Capital income taxation and the Atkinson-Stiglitz theorem

Firouz Gahvari
Department of Economics
University of Illinois at Urbana-Champaign, USA

Luca Micheletto*
Department of Law, University of Milan, Italy,
UCFS Uppsala, Sweden, and
CESifo Germany

July 2016

Abstract

Accounting for the role of financial system and money holdings in an optimal nonlinear income tax model, we argue that capital income taxation is a non-redundant policy tool even if individual preferences are separable between leisure and other goods.

JEL classification: H21.

Keywords: Nonlinear labor income taxation, capital income taxation, redistribution.

*Corresponding author. Department of Law "C. Beccaria", University of Milan, via Festa del Perdono 7, 20122 Milan, Italy. *E-mail address*: luca.micheletto@unimi.it (Luca Micheletto).

1 Introduction

In their classic contribution, Atkinson and Stiglitz (1976) prove that if preferences are weakly separable in leisure and other goods, an optimal nonlinear labor income tax is sufficient to implement any incentive compatible Pareto-efficient allocation. The reason is that, under separability, commodity taxes have no bite on the self-selection constraints faced by the government in the design of the nonlinear income tax. Treating current and future consumption as two different commodities, a corollary of the Atkinson and Stiglitz (1976) result is that, under separability, there should be no taxes on interest income.

In this paper, we argue that this result is due to the strategy of ignoring the role of financial system and money holdings in taxation models. This is quite understandable as accounting for financial institutions and money introduces many complications into the picture that are often best ignored. Including them often leads the researcher astray by having to address issues that are not pertinent to the main question at hand. However, when it comes to taxation of savings, this strategy can be quite misleading. It implies that all individual types, high- and low-productivity (rich and poor) earn the same rate of return on their savings. Yet, there is ample empirical evidence that lower income consumers tend to carry a higher percentage of their income in cash (which, ignoring inflation, bear no rate of return).[1] An immediate implication of this observation is that even with a unique rate of interest on *real* savings, the rate of return on *total* savings (that include cash) will not be the same for the poor and the rich. Indeed, the rate will be higher for the rich. Now given that earnings and earning abilities are positively correlated, one can only conclude that high-productivity persons are somehow able to earn a higher rate of return on their savings. This can be inferred from the recent work of Williamson (2008) who, in a monetary model, makes a distinction between "connected"

[1] See, e.g., Albanesi (2007).

and "unconnected" agents in terms of their access to financial institutions.[2] He argues that different agents may have to carry different levels of cash balances because of their sophistication and/or their different levels of access to other financial instruments. One would naturally expect that sophistication be positively correlated with earning abilities, and access to other financial instruments with one's actual income (as well as sophistication). Indeed, given that higher earning abilities are translated into higher earnings, it is no wonder why low-income groups carry a higher percentage of their income in cash (as compared to high-income groups). The upshot is that, as a rule, the more productive and the richer an individual, the higher would be the rate of return that he can secure on his savings.

We will show that incorporating these more realistic assumptions makes capital income taxation a useful policy tool even if individual preferences are separable between leisure and other goods. Of course, there are other reasons why, despite separability in preferences, one might still want to tax capital income (as recently surveyed by Stiglitz, 2015, pp. 245–246).[3] Yet the mechanism that we highlight in this paper is a different one as it simply incorporates a higher degree of realism into our modeling.

2 The model

Consider a two-period model wherein individuals work in the first period and consume in both. The economy is populated with two groups of agents who differ in labor

[2] His aim in doing so is to show that this source of heterogeneity causes monetary policy to have significant redistributive implications which, in turn, often leads to a negation of the Friedman rule.

