Hamiltonian for this problem[6] is

$$\begin{aligned}
H = {} & u(c, n, g) + \phi_k(f(k, n, g, t) - c - g) + \phi_\lambda \lambda(\rho - \bar{r}) \\
& + \mu(\bar{r}B - (f(k, n, g, t) - r\bar{k} - w\bar{n} - g)) + \phi_c(u_c - \lambda) + \phi_n(u_n + \lambda\bar{w}) \\
& + \nu\bar{r}
\end{aligned} \tag{7}$$

where ϕ_k is the social marginal value of private capital k, ϕ_λ is the social marginal value of λ, ϕ_c is the social marginal value of the requirement that the planner's consumption choice be on the representative agents intertemporal demand curve, ϕ_n is the social marginal value of the requirement that consumption, labor supply, and net wage be consistent with the representative agent's preferences, ν is the Kuhn–Tucker multiplier on the requirement that $\bar{r} \geq 0$, and μ is the social marginal utility value of debt B. We will sometimes refer to the capital and labor income tax rates, τ_K and τ_L, which are defined by $(1-\tau_K)f_k = \bar{r}$ and $(1-\tau_L)f_n = \bar{w}$.

The first-order conditions for c, n, $\bar{r}$, and $\bar{w}$ are

$$0 = u_c - \phi_k + \phi_c u_{cc} + \phi_n u_{cn} \tag{8}$$

$$0 = u_n + \phi_k f_n - \mu(f_n - w) + \phi_c u_{cn} + \phi_n u_{nn} \tag{9}$$

$$0 = \mu n + \phi_n \lambda \tag{10}$$

$$0 = -\phi_\lambda \lambda + \mu(B + k) + \nu \tag{11}$$

We assume several public goods; for each one we have the first-order condition

[6] I will follow the standard approach to this problem. The problem statement in (6) and the Hamiltonian in (7) are not correct since the revenue constraint in (6) really should be an integral constraint stating that the present value of B equals discounted future surpluses. It can be shown that the solution to the correct isoparametric formulation of (6) does reduce to the solution we display below; we illustrate that procedure later in our analysis of human capital.

$$0 = u_g + (\phi_k - \mu)(f_g - 1) + \phi_c u_{cg} + \phi_n u_{ng} \tag{12}$$

The costate equations are

$$\begin{aligned} \dot{\phi}_k &= \phi_k(\rho - f_k) - \mu(-f_k + \bar{r}) \\ \dot{\mu} &= \mu(\rho - \bar{r}) \\ \dot{\phi}_\lambda &= \phi_\lambda \bar{r} - \phi_c + \bar{w}\phi_n \end{aligned} \tag{13}$$

There are a few items which can be immediately determined. First, $\lambda > 0$ from (5). Second, $\mu \leq 0$ since the planner can always give bonds to agents[7]. Third, if some public good g has no effect on tastes or technology but must be nonnegative, then slackness condition complementary to (12) implies $\phi_k - \mu \geq 0$; this is essentially saying that the government can buy goods and costlessly destroy them, a free disposal assumption. All other public goods will be assumed to have interior solutions to (12), using an Inada condition if necessary. We summarize this in the Lemma 1.

Lemma 1. *If bonds can be given to private agents, and the government is capable of free disposal of goods, then at all times t,*

$$\lambda > 0,\ \mu \leq 0,\ \phi_k - \mu \geq 0 \tag{14}$$

We proceed under the assumptions made in Lemma 1.

3.4. Evolution of Multipliers

We next derive important intermediate results concerning the evolution of the critical shadow prices μ, λ, and ϕ_k. The first important item to note is the constancy of the ratio $-\mu/\lambda > 0$, which we denote m. Lemma 2 holds since (4) and (13) imply $\frac{\mathrm{d}}{\mathrm{d}t}(\frac{\mu}{\lambda}) = 0$.

Lemma 2. *At all t,* $m \equiv -\mu/\lambda = -\mu_0/\lambda_0$ *is constant on the solution to* (*6*).

The quantity $m \geq 0$ is the wealth equivalent of the cost to the social objective of one more dollar of initial debt; it is the marginal social cost of funds, also known as the marginal excess burden of taxation. We will use m and $-\mu/\lambda$ interchangeably.

We will proceed under the usual assumption that the optimal policy is deterministic. This is not an innocent assumption since (6) is not a concave

[7]Technically, this requires adding a policy instrument, bond transfers to agents, which we have not included above. Since they are unlikely to be used and their only function is to show $\mu \leq 0$, we avoid some notational clutter and do not include such bond transfers explicitly.

problem, and randomization is part of the solution for some optimal taxation problems. The assumption of a deterministic policy is reasonable in many cases, particularly if m is small: we know that the solution is deterministic if $m=0$ and one suspects that an implicit function theorem is valid for m near 0. We follow the usual practice of ignoring these technical details.

Most of our results concern the structure of the optimal policy when the $\bar{r}>0$ constraint is slack. At such times, (11) implies $\phi_\lambda = -m(B+k)$, the differentiation of which implies

$$\dot{\phi}_\lambda = -m\frac{\mathrm{d}}{\mathrm{d}t}(B+k) = -m(\bar{r}(B+k)+\bar{w}\bar{n}-c)$$

When this is combined with the costate Eqs. (13) and first-order conditions (10, 11) to eliminate $\dot{\phi}_\lambda$, we find

$$\begin{aligned} -m(\bar{r}(B+k)+\bar{w}\bar{n}-c) &= \dot{\phi}_\lambda \\ &= -\bar{r}m(B+k)+\phi_c-\phi_n\bar{w} \Rightarrow -m(\bar{w}\bar{n}-c) \\ &= \phi_c-\phi_n\bar{w} \end{aligned}$$

However, (10) shows that $\phi_n = mn$; therefore $\phi_c = mc$. Combining this solution for ϕ_c with (8) and (10) yields Lemma 3.

