PARETO EFFICIENT TAXATION AND EXPENDITURES:
PRE- AND RE-DISTRIBUTION

Joseph E. Stiglitz

ABSTRACT

This paper shows that there is a presumption that Pareto efficient taxation entails a positive tax on capital. When tax and expenditure policies can affect the market distribution of income in ways that cannot be directly offset, those effects need to be taken into account, reducing the burden imposed on distortionary redistribution. The paper extends the 1976 Atkinson-Stiglitz results to a dynamic, overlapping generations model, correcting a misreading of the result on the desirability of a zero capital tax. That result required separability of consumption from labor and that the only unobservable differences among individuals were in (fixed) labor productivities. In a general equilibrium model, one needs to take into account the effects of policy changes on binding self-selection constraints. In a simple model with time separability but with non-separability between consumption and leisure, capital taxation depends on the complementarity/substitutability of leisure during work with retirement consumption.

The final section constructs a simple two class model, capitalists who maximize dynastic welfare and workers who save for retirement, whose productivity can be enhanced by (publicly provided) education. It derives a simple expression for the optimal capital tax, which is positive, so long as the social welfare function is sufficiently equalitarian and the productivity of educational expenditures are sufficiently high.

Joseph E. Stiglitz
Uris Hall, Columbia University
3022 Broadway, Room 212
New York, NY 10027
and NBER
jes322@columbia.edu

Pareto Efficient Taxation and Expenditures: Pre- and Re-distribution[1]

Joseph E. Stiglitz

Beginning with my supervision of Tony Atkinson in Cambridge in 1965-1966 while I was a junior research fellow at Gonville and Caius College, Tony and I enjoyed a close collaboration and friendship. One of our early results that received a great deal of attention was that when there was separability in the utility function between consumption and leisure[2], if there existed an optimal income tax, it was optimal to have no commodity taxation (Atkinson and Stiglitz, 1976). An immediate corollary of that result was that, under the stipulated conditions, there should be no tax on interest income—treating consumption at different dates as different commodities. This and similar results from optimal tax theory were seized upon as a basis of policy by those critical of capital taxation.[3] Those who did so typically did not understand (or did not want to understand) the limitations of the model. As always, one has to look carefully at the assumptions going into a model to judge whether they provide an appropriate basis for policy. This paper argues that *even within the confines of a model in which differences in labor productivity are the only source of differences in income, the conclusion that there should be no capital taxation is, in general wrong, even in the presence of separability, and that in a plausible*

[2] That is, utility U could be written $U = U(u(\mathbf{C}), L)$ where $\mathbf{C}$ is the vector of consumption goods and L is labor supply. We do not require additive separability.

[3] See, for instance, Mankiw *et al* (2009) and Atkeson *et al.* (1999). Using a quite different framework, with individuals with infinite lives, Chamley (1986) has also concluded that the optimal tax rate on capital should be zero in the long run. In section II, we consider a social welfare maximization problem extending infinitely far into the future, but using an overlapping generations model, and show that his results obtains if and only if there is separability in the utility function between labor and consumption. Later sections, present still more general frameworks, showing that even with separability, results on asymptotic zero capital taxation are not general.

model where individuals differ as well in the amount of inherited capital, there is a presumption for a possibly quite high tax on income from capital.

This paper is divided into eight parts. The first presents some general reflections on our 1976 paper, our motivations in writing it, what we saw as some of the limitations, and some of the important extensions of the result in the subsequent literature. The second presents the basic model, based on Stiglitz (1982a), including some straightforward but important extensions to Pigouvian taxation and the provision of public goods. In both cases, in the case of separability, results much more akin to those that arise when government can impose out of lump sum taxes than to those generated when government relies on distortionary Ramsey results obtain. For instance, the sum of the marginal rate of substitution equals the marginal rate of transformation. In the standard Ramsey-Diamond-Mirrlees model with public goods with distortionary taxation, the Samuelson condition for optimal supply (1955, 1958), that the marginal rate of transformation (MRT) equal the sum of the marginal rates of substitution (MRS) is altered, and instead MRT < Σ MRS, the marginal rate of transformation is less than the sum of the sum of the marginal rate of substitution; and the level of consumption of public goods is reduced.[4] Section 2 shows that the Samuelsonian condition is restored with optimal non-linear taxation, provided there is separability between labor and consumption goods.[5] Similarly, first best Pigouvian taxation is optimal with separability either between leisure and consumption, or between the externality on the one hand and consumption and labor on the other, though in the latter case, the Pigouvian tax has to be interpreted as the tax in excess of the optimal commodity tax (the shadow price.) [6]

To analyze optimal interest taxation in general equilibrium, once has to construct a dynamic general equilibrium model. Section 3 does this, embedding optimal non-linear taxation in a general equilibrium overlapping generations model. While Atkinson and Sandmo (1980) had extended the Atkinson-Stiglitz (1972) model to a similar overlapping generations

---

[4] See Pigou (1947), Dasgupta and Stiglitz (1971) and Atkinson and Stern (1974). Atkinson and Stern make the important point that one cannot infer from the fact that the sum of the MRT < Σ MRS the impact of distortionary taxation on the supply of public goods.

[5] Boadway and Kean (1993), using the Stiglitz (1982b) framework for the analysis of optimal taxation, had noted this result earlier. See also Kaplow (1996).

[6] Kaplow (2006a) had noted the former result. Sandmo (1975) had established that one could obtain results that were closely akin to first best Pigouvian taxes even in a Ramsey-Diamond-Mirrlees model.

2

setting, this paper, a development of Stiglitz (1985), provides the corresponding extension of Atkinson-Stiglitz (1976). The conclusion of Atkinson and Sandmo (1980) was that "it is difficult to make a strong case *either* for the expenditure tax *or* for taxing interest income at the same rate as wage income." By contrast, we find that the basic results of our static model extend to this dynamic framework: (a) With separability, and fixed relative productivities, there should be no taxation of capital income *provided that relative wages are either observable or unaffected by taxation or investment (by the consumption/labor supplies which they induce)* ; but (b) in the absence of separability between consumption and leisure, but with intertemporal separability, whether there should be an interest income tax or subsidy depends on whether first period consumption is a complement or substitute for leisure.[7] But both separability and the italicized proviso are crucial.   While the latter result is reminiscent of that of Corlett and Hague (1953) (in a model with no income taxation and just commodity taxation), we establish it in the presence of an optimal non-linear income tax. (This is important, because in many causes, the presence of one tax—in particular an income tax—changes markedly the taxes that should be imposed elsewhere in the system; for reasons we explain, often with unconstrained non-linear taxes, results that hold only with lump sum taxes are stored.)As a prelude to the later discussion of public investment in the next section, the section also extends the analysis shows that the standard result on the desirability of productive efficiency holds.

The rest of the paper then provides a critique of the Atkinson-Stiglitz (1976) result. The fourth section focuses on the case where relative wages are not fixed. Even with separability between labor and consumption, (a) it is in general desirable to impose distortionary commodity taxation, even on high ability individuals; (b) there should be a distortionary income tax, even on high ability individuals; (c) there should be an interest income tax (or subsidy), even on high ability individuals; (d) productive efficiency of the economy (where there is no production taxes and the marginal rate of transformation of all goods is the same amongst all firms) is not desirable[8]; (e) even asymptotically, the rate of return on public capital goods may

---

[7] Similar in spirit to the Corlett and Hague (1953) analysis.

[8] One of the important results in Diamond-Mirrlees (1971) was establishing that optimal taxation entailed only consumption taxes, no production taxes, and that the economy was productively efficient.   Dasgupta and Stiglitz (1971, 1972) had established this was not so in the standard Diamond-Mirrlees model, but with restricted taxation (i.e. the government could not impose 100% taxes on pure rents or profits and could impose taxes on all goods and

not equal either the pure rate of time discount or the rate of return on private goods; the long run return to some kinds of public capital may be lower than the social discount rate.

Recent policy discussions have suggested the desirability of focusing on policies which improve the market distribution of income, leaving less of a burden on *redistribution.*[9] Standard models which assume that relative wages are fixed cannot address this issue. Our analysis establishes that this intuition is correct. In each of the cases listed above, taxes and expenditures are adjusted at the margin to increase wages of the low skilled relative to the high skilled.[10] Formally, the improvement in the distribution of income loosens the binding self-selection constraint, allowing for a Pareto improvement.

The optimal tax literature developed in the context of an economy of seemingly self-employed individuals, say making widgets. The only information problem was that of the government, who, seeing the output of widgets (or income from producing widgets) could not tell whether a high output was the result of hard work (long hours) or high ability. But most individuals in a modern economy work for others, and employers have an analogous information problem to that of government: they seek to discriminate among individuals, including by using self-selection mechanisms. But private screening is affected by public tax policy, and optimal taxation needs to take this into account. Section 5 does this, again showing that the Atkinson-Stiglitz (1976) result may not hold, even with separability.

Sections 6 and 7 consider optimal taxation with inheritances. Inheritances mean that individuals differ not just in their labor productivity, and optimal taxation ought to take that into account. This is so even in a world with dynastic families optimally sharing their wealth with their children. We show that the appropriate public policy response is taxation of inheritances and the return to capital. But the design of capital taxation is more nuanced than populist calls might suggest: careful attention has to be paid to incidence. A simplistic call for capital taxation with proceeds distributed to workers may actually leave workers worse off. In a simple two class model (workers and capitalists), we derive a simple formula for optimal

---

services at differential rates). The result here generalizes that of Naito (1999), whose analysis built on Stiglitz (1982a)

[9] See Hacker and Pierson (2011) and Stiglitz *et al* (2015).

[10] In Stiglitz (1982a), I showed that this was true for taxes on labor income, but did not explore the full range of issues examined here.

taxation when the proceeds are invested in human capital or in productive capital investments. The real argument behind the taxation of capital almost surely is related to disparities in inherited capital and in disparities in the ability to obtain returns out of capital—and luck—than with differences in productivities amongst workers, the subject of our 1976 paper.

Section 8 provides some concluding remarks, including putting the results presented here into some broader perspectives.

I. Reflections on the Atkinson-Stiglitz Theorem

Our analysis was motivated in part as a critique of Ramsey taxation (Ramsey, 1927), which seemed to justify not just high taxes on basic necessities like food (which had a low elasticity of demand) but also patterns of pricing by monopolies like A T & T.[11] Ramsey had established that optimal commodity taxes should be inversely related to the elasticity of demand and to the elasticity of supply in a model in which everyone was identical; hence there were no distributional concerns—and therefore a lump sum tax would have been presumably acceptable). Our 1972 paper (Atkinson and Stiglitz, 1972) had shown that in the absence of an income tax, optimal commodity taxes could be described by a simple Ramsey-like formula incorporating distributional effects, which suggested that when distributional concerns were given sufficient weight (for instance, in a society with a high level of both inequality and inequality aversion) goods like food with a low price elasticity of demand but a low income elasticity of demand would not be taxed at a high rate, but rather, that luxuries like perfume might face high rates, even though they have a higher price elasticity than food. We were, however, never satisfied with that result, since it was obvious that we had an income tax, the intent of which was at least partly redistributive.[12] The question was, given the existence of such a redistributive income tax, was commodity taxation still desirable? Our 1976 paper yielded stronger results than we had anticipated.

---

[11] See, e.g. Baumol and Bradford (1970) and Boiteux (1956). Elsewhere, David Sappington and Stiglitz (1987) explained while there was some similarities between the two problems, there were also some critical differences.
[12] Our 1972 results would, presumably, still be relevant for developing countries and emerging markets in which income taxation played little role, though in many such economies other considerations (such as migration and rural/urban income differences) play an important role. See Sah and Stiglitz (1992).

5

Another of the motivations behind our paper was to demonstrate two general results in the general theory of screening in the context of optimal taxation:

(a) *In general,* the optimal commodity taxes depended on the set of *other* taxes that were available[13]. Ramsey (1927) and Diamond Mirrlees (1971) analyses clearly depended on the absence of lump sum taxes. With lump sum taxes, they would be employed, and the optimal commodity taxes would be zero. Some of the Diamond-Mirrlees results, both on production efficiency (not using producer taxes) and the structure of taxation were shown to be undermined if there were profits (rents) which could not be taxed directly at 100% [14](Dasgupta and Stiglitz, 1971, 1972). Our 1976 paper (Atkinson and Stiglitz, 1976) thus explored the interaction between an (optimal) income tax and commodity taxation.

(b) The optimal income tax problem was at its heart a standard problem in asymmetric information with adverse selection. We assumed that individuals differed only in their abilities to perform work, and they all had the same endowment of time and the same utility functions. The Mirrlees' optimal tax problem could be thought of as extracting information about those differences efficiently as one raised tax revenues. Even though there was a single "dimension" in which individuals differed (ability), in general, it seemed possible that one could extract information about that difference more efficiently by looking not just at the individual's labor supply, but also at his consumption patterns. While that conjecture turned out to be true in general, in the special case of separability, Pareto efficient taxation required only an income tax; there was no benefit to be had by using information about consumption patterns.

We never thought that the separability assumption was plausible. It meant that there were no goods that were complementary to leisure. Clearly, the marginal rate of substitution between skis and say food depends on the amount of leisure. Someone with no leisure time simply doesn't value skiing much. Thus, we never attached much weight to the result that there should be no tax on capital with separability.

---

[13] There are a set of parallel questions in the more general theory of incentives and selection with asymmetric information. For instance, the design of the optimal sharecropping contract is affected by the ability of the landlord to affect credit, the purchase of inputs, and the choice of technology. See Braverman and Stiglitz (1982, 1986a, 1986b).

[14] See Dasgupta and Stiglitz (1971, 1972). They thus reconcile the Diamond-Mirrlees results, where supply elasticities play no role, and Ramsey's analysis, where they are central. Dasgupta and Stiglitz also show that other restrictions on taxation fundamentally change optimal tax structures, including the desirability of efficiency.

But this was not the only, or even the most important, reason that our theorem did not provide the basis of policy for capital taxation. For instance, if individuals differed in other ways—as in fact they do, in their level of inherited capital and their ability to earn returns on capital— it was possible that capital taxation was desirable, even if the separability assumption held. Indeed, it is not hard to generate more general plausible models in which capital taxation is clearly desirable. Showing this is one of the main objectives of this paper.

Thus, in section 8, we consider a simple model, which perhaps captures key aspects of the sources of inequality which give rise to the desirability of capital taxation far better than generalizations of Atkinson-Stiglitz (1976) discussed earlier in the paper. We assume that there are two classes of individuals—a group of capitalists with large endowments of capital who invest their capital, earn returns, consume, and leave bequests to their children, and workers-- with the capitalists sufficiently richer than the workers that the social marginal utility of a dollar to the capitalists is negligible.[15] It is obvious that in such a situation the optimal utilitarian[16] tax on capitalists is close to a Rawlsian tax, maximizing the revenue that one could obtain from a tax on capital[17]. It is clear that the optimal capital tax is significantly greater than zero, and considerable evidence that it is far higher than the current tax on capital, especially the tax on capital gains.[18]

There are other "thought" experiments that help us understand the role of capital taxation, and why in general it should be positive. Consider, for instance, a simple model in which all members of a group of individuals have an equal endowment of capital, but some individuals are better at transforming a unit input of capital into output (i.e., they are better at selecting good projects.) Assume that the more effort an individual exerts, the higher income;

---

[15] This model is a generalization of Stiglitz (2015) and Mattauch *et al* (2016, 2017). It many ways, it formalizes the model that is implicit in Piketty (2015), building on Stiglitz (1969a).