[3] For example, to restore the desirability of capital income taxes, Saez (2002) assumes that, besides differing in labor market productivities, individuals also differ in tastes, while Cremer et al. (2003) assume that they differ in inherited wealth; Pirttilä and Tuomala (2001) assume that relative wage rates are sensitive to savings via their effect on capital accumulation; Christiansen and Tuomala (2008) introduce the possibility of income shifting by assuming that labor income can be camouflaged as capital income for tax purposes; Aronsson et al. (2009) relax the assumption that the labor market is perfectly competitive. Finally, Piketty and Saez (2012) and Bastani et al. (2013) highlight the desirability of taxing capital income in models where agents face uninsurable uncertainty about, respectively, future returns and future labor market productivities.

productivity: some are high-skilled, and are paid a wage rate w^h, and some are low-skilled, and are paid a wage rate w^ℓ, with $w^\ell < w^h$. Production takes place through a linear technology with different types of labor as inputs. Total population is normalized to unity and the fraction of agents of type j is denoted by π^j (with $j = \ell, h$). Agents' preferences are separable between leisure and other goods and are represented by the concave utility function:

$$U = u(c_1, c_2) + v(l), \tag{1}$$

where l denotes leisure, c_1 consumption in the first period and c_2 consumption in the second period. All goods (including leisure) are assumed to be normal and the utility function is strictly increasing in each argument.

Agents can transfer resources across the two periods at an interest rate denoted by r^j, $j = \ell, h$. That the interest rate differs for different types reflects Williamson's (2008) notion of connectedness as discussed in the Introduction. Specifically, this notion can be captured through differences in productivities and incomes. Higher productivity goes hand in hand with a higher inherent degree of sophistication on the part of the individual. The more sophisticated an individual, the higher his knowledge of the financial sector and his ability to take advantage of financial institutions. At the same time, the more income an individual earns, the more "connections" he makes (higher incomes open more doors). This make r^j a function of income as well. Both of these factors appear to underlie Williamson's concept of financial connectedness allowing richer individuals to earn a higher rate of return on their savings. Denoting gross labor income by I, this means that:

$$r^j = r^j(I), \tag{2}$$

where $\partial r^j(I)/\partial I \geq 0$ and $r^h(I) > r^\ell(I), \forall I$.

The government uses labor- and interest-income taxation to redistribute among agents. Neither an agent's skill type (i.e. his wage rate w) nor his labor supply, denoted by L, are publicly observable. Labor income I $(= wL)$ is instead publicly observable.

Thus, while first-best type-specific lump-sum taxes/transfers are unfeasible, labor income can be taxed via a nonlinear income tax function $T(I)$. Regarding interest income taxation, we assume that savings transactions can only be observed *anonymously*, not at an *individual* level:[4] one can observe the amount in a savings account, but one can not know whether the savings are done by the official holder of the account or by someone else (like a close friend or a relative).[5] This means that interest income can only be taxed/subsidized at a proportional rate.

To characterize the (constrained) Pareto-efficient allocations, one has to account for the economy's resource balance, the standard self-selection constraints due to our informational structure, and the implementability constraints caused by the linearity of the interest income tax—itself due to informational constraint. To this end, we derive an optimal revelation mechanism. This will consist of a set of type-specific before-tax labor incomes, I^j's (with $j = \ell, h$), after-tax incomes, B^j's (with $j = \ell, h$), and a proportional interest income tax rate, τ.[6]

To proceed further, it is necessary to consider the optimization problem of an individual for a given mechanism (τ, B, I). This is necessitated by the fact that the mechanism determines personal consumption levels only indirectly, namely through prices. The mechanism assigns the triple $\left(\tau, B^j, I^j\right)$ to a young individual who reports his type as j. The individual will then allocate B^j between first- and second period consumption.

Denote by Θ the total time endowment of each individual. Given any vector (τ, B, I),

[4] The same observational assumption is made in Boadway et al. (2000), Cremer et al. (2003), Blomquist and Micheletto (2008), Bastani et al. (2013), and Gahvari and Micheletto (2014).

[5] If one tried to tax savings by a nonlinear function, the incentives for a person with a high marginal tax on savings to ask a friend, with a lower marginal tax, to do the savings for him would be large. Our assumption here parallels the assumption usually made about purchases of commodities. The anonymous transactions can be observed and taxed by a proportional tax, but the personal consumption levels are not publicly observable.