Lemma 3. *When the $\bar{r}\geq 0$ constraint in (6) is slack, then*

$$0 = u_c - \lambda_k + m(cu_{cc}+nu_{cn}) \tag{15}$$

$$\frac{\lambda-\phi_k}{\mu} = \frac{cu_{cc}+nu_{cn}}{u_c} \tag{16}$$

$$\frac{\phi_k-\mu}{\lambda} = 1+m\left(\frac{cu_{cc}+nu_{cn}}{u_c}+1\right) \tag{17}$$

Because of its critical importance, we define a composite multiplier

$$\Lambda \equiv \frac{\phi_k-\mu}{\lambda} \tag{18}$$

Λ is the marginal social value of government wealth *holding private wealth constant* at time t since it measures the social value of increasing the capital stock by one unit but decreasing the stock of bonds by an equal amount. With lump-sum or no taxation, $\Lambda=1$ at all t. Deviations in Λ from unity occur because of the distortionary cost of taxation.

4. Optimal Provision of g

In this paper we take the standard public finance approach to modelling government expenditure. Given the role they play in our analysis, we will now discuss the implications for g. Eq. (12) displayed the first-order condition for any public good and can be simplified in certain cases. We first display a simple result.

Lemma 4. *If a public good g affects only output, then* (*12*) *reduces to* $f_g - 1 = 0$.

Therefore, the optimal provision of productive public goods satisfies full productive efficiency independent of the efficiency costs of taxation. Even if the marginal cost of funds is infinite, one still finances intermediate public goods fully to the point of productive efficiency.

If g affects only utility and only in an additively separable fashion, then (12) implies

$$\frac{u_g}{u_c} = \frac{\phi_k}{\lambda} + \left(-\frac{\mu}{\lambda} \right) = \frac{\phi_k}{\lambda} + m = \Lambda \tag{19}$$

This optimality condition equates the direct benefits of g with the opportunity cost for the social problem of foregone investment, ϕ_k/λ, plus the social marginal cost of funds, $m = -\mu/\lambda$. It is conventional to assume that $u_g/u_c > 1$ since $m > 0$, but ϕ_k/λ may be less than unity. We make assumption concerning u_g/u_c other than it being positive.

The presence of at least one public good of a special but reasonable nature implies a useful property for Λ. Lemma 5 follows from (19).

Lemma 5. *If there is a public good such that* $u_{cg} = u_{ng} = 0$ *and* $u_g(c, n, 0) = \infty$, *then for that good*

$$0 < \frac{u_g}{u_c} = \Lambda \tag{20}$$

at all t along the optimal path. Furthermore, (*16*) *implies that*

$$0 < \Lambda = 1 + m\left(\frac{cu_{cc} + \nu_{cn}}{u_c} + 1 \right) \tag{21}$$

along an optimal path.

Including government expenditure in our analysis has important advantages even if we are interested only in the tax policy questions. Lemma 5 provides a simple sufficient condition for $\Lambda > 0$ on the optimal path. While this only states the intuitive property that the social value of government wealth is positive, it is easiest to establish by including an additively separable public consumption good in the analysis. Intuitively, this allows the planner to reduce the capital stock

through increasing expenditure on this public good without affecting the incentive constraints implied by (5).

5. A Bound on Cumulative Capital Income Taxation

We now derive our basic result on the convergence of the optimal average distortion to zero. The result in Judd (1985) just stated that the optimal tax rate in the steady state was zero assuming convergence of the optimal plan to a steady state. Our new proposition does not depend on any long run convergence assumption, and essentially says that the optimal tax on capital income is zero on average in the long run, generalizing Judd (1985); Bull (1993), and Jones et al. (1997)

We establish our basic result under a simple condition. We assume that the marginal social value of government wealth, Λ, is uniformly bounded below and above over time. More precisely, we assume that for some Λ^{∞}, $\Lambda_{\infty}>0$,

$$\Lambda^{\infty} > (\phi_k(t) - \mu(t))/\lambda(t) \equiv \Lambda > \Lambda_{\infty}, \tag{22}$$

for all time t in the optimal plan. This is much more general than convergence to a steady state level or growth rate, which implies an asymptotically constant Λ. Under (22), Λ can converge to a cycle or to any other path which lies in a compact interval bounded away from zero. Condition (22) sounds simple and forms the basis of our analysis, but is rather abstract; below we will present some sufficient conditions for (22).

The critical calculation is solving the differential equation for Λ. The costate equations for the individual (4) and the social planner (13) imply that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\Lambda) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\phi_k - \mu}{\lambda}\right) = \frac{\phi_k}{\lambda}(\rho - f_k) + \frac{\mu}{\lambda}(f_k - \bar{r}) - \frac{\phi_k}{\lambda}(\rho - \bar{r}) \\ &= \left(\frac{\phi_k - \mu}{\lambda}\right)(\bar{r} - f_k) = \Lambda(\bar{r} - f_k)\end{aligned} \tag{23}$$

holds at all times. The differential equation (23) which has the general solution

$$\Lambda = \Lambda_0 \mathrm{e}^{\int_0^t (\bar{r} - f_k)\mathrm{d}s} \tag{24}$$

Assumption (22) then implies

$$\ln\left(\frac{\Lambda(0)}{\Lambda^{\infty}}\right) \le \int_0^t (f_k - \bar{r})\,\mathrm{d}s \le \ln\left(\frac{\Lambda(0)}{\Lambda_{\infty}}\right) \tag{25}$$

The inequalities in (25) brings us back to the simple *MRS*/*MRT* analysis discussed above (1). Since f_k is the social rate of return to investment and r is the private return, the gap $f_k - \bar{r}$ is the difference between instantaneous *MRT* and instantaneous *MRS*, and $\int_0^t (f_k - r)ds$ is the gap between the *MRT* and *MRS* for the time 0

and time t consumption goods. The term $\int_0^t (f_k - \bar{r})\, ds$ is the cumulative tax on capital income from time 0 to time t. The inequality in (25) says that in the optimal tax policy there are uniform bounds on this distortion. This implies that the average distortion over any long interval of time must be close to zero. We summarize this in Theorem 6.