[16] In the discussion below, we often refer to results based on a utilitarian social welfare function, such as that employed by Mirrlees (1971). But virtually all of the results would hold for any individualistic inequality averse social welfare function, e.g. along the lines discussed in Rothschild and Stiglitz (1973), and indeed, for any Pareto efficient tax structure (Stiglitz, 1987; Brito *et al* 1990).

[17] It is easy to design capital taxes that effectively tax only the capital income of capitalists, e.g. exempting life cycle savings or taxing only the returns on capital above a certain level. (In this model, there is great inequality in the ownership and returns to capital.) The analytics sometimes turn out simpler for a general tax on capital.

[18] Especially in countries like the United States, where there is a step-up-in-basis at death, resulting in capital gains not being taxed at all. There are a wide range of ways by which through tax planning high income individuals can reduce their taxes. See Stiglitz 1985b.

and that there is disutility associated with the exertion of effort.[19]   Then the optimal tax structure for this group would be parallel to that which we analyzed in our paper: with separability between the utility of goods and effort, then the optimal tax is *only* a tax on capital income. In the absence of separability, both goods taxes/subsidies and a capital tax are desirable.

With individuals having equal endowments of both capital and labor, with full separability (of consumption from effort and labor, and effort and labor from each other), but differing in both the ability to transform labor and capital into "income," if one can separately identify income from labor and from capital[20], optimal taxation involves separate non-linear taxation of capital income and labor income, but no commodity taxation; but even with separability of consumption from effort and labor, but effort and labor not being separable from each other, then matters are far more complicated. While there would still be no indirect taxation, in general, taxes would not just be a function of total income.[21]

Of course, some individuals have larger endowments of wealth. If such wealth were observable, a utilitarian would want to impose a lump sum tax on it. If the supply of that wealth were elastic—parents' transfers to children could be affected by the taxes imposed on the transfer or assets or the after-tax income generated—then there might be a trade-off, with

---

[19] Note that if the higher returns of some individuals is *just* a result of luck, then "excess" returns (i.e. returns above the average) should be fully taxed, since there is no adverse incentive effects. The same thing is true if the higher returns are a result of just inherited ability or better connections. (Such taxes could, of course, affect the incentives of those providing "connections," if it is costly to do so or if the benefit to the "gift" of the giver depends on the amount the recipient receives after tax. See section 5 below.)   In each of these cases, the desirability of a highly progressive capital tax is obvious. Here, we make the equally obvious point: if there is an unobservable variable, effort, that together with ability, determines returns to capital, there should be a progressive tax along the lines of that analyzed in the optimal (labor) income tax literature. There is one difference: If we assume that total returns, R, are a function of capital, K, ability, a, and effort, e, $R=R(e, K, a, \epsilon)$, where ϵ is a random variable, and Utility is a function of consumption and effort (but not of K), then there is an additional variable we have to deal with. (It is as if individuals had different endowments of time.) If K is observable, then taxes paid would be a function of $\{K, R\}$. This model is formally identical to the standard one when K is the same for all individuals, there is no variability (ϵ is fixed) and e is not fixed. If everyone had the same K and a, we would have the standard moral hazard model. With K fixed (observable), but ϵ variable, individuals' differing in a, and e not fixed, then we have the standard moral hazard cum adverse selection model, where both self-selection and incentive compatibility constraints have to be taken into account. This changes some of the standard results that have been obtained in models with pure adverse selection. See Stiglitz and Yun (2013).

[20] Note that for small businesses, tax authorities cannot separately identify income from capital and labor. At best they can impute income from capital.

[21] Let W = wage income, C = capital income, and T be total taxes paid. The optimal tax function T(W, C) is not of the form T = T(W +C).

higher tax rates generating more tax revenue but leading to less accumulation and transfer of wealth. Even a Rawlsian would impose a tax that was less than 100%. It may be that the actual value of the wealth transfer is not observable, but only the income generated by it.   Then there is no presumption that the optimal tax (whatever the social welfare function) on the income generated by that wealth would be taxed at zero. Quite the contrary: there is a presumption that it would be taxed, and possibly at high rates.

Taxing capital goods and rents

Moreover, the result that one does not want to tax the returns to individuals is often confused with not taxing the returns to particular assets or corporations. If land is inelastically supplied, there is nothing in our analysis that undermines the conventional result that one would want to tax the returns on that asset at a very high rate (100%) (George, 1879).[22] The price of the asset would fall, to reflect the tax. Indeed, more recent results—extending the overlapping generations model as exposited in Atkinson-Stiglitz (1980) to include land as an asset—note that since land crowds out capital, a tax on land leads to a higher steady state level of capital (per capita) and thus to higher levels of income.[23] All that the Atkinson-Stiglitz theorem said was that (under the stipulated conditions) one should not tax the returns received by the individual.

Risk

We (along with most of the subsequent literature) ignored uncertainty. This was an important omission. In a world in which there are imperfect risk markets, Domar and Musgrave (1944)[24] had already shown that an income tax with appropriate loss offsets could encourage risk taking. With, say, expensing of capital goods, the government entered as a silent partner. The government was in a better position to absorb risk than the firm or individual: it could

---

[22] For a more modern rendition, see Arnott and Stiglitz (1979, 1981).

[23] This is a result obtained earlier in models with money and capital accumulation (see, e.g. Shell, Sidrauski, and Stiglitz (1969)). For more recent and general results, see Stiglitz (2015, 2017).

[24] Their analysis was put into modern expected utility form by Stiglitz (1969b). See also Piketty and Saez (2013) who explain how if there is uninsurable uncertainty about future returns, capital taxation can be viewed as providing insurance against returns. (The older literature emphasized the importance of loss offsets.)

spread risk over all of society. (Arrow and Linn, 1970, 2014).[25] If the tax rate was τ, the

government picked up τ per cent of the costs, taking in turn τ percent of the returns. The

expected marginal return per dollar invested was thus unchanged. But taking account of risk,

the tax actually led to more investment.[26] A utilitarian government would again want to

impose a capital income tax. Indeed, it would be part of Pareto efficient taxation.

Extending Atkinson-Stiglitz

There were two natural extensions to our analysis: to ask whether there were more

general conditions under which commodity taxation was not necessary, and to ask whether the

result remained true with an income tax that was not the optimal non-linear tax, e.g. with an

(optimal) linear income tax. Deaton (1979, 81) showed that with an optimal *linear* income tax,

weak separability between goods and leisure together with linear Engel curves for goods is a

sufficient condition for no indirect taxation. But when these restrictive conditions are not

satisfied, in general commodity taxation is required, even with an optimal linear tax (Stiglitz

(2009a), a revision and extension of Stiglitz (1976a)).[27] Gauthier and Laroque (2009) provide a

more general account of the role of separability in optimal taxation and expenditure.

Sections 2 and 3 explore some of the more general results (some, as we note, already in the

literature) in which separability suffices to ensure that commodity taxation is not needed with

---

[25] A fuller analysis would address why the government is better able to spread risks than the market. Well-developed theories based on information asymmetries have explained imperfections in capital markets and their ability to share risk.

[26] Musgrave and Domar (1944) used mean variance analysis. Stiglitz (1969b) showed that this was true more generally, for risk averse individuals.

[27] It is worth noting, though, that the structure of optimal commodity taxation does not look like that described by Ramsey (1927) (where tax rates were simply related to elasticities of demand and supply) or Diamond and Mirrlees (1971). At the same time, one can still derive a simple formula, given the optimal commodity taxes, for the optimal linear tax rate, based on an appropriately weighted average of the labor supply elasticities. The optimal tax can be related to a measure of the *marginal* benefit associated with inequality reduction (this marginal measure differs from the Atkinson measure, which identified how much individuals would be willing to pay to get rid of all inequality. Obviously, at the *margin,* there is greater social cost of inequality. Stiglitz (1976a) relates this to a societal measure of inequality aversion (or risk aversion), obtaining a remarkably simple approximation for the optimal tax: $\frac{\tau}{1-\tau} = \frac{R\, s_Y{}^2}{u\bar{n}_L(1-g)}$ where as usual τ is the (income) tax rate, R is the measure of risk (inequality) aversion, $s_Y{}^2$ is the variance of income inequality, g is the share of government revenue spent on public goods, and $_u\bar{n}_L$ is a welfare weighted averaged compensated labor supply elasticity. (Earlier, Sheshinki (1972) analyzed the optimal income tax, but did not derive the above simple expression for the optimal tax rate. Subsequently, Piketty and Saez (2012) extended these earlier papers to provide tax formulae based on observable sufficient statistics.)

10

an optimal income tax. Later sections show that, in particular, an interest income tax is

needed.

II.   The Basic Model

We begin our discussion with the standard model, with two types of individuals,

differing in ability, but having the same utility function, reviewing the results of Stiglitz (1982a),

as a prelude to the subsequent analysis. The ith individual faces a before-tax wage (output per

hour) of w_i, and thus, in the absence of taxation, his budget constraint is simply

(1) $\sum_j C_{ij} = w_i L_i$

where C_{ij} = the ith individual's consumption of good j, all goods are normalized to have a price

of unity, L_i = number of hours worked by the ith individual. Neither w_i nor L_i is observable,

but the product, the ith individual's income, $Y_i = w_i L_i$ is. The ith individual receives utility from

consuming goods and disutility from work:

(2) $U^i = U^i(\boldsymbol{C_i}, L_i)$

where $\frac{\partial U^i}{\partial C_i} > 0, \frac{\partial U^i}{\partial L_i} < 0$, U is quasi-concave, and $\boldsymbol{C_i}$ represents the ith individual's consumption

vector: $\boldsymbol{C_i} = (C_{i1}, \dots, C_{ij}, \dots)$.   Assume the government imposes a tax, T, as a function of

income: $T_i = T(Y_i)$. The individual's consumption now is his income minus his tax payment,

$\Sigma C_{ij} = Y_i - T(Y_i)$. The individual maximizes his utility subject to his budget constraint:

(3) $\max_{\{C_i, Y_i\}} U^i(\boldsymbol{C_i}, L_i)$ $s.t$ $\sum_j C_{ij} \leq w_i L_i - T(w_i L_i)$

yielding the first-order condition (assuming differentiability, etc.):

(4) $\frac{dU^i/dL_i}{dU^i/dC_{ij}} = -w_i(1 - T')$

The left-hand side is the individual's marginal rate of substitution. The right-hand side is the

after-tax marginal return to working an extra hour.

The problem of the government concerned with Pareto efficiency[28] is to maximize the

utility of, say, individuals of type 2, subject to (a) individuals of type 1 having at least a given

---

[28] The notion of Pareto efficient tax structure is a slight generalization of the "optimum" tax analysis of Mirrlees
and Diamond and Mirrlees. Pareto efficient tax structures are those (given the admissible set of taxes and the

11

level of utility and (b) raising a given amount of revenue. It does this by offering two $\{C, Y\}$ packages, one of which will be chosen by the first group, the other of which will be chosen by the second group.[29] We write $U^i(C_i, L_i) = U^i\left(C_i, \frac{Y_i}{w_i}\right) \equiv V^i(C_i, Y_i)$. The government thus solves the following program:

(5) $\quad \max\limits_{\{C_1, C_2, Y_1, Y_2\}} V^2(C_2, Y_2)$

s.t $\quad V^1(C_1, Y_1) \geq \overline{U^1}$;

$\quad\quad V^2(C_2, Y_2) \geq V^2(C_1, Y_1); \; V^1(C_1, Y_1) \geq V^1(C_2, Y_2)$ (the self-selection constraints)

and $\quad R = (Y_1 - C_1)N_1 + (Y_2 - C_2)N_2 \geq \bar{R}$ (the revenue constraint)

where $C_i = \sum_j C_{ij}$, the ith consumer's aggregate consumption, R is government revenue; $\bar{R}$ is the revenue requirement, and N_i the number of individuals of type i. If we let μ be the shadow price associated with the utility constraint, λ_i be the shadow prices associated with the self-selection constraints, and γ be the shadow price associated with the revenue constraint, then without loss of generality, the Lagrangian can be written as[30]:

(6) $\quad \Lambda = N_2 V^2 + N_1 \mu (V^1 - U^1) + N_2 \lambda_2 (V^2(C_2, Y_2) - V^2(C_1, Y_1)) + N_1 \lambda_1 (V^1(C_1, Y_1) - V^1(C_2, Y_2)) + \gamma \big((Y_1 - C_1)N_1 + (Y_2 - C_2)N_2 - \bar{R} \big)$

Then, the first order conditions are:

(7a) $\quad \dfrac{d\Lambda}{dC_{1j}} = N_2 \mu \dfrac{dV^1}{dC_{1j}} - N_2 \lambda_2 \dfrac{dV^2}{dC_{1j}} + N_1 \lambda_1 \dfrac{dV^1}{dC_{1j}} - \gamma N_1 = 0$

(7b) $\quad \dfrac{d\Lambda}{dC_{2j}} = N_2 \dfrac{dV^2}{dC_{2j}} + N_2 \lambda_2 \dfrac{dV^2}{dC_{2j}} - N_1 \lambda_1 \dfrac{dV^1}{dC_{2j}} - \gamma N_2 = 0$

(7c) $\quad \dfrac{d\Lambda}{dY_1} = N_1 \mu \dfrac{dV^1}{dY_1} - N_2 \lambda_2 \dfrac{dV^2}{dY_1} + N_1 \lambda_1 \dfrac{dV^1}{dY_1} + \gamma N_1 = 0$

(7d) $\quad \dfrac{d\Lambda}{dY_2} = N_2 \dfrac{dV^2}{dY_2} + N_2 \lambda_2 \dfrac{dV^2}{dY_2} - N_1 \lambda_1 \dfrac{dV^1}{dY_2} + \gamma N_2 = 0$

It is easy to show that at most one of the two self-selection constraints will be binding. We focus on the case where $\lambda_1 = 0$ and $\lambda_2 > 0$: only the second self-selection constraint is binding.

---

required public revenue) which are such that no one can be better off without making someone worse off. We identify properties of Pareto efficient tax structures which hold regardless of the social welfare function. See, e.g. Stiglitz (1982a, 1987) and Brito *et al* (1990).

[29] Obviously, the government can offer a continuum of {C,Y} packages (i.e., an entire tax function), but at most two will be chosen, and therefore we need be concerned with at most two.

[30] Recall that **C_i** (bold) represents the vector of consumption by the ith individual, C_i represents the total value of consumption of the ith individual, valued at producer prices (unity.)