[6] This procedure determines τ from the outset; a complete solution to the optimal tax problem per se, i.e. determination of I^j by the individuals via utility maximization, then requires only the design of a general income tax function $T(I)$ such that $B^j = I^j - T\left(I^j\right)$.

an individual of type j chooses c_1 and c_2 to maximize

$$U = u\left(c_1, c_2\right) + v\left(\Theta - \frac{I}{w^j}\right), \quad j = \ell, h, \tag{3}$$

subject to the intertemporal budget constraint

$$c_1 + \frac{1}{1 + (1-\tau) r^j\left(I\right)} c_2 = B. \tag{4}$$

Substituting (4) into (3), the first-order condition of the problem above is:

$$\frac{\partial u / \partial c_2}{\partial u / \partial c_1} = \frac{1}{1 + (1-\tau) r^j\left(I\right)}. \tag{5}$$

For a given bundle (I, B), condition (5), and the individual's intertemporal budget constraint (4), yield the conditional demands for the j-type's first- and second-period consumption, c_1^j and c_2^j. Substituting these conditional demands into (3) provides the conditional indirect utility of a j-type agent for a given (I, B)-bundle. Denoting this conditional utility by $V\left(\tau, B, I; w^j\right)$, and using $V^j\left(\tau, B, I\right)$ as a shorthand for $V\left(\tau, B, I; w^j\right)$, the government's problem can then be represented as,

$$\max_{\tau, I^\ell, B^\ell, I^h, B^h} \quad \sum_{j=\ell,h} \delta^j V^j\left(\tau, B^j, I^j\right)$$

subject to:

$$\begin{aligned} V^h\left(\tau, B^h, I^h\right) &\geq V^h\left(\tau, B^\ell, I^\ell\right), \qquad (\lambda) \\ \sum_{j=\ell,h} \pi^j \left[I^j - B^j + \tau r^j\left(I^j\right) s^j\right] &\geq \bar{R}, \qquad (\mu) \end{aligned}$$

where δ^j's are positive social welfare weights (with the normalization $\sum_{j=\ell,h} \delta^j = 1$), Lagrange multipliers are within parentheses, $s^j \equiv B^j - c_1^j$ denotes the savings of a j-type agent, and $\bar{R}$ is an exogenous revenue requirement.

The λ-constraint is the self-selection constraint requiring that a high-skilled agent is (weakly) better off by choosing the (I, B)-bundle intended for him by the government

rather than behaving as a "mimicker" and choosing the bundle (I^ℓ, B^ℓ) intended for low-skilled agents.[7] The μ-constraint is the government's budget constraint.

Denote the indirect utility of a high-skilled agent mimicking a low-skilled agent by $V^{h\ell} \equiv V^h\left(\tau, B^\ell, I^\ell\right)$, and denote by $\partial\widetilde{s}^j/\partial\tau$ the compensated change in savings for a type-j agent following a marginal increase in τ: $\partial\widetilde{s}^j/\partial\tau \equiv \partial s^j/\partial\tau + r^j s^j\left(\partial s^j/\partial B^j\right) < 0$. Proposition 1 states our main result.

Proposition 1 *In the presence of an optimal nonlinear labor income tax, and despite the separability of preferences in labor supply and goods, a capital income tax is a useful policy tool. Its optimal value is given by*

$$\tau = \frac{\lambda\left(r^\ell s^\ell - r^{h\ell} s^{h\ell}\right)\left(\partial V^{h\ell}/\partial B^\ell\right)}{\mu\sum_{j=\ell,h}\pi^j r^j\left(\partial\widetilde{s}^j/\partial\tau\right)} > 0. \tag{6}$$