Theorem 6. *If for some* Λ^∞, $\Lambda^\infty > 0$, (*22*) *holds then*

$$\ln\left(\frac{\Lambda(0)}{\Lambda^\infty}\right) \leq \int_0^t (f_k - \bar{r})\, ds \leq \ln\left(\frac{\Lambda(0)}{\Lambda_\infty}\right) \tag{26}$$

holds at all t. Furthermore, the average distortion $f_k - \bar{r}$ over any long interval goes to zero; more precisely, for all $t_1 > 0$,

$$\lim_{t_2 \to \infty} \frac{\int_{t_1}^{t_1+t_2} (f_k - \bar{r})\, ds}{t_2} = 0 \tag{27}$$

Therorem 6 gives us a restatement of the usual theorem. When we assume convergence of a steady state, then Therorem 6 implies a zero tax rate in the steady state. Our alternative approach provides us with a cumulative limit on the total tax distortion $f_k - \bar{r}$ independent of convergence assumptions. We suspect that this approach is robust, applicable to many models.

We next want to focus on those times when the $\bar{r} \geq 0$ constraint is not binding. Corollary 7 follows directly from Theorem 6 since $\int_0^t (f_k - \bar{r})\, ds$ is an undiscounted summation of the tax distortion.

Corollary 7. *If for some* Λ^∞, $\Lambda^\infty > 0$, (*22*) *holds then* $\bar{r} = 0$ *for only a finite amount of time.*

Corollary 7 says that the $\bar{r} \geq 0$ constraint cannot be important outside of some initial period. Therefore, we now focus on times when $\bar{r} > 0$. Combining the definition of Λ with (23, 21) yields Theorem 8.

Theorem 8. *At all times when the* $r \geq 0$ *constraint is slack, the optimal tax rate is given by*

$$f_k - \bar{r} = -\frac{\dot{\Lambda}}{\Lambda} = -m \frac{d}{dt}\left(\frac{c u_{cc} + n u_{cn}}{u_c}\right) \Lambda^{-1} \tag{28}$$

The form for the optimal tax in (28) tells us some important facts. First, the distortion $f_k - \bar{r}$ is proportional to m, the marginal excess burden. Second, if $(e u_{cc} + n u_{cn})/u_c$ is constant, then $f_k - \bar{r} = 0$ and there is no tax on capital income. If $u_{cn} = 0$, then $(c u_{cc} + n u_{cn})/u_c$ is just the inverse of the elasticity of consumption demand, and the sign of the tax is the opposite of the rate of change in this

elasticity whenever $\bar{r}>0$. This corresponds to the general inverse elasticity result. If c and n converge to steady state levels, as they typically do in simple growth models, then $(cu_{cc}+nu_{cn})/u_c$ must also converge to a constant and the optimal average distortion $f_k-\bar{r}$ converges to zero.

We next work to establish conditions sufficient for (22) to hold, which then provide conditions for Theorems 6 or 8 to apply. When we combine them with Lemma 5, we obtain Corollary 9.

Corollary 9. *If there is a pure utility public good g such that* $u_{cg}=u_{ng}=0$ *and* $u_g(c, n, 0)=\infty$ *and the optimal policy paths for c, n, and g imply that* $u_g/u_c\in[\Lambda_\infty, \Lambda^\infty]$ *for some* $\Lambda_\infty, \Lambda^\infty>0$, *then the zero average distortion expression* (*27*) *applies. In particular, if the optimal policy converges to a steady state growth path where all shadow prices grow at the same rate* (*possibly zero*), *then the steady state capital distortion,* $f_k-\bar{r}$, *is zero.*

This is a general result, making no requirement on the dynamic behavior of any level, only bounds on the marginal rate of substitution between g and c. As long as the marginal rate of substitution between g and c oscillates between two positive bounds, the long-run average distortion $f_k-\bar{r}$ must be zero.

Straightforward applications of Theorems 6 and 8 are stated in the following corollaries. We first identify a form for the utility function which causes the capital income tax to be zero except for the short run.

Corollary 10. *If* $u(c, n, g)=v^1(g)c^{1+\gamma}/(1+\gamma)+v^2(n, g)$, *then* Λ *is constant and* $f_k=\bar{r}$ *at any time when* $\bar{r}\neq 0$.

A popular functional form in the growth literature is the Cobb–Douglas utility function. Since the Cobb–Douglas utility function is not separable within a period, movements in labor supply will affect the intertemporal elasticity of substitution in consumption and lead to a nonzero tax rate. The next corollary shows that the long-run zero average tax rate result is equivalent to a plausible bound on labor supply.

Corollary 11. *If* $u(c, n, g)=c^\alpha(1-n)^\beta v(g)$ *then*

$$\frac{cu_{cc}+nu_{cn}}{u_c}=-1+\alpha+\beta(1+1/(n-1))$$

and

$$f_k-\bar{r}=-m\frac{\beta}{(1-n)(1-\alpha)+\beta n}\frac{\dot{n}}{1-n}$$

If $\bar{r} \neq 0$ *after some initial period and labor supply n is bounded uniformly above away from 1 on the optimal path then* (*27*) *holds.*

The next corollary is a simple case where the intertemporal elasticity of substitution varies in a simple way. In this case, there is always some tax or subsidy on capital.