Note that the standard case of *optimal utilitarian tax policy*, where the government maximizes $\sum_i N_i V^i$ implies that $\mu = 1$.[31]

From (7a-d) we obtain:

(8a) $\quad \dfrac{dV^2/dC_{2j}}{dV^2/dC_{2k}} = 1,$ $\qquad$ (8b) $\dfrac{dV^2/dC_{2j}}{dV^2/dY_2} = 1$

(9a) $\quad \dfrac{dV^1/dC_{1j}}{dV^1/dC_{1k}} = \dfrac{\gamma + \lambda_2 dV^2/dC_{1j}}{\gamma + \lambda_2 dV^2/dC_{1k}}$ (9b) $\quad \dfrac{dV^1/dC_{1j}}{dV^1/dY_1} = \dfrac{\gamma + \lambda_2 dV^2/dC_{1j}}{\gamma + \lambda_2 dV^2/dY_1}.$

We have formulated the problem as a directly control problem on $\{C_i, L_i\}$, but the solution can be implemented by a simple non-linear tax system, i.e. individual 1 is confronted with indirect taxes giving relative prices corresponding to the LHS of (9a), a marginal income tax rate given by the LHS of (9b), and a non-linear income tax giving the equivalent of a lump sum tax combined with a linear income tax with marginal rate given by (9b) yielding utility level for individual 1 of $\overline{U^1}$ (V^{1*}). Similarly, the high ability individual has no marginal income tax or consumption tax, but faces a non-linear income tax giving the equivalent of a lump sum tax yielding V^{2*}, where $\{V^{1*}, V^{2*}\}$ are the utility levels given by the solution to the optimal tax problem. Thus, we have the familiar result:

<u>Proposition 1a.</u> *There should be no distortionary taxation on the individual with the highest ability while the labor supply of the less able is distorted.*

This result goes beyond that of Mirrlees (1971) which said that the top marginal income tax rate should be zero[32]. (8a) says that there should be no commodity taxation on the high ability as well.[33]

Figure 1 provides a simple diagrammatic illustration with a single consumption good. In Figure 1A, the two types are quite different, so different that the high ability does not want to mimic the low ability if we impose non-distortionary taxation. Point E_1^* is the point on the low

---

[31] With $\overline{U^1}$ becoming, in effect, an endogenous variable.

[32] See also Sadka (1976), Seade (1977), and (for an overview) Tuomala (1990). Saez (2001), following Mirrlees (1971), shows that these results are not robust to a distribution with no top income (e.g. a Pareto tail) and derives simple asymptotic results. See also Diamond (1998). But the Pareto tail is used as a good approximation of a *description* of the probability distribution of the tail. There is, in practice, an identifiable person with the top income. Thus, we would have no problem in the United States identifying who that person is (assuming honest reporting of income), and setting a top marginal tax rate for that individual at zero. That income will be going up over time—as the incomes of all individuals will be changing, implying a change in the optimal structure. But there is no reason to impose a distortionary tax on that individual simply because we could conceive of there being a higher income individual.

[33] This result does not depend on separability of the utility function.

ability indifference curve U_1* which maximizes revenues, i.e. in which there is no distortionary taxation, so the slope of the indifference curve is unity. (Government revenue is just the difference between Y and C). Knowing R_1, the revenue raised from the low ability individuals (or, as illustrated in this figure, the subsidy provided to them) and the required government revenue, we know the revenue that has to be raised by from the high ability individuals, and $V^{2}*$ is the highest level of utility of the high ability consistent with raising that revenue, E_2* the maximal revenue raising point on the corresponding indifference curve. It is clear that the high ability individual does not want to imitate the low ability individual. But Figure 1B shows the more typical case, where they would. The maximal revenue consistent with U^1 being equal to $\overline{U^1}$ requires a sufficiently high tax on the high ability that they would mimic the low risk. In the figure, we depict the tax on the high ability that would make them indifferent between mimicking, and we assume it does not suffice. For the self-selection constraint to be satisfied, E_1* now must entail lower levels of income and consumption (labor supply) and lower tax revenues from the low ability, and hence a higher level of revenues raised from the high ability. At the optimum, the high ability is just indifferent to imitating the low.

[FIGURE 1 ABOUT HERE]

The interpretation of (9a) is, however, somewhat more subtle. For simplicity, denote

$$\alpha_{jk}^i = MRS_{ijk}(\boldsymbol{C}_1, Y_1)$$

i.e., i's marginal rate of substitution between j and k at the bundle $\{\boldsymbol{C}_i, Y_i\}$. Individual 1 and 2 differ only in their abilities. Let $\lambda_2 \frac{\partial V^2}{\partial C_{1k}} = b$. Then, (9a) can be rewritten as: $\alpha_{jk}^1 = \frac{\gamma + b\, \alpha_{jk}^2}{\gamma + b}$ and it immediately follows that $\alpha_{jk}^1 - \alpha_{jk}^2 = \{\gamma[1 - \alpha_{jk}^2]\}/\{\gamma + b\}$. The denominator of the preceding expression is positive. In the case of separability, the marginal rate of substitution between j and k is unaffected by the amount of leisure, so $\alpha_{jk}^1 = \alpha_{jk}^2$ which in turn implies $\alpha_{jk}^1 = \alpha_{jk}^2 = 1$, thus establishing:

Proposition 1b.   *There should be no commodity taxation on either high or low ability individuals if leisure and consumption are separable.*

Whether commodity *j* should be taxed (relative to *k*) depends simply on the impact of an increase in leisure on the marginal rate of substitution between commodities *j* and *k* -- a

property that has nothing to do with Ramsey's focus on deadweight loss, but tax structure is related to the deadweight loss associated with the self-selection constraint which come into play if government needs to raise a given revenue, cares about distribution, but is in an environment in which governments cannot tell who is able and who is not. The self-selection constraints that the government employs to achieve fair taxation create first order distortions. Commodity taxation can, under certain conditions, reduce the force of those distortions.

The intuition is simple: consider a good which is complementary with leisure. If a high ability individual tries to mimic a low-level individual, he has more leisure, and his demand for that good will accordingly be higher. Hence, taxing that good and using the proceeds to subsidize a good which is a substitute for leisure will make mimicking the low ability individual less attractive. With a weakened self-selection constraint, the low ability individual can work more, moving along his indifference curve—and in doing so generate more revenue for the government. Thus, the utility of both individuals can be increased and still satisfy the revenue constraint: indirect taxation is a Pareto improvement. However, if there is separability, then the levels of consumption of the goods is identical when the high ability imitates the low ability, and hence there is no way to use differential taxation to weaken the self-selection constraint.

There is another intuition which may prove helpful. In the case of separability, where $V^i = V^i(\phi(C_i), L_i)$, one can think of ϕ itself as an aggregate consumption good. Individuals take their wage income and buy "ϕ". We know from Diamond and Mirrlees (1971) that we want production efficiency[34]. Productive efficiency in producing φ implies no commodity taxation.

Taxation of Interest Income

An immediate implication of these results is that *in the case where consumption in different periods is separable from leisure--typical of many of the simplified models explored in the macroeconomic literature--there should be no interest income taxation.*

There is no strong presumption whether in the absence of separability there should be an interest income tax or subsidy. Consider, for instance, the standard life cycle formulation

---

[34] Later, we will qualify this result. See also Dasgupta and Stiglitz (1971, 1972).

(dropping for the moment subscripts describing the individual, and simplifying by assuming consumption is uni-dimensional).

(10) $U^t = u_t(C_t, L_t) + u_{t+1}(C_{t+1})$,

where there is time separability, but not separability between consumption and leisure and where C_t is consumption of the individual when he is young and working and C_{t+1} is consumption in retirement. Then the marginal rate of substitution between second period consumption and first period consumption is $\frac{\partial u_t/\partial C_t}{\partial u_{t+1}/\partial C_{t+1}}$. It follows that:

Proposition 2. *With the time separable utility function (10), if an increase in leisure increases the marginal utility of consumption in the first period, i.e. if leisure and consumption are complements, then there should be an interest income tax. Conversely, if an increase in leisure decreases the marginal utility of consumption in the first period, i.e. if they are substitutes, then there should be an interest income subsidy. Either is plausible and individuals may differ.*

The intuition provided earlier helps understand what is going on. We want to make it more expensive (less advantageous) for the high ability person to pretend to be the low ability person. In doing so, he increases his leisure. Skiing is a complement to leisure, so we want to tax skis; household help is a substitute for leisure, so we want to subsidize domestic help. High income people who work long hours with high savings rates, in anticipation of an enjoyable retirement, are evidencing complementarity between leisure and consumption; a subsidy on the return to savings increases the cost of cutting back on savings (as would happen if the high ability pretends to be a low ability). [35]

Atmospheric externalities

Still another generalization that is easy to incorporate into the model is atmospheric externalities[36], i.e. there is some variable *B* (for Bad) which is a function of the aggregate consumption variables. Both production and utility can be adversely affected by *B:*

---

[35] It should be clear from the structure of the analysis that if individuals supplied labor at different dates, so U^i = $U^i(C_t, C_{t+1}, \dots C_T, L_t, L_{t+1}, \dots L_T)$, but the utility function remained separable between consumption and leisure (work) $U^i = u^i(C_t, C_{t+1}, \dots C_T) + v^i(L_t, L_{t+1}, \dots L_T)$ it would remain true that no interest income tax should be imposed, though, in general, the tax an individual would pay on wage income in one year would depend in wage income earned in other years.

[36] An atmospheric externality is an externality the value of which depends on aggregate consumption. In the notation below **C,** is the vector aggregate consumption of all commodities.

(11) $$V^i = V^i(\boldsymbol{C_i}, L, B) \quad \text{where } V_B^i < 0$$

and $Q = F(K, L, B)$, where $F_B < 0$. It is useful if we reformulate our self-selection constraints, to represent the self-selection constraint as:

(12a) $$V^2(\boldsymbol{C_2}, L_2, B) \geq V^2(\boldsymbol{C_1}, \hat{L}_2, B)$$

where $\hat{L}_2$ is the labor a type 2 individual would have to put forth to imitate the income of a type 1 individual, i.e,

(12b) $$\hat{L}_2 = \frac{w_1 L_1}{w_2} = vL_1$$

where v is the ratio of the productivity of the low ability individual to the high ability individual, assumed now fixed. Then we have:

(13) $$\frac{d\Lambda}{dC_i} = [\ldots] + \left\{ \underbrace{\sum \mu_i \frac{dV^i}{dB}}_{A} + \underbrace{\gamma \frac{dF}{dB}}_{B} + \underbrace{\lambda_2 \left(V_B^2 - \hat{V}_B^2 \right)}_{C} \right\} \frac{dB}{dC_i}$$

where [...] represents the derivative of the Lagrangian in the absence of externalities. There are three additional terms: {A} represents the direct consumption externality; {B} represents the production externality; and {C} represents the impact on the self-selection constraint.

In the case of a separable utility function, which can take on the form of: [37]

(11a) $\quad V^i = v(\boldsymbol{C}, L) - Z(B)$,

the third term disappears, and there are only the direct externality effects of the kind that have been analyzed by Pigou (1920) and Sandmo (1975). The results for (11a) are obvious: because of the structure of preferences, the level of B can't affect any choices made by anyone, and therefore can't affect the structure of taxation. However, in the case of (11a), the very presence of self-selection constraints changes the social marginal valuation of consumption (in effect, the shadow price on different consumption goods). The additional terms A and B represent alterations to these shadow prices, as opposed to the market prices. Moreover, the presence of atmospheric externalities (under 11a) does not affect choices, and so we can ignore

---

[37] Obviously, it can also be fully separable, and the analysis does not require additive separability.

the indirect income effects *arising from B,* though obviously, not those arising from the taxation of labor or goods.

Alternatively, separability could take the form of

(11b) $V^i = v(C, B) - Z(L)$.

The results for (11b) follow directly from our earlier analysis. A change in B simply shifts the common utility function for goods.

Proposition 3A. If labor is separable from consumption then a straightforward Pigouvian tax/subsidy is optimal. If the externality is separable from consumption and labor (but consumption and labor are not necessarily separable), then a first-best Pigouvian tax/subsidy— over and above the shadow prices (prices with optimal commodity taxes/subsidies)—is desirable.

In other words, in these cases, the complicated formulae underlying the literature on corrective taxation that focus on interactions between the corrective taxes and other taxes are not relevant. What has been called "first best Pigouvian taxation" –where the tax is imposed to reflect marginal harm, is desirable. (Kaplow, 2006a). On the other hand, if the utility functions are not separable, simple Pigouvian corrective taxation is not appropriate, but the deviations from simple Pigouvian corrective taxation do not depend (just) on indirect revenue effects, but (also) on the impacts on self-selection constraints.

Proposition 3B. With non-separable utility functions, Pigouvian corrective taxation has to be modified to take account not just of the second order effects on government revenue but of the first order effects on the self-selection restraints.[38]

---

[38] Sandmo (1975) provided the classic modern discussion of optimal taxation in the presence of externalities, but he conducted the analysis in the standard Ramsey-Diamond-Mirrlees framework with no income tax. The contribution here is simply to add the self-selection constraint. Kaplow (2006a) analyzes the problem posed here, using a very clever approach which he applies to the analysis of public goods and to the general problem of indirect taxation discussed above (Kaplow (1996, 2006b)), which highlights the role of separability. He shows that if there were a differential tax on commodities, there would always exist a perturbation in the tax function that would lead to a Pareto improvement. In some respects, his result is stronger than that here, for he shows that differential commodity taxation is not desirable regardless of whether the imposed income tax is optimal or not, *so long as it is possible to make any perturbation in the income tax function.* The last observation, though, is critical. For instance, Stiglitz (1976a, 2009a) considers differential commodity taxation in the presence of an optimal linear income tax. There are both administrative, economic, and political reasons for a linear, or a piece-wise linear, tax system, implying that arbitrary deviations from (quasi-) linearity may not be feasible or will entail costs not incorporated into the analysis. Kaplow's analysis focuses on the case of separability between consumption and labor (11b).

In particular, if those with more leisure are more sensitive to the quality of the environment than those with less leisure, then the (Pareto) optimal level of consumption of negative externality generating consumption is higher, because by increasing the level of the atmospheric externality, it enhances the ease of separation.

Public Goods

The previous subsections' analysis can be directly applied to public goods. A public good, like a public bad, enters everyone's utility, and since we are assuming all individuals have the same preferences: $V^i = V^i(\boldsymbol{C_i}, L, G)$. Assume the production function is of the form $\sum_i C_i + G = \sum_i w_i L_i$, and let γ now represent the overall resource constraint. Then optimal G entails:

$$\sum_i \mu_i N_i V_G^i + \lambda_2\left(V_G^2 - \hat{V}_G^2\right) = \gamma$$

In the utilitarian formulation, $\mu^i = 1$, for all i. With separability (of either the form of (11a) or (11b), with G replacing B), the sum of the marginal utilities of the public good equals the marginal social cost, γ. But γ is normally itself influenced by the self-selection constraint. Consider first the case where neither self-selection constraint is binding (Figure 1A). Then we have $\lambda_2 = 0$ which implies that $(\mu_i / \gamma)\dfrac{dV^i}{dC_{ij}} = 1$ or $\sum_i(N_i V_G^i / \dfrac{dV^i}{dC_{ij}}) = 1$: we obtain the standard result that the sum of the marginal rate of substitution equals the marginal rate of transformation, here unity. But more generally, so long as we have separability, we have

$\sum_i(N_i V_G^i / \dfrac{dV^i}{dC_{ij}}) = 1$: Even with distortionary taxation, the Samuelson rule holds. In a sense, the result should not be too surprising. With unconstrained optimal non-linear taxation, there is no distortion associated with the top person, and with separability, the self-selection constraint is effectively not binding. Then there is, in effect, no *marginal* distortion arising from raising *additional* revenue—even though taxes are, as a whole, distortionary. In the absence of separability, the sum of the marginal rates of substitution may be either greater or less than the marginal rate of transformation. It depends on the effect of G and of the numeraire (the "private good") on the self-selection constraint. In the case of separability between G and {goods and labor}, G doesn't affect the self-selection constraint, but the private consumption good may either tighten the constraint (leading to lower consumption and a higher valuation of

the private good for the low ability individual) or loosen it.[39] Note that since different public

goods may enter into preferences differently, for some goods the sum of the MRS's may be less

than the MRT, for others greater.