Proof. We derive eq. (6) in the Appendix. To prove that $\tau > 0$, first observe that, since the denominator of the right hand side of (6) is negative, interest income ought to be taxed (subsidized) when $r^\ell s^\ell < (>)\, r^{h\ell} s^{h\ell}$. Given our assumption that, $\forall I$, $r^h(I) > r^\ell(I)$, and given that a high-skilled mimicker earns the same labor income as a low-skilled agent, we can conclude that $r^{h\ell} > r^\ell$. Thus, if $s^{h\ell} \geq s^\ell$, one immediately gets $r^{h\ell} s^{h\ell} > r^\ell s^\ell$. But the same result is obtained also if $s^{h\ell} < s^\ell$. This is due to the fact that $r^{h\ell} > r^\ell \Longrightarrow c_2^{h\ell} > c_2^\ell$;[8] with $c_2 = [1 + r(1-\tau)]\, s$, the last inequality implies that $\left[1 + r^{h\ell}(1-\tau)\right] s^{h\ell} > \left[1 + r^\ell(1-\tau)\right] s^\ell$ or, rearranging, that $(1-\tau)\left(r^{h\ell} s^{h\ell} - r^\ell s^\ell\right) > s^\ell - s^{h\ell} > 0$. ■

Intuitively, despite separability in preferences, a high-skilled mimicker and a low-skilled agent end up earning different rates of return on their savings and will have different (intertemporal) consumption profiles. Interest income taxation can then be

[7] Restricting attention to the so called "normal" case when the direction of redistribution goes from the high-skilled to the low-skilled agents, we can safely disregard the self-selection constraint requiring low-skilled agents not to mimic high-skilled agents.

[8] A high skilled behaving as a mimicker and a low-skilled agent both choose the bundle (I^ℓ, B^ℓ) intended by the government only for the latter. For a given value of B^ℓ, a higher r implies substitution and income effects that both push towards increasing second-period consumption.

used as an instrument to relax the binding self-selection constraint faced by the government in the design of the nonlinear income tax.

Before concluding, three remarks are in order.

First, allowing for individual preferences that are non-separable between leisure and other goods might change the result that it is optimal to tax interest income only if leisure and second-period consumption are sufficiently strong substitutes.[9]

Second, notice that the fact that r has been assumed to also be a function of I (see (2)) has really no bite for our result. What drives the result that $\tau > 0$ is the assumption that, $\forall I$, $r^h(I) > r^\ell(I)$, capturing the idea that agents differ in terms of financial sophistication.

Finally, notice that the same qualitative result, *viz.* the fact that $\tau > 0$, would be obtained in a setting where r can be affected by the time-effort exerted by the saver searching for better investment opportunities. Consider for instance the case where, denoting by e the time devoted by agents searching for more profitable investment opportunities, r is defined as $r \equiv r(e)$, with $r' > 0$ and $r'' < 0$, and individual preferences are represented by $U = u(c_1, c_2) + v\left(\Theta - \frac{I}{w^j} - e\right)$. Even when the $r(\cdot)$ function is the same for all agents, it is straightforward to show that, with a high-skilled mimicker and a low-skilled agent choosing the same (I^ℓ, B^ℓ)-bundle, the former finds it optimal to spend more time than the latter searching for better investment opportunities, i.e. $e^{h\ell} > e^\ell$.[10]

[9] Complementarity between leisure and second-period consumption would strengthen the argument for using interest income taxation as a mimicking-deterring device. This is due to the fact that, since a high-skilled mimicker, being more productive than a low-skilled agent, works less than a low-skilled agent, complementarity between leisure and second-period consumption implies that a mimicker would like to consume more in the second period (and therefore save more) than a low-skilled. This introduces a second mechanism, acting on top of the one due to the difference in financial sophistication, that works in the direction of making the difference $r^\ell s^\ell - r^{h\ell} s^{h\ell}$ negative. If instead leisure and second-period consumption were substitutes, the sign of the optimal interest income tax would be a priori ambiguous. When leisure and second-period consumption are substitutes, a mimicker would like to consume less in the second period (and therefore save less) than a low-skilled. In this case the sign of $r^\ell s^\ell - r^{h\ell} s^{h\ell}$ is ambiguous. On one hand, the difference in financial sophistication tends to make it negative; on the other hand, the fact that leisure and second-period consumption are substitutes works in the opposite direction.