Corollary 12. *If* $u(c, n, g) = v^g(g)(c+a)^{1+\gamma}/(1+\gamma) + v^n(g, 1-n)$ *then* $(cu_{cc} + nu_{cn})/u_c = \gamma \frac{c}{c+a}$ *and whenever* $\bar{r} \neq 0$

$$f_k - \bar{r} = -\gamma \frac{\dot{c}}{c} \frac{ac}{(c+a)^2} \frac{m}{\Lambda}$$

Corollary 12 is a case where the capital income tax is never zero, but for reasons which are consistent with the inverse elasticity rule. If $a > 0$ (< 0) and c is increasing over time, then the elasticity of demand for time t consumption is decreasing (increasing) and the tax rate on capital is positive (negative). Since we allow f to depend on t, the exogenous growth factors could result in any pattern for the consumption growth rate, $\frac{\dot{c}}{c}$. However, as long as Λ is bounded above and below, then (28) implies that consumption growth is sufficiently well-behaved so that the tax rate converges to zero fast enough so that (27) still holds. In particular, if the consumption growth rate converges to a constant then (27) holds.

These corollaries are all economically interesting cases. Assumptions which produce steady states and steady state growth rates are not necessary for our results to apply. Instead we assumed the existence of a couple particular kinds of public goods, and avoided making special asymptotic dynamic assumptions.

6. Interpretation and Robustness of the Efficiency Rule

In the analysis leading to the bound (26) above we only specified the aggregate production function. We did not make any assumptions about the firm level production function; we only assumed that factors were traded in competitive markets. Our results were stated in terms of the gap $f_k - \bar{r}$, the gap between the social marginal product of capital and the private after-tax return. In this section we explore the meaning of $f_k - \bar{r}$.

In general, the firm level production function is $Y = F(K, N, k, n, g, t)$ and must be CRTS in (K, N), the private inputs of capital and labor, for competitive equilibrium to exist. This implies $r = F_K$, $w = F_N$. However, the aggregate production function f is defined by $y = F(k, n, k, n, g, t) = f(k, n, g, t)$ and can display externalities and global economies or diseconomies of scale. The efficiency condition $\bar{r} = f_k$ implies $\bar{r} = F_K + F_k = f_k = F_K(1-\tau)$ where τ is a tax or

subsidy. Hence, full productive efficiency may require that we tax or subsidize capital investment to correct externalities.

One example would be congestion on highways. Suppose that g is highway expenditures (highway policeman, construction, repairs, etc.). Then g is privately valuable, but an increase in the aggregate capital stock, k, would imply increased use by other firms and would reduce other firms' output through increased congestion on the highways. For example, the firm's production function may have the Cobb–Douglas form

$$Y = K^{\alpha}N^{1-\alpha}k^{-\beta}n^{-\gamma}g^{\beta+\gamma} = F(K, N, k, n, g, t)$$

Note that in this case both the private firm's and society's production functions are CRTS in the private and social inputs, respectively. The negative externality of congestion is just balanced by the government expenditure. In this case, the aggregate production function is $y = f(k, n, g, t) = k^{\alpha-\beta}n^{1-\alpha-\gamma}g^{\beta+\gamma}$, which has constant returns to scale in the three factors, k, n, and g. Productive efficiency would impose, in the long run,

$$\bar{r} = f_k = F_1(k, n, k, n, g) + F_3(k, n, k, n, g) = F_1\left(1 - \frac{\beta}{\alpha}\right)$$

implying an optimal tax of β/α. Alternatively, we may have global increasing returns to scale in capital, labor, and/or government inputs. Consider a Cobb–Douglas example $Y = K^{\alpha}N^{1-\alpha}k^{\beta}n^{\gamma}g^{\delta}$. In this case, the optimal policy is a capital subsidy of β/α. Notice that the scale factors from labor and g do not affect the optimal tax on capital income. These results have no essentially dynamic flavor, following from basic ideas of corrective Pigouvian taxation.

The bound (27) is essentially a dynamic productive efficiency requirement. In the absence of externalities, it implies a zero long-run average distortion $f_k - \bar{r}$. This result should come as no surprise once we take a commodity tax perspective. Diamond–Mirrlees argues against any distortion in the allocation of intermediate goods, such as capital. The obvious counterargument to this is that capital in place at $t = 0$ represents quasi-rents, which Diamond–Mirrlees assumed to be taxed away. We do not assume that the quasi-rents of capital are taxed away. The limitations on appropriating these rents imply some initial period of capital income taxation. However, in the long run these initial quasi-rents disappear and the Diamond–Mirrlees prohibition against intermediate good taxation takes over. There may be variation around the zero rate to accommodate variation in the elasticity of substitution in consumption, but these variations must be limited.

Comparing (20) with Jones et al. (1997) shows that it is important how we model government expenditures. Jones et al. imposed a requirement that the government spend a constant fraction of output on g, whereas we have g determined endogenously. Jones et al. found a positive tax on capital in the steady state, violating our result in (27). The difference arises because economic growth in Jones et al. forces the government to spend more on g, which generates no

benefits but does increase revenue requirements; therefore, growth-reducing tax policy, such as a positive tax on capital income, is good because it reduces future worthless expenditures. We make g a choice variable, and find that we still have the zero long-run distortion result; hence, rational government expenditure does not imply positive taxation of capital income. Furthermore, for many utility functions, such as the Cobb–Douglas case, the optimal g/c ratio will not be zero in the long run; therefore, one does not need ad hoc relations between g and c in order to get growth in both.