III.   Extension to Many Periods with Overlapping Generations

If we are to analyze meaningfully the desirability of an interest income tax, we have to

do so within a model in which the return to capital is determined endogenously. We do so in

the standard overlapping generations model. Each generation is assumed to live for two

periods, working in the first, and consuming in both periods. Within each generation, there are

two types (the able and the less able), but the government does not know who is who. The

utility function of the ith type in the tth generation is denoted by $V_t^i(C_{it,t}, C_{it,t+1}, L_{i,t})$

where $C_{it,t}$ is the tth generation's vector of consumption during the first (working) period, and

$C_{it,t+1}$ is the tth generation's consumption during the second (retirement) period of this life. As

before, the self-selection constraint can be written as :

$$V_t^2(C_{2t,t}, C_{2t,t+1}, L_{2,t}) \geq V_t^2(C_{1t,t}, C_{1t,t+1}, \hat{L}_{2,t}).$$

Output at the time t is given by a neo-classical production function of the form

$Q_t = F(K_t, E_t)$, where $E_t = v N_1 L_{1,t} + N_2 L_{2,t}$ is the effective labor supply. If there are no

constraints on the government's ability to use social security or debt policy to control the

capital stock (see Atkinson and Stiglitz [1980]) then the only macro-constraint that needs to be

taken into account are the period resource constraints, which we write as

(14) $Q_t = K_{t+1} + \sum_i N_i C_{it,t} + \sum_i N_i C_{it-1,t} + G = F_t(K_t, E_t)$

where we have assumed that capital wears out each period. Forming the Lagrangian as

before[40] , but with separate constraints for each period (including for the self-selection

---

[39] Boadway and Keen (1993) provider a fuller analysis of the optimal provision of public goods. Note that the
question of whether the sum of the MRS's is greater or less than the marginal rate of transformation does not
translate directly into whether the supply of public goods with distortionary taxation is less than or greater than
with lump sum taxation, since this is a matter of general equilibrium, in which all the variables are endogenous.
See Atkinson and Stern (1974) for a discussion of the matter in the context of the Ramsey-Diamond-Mirrlees
model.

[40] Where μ_{it} is the Lagrange multiplier associated with the constraint $V_t^i \geq \bar{V}_t^i$, $\mu_{1,0} = 1$

constraint and the resource constraint (14) which replaces the budget constraint), and differentiating with respect to the consumption vector, we obtain

(15a) $\quad \dfrac{dV^2/dC_{it,t}}{dV^2/dC_{it,t+1}} = \dfrac{\gamma_t}{\gamma_{t+1}}$ (15b) $\dfrac{dV^1/dC_{it,t}}{dV^1/dC_{it,t+1}} = \dfrac{N_1\gamma_t + \lambda_{2,t}dV^2/dC_{1jt,t}}{N_1\gamma_t + \lambda_{2,t}dV^2/dC_{1jt,t+1}}$

(15c) $\quad \dfrac{dV^2/dC_{i,t,t}}{dV^2/d\,L_{2,t}} = 1$ (15d) $\dfrac{dV^1/dC_{it,t}}{dV^1/dL_{1,t}} = \dfrac{N_1\gamma_t + \lambda_{2,t}dV^2/dC_{it,t}}{N_1\gamma_t + \lambda_{2,t}dV^2/dL_{1,t}}$

where γ_t is the Lagrange multiplier associated with the tth period government budget constraint (14).   By differentiating the Lagrangian with respect to K_t , we obtain

(16) $\quad \gamma_{t+1}F_K(K_{t+1}, E_{t+1}) = \gamma_t$

Thus $\dfrac{\gamma_t}{\gamma_{t+1}}$ $= F_K(K_{t+1}, E_{t+1}) =$ the (gross) marginal return on capital.

<u>Proposition 4a.</u> *In the overlapping generations model with two types of labor at every t, there should be no interest income tax on the savings of the high ability individuals, a zero marginal income tax, and no commodity taxation; while, as in the static model, low ability individuals face distortionary taxation on labor, commodities, and savings. With separability between leisure and consumption, there should be no commodity or interest taxation on the low ability. Without separability, there should be commodity and interest income taxes (or subsidies.)*

Rather than focusing on Pareto optimality, we could have maximized intertemporal social welfare, say using a utilitarian social welfare function: max $\sum_t \delta^t(N_1\, V_t^1 + N_2\, V_t^2)$, where the welfare of future generations are discounted at the rate δ (δ is the pure rate of social time preference). In steady state $\dfrac{\gamma_t}{\gamma_{t+1}} = F_K(K^*, E^*) = r^* = 1/\delta$. (15) and (16) then imply that the economy with distortionary taxation converges to the modified golden rule, where the rate of return equals the pure rate of social time preference. With separabilty, both individuals' MRS is equal to that rate, but in general, the low ability individual's may not be.

<u>Proposition 4b</u>. *In the steady state of the overlapping generations utilitarian optimal tax problem, with, and, in general, only with separability, the rate of interest (both the producer and consumer rate of interest, for both the low and high ability individuals) equals the pure rate of social time preference.*

Public and Private Investment

It is easy to extend the analysis to incorporate public as well as private investment. Let K_g represent public investment, K_p private investment, and let the production function now be $Q = C + K_g + K_p = F(K_g, K_p, E)$, where C is aggregate consumption . It immediately follows that $F_{K_g} = F_{K_p}$.

Proposition 4c. *Pareto efficient taxation-cum-expenditures requires that the return on public capital equal the return on private capital, which asymptotically, is just equal to the pure rate of time preference.*

These results are clearly parallel to those of Diamond and Mirrlees (1971), who argued for production efficiency.

IV. Distributive effects of tax and expenditure policy

The analysis so far has established that the results derived earlier, both with respect to taxation and public production, in simple partial equilibrium models hold more generally in a dynamic, overlapping generations, general equilibrium model; it has shown that Ramsey's analysis provides little guidance either for the design of efficient or redistributive commodity taxation; it has confirmed that the Diamond-Mirrlees result on the desirability of efficiency holds in this setting; and it has shown that the analysis can be extended to incorporate externalities, with modifications from Pigouvian corrective taxation based not on Ramsey-like indirect impacts on tax revenues but on impacts on self-selection constraints.

Unfortunately, several of these results are not robust, if government tax or expenditure policies can affect the distribution of income in ways which cannot be directly offset by tax policy: in our model, relative wages, v. If the two types of labor are not perfect substitutes, then in general v will be affected by levels of consumption of each of the goods and labor supply of the two types of labor. Optimal tax theory is predicated on the inability to observe either output per hour or the number of hours worked, because if one could do so, one could infer ability, and thus impose lump sum taxation.[41]   Thus, dropping the time subscripts (focusing on the basic model of section II), the first order condition for C_{2j} is:

$$\frac{d\Lambda}{dC_{2j}}/N_2 = \frac{\partial V^2}{\partial C_{2j}} + \lambda_2 \left\{ \frac{\partial V^2}{\partial C_{2j}} - \left(\frac{\partial V^2}{\partial L} \right) L_1 \left(\frac{\partial v}{C_{2j}} \right) \right\} - \gamma = 0.$$

Any tax or expenditure which changes v affects the self-selection constraint, and hence all of the previous conditions are altered. Note that $\frac{\partial V^2}{\partial L} < 0$. If follows directly that

Proposition 5. *If relative wages are affected by consumption levels of different commodities by the high skilled, then there should be indirect taxation on the high skilled; if relative wages are affected by levels of different commodities by the low skilled, then there should be indirect taxation on the low skilled even with separability between consumption and labor; if labor supply by high skilled individuals affects relative wages, then high wage individuals should face a marginal income tax (or subsidy). Any change in C_{ij} or L_i that improves the distribution of*

---

[41] Of course, that would not be the case if output per hour was itself a function of some unobservable variable besides ability, e.g. effort or education. We cannot pursue these ideas here.

23

income (i.e. increases the low ability individual's wage relative to that of the high ability)
improves welfare, and hence taxes should be designed to reflect this.

This means that the consumption of unskilled labor-intensive goods should be encouraged (their consumption prices reduced). In particular, we know from Samuelson's (1948) analysis of the two-factor two good international trade model, where output in each sector is a function of skilled and unskilled labor, that a movement along the production possibilities schedule (with fixed inputs of skilled and unskilled labor) towards the production of the unskilled-intensive good leads to an increase in relative wages of unskilled workers. The implications for labor supply (taxation of labor income) are, in this model where there are no other factors of production, disquieting (Stiglitz, 1982b). Since in such a model, the two types of labor have to be complements ($F_{12} > 0$), an increase in skilled labor has to increase the relative wage of the unskilled. This means that we have to encourage more labor supply by the skilled (from what it would be in the absence of an effect on relative wages), i.e. the top marginal tax rate should be negative. This result, however, is not general: if we have a three factor production function (output is a function of the two types of labor and land), then skilled and unskilled labor can be substitutes, in which case the top marginal tax rate should be positive, in contrast to the standard result in optimal (income) taxation, that it should be zero. (See Mirrlees (1971) and Saez (2001).)

Similarly, the optimal Pigouvian tax formulae for correction of externalities and the formulae for the optimal supply of public goods have to be corrected for induced wage effects: the externality or public good may have an effect on relative wages, and thus on the self-selection constraints. The required modifications in the optimal formulae are straightforward. Note, in particular, that the simplifying results that hold with separability (that the effects on self-selection constraints can be ignored) no longer hold.

Production efficiency

Soon after Diamond and Mirrlees established their efficiency result—that optimal commodity taxation should ensure production efficiency—Dasgupta and Stiglitz (1971, 1972) showed that this would not be so in general, in particular if there were restricted taxation, for instance if there were rents (pure profits) that could not be taxed at 100%, or if there were

24

some factors or goods that had to be taxed at the same rate. A direct implication of the Diamond Mirrlees efficiency result was that it was not optimal to impose tariffs. Correspondingly, an implication of the Dasgupta-Stiglitz results was that it often was desirable to impose tariffs. (Dasgupta and Stiglitz, 1975.) Not surprisingly, the same result holds in the current context. Here, the critical issue is distribution. In the previous paragraphs, we assumed that government could not directly control v by the imposition of (optimal) indirect taxes on the labor supply of skilled and unskilled labor—an assumption we noted that was fully consistent with the standard optimal tax model, where only the product of hours worked and wages per hour is observable. But even if the wages received by or paid to any individual is not observable to the government, the government can affect the relative wage paid to unskilled workers.

Thus, if the government imposes a tariff on the relative price of the unskilled labor intensive good, it increases v, relaxing the self-selection constraint.[42] The benefit of relaxing that constraint has to be compared with the cost of the resulting inefficiency, which is captured by the impact on the resource constraint of an increase in p, the domestic producer price, over the international price, which we assume is unity. With international trade[43], the resource constraint can be written $\sum_j X_j(p; L_1, L_2) = \sum_i \sum_j C_{ij}$, where X_j is the output of the jth commodity, a function of relative prices and the aggregate supplies of the two types of labor. Specifying $\{L_1, L_2\}$ determines the production possibilities curve. Specifying p determines where on the production possibilities curve the economy operates. In Figure 2, AB measures the inefficiency at the equilibrium of aggregate consumption of one good in terms of the other good: how much more aggregate consumption could have been had a tariff not been imposed. We impose a tariff on the unskilled labor intensive good, which increases unskilled relative wages. The optimal tariff satisfies the first order condition[44]:

$$\lambda_2 \left\{ - \left(\frac{\partial V^2}{\partial L} \right) L_{1t} \left(\frac{\partial v}{\partial p} \right) \right\} + \frac{\gamma \partial X_1}{\partial p} (1 - p) = 0$$

---

[42] See also Naito (1999)
[43] Assuming all goods are tradable.
[44] The effect on the budget constraint is $\frac{\partial X_1}{\partial p} + \frac{\partial X_2}{\partial p} = \frac{\partial X_1}{\partial p}$ [1 − MRT].

25

where p = MRT = $\frac{\partial X_2}{\partial p} / \frac{\partial X_1}{\partial p}$ (where MRT stands for the marginal rate of transformation). The first term is the benefit from the relaxation of the self-selection constraint; the second term the resource cost of the tariff.