[10] Intuitively, this is due to the fact that a high-skilled mimicker has more time available, net of

With $\partial r/\partial e > 0$, this in turn would imply $r^{h\ell} > r^{\ell}$, and one could then proceed, as done in the proof of Proposition 1, to prove that $r^{\ell}s^{\ell} - r^{h\ell}s^{h\ell} < 0$. Effectively, assuming that $r = r(e)$ (with $r' > 0$ and $r'' < 0$) allows to endogenously derive a difference in financial sophistication based on a model where the only dimension of (exogenous) heterogeneity is labor productivity.

3 Concluding remarks

This paper has developed a simple model to show that the widely-held belief that the Atkinson and Stiglitz (1976) theorem proves capital incomes should not be taxed is unwarranted. It has argued that this result is in fact due to the strategy of ignoring the role of financial system and money holdings in taxation models. Incorporating these realistic features makes capital income taxation a useful policy tool even if individual preferences are separable between leisure and other goods. The underlying reason is that in an economy with financial institutions and money holdings, the more productive and richer individuals are also able to secure a higher rate of return on their savings than the less productive and poorer individuals. Consequently, interest income taxation relax the binding self-selection constraint the government faces in the design of the optimal nonlinear labor income tax.

working hours I/w, than a low-skilled agent.

Appendix

Derivation of eq. (6): The first-order conditions to the government's problem are:

$$\left(\delta^h+\lambda\right)\frac{\partial V^h}{\partial I^h}+\mu\pi^h\left\{1+\tau\left[s^h+r^h\frac{\partial s^h}{\partial r^h}\right]\frac{\partial r^h\left(I^h\right)}{\partial I^h}\right\}=0, \tag{A1}$$

$$\delta^\ell\frac{\partial V^\ell}{\partial I^\ell}-\lambda\frac{\partial V^{h\ell}}{\partial I^\ell}+\mu\pi^\ell\left\{1+\tau\left[s^\ell+r^\ell\frac{\partial s^\ell}{\partial r^\ell}\right]\frac{\partial r^\ell\left(I^\ell\right)}{\partial I^\ell}\right\}=0, \tag{A2}$$

$$\left(\delta^h+\lambda\right)\frac{\partial V^h}{\partial B^h}+\left(-1+\tau r^h\frac{\partial s^h}{\partial B^h}\right)\mu\pi^h=0, \tag{A3}$$

$$\delta^\ell\frac{\partial V^\ell}{\partial B^\ell}-\lambda\frac{\partial V^{h\ell}}{\partial B^\ell}+\left(-1+\tau r^\ell\frac{\partial s^\ell}{\partial B^\ell}\right)\mu\pi^\ell=0, \tag{A4}$$

$$\left(\delta^h+\lambda\right)\frac{\partial V^h}{\partial \tau}+\delta^\ell\frac{\partial V^\ell}{\partial \tau}-\lambda\frac{\partial V^{h\ell}}{\partial \tau}+\mu\sum_{j=\ell,h}\pi^j\left(r^js^j+\tau r^j\frac{\partial s^j}{\partial \tau}\right)=0. \tag{A5}$$

Using Roy's identity, we can rewrite eq. (A5) as,

$$-\left(\delta^h+\lambda\right)r^hs^h\frac{\partial V^h}{\partial B^h}-\delta^\ell r^\ell s^\ell\frac{\partial V^\ell}{\partial B^\ell}+\lambda r^{h\ell}s^{h\ell}\frac{\partial V^{h\ell}}{\partial B^\ell}+\mu\sum_{j=\ell,h}\pi^j\left(r^js^j+\tau r^j\frac{\partial s^j}{\partial \tau}\right)=0. \tag{A6}$$