7. Human Capital versus Physical Capital

In this section we ask how human and physical capital are treated in the optimal plan. This is itself a major question that deserves a separate treatment. We limit ourselves here to a simple example which illustrates some basic points. We use a common simplification in multisector models and examine the case where we have a single capital stock which is allocated among alternative uses in each period. This is appropriate since we are interested in the long-run character of our problem. More specifically, we examine the model

$$\begin{aligned} \max \quad & \int_0^\infty \mathrm{e}^{-\rho t} u(c, n, H, g)\, \mathrm{d}t \\ \text{s.t.} \quad & \dot{A} = \bar{r}A + \bar{w}\bar{L}(H, n) - c - x - \tau_H H \\ & \dot{H} = x \end{aligned} \tag{29}$$

where H is human capital, $L(H, n)$ is effective units of labor given n hours of labor and H in human capital, A is financial assets of the individual, and τ_H is a tax on human capital holdings. We can reformulate the problem in terms of a single state variable, $W = A + H$, resulting in the new problem

$$\begin{aligned} \max \quad & \int_0^\infty \mathrm{e}^{-\rho t} u(c, n, H, g)\, \mathrm{d}t \\ \text{s.t.} \quad & \dot{W} = \bar{r}(W - H) + \bar{w}\bar{L}(H, n) - c - \tau_H H \end{aligned}$$

where W now is all individual wealth which is allocated at each instant between A and H. The individual's Hamiltonian is

$$u(c, n, H) + \lambda(\bar{r}(W - H) + \bar{w}\bar{L}(H, n) - c - \tau_H H)$$

The costate equation for λ, now the shadow price of W, is again $\dot{\lambda} = \lambda(\rho - \bar{r})$, and the first-order conditions for c, H, and n are

$$0 = u_c - \lambda = u_n + \lambda \bar{w}\bar{L}_n = u_H + \lambda(-\bar{r} + \bar{w}\bar{L}_H - \tau_H)$$

We next formulate the social problem. Private agents own total wealth, denoted by

k. At each instant, the planner allocates some portion of k to human capital use, H, and uses the rest, $k-H$, as physical capital where the production function is $f(k-H,\ L(H,\ n),\ g,\ t)$. The optimal tax problem becomes the isoparametric problem

$$\begin{aligned}
&\max_{c,\ n,\ H,\ \tau_H} && \int_0^\infty e^{-\rho t}u(c,\ n,\ H,\ g)\,dt\\
&\text{s.t.} && \dot{k} = f(k-H,\ L,\ g,\ t) - c\\
& && \dot{\lambda} = \lambda(\rho - \bar{r})\\
& && B_0 = \int_0^\infty e^{-\int_0^t \bar{r}\,ds}\left(f(k-H,\ L,\ g,\ t) - \bar{r}(k-H) - w\bar{L} + \tau_H H\right)dt\\
& && 0 = u_c - \lambda = u_n + \lambda w\bar{L}_n = u_H + \lambda(-\bar{r} + w\bar{L}_H - \tau_H)\\
& && \bar{r} \geq 0
\end{aligned} \tag{31}$$

The optimal tax problem has a structure similar to that examined in (31); the same techniques show that the $\bar{r}\geq 0$ constraint is important only for a finite amount of time. For time periods when the $\bar{r}\geq 0$ constraint does not bind[8], we can rewrite the optimal tax problem in a direct form similar to the direct form in Atkinson and Stiglitz (1972). This approach integrates the bond equation to produce an integral constraint, which reduces (31) to[9]

$$\begin{aligned}
&\max_{c,\ g,\ n,\ H} && \int_0^\infty e^{-\rho t}\tilde{u}(c,\ n,\ H,\ g)\,dt + \mu_0(B_0 - k_0)\\
&\text{s.t.} && \dot{k} = f(k-H,\ L(H,\ n),\ g,\ t) - c
\end{aligned} \tag{32}$$

where μ_0 is the shadow price of initial debt in (31),

$$\tilde{u}(c,\ n,\ H) = u(c,\ n,\ H) + m\left(cu_c + \left(\frac{L - HL_H}{L_n}\right)u_n + Hu_H\right)$$

is the *virtual utility function*, and $m = -\mu_0/\lambda_0$ is the marginal cost of debt in wealth terms. The first-order conditions become

$$\begin{aligned}
&0 = \tilde{u}_c - \phi_k = \tilde{u}_n + \phi_k f_2 H = \tilde{u}_H + \phi_k(-f_1 + f_2 L_H)\\
&0 = \tilde{u}_n + \phi_k f_2 L_n\\
&0 = \tilde{u}_H + \phi_k(-f_1 + f_2 L_H)
\end{aligned} \tag{33}$$

[8]It is difficult to implement the $\bar{r}\geq 0$ constraint in (32). To handle it explicitly, we must replace $\bar{r}\geq 0$ with expressions including the derivatives of c and n, which turns c and n into state variables. We forego the details here since we will use the direct form only when $\bar{r} > 0$.

[9]See the Appendix A for a proof of this assertion.

Substituting out ϕ_k, we find

$$\frac{\tilde{u}_n}{\tilde{u}_c} = -f_2 L_n, \frac{\tilde{u}_H}{\tilde{u}_c} = f_1 - f_2 L_H$$

If $\tilde{u}_H = 0$, the distortion between the marginal product of k and H, $f_1 - f_2 L_H$, must be zero. One way for $\tilde{u}_H = 0$ to hold is if $u_H = 0$ and if $(L - HL_H)/L_n$ is independent of H. Theorem 13 states the crucial result.

Theorem 13. *If $u_H(c, n, H, g) \equiv 0$, and either L is CRTS in H and n, or L is Cobb–Douglas (possibly with IRTS or DRTS), then $f_1 = f_2 L_H$ in (31) whenever $\bar{r} > 0$.*

Theorem 13 shows that we have productive efficiency across human and physical capital allocation if human capital is only an intermediate good, and if H and n are aggregated in a CRTS or Cobb–Douglas fashion. The CRTS case corresponds to sufficient conditions for the Diamond–Mirrlees analysis of productive efficiency. This holds at any time when the tax rate on capital income is less than 100%, not just in the steady state. In the Cobb–Douglas case, the special functional form causes $(L - HL_H)/L_n$ to be independent of H, even if there is IRTS or DRTS in (H, n).