[FIGURE 2 ABOUT HERE]

There are similarly many ways within the closed economy to show the (potential) desirability of distortionary production taxation in a world where we cannot tax separately skilled and unskilled work. For instance, assume we have two sectors in the economy, each producing all the goods (a dual economy), one of which is more unskilled labor intensive. For each we have a transformation function $\mathfrak{I}^k \left(\boldsymbol{C_1^k}, \boldsymbol{C_2^k}, \ G^k, L_1^k, L_2^k \right) = 0$, where the jth element of $\boldsymbol{C^k_1}$ is output of consumption good j in the kth sector, so $C_{1j}^k + C_{2j}^k =$ total production of commodity j in the kth sector, which in equilibrium must be equal to the total consumption of that product. We can tax differently the consumption of good j depending on the producer. Doing so is equivalent to imposing a producer tax. In our direct control problem, where we specify the level of consumption of each good by each type of individual, we now specify the level of consumption of each good by sector. Since what we are concerned about is relative wages, what matters is aggregate consumption across the two individuals of output from the two sectors, which can variously be thought of as the advanced or large corporate sector, and the informal or SME sector. In equilibrium, there are additional constraints imposed by the fact that while the government cannot observe wages, individuals do, and they allocate themselves so that the wages in the two sectors for each type of labor are the same. (A more realistic model would incorporate labor market frictions.) The Lagrangian now needs to be rewritten, with additional terms representing resource constraints $\gamma_k [\mathfrak{I}^k - 0]^{45}$; and the free labor mobility constraint, ensuring equal marginal products of labor in the two sectors relative to say the numeraire, 0: $\zeta_i [\mathfrak{I}_{L_i}^a / \mathfrak{I}_o^a - \mathfrak{I}_{L_i}^u / \mathfrak{I}_o^a]$, where the superscript a stands for the advanced sector, and u for the less advanced sector, and $\mathfrak{I}_{L_i}^k / \mathfrak{I}_o^k$ is the wage in the kth sector of type i labor in terms of the numeraire, denoted "o". The optimization equation (the derivative of the Lagrangian) for C_{ij}^k now contains an extra set of terms, one focusing on the effect of the

---

[45] Obviously, without loss of generality, we can drop the term -γ^{k}0

consumption of the jth good from the kth production sector on the self-selection constraint, another simplifying the resource constraint itself, the third embedding the free movement of labor across sectors. Production of a good in one production process may affect the self-selection constraint differently from that of the other.

$$\lambda_2\left\{-\left(\frac{\partial V^2}{\partial L}\right)L_{1t}\left(\frac{\partial v}{\partial C_{ij}^k}\right)\right\}+\ \partial\{\textstyle\sum\gamma_k\,[\mathfrak{J}^k]+\sum\zeta_i\,[\,\mathfrak{J}_{L_i}^a\,/\,\mathfrak{J}_o^a\ -\ \mathfrak{J}_{L_i}^u\,/\,\mathfrak{J}_o^a\,]\}\partial C_{ij}^k=0.$$

With an increase in production a of say good j in the advanced and SME sectors having different effects on the demand for unskilled labor, and thus affecting the value of v and thus the self-selection constraint, there is a strong presumption *against* production efficiency.[46]

Proposition 6. *In general, production efficiency is not desirable.   If there are two (or more) sectors of the economy with different production functions, and in particular, different responses of relative demands for different types of labor (and thus relative wages) to demands for different goods, then production efficiency is not desirable. In particular, distortionary tariffs are in general desirable.*

Distortionary production tax in effect reduces the burden imposed on distortionary consumption taxes.   The freedom that Diamond and Mirrlees had in setting taxes on skilled and unskilled labor—different factors of production that to them were observable—meant that these distributive effects could always be "undone" by changes in factor taxation. The assumption here is that this is not possible: if we could observe the type of labor (skilled or unskilled), we could observe who was providing that type of labor, and thus infer directly the type of the individual, and thus impose lump sum taxation.[47]

Extending the life cycle model

In this sub-section, we extend the analysis of the previous sections to include distributive effects of taxation within a life cycle model. As before, we assume that the two types of labor are not perfect substitutes, i.e., the aggregate production is of the form $Q_t = F(K_t, L_{1t}, L_{2t})$ with $w_{i,t} = F_{L_{i,t}}$, where $Q_t = K_{t+1} + \sum_i \sum_j C_{ijt}$.   Now, relative wages depend on investment.

---

[46] Production efficiency requires that the marginal rate of transformation of all goods that are produced in both sectors be the same in the two sectors, i.e. $\mathfrak{J}^a_j\,/\,\mathfrak{J}^a_m = \mathfrak{J}^u_j\,/\,\mathfrak{J}^u_m$, where j and m are any two goods.

[47] It is thus *conceivable* that the government observe the type of labor input, but no the "name" of the individual providing that type of labor. Especially in the unincorporated sector, our assumption is more plausible; and so long as there are some sectors of the economy where the type of labor being provided is unobservable, our results on the desirability of distortionary production taxation hold.

If we assume constant returns to scale, $v_t = \frac{w_{1t}}{w_{2,t}} = \Phi\left(\frac{L_{1t}}{K_t}, \frac{L_{2t}}{K_t}\right)$. Given this reformulation, we

can again form the Lagrangian, obtaining the same first order conditions for $C_{it,t+1}$ and $C_{it,t+1}$

as before but now there is an extra term in each equation to reflect the effect of relative wages

on the self-selection constraint through K, so

$$(17) \qquad \gamma_{t+1}F_K = \gamma_t + \lambda_{t+1}^2 L_{1t}\frac{dV^2}{dL_{it}}\left(C_{it+1,t+1}, C_{it+1,t+2}, vL_{it+1}\right)\frac{dv}{dK}$$

In turn $\frac{dlnv}{dK}$ = ($F_{13}/F_1 - F_{23}/F_2$). While constant returns to scale implies $F_{13} + F_{23} + F_{33}$ = 0, there

are no implications about $\frac{dlnv}{dK}$. If, for instance, capital is a substitute for unskilled labor and a

complement to skilled labor, then an increase in K decreases v. Substituting into the first order

conditions for consumption of the high ability at t and t+1, we obtain

$$\frac{\frac{dV^2}{dC_{2jt,t}}}{\frac{dV^2}{dC_{2jt,t+1}}} = \frac{\gamma_t}{\gamma_{t+1}} = F_K - \lambda_{t+1}^2 L_{1t}\frac{dV^2}{dL_{it}}\left(C_{it+1,t+1}, C_{it+1,t+2}, vL_{it+1}\right)\frac{dv}{dK}/\gamma_{t+1}$$

where $C_{2jt,t}$ is the high ability individual of generation t's consumption of commodity j in time t,

$C_{2jt,t+1}$ is that at time t + 1. A similar calculation applies to type 1 individuals. So too, consider

two types of capital goods, say K_g and K_p using the same amount of resources to construct, Q_t =

$K_{gt+1} + K_{pt+1} + \Sigma\Sigma C_{ijt}$, so production efficiency requires $F_{Kgt} = F_{Kpt}$, the marginal returns to these

two types of capital should be the same. It thus immediately follows that:

Proposition 7a: If the level of capital affects relative wages, the marginal rate of substitution is

in general not equal to the marginal rate of transformation even for the high ability individual;

even in the presence of separability, it may be optimal to have an interest income tax or subsidy

on both high and low ability individuals. Production efficiency is not desirable if different kinds

of capital affect the distribution of income differently.

Increases in capital (and in different types of capital goods) change wage inequality. We

impose a tax on interest income if an increase in K increases wage inequality, i.e. if capital (like

robots) lowers unskilled wage relative to skilled wages.[48] The reason we do so is related to how

the self-selection constraints are affected. More generally, whether there should be a tax or

---

[48] Hicks (1932) defined an innovation as labor saving if at the given levels of inputs the marginal return to labor is reduced. Labor saving innovations lead to a reduction in wages. The vocabulary extends directly to the context here: an increase in capital is unskilled labor saving if at the given levels of inputs, the marginal return to unskilled labor decreases. We are particularly concerned here how a change in K affects *relative* wages, v.

28

subsidy on capital depends on whether an increase in the capital stock increases or decreases relative wages, i.e., makes the self-selection constraint more or less binding.[49]

If we can differentiate among different kinds of capital, i.e. **K** is a vector, with some components labor saving or using for different kinds of labor, then there should be a tax on those kinds of capital goods which are unskilled labor saving and a subsidy on those which are unskilled labor using. The returns on different kinds of capital goods should, accordingly, differ. There is no reason that the effect on inequality of public and private capital goods will be the same. In general, government can have more control (directly) in the choice of the extent to which public capital goods are complementary to unskilled labor (i.e. increase v) then public capital goods should be advantaged over private capital goods; the marginal return ("F$_K$") to public investment goods should be less than that for private capital goods, and in a utilitarian dynamic optimal growth model, in steady state, F$_{Kg}$ may thus be less than $1/\delta$.

More generally, the economy's transformation curve can be written as before as

(18) $\mathfrak{I}(C_1, C_2, L_1, L_2, K_{t+1}, K_t) = 0$, with $v = \dfrac{\mathfrak{I}_{L_2}}{\mathfrak{I}_{L_1}}$.

(18) describes an economy in which different consumption or capital goods are produced with different production functions. A movement along the transformation curve changes the relative demand for high and low skilled workers, i.e. affects relative wages. Now, in general, even the basic formulae for consumption tax (15), which we have shown hold in the more general overlapping generations model, have to be changed; it follows that

Proposition 7b: _If relative wages depend on consumption or public expenditure patterns, then in general, even with separability_ (i) _the rate of return should not equal the intertemporal marginal rate of substitution, (ii) the rate of return in the public sector should not equal that in the private one, and (iii) even asymptotically, with utilitarian intertemporal optimization, it will not equal the pure rate of time preference._ [50]

Consumption at date t affects the self-selection constraint at date t, while consumption at date t+1 _of the same commodity_ affects the self-selection constraint at t+1. But individuals

---

[49] Conditions under which either result obtains may easily be derived, simply by differentiating $\dfrac{F_{L_1}\left(\frac{N_1 L_1}{K}, \frac{N_2 L_2}{K}\right)}{F_{L_2}\left(\frac{N_1 L_1}{K}, \frac{N_2 L_2}{K}\right)}$ with respect to K

[50] In contrast to the earlier result where F$_K$ = $1/\delta$.

consume different commodity bundles; if the elderly on average consume more unskilled labor-intensive services (e.g. home care), then we want to encourage individuals to consume more later in life, because that drives up the relative wages of unskilled workers, reducing before tax inequality and the extent to which the self-selection constraint binds, i.e. we would want to subsidize high income individual's savings. Conversely, if the elderly on average consume more skill-labor intensive goods and services, like complex medical procedures. (As we explain in the next section, we do not believe that this model captures accurately the most important aspects of capital taxation.)

Thus, in the utilitarian dynamic optimization problem, even with separability, the consumer rate of discount (for neither high nor low ability individuals) may not converges to δ, and the producer rate of return is greater or less than δ as increasing K decreases or increases v, the relative wage. If an increase in capital accumulation increases the wage differential (if capital is a complement to high skill labor and a substitute for low skilled workers) the optimal level of F_k (assuming separability in consumption) is higher than the pure discount rate, i.e. the optimal level of capital accumulation is lower.

Distortions in the overlapping generations model with unskilled labor replacing robots

We have already seen that the level of investment in different kinds of investment goods, whether in the private or public sector, depends in part on the impact of that good (at the margin) on the distribution of income. Recently, there has been some controversy over the taxation of robots.[51]   Assume we have robots which are perfect substitutes for unskilled labor.

Consider robots (a particular type of machine), each costing one unit of ordinary capital goods, replacing unskilled labor. Our production function is now $Q = F(L_1 + R, L_2, K - R)$. Production efficiency requires $F_{L_1} = F_K$. But maximizing social welfare entails:

$$(19) \quad \lambda_2 \left(\frac{\partial V^2}{\partial L}\right) L_{1t} \left(\frac{\partial v}{\partial R}\right) = \gamma(F_{L_1} - F_K)$$

---

[51] Bill Gates suggested that there should be such a tax (Delaney, 2017). There was the predictable backlash among those who thought that that was foolish, since it would interfere with efficiency and impede innovation. This paper explains why the former charge has no validity. The second might. If there is learning by doing with spill-overs to other sector, there is a presumption that the government should be subsidizing robot development. (See Stiglitz and Greenwald, 2014), and tax or subsidy discussed here should be viewed as an increase or decrease in that tax.

where $\frac{\partial \ln v}{\partial R_1} = \frac{F_{11}-F_{13}}{F_1} - \frac{F_{21}-F_{23}}{F_2}$. (19) implies that when there is production efficiency ($F_{L1} = F_K$)

there should be further investments in unskilled labor replacing robots if it increases v; and if it

decreases v, the level of investment in robots should be decreased (there should be a tax on

that kind of robot.) Thus, so long as the direct effect of "competition" from robots driving down

wages ($F_{11} < 0$) dominates any secondary effects from the decrease in K on wages of high skilled

workers, it is desirable to tax unskilled labor replacing robots. It is easy to see that that is the

case for the standard Cobb-Douglas production function $Q = L_1^{\alpha_1} L_2^{\alpha_2} K^{1-\alpha_1-\alpha_2}$. [52] [53]

Central message

The central message of this section of the paper is simple: tax and expenditure policies

can have effects on the distribution of income that cannot be fully offset by other tax and

subsidy policies. Measures which improve the market distribution of income (here increase v)

are desirable. Heuristically, we can think of such measures as reducing the burden imposed on

distortionary consumption taxation. This result provides theoretical underpinnings for recent

policy perspectives, noted in the introduction, which have emphasized the importance of

achieving greater equality of market incomes. Thus, tax and expenditure policies need to be

sensitive to the general equilibrium distributional consequences. A tax which has the effect of

reducing the demand for unskilled labor should be reduced from the level which it otherwise

would have been; conversely if it increases the demand for skilled labor. Public investments

should be targeted to increase the productivity of unskilled workers. Private expenditures

aimed at replacing unskilled labor with machines should be discouraged (through taxation.)

The mechanism by which the welfare benefits are generated in this model is not just the

direct effect on, say unskilled workers. In the context of a government already engaged in

welfare maximizing redistributions with lump sum taxes, all that would matter is what happens

to aggregate output. But such taxes and redistributions are not feasible. Hence, changes in

relative wages also matter because they may affect the constraints on redistribution. Here, the

central constraint is the self-selection constraint. In our model, an "improvement" in the

distribution of income through an increase in the relative wage of unskilled workers loosens the

---

[52] where $\{(F_{11} - F_{13})/F_1\} - \{(F_{21} - F_{23})/F_2\} = -Q/(L_1 + R^*) < 0$.
[53] A similar analysis applies to skilled labor replacing robots.

self-selection constraint and thus enables the government to engage in further redistributions which improve societal welfare. Our analysis formalizes the notion that improving the before tax distribution of income makes it easier to get a better after tax distribution of income.

More generally, there are a broad set of hard-to-model political constraints. It is apparent that tax policy does not quickly (if at all) act to offset changes in the market distribution of income. Thus, an improvement in the before tax distribution of income will typically lead to an improvement in the after tax distribution of income and in societal welfare.

Perhaps the most striking result is that these concerns over distribution are so strong that they justify policies that lead to production inefficiency. In Diamond-Mirrlees, any distributive benefits achieved through production inefficiency could be achieved at lower cost simply by adjusting tax rates. But if there are restraints on the set of feasible taxes that go beyond just limitations on lump sum taxes, then government cannot necessarily achieve the general equilibrium effects on distribution that are obtained through distortionary production taxes (including tariffs).

V. Layered self-selection

There is one more important set of circumstances requiring significant modification of the Atkinson-Stiglitz (1976) framework. The standard Mirrlees model, of which our analysis can be viewed as an elaboration, is based on self-employed individuals producing widgets. The government can only observe the number of widgets produced, not the hours worked, and therefore it cannot infer abilities. But in fact, in modern economies, most individuals work for others, and their employers also need to identify abilities. Individuals too will seek to demonstrate to their employers that they are more able, so as to obtain higher wages.[54] While screening and signaling may be done in a number of ways, the government may be able to partially free ride on this information; and in any case, government has to take into account that tax policies affect screening and signaling.