Multiplying eq. (A3) by $r^hs^h\left(\partial V^h/\partial B^h\right)$ and eq. (A4) by $r^\ell s^\ell\left(\partial V^\ell/\partial B^\ell\right)$, and adding up the resulting equations to eq. (A6) gives,

$$\begin{aligned}
&\left(\delta^h+\lambda\right)r^hs^h\frac{\partial V^h}{\partial B^h}+\delta^\ell r^\ell s^\ell\frac{\partial V^\ell}{\partial B^\ell}-\lambda\frac{\partial V^{h\ell}}{\partial B^\ell}r^\ell s^\ell\\
&+\left(-1+\tau r^h\frac{\partial s^h}{\partial B^h}\right)\mu\pi^hr^hs^h+\left(-1+\tau r^\ell\frac{\partial s^\ell}{\partial B^\ell}\right)\mu\pi^\ell r^\ell s^\ell\\
&+\mu\sum_{j=\ell,h}\pi^j\left(r^js^j+\tau r^j\frac{\partial s^j}{\partial \tau}\right)\\
&+\lambda r^{h\ell}s^{h\ell}\frac{\partial V^{h\ell}}{\partial B^\ell}-\left(\delta^h+\lambda\right)r^hs^h\frac{\partial V^h}{\partial B^h}-\delta^\ell r^\ell s^\ell\frac{\partial V^\ell}{\partial B^\ell}=0.
\end{aligned}$$

Simplifying and collecting terms yields,

$$\mu\tau\left[\sum_{j=\ell,h}\left(r^j\right)^2\pi^j\frac{\partial s^j}{\partial B^j}s^j+\sum_{j=\ell,h}\pi^jr^j\frac{\partial s^j}{\partial \tau}\right]+\lambda\frac{\partial V^{h\ell}}{\partial B^\ell}\left(r^{h\ell}s^{h\ell}-r^\ell s^\ell\right)=0. \tag{A7}$$

Using a "tilde" symbol to denote a compensated variable and exploiting a Slutsky-type decomposition to write $\partial s^j/\partial \tau = \partial \widetilde{s}^j/\partial \tau - r^j s^j \left(\partial s^j/\partial B^j\right)$, we can finally rewrite eq. (A7) as,

$$\mu\tau \left(\sum_{j=\ell,h} \pi^j r^j \frac{\partial \widetilde{s}^j}{\partial \tau} \right) + \lambda \frac{\partial V^{h\ell}}{\partial B^\ell} \left(r^{h\ell} s^{h\ell} - r^\ell s^\ell \right) = 0.$$

Rearranging terms yields eq. (6).

References

[1] Albanesi, S., 2007. Redistribution and optimal monetary policy: results and open questions. *Rivista di Politica Economica* 97, 3–47.

[2] Aronsson, T., Sjögren, T., and T. Dalin, 2009. Optimal taxation and redistribution in an OLG model with unemployment. *International Tax and Public Finance* 16, 198-218.

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

[4] Bastani, S., Blomquist, S., and L. Micheletto, 2013. The welfare gains of age-related optimal income taxation. *International Economic Review 54*, 1219–1249.

[5] Boadway, R., Marchand, M., and P. Pestieau, 2000. Redistribution with unobservable bequests: A case for taxing capital income. *Scandinavian Journal of Economics 102*, 253–267.

[6] Christiansen, V., and M. Tuomala, 2008. On taxing capital income with income shifting. *International Tax and Public Finance 15*, 527-545.

[7] Cremer, H., Pestieau, P., and J.C. Rochet, 2003. Capital income taxation when inherited wealth is non observable. *Journal of Public Economics 87*, 2475–2490.

[8] Gahvari, F. and L. Micheletto, 2014. The Friedman rule in an overlapping-generations model with nonlinear taxation and income misreporting. *Journal of Public Economics 119*, 10–23.

[9] Piketty, T. and E. Saez, 2012. A theory of optimal capital taxation. NBER working paper no. 17989.

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

[11] Saez, E., 2002. The desirability of commodity taxation under nonlinear income taxation and heterogeneous tastes. *Journal of Public Economics 83*, 217–230.

[12] Stiglitz, J. E., 2015. In praise of Frank Ramsey's contribution to the theory of taxation. *Economic Journal 125*, 235–268.

[13] Williamson, S. D., 2008. Monetary policy and distribution. *Journal of Monetary Economics 55*, 1038–1053.