We next illustrate the general optimal tax rules with the case of separable, isoelastic utility. Specifically, assume

$$u(c, n, H, g) = c^{1+\gamma}/(1+\gamma) - n^{1+\eta}/(1+\eta) + \theta H^{1+\delta}/(1+\delta) + v(g) \qquad (34)$$

where $\delta < 0$ is the inverse of the elasticity of demand for human capital services for final consumption purposes and θ is an intensity parameter. Assume also that $L(H, n)$ displays constant returns to scale in H and n. Then the virtual utility function becomes

$$\tilde{u}(c, n, H, g) = \frac{1+m_c}{1+\gamma} c^{1+\gamma} - \frac{1+m_n}{1+\eta} n^{1+\eta} + \theta \frac{1+m_H}{1+\delta} H^{1+\delta} + v(g)$$

where $m_c = m(1+\gamma)$, $m_n = m(1+\eta)$, and $m_H = m(1+\delta)$. We assume that $1+m_c$, $1+m_n$, $1+m_H > 0$ so as to assure the concavity of $\tilde{u}$. Corollary 7 shows that the long-run tax on capital income is zero in this model, in which case the optimal labor and human capital tax policies imply the first-order conditions

$$f_2(1-\tau_L) = w = -\frac{1}{H}\frac{u_n}{u_c} = -\frac{\tilde{u}_n/(1+m_n)}{\tilde{u}_c/(1+m_c)} = f_2 \frac{1+m_c}{1+m_n}$$

Combining these first-order conditions in $\tilde{u}$, (33), with the first-order conditions on u, (30), implies Theorem 14.

Theorem 14. *If* $L(H, n)$ *is CRTS, utility is* (*34*), *and* $\tilde{u}(c, n, H, g)$ *is concave in* (c, n, H), *then the optimal labor tax rate is*

$$\tau_L = \frac{m_n - m_c}{1 + m_n} = m\frac{\eta - \gamma}{1 + m_n} > 0 \tag{35}$$

and the optimal tax on human capital formation is

$$\tau_H = m\frac{\gamma - \delta}{1 + m_c}\frac{H^{\delta}}{c^{\gamma}}\theta + \tau_K f_1 - \tau_L f_2 L_H \tag{36}$$

whenever $\bar{r} > 0$. In particular, if $\tau_K = 0 = \theta$, then the optimal tax system is a rate of τ_L applied to $f_2 L - f_2 L_H H$, which by Theorem 13 equals $f_2 L - f_1 H$, labor income minus the opportunity cost of human capital.

The labor tax rate rule (35) rule says that the labor tax is positive, proportional to the shadow price of funds, m, and the sum of the inverse elasticities of consumption demand and labor supply, $\eta - \gamma$. The human capital investment tax rate rule (36) says that any tax on labor income reduces τ_H, a rational response to prevent the tax on labor income from distorting human capital investment incentives. When $\tau_K = 0$ the wage tax taxes both the hours choice and human capital investment but τ_H taxes only human capital. If τ_H is negative it subsidizes human capital investment, and if sufficiently negative the net result will distort only the hours choice. In the initial phase of an optimal policy when $\tau_K > 0$, there may be a net positive tax on human capital formation set to avoid misallocation between investment in physical and human capital. However, when τ_K disappears, so will the net distortion on human capital formation.

If $\theta > 0$ then human capital is also a final good. Concavity of u implies that $\delta < 0$. Here (36) is altered by a factor proportional to θ, m, and $\gamma - \delta$. The difference $\gamma - \delta$ is a measure of the difference in elasticities of demand for c and H. If $\gamma = \delta$ then the elasticity of demand for H equals the elasticity of demand for c. Therefore, a uniform consumption tax on both final goods is optimal, can be implemented by a flat wage tax with $\tau_H = 0$. Only if the final good demand elasticity for H is less than the elasticity of demand for c will τ_H be positive. There is the possibility that human capital is a final ''bad.'' If $\theta < 0$ then human capital is a bad. Concavity of u implies $\delta > 0$ and $(\gamma - \delta)\theta > 0$. Therefore, τ_H will be positive in the final ''bad'' case.

We next examine the utility and technology functions frequently used in the endogenous growth literature. Comparing Theorem 15 to the other results shows that the zero tax results of Jones et al. and Bull are due to the particular utility function they use.

Theorem 15. *If* $L(H, n) = Hn$ *and* $u(c, n, H) = c^{\alpha} v(n) H^{\delta}$ *then* $\tilde{u}(c, n, H) = (1 + m(\alpha + \delta))u(c, n, H)$ *and the optimal tax policy sets all taxes to zero whenever* $\bar{r} > 0$.

Comparing the last two theorems shows just how sensitive labor tax results are to the specification of $L(H, n)$. Recent economic growth studies differ in their choice. Mankiw et al. (1992) assume output is CRTS Cobb–Douglas in physical capital, human capital, and hours, a specification consistent with the assumptions in Theorem 13, whereas Jones et al. and Bull assume the specification of Theorem jonesthm so that the equilibrium has a positive steady state growth rate. Assuming $L(H, n)=Hn$ causes production functions such as $K^{\alpha}(Hn)^{1-\alpha}$ to have increasing returns to scale in the inputs (K, H, n). This increasing returns to scale property does not prevent the existence of competitive equilibrium in factor markets since an individual must sell Hn, not H and n separately. However, tax policy can distinguish among the factors, and these results show us that our standard intuitions from constant returns to scale models fail us in general.

Since there is no strong empirical case favoring any particular form for $L(H, n)$, it is difficult to make precise statements about the optimal labor and consumption taxes. In particular, earlier results arguing for no taxation in the steady state rest on special assumptions. However, we always find a zero long-run average tax rate, as expressed in (27), on physical capital. Furthermore, positive taxation of human capital is optimal only when human capital is a final good or $L(H, n)$ is neither CRTS nor Cobb–Douglas. Therefore, the simplest specifications argue for productive efficiency between human and physical capital along with labor income (or, equivalently, consumption) taxation in the long run.