Consider the simplest case where an employer cannot observe either an individual's output, Y, or his ability but can observe his level of education. But Y itself is a function of

---

[54] See Stiglitz (1975), where it is established that under certain conditions, there may exist a pooling equilibrium. The only competitive *self-selection* equilibria entail separation.

education, E, and ability, A. High and low ability individuals differ in the cost of education, so we can write utility as $V^i = V^i(w, E)$, with $i = 1,2$, where w is the wage paid to the individual with education E, with a low ability individual's indifference curve in {E,w} space being steeper: he requires a greater increase in wages to compensate him for getting more education. The standard competitive equilibrium entails a separating equilibrium, with $\{E_1^*, w_1^*\}$ being the point on the low ability production function (or more precisely, $w_1 = f_1(E_1)$) which maximizes V^1. $\{E_2^*, w_2^*\}$ is the point along the high ability's productivity curve $w_2 = f_2(E_2)$) satisfying the self-selection constraint: $V^1\{E_1^*, w_1^*\} \geq V^1\{E_2^*, w_2^*\}$. If w is observable *to the government*, even if E is not, the government knows the ability of the individual. It effectively free rides off of the private sector. By imposing effectively a lump sum tax on high wage individuals used to pay a lump sum subsidy to low wage individuals, the government can achieve another Pareto efficient self-selection equilibrium, one more in accord with its social welfare function.[55] In this case, optimal income taxation is far simpler than that modelled by Mirrlees.

But there is a more difficult case, where there are binding self-selection constraints where the outcomes of the private self-selection processes are not observable. Consider, in particular, the above education model, but where education affects individual's productivity. Firm's observe this, and pay a wage accordingly; but, as in Mirrlees, the government cannot observe the wage (output per hour) or education level. In the absence of taxation, individual's utility can be written as a function of their wage (= productivity) and E: $V^i = V^i(w, E)$, where now individuals can choose how much to work. We now need to embed the private self-selection constraint inside our social welfare maximization problem, adding a term to our Lagrangian $\zeta\ [V_1(\boldsymbol{C}_1,\ L_1, E_1) - V_1(\boldsymbol{C}_2, L_2, E_2)]$, where we have reverted to our usual notation where 1 is the low ability individual and 2 is the high ability. More specifically, we assume that education level E requires a vector of consumption $\boldsymbol{C}_i(E)$ and time $L_i(E)$ which yield no direct utility, leaving $\boldsymbol{C} - \boldsymbol{C}_i(E)$ and $L - L_i(E)$ for "real" consumption and leisure. While E is observable, and the functions $\boldsymbol{C}_i(E)$ and $L_i(E)$ are known, the type of the individual is not, so that if one type mimics another, the government can only infer the effects, e.g. on the marginal rates of substitutions. The resource constraint also needs to be rewritten to include the costs

---

[55] The taxes will, of course, affect equilibrium wages and education levels, because of income effects.

and benefits of education[56]: $C_1 + C_2 + E_1 + E_2 = F(L_1 f(E_1) + L_2 f(E_2))$. Note that while the binding self-selection in the usual optimal tax problem entails the high ability pretending to be low ability—working less and getting a lower income—in the usual competitive market equilibrium, the binding self-selection constraint entails the low ability pretending to be high ability. Thus, as he increases his education level (to imitate the high ability) and so gets a high wage (w_2), he also may be required to work more. It is this which upsets the standard results of optimal tax theory.

Focusing in particular on the first order conditions for L_2 and C_{2j}, there is an additional term $-\zeta V_L^1$ and $-\zeta V_{C_{2j}}^1$.[57] If $\zeta = 0$, (i.e. the private self-selection constraint was not binding), we would have much as before (letting $\lambda_1 = 0$)

$$\frac{d\Lambda}{dL_2} = N_2 \frac{dV^2}{dL_2} + N_2 \lambda_2 \frac{dV^2}{dL_2} + \gamma N_2 F' w_2 = 0$$

$$\frac{d\Lambda}{dC_{2j}} = N_2 \frac{dV^2}{dC_{2j}} + N_2 \lambda_2 \frac{dV^2}{dC_{2j}} - \gamma N_2 = 0$$

where $w_2 = f(E_2)$ yielding $MRS = -\dfrac{\frac{dV^2}{dL_2}}{\frac{dV^2}{dC_{2j}}} = F' w_2 = MRT$: There should be no marginal tax

on the high ability individual. But with $\zeta > 0$, we have

---

[56] In this formulation, E and C occur simultaneously. A more natural formulation (especially in our dynamic model) is that there are three periods in an individual's life, in the first of which he gets educated, in the second of which he works, and in the third of which he is retired. None of the results are dependent on this formulation.

Moreover, this formulation assumes that skilled and unskilled labor are perfect substitutes. The more general case is considered in section 4.

[57] In general, there are several self-selection constraints, all of which could be binding: (a) the low ability individual is indifferent between mimicking the high ability in the eyes of the firm; (b) the high ability individual who has obtained a high level of education is indifferent between mimicking the low ability individual who has obtained a low level of education in the eyes of the government; and (c) the low ability individual who has mimicked a high ability individual in obtaining a high level of education (and thus has become more productive) is indifferent between mimicking a low ability individual who has not obtained a high level of education. But in this model, if a high ability individual is indifferent between mimicking a low ability individual in the eyes of the government, then a low ability individual would not want to simultaneously pretend before his employer to be high ability (and thus get a high level of education) and pretend before the government to be low ability, for in doing so, he reduces the value of education, since in imitating the low ability, he must restrict the number of hours worked. Thus, here we can focus simply on the single additional self-selection constraint. Whether we have to include a still further self-selection constraint, the more general point that the addition of the new self-selection constraint means that only when certain separability conditions are satisfied will it be desirable not to impose distortionary taxation on the high ability individual, remains valid.

$$-\frac{dV^2}{dL_2} \bigg/ \frac{dV^2}{dC_{2j}} = = \{\gamma F' w_2 - \zeta V_L^1\}/\gamma + \zeta V_{C_{2j}}^1$$

so that in general, the MRS is not equal to the MRT: there should be intervention even in the labor market of high ability individuals. By the same token the marginal rate of substitution between two commodities j and k is given by $\frac{\gamma + \zeta V_{C_{2j}}^1}{\gamma + \zeta V_{C_{2k}}^1}$ which equals unity only if $\zeta V_{C_{2j}}^1 = \zeta V_{C_{2k}}^1$, which is true if $\zeta = 0$ or if $V_{C_{2j}}^1 = V_{C_{2k}}^1$. As earlier, we can write :

$\text{MRS}^2 - \text{MRS}^1 = \gamma (\text{MRT} - \text{MRS}^1)$.

Thus, if the low ability is mimicking the high ability in the acquisition of education, he is spending more on, say, tutoring services and more time doing homework, leaving less income and less leisure. If, somehow, the result is that the marginal rate of substitution between good j and k is the same, that would mean that $\text{MRT} = \text{MRS}^1 = \text{MRS}^2 = 1$, i.e. there should be no consumption taxes; in the case of consumption at different dates, no interest income tax. But education takes away from "net" consumption and leisure when the individual is young, and thus inevitably changes the MRS between consumption at t and t + 1: it will differ between high ability and low ability individuals. Taxing interest income—making the future less valuable at the margin relative to today—discourages low ability individuals from imitating high ability, loosening the private self-selection constraint, and thus increasing welfare.

The general point is that tax structures affect private mechanisms for sorting individuals, and government needs to take into account these impacts. Greenwald and Stiglitz (1986) and Arnott, Greenwald, and Stiglitz (1994) established that, in general, whenever there are self-selection constraints, the competitive equilibrium is not (constrained) Pareto efficient. The distortions are first order. Small taxes have a second order deadweight loss associated with them, but can have a first order effect in correcting distortions associated with self-selection constraints, e.g. in improving the efficiency of market sorting. Thus, we can think of optimal taxation as entailing two parts: (a) corrective taxation, to ensure that the market is Pareto efficient; (b) redistributive taxation, to move us towards a "better" Pareto efficient allocation. Not surprisingly, when both of these are taken into account, the presumption of AS 76 that there should not be intertemporal taxation no longer holds.

VI. Inheritances

The most critical assumption underlying Atkinson and Stiglitz (1976) was that all individuals were identical except for their abilities. But there are huge differences in inherited wealth and huge differences in returns to capital of those with wealth.[58] Much of the argument for capital taxation is that taxation of capital and returns to capital can achieve greater equity at a lower cost (say in terms of the deadweight losses associated with the tax) than just reliance on an earned income tax. In section 2, we commented on the taxation of capital in the context of a world in which individuals differ in their ability to generate returns on capital. Here, we focus on inheritances.

It is easy (in principle) to separate the taxation of inheritances and the return on inheritances from the life cycle savings discussed in previous sections.   The latter can be thought of as put into Income Retirement Accounts (IRA's), and the analysis of earlier sections argued that in the absence of e.g. distributive effects and with separability, there should be no taxation of returns in IRA's. But that says nothing about the taxation of capital in general, which we will argue should be taxed, and possibly at a high rate. If some inherited savings slips into IRA's and not all life cycle savings is included in tax-exempt IRA's, there would be some presumption for taxing the return to capital in IRA's, but still taxing the return to capital not in IRA's.

An analysis of optimal/Pareto efficient taxation of inheritances entails two complications, beyond the general equilibrium distributive effects upon which section 4. focused. (a) A full analysis of taxation and inequality in the presence of bequests has to come to terms with the drivers of financial and human capital bequests—the determinants of the intergenerational transmission of advantage and disadvantage.[59] (b) It is not obvious what the appropriate social welfare function should be.

Determinants of inheritances and the intergenerational transmission of advantages

---

[58] In a world without uncertainty in returns, the taxation of capital and the return to capital are equivalent, but this is not so in general.

[59] See Stiglitz (1966, 1969a, and, most related to the discussion here, 2015 and 2017) and Becker and Tomes (1979, 1986) for a more extensive discussion.

Individual's life time income (wealth) is determined by their endowments--inherited ability, education provided by their parents and the State, and inherited financial wealth—and the returns on those endowments, itself a function of those endowments, effort and luck. Thus, there are three sources of high levels of lifetime wealth: children who have the good fortune of having parents who bequeath them good endowments, either high abilities or substantial amounts of financial and human capital; high levels of effort; and those who get high earnings on their endowments, often as a result of luck. The latter include the inventors and entrepreneurs who earn outsized incomes, often through the exploitation of market power. Gates and Rockefeller are examples. The "successful" also include individuals who were particularly effective in receiving favors from the state, e.g. in the form of land grants. In many cases, wealth begets wealth through a variety of mechanisms, including the exercise of market and political power and returns to information asymmetries.

While some of the determinants are under the control of the individual, others are not: most importantly, inherited ability. (Some important aspects of the transmission process, besides taxation, are matters of public policy, most importantly, education.) This implies that any model of equilibrium wealth distribution must entail at least two (linked) stochastic difference or differential equations—one for ability and one for (financial and human) capital (Stiglitz, 1966, 2015; Bevan 1974, 1979; Bevan and Stiglitz, 1979). While some receive a high inheritance because of the luck of having parents with a high "taste" for bequests, others because of the luck of having very rich parents, others because they have fewer siblings among whom the wealth of the parents have to be divided, there is still a fourth source of differences in inherited wealth (more important among "life cycle savers"): in the absence of good annuities and reverse mortgages, upper individuals who do not wish to turn to the government for support in their final years have to hold more than what they actuarially need to live; on average, they die before eating up all of their savings. The children of those who die early may receive a small fortune, especially if there are no siblings.[60] While it is natural that economists begin with a discussion of *choices and tastes* (some individuals or dynasties might give greater

---

[60] This theory supports a Pareto tail distribution. See Stiglitz (1978).   Flemming (1979) and Kupczuk and Lupton (2007) show the empirical importance of such bequests.

weight to their descendants than others),[61] almost surely, these other factors are as or more important.

In most of these cases there is little reason to believe that, in general, taxing transfers at the top at a high marginal rate would have significant adverse effects on economic activity. For instance, the amount of money that individuals hold for their old age as a result of the absence of annuity markets is likely determined largely by the individual's risk aversion, the extent of public support for the aged and expectations of support from one's children, and uncertainties about life expectancy. Fertility decisions are similarly unlikely to be significantly affected by inheritance taxation. The efforts of those like Gates and Rockefeller (either at innovative and entrepreneurial activities or at exerting market power) are unlikely to be affected much by high inheritance or capital taxation at levels of wealth and income exceeding some billions of dollars. Indeed, there is some positive evidence and theory that high marginal tax rates might reduce incentives for rent seeking and exploitation, thereby increasing economic efficiency.[62]

Thus, there is some presumption that there should be high tax rates on very large bequests, the main concern being not that individuals won't work as hard, but that they would engage more in tax avoidance activities (in the extreme cases, changing residence to avoid taxation).[63] Concern about inherited wealth inequality would center progressive taxation on inheritances, not bequests.

Whatever the reason, in practice, in most countries there is still substantial intergenerational transfer of wealth, implying that even with bequests taxes, there will be individuals with substantial incomes from inherited capital. Thus, (progressive) taxation on capital arises naturally within any utilitarian framework. It can be viewed as an indirect (and imperfect) way of taxing bequests, as a second-best response to an important source of inequality.[64]

---

[61] See, for instance Piketty and Saez (2013).

[62] See Saez, Piketty, and Stantcheva (2014) and Stiglitz (2017).

[63] To the extent that individuals are engaged in intertemporal income smoothing with their children, inheritances lead to less variability in consumption. But taxing inheritances and using the proceeds for investments in human capital and redistributive transfers can more than offset both these and other adverse general equilibrium effects. See Stiglitz (1978), Bevan and Stiglitz (1979) and the discussion below.

[64] See Cremer and Pestieau (2006) and Cremer, Pesteiau and Rochet (2003), Farhi and Werning (2010), Piketty and Saez (2013), and Stiglitz (2017).

Objectives

While Pareto efficiency and utilitarianism provide a natural framework for evaluating alternative taxes on wages, the welfare evaluation of inheritance taxes is more complicated. As Stiglitz (1987) noted, bequests are doubly blessed: while they give utility to the giver, they also give utility to the receiver. There is a positive consumption externality. There is thus a presumption that there will be an undersupply of bequests, all else being equal, and that they should be subsidized.[65] In the context of models such as those considered in previous sections, apart from this externality effect—and possible general equilibrium distributive effects of the kind discussed in the previous section-- bequests are just another form of utility generating expenditure, and thus there should be no bequest taxation if (and only if) if there is separability between consumption (now including bequests) and labor. (Kaplow 2001).[66]

Trade-offs between consumption inequality and wealth inequality

But there are two reasons for thinking that this analysis does not fully capture societal concerns about inequality. First, note that there may be a trade-off: the imposition of inheritance taxes, in the absence of offsetting taxes, may increase consumption inequalities, in a world in which bequests serve to smooth consumption *within* a dynasty, i.e. where rich parents share some of their wealth with their children. (Stiglitz, 1976b) Consider a model (such as Bevan and Stiglitz, 1979) where individuals themselves think about their children and, when they have good luck, either being born with high ability or getting an unusually large return on their investments, set aside an optimal amount to share with their descendants. Ignoring general equilibrium effects (impacts on the distribution of income and self-selection constraints), each individual is optimally smoothing income over generations. It might seem that so long as the family's intertemporal discount rate is the same as the social welfare

---

[65] This idea has been incorporated into and commented upon by many of the more recent models of inheritance taxation. See, e.g. Kaplow (2001) and Piketty and Saez (2013). At the same time there is another important externality, that associated with an increase in inequality. There is a large literature emphasizing how inequality affects economic performance and society more generally. Most of the economics literature on bequests has not incorporated these important externalities.