8. Consumption Taxation versus Consumption Tax Proposals

The analysis above focussed on a model far simpler than the real world. However, it can be used to discuss basic aspects of proposed tax reforms in the U.S.[10]. This discussion will help us understand how to interpret the results above. This is essential to do since the complex financing of education makes it less clear how the critical elements in our model relate to actual tax systems.

The Flat Tax (see Hall and Rabushka, 1995), consumption tax (see Bradford, 1986, and McClure and Zodrow, 1996), the USA tax (see Weidenbaum, 1997, for a description) and VAT proposals all argue that the tax base should be consumption—the differences among these proposals are primarily accounting differences, not economic differences. The principle advantage of a consumption tax is that it would eliminate the bias against investment and savings in the current tax system. While these writers do not present in detail their assumptions concerning tastes, technology, and demographics, the model we present here is consistent with their

[10]I have little idea how these comments apply to tax systems in other countries. Hopefully, the discussion below will highlight the critical tax and institutional details so that the reader can deduce the implications for his country.

apparent assumptions. In fact, very similar models are used by some of these writers[11].

However, these proposals do not actually propose a true consumption tax which eliminates biases against all investment. These major tax reform proposals define ''consumption'' as income minus investment in physical capital only. The various tax proposals differ little on their treatment of human capital investments. The Hall-Rabushka-Armey-Forbes Flat Tax proposals clearly allow few deductions for educational investments; the sales tax and VAT proposals are similar. The USA tax allows limited deductibility of some educational expenses. Tax reform advocates ignore the educational issues we analyzed above in their discussions, and offer no rationale for their proposal to favor physical capital investments over human capital investments. In any case, these proposals are not true consumption taxes.

The real picture is more complex. Both the current tax system and ''consumption tax'' proposals essentially expense the foregone wages of any student, costs which comprise roughly half of the total cost of education[12]. This expensing feature has been emphasized in much of the literature, such as Boskin (1977) and Heckman (1976). Under the current tax system, firms can deduct expenditures on employee training; that would presumably continue under most consumption tax proposals. Some have even argued that the current income tax system is biased in favor of human capital investment because the foregone wages are expensed; see Hamilton (1987) for a discussion of this issue.

The difference between the current tax system and ''consumption tax'' proposals is the treatment of some of the marketed goods and services used in formal education. The current tax system does somewhat better that these ''consumption tax'' proposals. For example, charitable contributions to educational institutions are currently deductible, but not under these consumption tax proposals. Currently, the U.S. tax system allows some deductibility of educational investments through the state and local tax deduction—parents pay taxes to their local governments for public schools and then deduct these taxes from their Federal income tax if they itemize. In many communities, the majority of voters are in households where the primary income earner itemizes; in particular, those in upper income brackets in many states itemize simply because of state and local property and income taxes. The elimination of the state and local tax deduction would increase the price they pay for educational services, presumably reducing educational expenditures. While own-time may comprise most of the direct personal costs of education, those costs which are paid indirectly through taxation are also important, particularly if one invokes Tiebout-style arguments. In any case, the key question for our purposes is whether those expenditures are affected

[11]For example, Hall and Rabushka cite similar representative agent analyses.

[12]Of course, one reason why the foregone wages comprise half of the cost may be that it is the one input which is always expensed. Therefore, we cannot conclude that the current system gets it ''half right.''

by changes in the Federal Income tax treatment of local taxation. Feldstein and Metcalf (1987) offer strong evidence that local expenditures are affected by Federal Income tax rules, supporting the approach we take here.

The problem with the current U.S. system is that only some taxpayers get to deduct educational expenditures. A true consumption tax consistent with the theory outlined above would allow *everyone* to subtract *all* human capital investment expenses from the tax base, whether they are paid directly or indirectly through local taxation. An optimal tax system may want to tax the consumption component of education, but that would be difficult to implement. This problem is not unique to human capital investments. Corporate executives need chairs, but do they need luxurious leather chairs? Again, there seems to be no essential difference between human and physical capital investments which justifies differential treatment.

Furthermore, there is no evidence that there is a significant consumption component to education. If significant educational expenditures were consumption goods and capital markets were perfect, then the average return on all educational expenditures would be below the return on alternative investments. I know of no evidence for this; in fact, the mean return on education is similar to that on physical capital investments and the conventional view is that they do not differ in terms of riskiness (see Becker, 1976). There appears to be no reason to reject $\theta = 0$ in our model, in which case human capital is essentially an intermediate good and an optimal tax policy would treat human and physical capital identically.

Many tax reform advocates apparently want to shift investment towards physical capital and away from educational investments, but never explain why. Our analysis shows that there is no aggregate efficiency reason for favoring physical capital investments over human capital investments.

9. Conclusions

We have shown that the optimal long-run average tax on capital income is zero in a wide variety of conditions. The key assumptions here are competitive factor markets, a flexible set of tax policy instruments, and the presence of some public goods. We substantially generalize previous results by replacing steady-state convergence assumptions with much looser compactness assumptions. We also avoid special functional forms for tastes and technology. We show that the nature of the optimal tax system in representative agent models do not depend on the presence or stability of steady state growth.

This analysis is much closer to simple intuitions from the commodity taxation literature. In particular, since both physical and human capital are intermediate goods, the Diamond–Mirrlees analysis clearly argues against capital income taxation, human and physical. We also show that the inverse elasticity rule from optimal commodity tax theory implie a zero tax on capital income in the long run.

The zero tax result is surely not a universal truth. In particular, Hubbard and Judd (1986) show that asset income taxation is desirable if individuals face binding borrowing constraints. Also, Atkinson and Sandmo (1980) show that restrictions on bond policy may affect the results. However, the purpose of this paper is to outline the basic intuition for some zero tax results, and how to properly extend it to alternative forms of capital, such as human capital. We further point out that the results have implications for tax reform proposals, showing that supposed consumption tax proposals are not true consumption taxes since they do not allow deductions for many human capital investments and are, therefore, not consistent with optimal tax theory.

Acknowledgements

The author thanks anonymous referees, and seminar participants at the NBER, the Hoover Institution, and the University of Wisconsin for their comments. Special thanks go to Rodolfo Manuelli for his comments and finding an error in Theorem 13. I gratefully acknowledge the support of NSF grant SBR-9309613.

Appendix A

Proof of (32)

The individual problem is

$$\max \quad \int_0^\infty e^{-\rho t} u(c, n, H)\, dt$$

$$\text{s.t.} \quad \dot{W} = \bar{r}(W - H) + w\bar{L}(H, n) - c - \tau H$$

with Hamiltonian $u(c, n, H) + \lambda(\bar{r}(W-H) + w\bar{L}(H, n) - c - \tau H)$. The first-order conditions are

$$\begin{aligned} 0 &= u_c - \lambda \\ 0 &= u_n + \lambda w \bar{L}_n \\ 0 &= u_H + \lambda(-\bar{r} + w\bar{L}_H - \tau) \end{aligned} \tag{37}$$

When we use the costate equation $\bar{r} = \rho - \frac{\dot{\lambda}}{\lambda}$, the present value bond constraint becomes

$$B_0 = \int_0^\infty e^{-\int_0^t (-\frac{\dot{\lambda}}{\lambda} + \rho)\, ds} (f(k - H, L(H, n)) - \bar{r}k - w\bar{L}(H, n) + (\bar{r} + \tau)H)\, dt$$

Computing $\exp(-\int_0^t(-\frac{\dot{\lambda}}{\lambda}+\rho)\,\mathrm{d}s)$ and applying the individual first-order conditions implies

$$B_0=\int_0^\infty \mathrm{e}^{-\rho t}\frac{1}{\lambda_0}(u_c c+\lambda\dot{k}+\dot{\lambda}k-\rho\lambda k+u_n\frac{L}{L_n}+\left(u_H-\frac{u_n}{L_n}L_H\right)H\Big)\,\mathrm{d}t \tag{38}$$

which, since $\int \mathrm{e}^{-\rho t}\,(\lambda\dot{k}+\dot{\lambda}k-\rho\lambda k)=\lambda_0 k_0$ by integration by parts, reduces to

$$B_0=\int_0^\infty \mathrm{e}^{-\rho t}\lambda_0^{-1}\left(cu_c+\left(\frac{L-HL_H}{L_n}\right)u_n+Hu_H\right)\mathrm{d}t-k_0 \tag{39}$$

This new expression of the isoparametric constraint implies the optimal tax problem (32).

References

Atkinson, A.B., Sandmo, A., 1980. Welfare implications of the taxation of savings. The Economic Journal 90, 529–549.

Atkinson, A., Stiglitz, J., 1972. The structure of indirect taxation and economic efficiency. Journal of Public Economics 1, 97–119.

Auerbach, A.J., 1979. The optimal taxation of heterogeneous capital. Quarterly Journal of Economics 93, 589–612.

Becker, G., 1976. Human Capital: A Theoretical and Empirical Analysis. Chicago and London, University of Chicago Press.

Boskin, M., 1977. Notes on the tax treatment of human capital. In: U.S. Department of Treasury Office of Tax Analysis, Conference on Tax Research 1975. U.S. Government Printing Office, Washington, D.C, pp. 185–195.

Bradford, D., 1986. Untangling the Income Tax. Harvard University Press.

Bull, N., 1993. When all the optimal dynamic taxes are zero. Board of Governors of the Federal Reserve System. Mimeo.

Eaton, J., 1981. Fiscal policy, inflation and the accumulation of risky capital. Review of Economic Studies 48, 435–445.

Diamond, P.A., Mirrlees, J.A., 1971. Optimal Taxation and Public Production. American Economic Review 61, 8–27 and 261–278.

Diamond, P.A., 1973. Taxation and public production in a growth setting. In: Mirrlees, J.A., Stern, N.H. (Eds.) Models of Economic Growth. Macmillan, London, pp. 215–234.

Hall, R.E., Rabushka, A., 1995. Low Tax, Simple Tax, Flat Tax, 2nd ed. McGraw-Hill, New York.

Feldstein, M.S., Metcalf, G.E., 1987. The effect of federal tax deductibility on state and local taxes and spending. Journal of Political Economy 95, 710–736.

Hamilton, J.H., 1987. Taxation, savings, and portfolio choice in a continuous time model. Public Finance 42, 264–282.

Heckman, J., 1976. A life-cycle model of earnings, learning, and consumption. Journal of Political Economy 84, S11–S44.

Hubbard, R.G., Judd, K.L., 1986. Liquidity constraints, fiscal policy, and consumption. Brookings Papers on Economic Activity 0, 1–50.

Jones, L.E., Manuelli, R., Rossi, P., 1997. On the optimal taxation of capital income. Journal of Economic Theory 73, 93–117.

Judd, K.L., 1985. Redistributive taxation in a simple perfect foresight model. Journal of Public Economics 28, 59–83.

McClure, C., Zodrow, G.R., 1996. A hybrid approach to the direct taxation of consumption. In: Boskin, M. (Ed.), Frontiers of Tax Reform. Hoover Press, Stanford, CA.

Mankiw, N.G., Romer, D., Weil, D.N., 1992. A contribution to the empirics of economic growth. Quarterly Journal of Economics 107, 407–437.

Weidenbaum, M., 1997. The Nunn–Domenici USA tax. In: Boskin, M. (Ed.), Frontiers of Tax Reform. Hoover Press, Stanford, CA.