[66] Much of the literature models the utility of bequests in a reduced form, rather than as a result of increasing the utility of one's descendants, the benefits of which are incorporated into one's well-being. Standard structural modelling (e.g. a dynastic utility function) also naturally entails separability between labor and bequests.

function[67], there is no reason for government to engage in intertemporal distribution. But that is not correct: it is still the case that *at this moment* some families have a possibly substantially larger endowment of "ability" and capital than others, and therefore a redistributive tax would be desirable.[68] Moreover, each period in the future, some individuals get a better roll of the dice—given their parents, they have higher wages or higher returns on their inherited capital. While they optimally share their good fortune with their descendants, they don't share their good fortune at all with others who have had a bad roll of the dice, ending up with low wages or low returns on capital. A utilitarian would wish to redistribute income between the lucky individual and those who had a bad roll of the dice. (Indeed, so would an individual, behind a veil of ignorance, not knowing how the dice would roll.)

If the government can observe inherited wealth I_t, then, at time 0, there should obviously be a tax on the inheritance and on returns to inherited capital. Indeed, taxing past bequests (capital that individuals have already inherited) at 100% is non-distortionary and would thus be part of a utilitarian optimal tax program. Of course, this is just the standard paradox of capital taxation: in the absence of some form of commitment, savers would know this, and there would be no savings and investment. There has to be some way that governments can make a credible commitment not to impose such taxes. We do not answer that question here, but rather ask, if the government could make a credible commitment to say a fixed inheritance tax and/or return to capital, how would those taxes be determined?[69] With such a commitment, the government must balance the benefits of greater equality across families with reduced intertemporal smoothing of consumption within the dynastic (and more broadly, other aspects

---

[67] And putting aside the fact noted above that in a utilitarian framework, a bequest gives utility to both the giver and the receiver.

[68] One obviously cannot study this effect in a model in which, starting at time 0, all individuals begin with the same wealth and ability, and making random draws going forward. In that world, at time 0, there would be no need for taxation; but subsequent periods would begin with wealth and ability inequalities and redistributive taxation would be desirable. Of course, at time 0, individuals would know this, and that would affect behavior at time 0

[69] No democratic government can make such commitments, and it is apparent that different parties have different views about the appropriate tax rate. Much political campaigning centers around such issues. Each party *should* however take into account the likelihood that a successor government would change tax rates and the effect of perceptions that this might be so on behavior. One can reconcile observed behavior, e.g. of the Republican Party in the US, as they adopt regressive taxation—which are more likely to be reversed with the next Democratic period of control-- with this perspective by assuming that they have become increasingly short sighted, perhaps because of a belief in declining probabilities of success in future elections; see Korinek and Stiglitz (2008,2009).

of deadweight loss associated with distortionary taxation.)   In principle, the optimal tax function would relate the tax paid by individual at time t to all observables related to his current behavior and past history (i.e. to Y_t, I_t for all t ≤ t₀, where t₀ is today's date)), since such data (together with knowledge of the stochastic processes) convey information about relevant current period variables, in particular w_t (the individual's ability). For instance, if ability were perfectly inherited, information from previous periods' incomes of the parents perfectly conveys information about the individual's current ability. If there is a high correlation, past data would still convey considerable information. (Since the standard models assume markovian processes for ability, only the previous period income is relevant.[70])

Of course, individuals know this, and if they care about their descendants, as assumed here, their behavior would correspondingly adjust. This can have profound consequences, including that the equilibrium is not a separating equilibrium (as in the previous sections of this paper) but a pooling one.[71] We thus have to compare the level of social welfare with the optimal tax of the form $T(Y_t, Y_{t-1}, I_t, I_{t-1})$ with that generated simply by a tax of the form: $T(Y_t, I_t)$. In the latter case, taxes would have to take into account the self-selection constraints described in previous sections of the paper.[72]

Societal implications of inequalities of wealth and inheritance

Should society be concerned with consumption inequalities within a generation, with inequalities in dynastic utility, or with inequalities in wealth?   Societies with large wealth inequalities are different; large wealth inequalities are associated with inequalities in political power and access to privilege. If society is concerned with excessive inequalities in wealth, then progressive inheritance taxes may be desirable, even if they cause "dead weight" losses in utility, in inducing excessive consumption of an individual within his lifetime. Concerns about extremes of inequalities in lifetime consumption can then be addressed separately by a progressive lifetime consumption tax.

---

[70] There is some evidence that the ability generating process is not so simple, e.g. that characteristics of the grandparents may affect the grandchildren independently of the state of the child.

[71] See Roberts (1982) and Stiglitz (2009b)

[72] In our earlier analysis, we implicitly assumed that the government did not make use of the information implicit in the inheritability of ability. But while using such information would almost surely be viewed as politically unacceptable, inheritances are correlated with income of previous generations.

Bequests as a source of inequality may be viewed as particularly iniquitous, because such bequests mean that there is not a level playing field; there is no equality of opportunity.[73] While society cannot create a perfectly level playing field, there are some forms of inequality—those associated with the transmission of financial wealth—which can be redressed. The question again are the costs of doing so. Since the marginal costs of small distortions is small, and, given the large inequities which currently exist, the marginal benefit of the corresponding reduction in inequalities arising from inheritance taxes and taxes on non-life-cycle capital income are large, there is a presumption in favor of such taxes.

VII. Optimal Taxation in a Two-Class Model

A simplification in which these perspectives can be investigated more closely and which captures well key elements of a modern capitalist economy entails workers in an overlapping generations model and a second group with so much inherited wealth that their wealth dwarfs their wage income, and so we ignore it. The latter group we refer to as capitalists; the former as workers. We focus on the differences in well-being between these two groups, rather than differences within each group. [74].

Total capital in the economy is workers' capital and capitalists' capital, generating a two class model along the lines studied earlier by Pasinetti (1962), Modigliani and Samuelson (1966) and Stiglitz (1967, 2015). We extend the model here by introducing utility maximizing workers and capitalists.[75] Capitalists maximize a standard intertemporal utility function with discount factor $1/\delta$. We normalize by assuming workers have a (fixed) population of unity, capitalists of

---

[73] It is worth noting that equality of opportunity is a distinctively different norm that that associated with, say, maximizing an inequality averse dynastic social welfare function. (Kanbur and Stiglitz (2015, 2016)). Equality of opportunity says the probability that an individual will be in any given decile during his life is independent of the decile of his parents. A two period egalitarian family social welfare function $W = W(U_t, U_{t+1})$ would seek to compensate an individual whose father is poor (and therefore whose childhood has more likely been spent in poverty) with higher income, calling for a quite different transition matrix.

[74] With stochastic wages and returns to capital, these can be analyzed along the lines of Stiglitz (2015). As discussed in the previous section, Pareto efficient taxation within each group can be analyzed in standard models such as those presented in earlier sections of this paper.

[75] Thus, we generalize Mattauch *et al* (2016) and Mattauch *et al* (2017). Their models used logarithmic utility functions for both capitalists and workers, which naturally generate the kinds of fixed savings rates assumed by the earlier literature. Mattauch *et al* 2017 focused on ascertaining the conditions under which in equilibrium all capital was held by workers. Here, we assume that the tax rates on capital (or on capitalists) are low enough that this is not so.

N, so capitalists' consumption is N C_c. We simplify the analysis by assuming the government simply wishes to maximize workers' steady state consumption.[76] We assume, for simplicity, that the labor supply is fixed (at unity as well).[77] Here, we assume all revenues from taxation are spent on education. Later subsections will generalize the results. Then output, Q, is a constant returns to scale function of (private) capital, K, and effective labor, which is a function of investment in human capital per capita, which, noting that labor supply has been normalized at unity, can be written as $\phi(K_g)$, with $\phi'(K_g) \geq 0$ and $\phi''(K_g) \leq 0$. So $Q = F\left(K, \phi(K_g)\right)$. Investments in education increase productivity, but there are diminishing returns. Because of constant returns to scale in capital and effective labor, we can write:

(20) $Q = \phi\, f(K/\phi) = \phi\, f(k)$

where $k = \frac{K}{\phi}$ the capital-"effective labor" ratio. We assume a tax rate of τ on the net return to capital.[78] This implies that in long run equilibrium k is given by:

(21) $1 + r = (1 - \tau)(f' - \eta) = \delta$

where r is the (after tax net) rate of return on capital, η is the rate of depreciation and the tax is applied only to net returns. The after tax net return to capital is simply related to the pure rate of time discount. (22) determines the equilibrium value of k as a function of τ (given η and δ).[79]

(22) $k = \psi(\tau)$, with $\psi'(\tau) = \delta/(1-\tau)^2 f'' < 0$.

We denote by k_o the value of k when $\tau = 0$: $k_o \equiv \psi(0)$. Tax revenue is $\tau\,(f' - \eta)K$, all of which is spent on public education. Hence

(23) $K_g = \tau\,(f' - \eta)K = \frac{\tau\delta}{1-\tau}K = \frac{\tau\delta}{1-\tau}\,\psi\phi$

[76] Similar results obtain if we analyze more generally Pareto efficient tax structures, or tax structures which maximize a more general inequality averse social welfare function, so long as the disparity between the two groups is large enough, so that the marginal social utility of consumption of capitalists relative to that of workers is small enough.

[77] Thus, while earlier sections focused on the issue of the labor supply elasticity, here our concern is on savings. Results on the desirability of capital taxation would be strengthened by allowing a positive labor supply elasticity.

[78] Again, similar results obtain if one can differentiate between taxes on capital for workers and that of capitalists, or if capitalists engage in some tax avoidance activities, to convert some of their capital income into what appears as wage income.

[79] Throughout the analysis below, we assume we are in an equilibrium in which there are capitalists. See Pasinetti (1962), Modigliani and Samuelson (1966) and Stiglitz (1967, 2015, 2016) and Mattauch *et al* 2017.

Assuming workers have homothetic indifference curves between consumption when young and old[80], worker savings are given by $s_w(r)\,w$,

where w is the wage:

(24) $w = \phi(K_g)\,g(k) = \phi(K_g)\,(f - kf')$

where $g(k) = f - kf'$ is the return per unit of effective labor. Hence

(25) $K_w = s_w(r)\,\phi(K_g)\,g(k) = s_w(\delta)\phi(K_g)\,g(\psi(\tau))$

using (22). (23) can be used to solve K_g as a function of τ. There is a unique solution if τ is small or if $\zeta < 1$, where $\zeta = \dfrac{d\ln\phi}{d\ln K_g}$, the elasticity of productivity with respect to investments in human capital. We can also solve for workers' steady state utility, $V^w = V^w(w, r)$, which depends just on $\{w, r\} = \{w, \delta\}$ which from (23) and (24) depends just on τ.

Taking the logarithmic derivative of w (from (24)), we can establish the tax rate that maximizes workers' welfare (and, given social weights that put little weight on capitalists well-being approximately societal welfare). Increasing τ increases K_g but decreases k and hence g, the wage per efficiency unit. The optimal tax balances the two effects:

(26) $\quad 0 < \tau^* = \dfrac{a}{1+a} < 1$

where $a = \dfrac{c}{d(b+c)}$, $c = \dfrac{\zeta}{1-\zeta}$; $b = S_K/\sigma$; and $d = \dfrac{\sigma}{(1 - S_K)\,(\delta/r)}$, where σ = elasticity of substitution and $S_K = f'k/f$. The larger is the elasticity of the return to human capital ζ, the larger τ^*; and the larger the elasticity of substitution, the smaller is the adverse effect on wages of an increase, and thus the larger is τ^*. Hence, we have:

<u>Proposition 8a:</u> *Provided $0 < \zeta < 1$, there is a strictly positive positive tax rate on capital taxation to finance education.*[81]

These results assume that there is no productive public investment to augment the productivity of capital goods and no transfers. More generally, the total income of a worker with transfers but net of labor taxes t are given by

$$W = \phi(f - kf') + \tau\,rK - K_g - K_{PI} - t$$

---

[80] If they don't, then $K_w = S(w, r)$, and the analysis proceeds much as below.
[81] A tax levied only on the capital earnings of capitalists would be even more favorable to workers.

where K_{PI} is investment in public capital goods allocated to augmenting capital productivity, and K represents as before just private capital. Output is a CRTS function of effective labor, K_{PI} and K: $Q = F(K, K_{PI}, \phi(K_g)L)$.

If the tax were only used to finance lump sum transfers to workers, then workers' (after transfer) wage is given by $W = g(k) + \tau(f' - \eta)k$. At $\tau = 0$, $\frac{dW}{d\tau} = -k\delta + (f' - \eta)k = 0$. But straightforward differentiation shows that for $\tau > 0$, $\frac{dW}{d\tau} < 0$, implying:

Proposition 8b: *A tax on capital used to finance a lump sum transfer to workers is welfare decreasing.*

If all tax revenues go to public capital that *augment*s the productivity of private capital, then $Q = F\left(H\left(\frac{K_{PI}}{K}\right)K, L\right)$ [82], and in equilibrium the effective capital labor ratio is determined so that:

(27) $(Hf'(k) - \eta)(1 - \tau) = \delta$.

Since public investment is financed by a tax on capital: $K_{PI} = \tau K H f'$, implying that $z \equiv \frac{K_{PI}}{K}$ is given by

(28) $z = \tau H(z)f'(k)$

Substituting into (27), we obtain: $\frac{z}{\tau - \eta}(1 - \tau) = \delta$, implying that $z = \tau\left(\eta + \frac{\delta}{1-\tau}\right)$. This can be solved for z as a function of τ, substituted into (28), so that we can solve for k as a function of τ. Workers' wages increase as k increases, i.e. as $(1 - m)\left(1 + \frac{\delta}{\eta + \delta}\right) > 1$, where $= \frac{H'z}{H}$. Thus, if public capital investments are not too "capital augmenting," then workers' incomes are increased by an increase in capital taxation with revenues used for capital augmenting public capital (even with capital goods capturing the direct returns.)

On the other hand, if the funds are spent on capital goods that substitute for private capital goods, then the aggregate capital labor ratio need not even fall; even with limited productive human capital investments, workers' income (through an increase in wages and/or transfers)

---

[82] This formulation is chosen so that the production function is constant returns to scale. It is assumed that the government does not appropriate directly the returns to public capital goods, but rather, the returns are appropriately by the owners of capital goods (just as the returns to human capital are appropriated by workers)

may increase, and so will their welfare.[83]   With the return to capitalists' capital below the critical threshold, they eventually consume their wealth. The economy is supported then by public investment. This result holds even if public investment is not as productive as private investment, so long as it is not too much less.

The central point is simple: while a capital tax the proceeds of which are simply distributed to workers hurts workers because of adverse incidence effects, under quite general conditions[84], with an equalitarian social welfare function it is desirable to impose a tax on capitalists' return to capital, and with large disparities in income between capitalists and workers, strong equalitarianism, and high productivity of investments in human capital, the optimal tax on capital may be very high.   If tax revenues are invested in capital goods which either increase the productivity of private capital good (but not too much) or substitute for private capital goods, then again taxation on capital is desirable. Indeed, in the latter case, a tax on capital over the long run may result in capitalists consuming all of their capital.

VIII. Concluding Remarks

It is remarkable that so much of the policy literature has focused on simplistic models in which differences among individuals were limited, arising mostly out of differences in wage incomes. The Atkinson Stiglitz 1976 model was useful in reminding us that the role of commodity taxation had to be seen in conjunction with other taxes that were in place as well as taxes that we might have wanted to have in place, but which are not there—a general principle of considerable importance. When there was an optimal non-linear income tax, the role of commodity taxation was limited.   Its role could be seen in two ways. First, as improving the before tax (market) distribution of income, putting less of a burden on distortionary redistributive taxes. Alternatively, it could be seen as part of the general theory of corrective

---

[83] That is, if Q = F(K + K_{PI}, L), then the expenditures on public capital goods drives down the after tax rate of interest so that (f' − η)(1 − τ) < δ, in which case capitalist start to "eat" their capital. To ensure that capital taxation does not decrease wages, it must be the case that π > s_p, where π = relative productivity of public investment and 1 - s_p is the fraction of profits consumed ("wasted" on capitalists).

[84] The result can be shown to hold with more general utility functions for capitalists (e.g. recursive Koopmans functions, so that the long run interest rate need not be invariant to tax policy.) It also holds with different specifications of taxation, e.g. a wealth tax on capitalists, or a tax on intergenerational transfers, rather than a tax on their net income.

taxation. Whenever there are self-selection constraints in an economy, there will be a first order distortion, and commodity taxation can be used to reduce the magnitude of that distortion. This is true whether the self-selection constraint arises purely within the private public sector – from the government attempting to engage in redistributive taxation – or also within the private sector – from the more-able individuals attempting to appropriate their ability rents. The presence of self-selection constraints represents a big wedge, a big distortion, in the economy. It is worth creating a small distortion elsewhere in the economy if it can reduce the size of this wedge.[85]

Public investment too needs to take into account its effects on the market distribution of income, and how it alters the distribution of income and self-selection constraints within the economy. In general, even asymptotically and even with separability, the rate of discount for public investment varies across public projects, and may be greater or less than the social rate of discount.

There are many other limitations to the basic Atkinson-Stiglitz model that we have not explored here. We have, in particular, assumed the absence of tax evasion and avoidance, and in particular the ability to shift income from one category to another. In practice, it is often difficult to distinguish between capital and labor income, especially for self-employed individuals; and if that is the case, large differences in tax rates provides large incentives to convert wage income into capital income.[86]

This is perhaps related to one of the deepest criticisms of the literature on optimal non-linear taxes: in practice, there are many restraints. We have noted, for instance, if government is restrained to using a linear income tax, commodity taxation is still desirable. The rationale is

---

[85] Thus, it is desirable to introduce random taxation, even though all individuals are risk averse, if by doing so the self-selection constraints are relaxed. See Stiglitz (1982b) and Brito *et al*. (1995).

[86] There are, of course, several other widely noted limitations on the model, which limit the direct applicability of the result: (a) Hours may be partially observable, in which case government would want to use that information; and (b) there may be large stochastic elements in the translation of labor, given an individual's ability and effort, into income. In dynamic models, as we noted earlier, observations of an individual in time t are relevant to inferences at subsequent dates (and this may even be true of observations of an individual's parents). Optimal taxation ought to make use of that information, but if individuals know this, it affects behavior. In effect, the models here entail a commitment on the part of government not to use such information. The models here are adverse selection models. There are also important incentive effects, e.g. in effort and education. Thus, a full analysis would include both effects, with incentive compatability constraints added to the self-selection constraints. (For a discussion of optimal taxation with moral hazard, see Arnott and Stiglitz, 1986.)

somewhat different from that presented above. Such taxation may improve the before tax distribution of income, as we have emphasized, but even if does not do this, it may enable us, with little social deadweight loss, to improve the after tax and transfer distribution of income, and thus social welfare. A tax on perfume may generate significant revenues from a subset of high income individuals—individuals who are such that with even an optimal linear income tax have a much lower social marginal utility than poor individuals. Similarly, for a bread subsidy. Though as Deaton noted, with linear expenditure systems, one can't achieve improvements through perfume taxes and bread subsidies, there may be significant departures from those patterns of expenditure.

This leads to the question: why the restraints. There are obvious advantages to simplicity, but when one looks at the tax code, it is clear that simplicity itself was not the guiding principle. Making the tax system non-linear, with several brackets, adds virtually no cost in a modern computer driven tax system, raising a key point: what is costly or feasible depends on technology, and that changes over time—with some changes endogenously driven. When Vickrey advocated the use of congestion taxes, a good congestion tax could not have been implemented. But advances in technology now make the administrative costs associated with imposing taxes low. Economists often refer to political costs, but my experience with "political constraints" is that they are soft and ambiguous. What seems like an insurmountable barrier at one time can be suddenly overcome, especially as economists explain the high costs of such a constraint. I had argued some quarter a century ago that there would be substantial benefits from substantially increasing the standard deduction, moving many individuals out of the income tax net and reducing the importance of many key distortionary deductions and credits but it was claimed doing so was politically infeasible; suddenly, in 2017, in the US, it happened. I would have thought that in a democracy, a tax cut for billionaires financed by tax increases for lower income individuals would be infeasible; again, in 2017, it happened.

There are two sets of constraints for which there are important rationale. One is that large differential tax rates, say between capital and labor, open up large opportunities for arbitrage. Similarly, for large differential rates among classes of similar activities. More broadly, there are limitations in what government can observe, and what it can define

precisely, and those limitations provide openings for arbitrage and distortions. When there are large differences in marginal tax rates, money can be shifted from one person (say in a family) to another. Much of the complexity of the tax system arises from these forms of differentiation, and any tax structure needs to take this into account.[87]

Most importantly, there is a simple reason to tax the returns to capital of the very rich: it is equitable. With any equality-preferring social welfare function with sufficient weight on equity, the benefits in terms of social equity would outweigh the distortionary effects, given the disparities in incomes observed in advanced countries such as the United States. The intertemporal distortions in consumption patterns affect the very rich, when the tax is levied only on very high incomes. But even if a capital tax is also levied on workers it may be desirable. If the proceeds of the tax are invested in ways which sufficiently enhance the productivity of workers and/or provide sufficiently highly valued public goods, workers are better off, and the gains to the workers more than offset the losses to the "capitalists" who bear the tax. We have shown that, in general, the adverse incidence effect on workers as a result of capitalists' reduced savings can be offset. With the standard utilitarian social welfare functions, with the degree of concavity conventionally assumed, there can be even considerable inefficiency in public investments, and still capital taxes would be desirable.

All of this has been predicated on individualistic inequality averse social welfare functions. But there is increasing awareness that inequality itself may have an adverse effect on societal well-being, and even productivity.[88] There is an increasing awareness that living in an unequal society changes the nature of society and that of those that live within it, with effects on preferences and behavior of both those at the top and the bottom that would widely be agreed

---

[87] In 1985, I set out a general theory of tax avoidance. It is particularly difficult for the government to impose different taxes on things which are similar (but still different). Defining the relevant differences can be extraordinarily difficult, and thus it may be necessary to tax the different commodities at the same rate.

All of modern tax theory begins, of course, with the hypothesis that certain things (like ability) are not observable. In fact, different things are observed with different degrees of accuracy.

[88] Standard theory takes individuals' preference and behavior as given, but recent advances in behavioral economics have shown that these are endogenous and have identified many ways in which these are affected by the nature of the society in which individuals are embedded, including the magnitude and nature of the inequalities. See Hoff and Stiglitz (2016, 2017).

to be adverse.[89] Indeed, there is increasing evidence that inequality itself may affect productivity.[90] Once we take these consequences of inequality into account, the case for a progressive tax on the income from capital, and in particular, taxing the return to what we have identified as "capitalists' capital" at a high rate becomes even more compelling.

I believe that Tony would have agreed with these conclusions.

References

Arnott, Richard and J. E. Stiglitz, 1979, "Aggregate Land Rents, Expenditure on Public Goods and Optimal City Size,", *Quarterly Journal of Economics*, 93(4), 471-500.

_____, 1981, "Aggregate Land Rents and Aggregate Transport Costs," *Economic Journal*, 91(362), 331-347

_____, 1986, "Moral Hazard and Optimal Commodity Taxation," *Journal of Public Economics*, 29, pp. 1-24.

Arnott, R., B. Greenwald, and J. E. Stiglitz, 1994, "Information and Economic Efficiency," *Information Economics and Policy,* 6(1), 77-88.

Arrow, K. J. and R. C. Lind, 1970, "Uncertainty and the Evaluation of Public Investment Decisions," Amer. Econ. Rev., June 1, 60, 364-78

_____2014, "Uncertainty and the Evaluation of Public Investment Decisions," *Journal of Natural Resources Policy Research* , 6(1), 29-44.

Atkeson, A., V. V. Chari, and P. J. Kehoe, 1999, "Taxing Capital Income: A Bad Idea," *Federal Reserve Bank of Minneapolis Quarterly Review,* 23(3): 3-17.

Atkinson, A., 1980, "Inheritance and the Redistribution of Wealth", in G.M. Heal and G.A. Hughes, eds., *Public Policy and the Tax System*, London: Allen and Unwin, 26-66.

_____, 2015, *Inequality*. Harvard University Press.

---

[89] With endogenous preferences, there are well-known difficulties in formulating appropriate social welfare functions.

[90] There is a large theoretical and empirical literature relating inequality to overall output and productivity. See, e.g. Ostry *et al* (2014) and Stiglitz (2012) and the references cited there.

Atkinson, A. and A. Sandmo, 1980, "Welfare Implications of the Taxation of Savings," *Economic Journal,* 90, 529-49.

Atkinson, A. and N. Stern, 1974, "Pigou, Taxation and Public Goods," *Review of Economic Studies*, 41, 119-128.

Atkinson, A. and J. E. Stiglitz, 1972, "The Structure of Indirect Taxation and Economic Efficiency," *Journal of Public Economics*, 1(1): 97-119

_____, 1976, "The design of Tax Structure: Direct Versus Indirect Taxation," *Journal of Public Economics*, 6: 55-75.

_____, 1980, *Lectures on Public Economics*, New York and London: McGraw-Hill Book Company.

Baumol, W. J. and D. F. Bradford, 1970, "Optimal Departures from Marginal Cost Pricing," *American Economic Review*, 60(3): 265-83.

Boadway, R. and M. Keen, 1993, "Public Goods, Self-Selection and Optimal Income Taxation." *International Economic Review*, 34, 463–478.

Bevan, D., 1974, "Savings, inheritance and economic growth in the presence of Earnings Inequality," Oxford University mimeo.

_____, 1979, "Inheritance and the Distribution of Wealth*", Economica* 46(184): 381-402

Bevan, D. and J. E. Stiglitz, 1979, "Intergenerational Transfers and Inequality," *The Greek Economic* Review, 1(1): 8-26.

Boiteux, M., 1956, "Sur la gestion des monopoles publics astreints à l'équilibre budgetaire," *Econometrica*, 24(1): 22-40.

_____, 1971, "On the management of public monopolies subject to budgetary constraints," *Journal of Economic Theory*, 3(3): 219-240.

Brito , D.L., J.H. Hamilton S.M. Slutsky and J. E. Stiglitz, 1990,"Pareto Efficient Tax Structures," *Oxford Economic Papers*, 42, pp. 61-77.

51

_____, 1995, "Randomization in Optimal Income Tax Schedules," *Journal of Public Economics*, 56(2): 189-223.

Chamley, C., 1986, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica,* 54(3): 607-622.

Christiansen, V., 1984, "Which Commodity Taxes Should Supplement the Income Tax?" *Journal of Public Economics* 24: 195-220.

Corlett, W.J. and D.C. Hague, 1953, "Complementarity and the Excess Burden of Taxation," *Review of Economic Studies* 21(1): 21-30.

Cremer, H. and P. Pestieau, 2006, "Wealth Transfer Taxation: A Survey of the Theoretical Literature," in S.C. Kolm and J. Mercier Ythier, eds, *Handbook of the Economics of Giving, Altruism, and Reciprocity* 2(2): 1107-1134.

Cremer, H., P. Pestieau, and J.-C. Rochet, 2003, "Capital Income Taxation When Inherited Wealth Is Not Observable," *Journal of Public Economics* 87(11): 2475-90.

Dasgupta, P. and J. E. Stiglitz, 1971, "Differential Taxation, Public Goods, and Economic Efficiency," *Review of Economic Studies* 38(2): 151-174.

_____, 1972, "On Optimal Taxation and Public Production," *Review of Economic Studies* 39(1): 87-103.

_____, 1975, "Benefit-Cost Analysis and Trade Policies," *Journal of Political Economy*, 82(1), January-February, pp. 1-33

Deaton, A., 1979, "Optimally Uniform Commodity Taxes," *Economics Letters* 2: 357-61.

_____,1981, "Optimal Taxes and the Structure of Preferences," *Econometrica,* 49(5):1245-59.

Delaney, Kevin J., "The robot that takes your job should pay taxes, says Bill Gates**,"** *Quartz*, February 17, 2017

Diamond, P., 1975, "A Many-Person Ramsey Tax Rule," *Journal of Public Economics* 4(4): 335-42

_____,1998, "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, 88: 83-95.

Diamond, P. and J.A. Mirrlees, 1971, "Optimal Taxation and Public Production," *American Economic Review* 61(1):8-27 and 261-278.

Domar, E. D., and R. A. Musgrave, 1944, "Proportional income taxation and risk-taking". *The Quarterly Journal of Economics*, 58(3), 388-422.

Farhi, E. and I. Werning, 2010, "Progressive Estate Taxation," *Quarterly Journal of Economics* 125(2): 635-673.

Flemming, J. S., 1979. "The effects of earnings inequality, imperfect capital markets, and dynastic altruism on the distribution of wealth in life cycle models," *Economica*, 46(184): 363-380.

Gauthier, S., and G. Laroque. "Separability and public finance." *Journal of Public Economics* 93(11): 1168-1174.

George, H., 1879. *Progress and poverty: An inquiry into the cause of industrial depressions, and of increase of want with increase of wealth, the remedy*. W. Reeves.

Greenwald, B. and J. E. Stiglitz, 1986, "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics*, 101(2), 229-264.

Hacker, J. and P. Pierson, 2011, *Winner-Take-All Politics: How Washington Made the Rich Richer—and Turned its Back on the Middle Class*, New York: Simon and Schulster.

Hicks, J., 1932, *The Theory of Wages*. London: Macmillan.

Hoff, K. and J. E. Stiglitz, 2016, "Striving for Balance in Economics: Towards a Theory of the Social Determination of Behavior," *Journal of Economic Behavior and Organization*, Issue 126 (June): 25–57

_____, 2017, "How Inequality Weakens our Economy and Divides our Society," paper presented to the American Sociological Asssociation Annual Meeting, Montreal.

Kanbur, R. and J.E. Siglitz, 2015, "Wealth and Income Distribution: New Theories Needed for a New Era" *Vox*, August 18, 2015.

_____, 2016, "Dynastic Inequality, Mobility and Equality of Opportunity," *Journal of Economic Inequality*, 14(4): 419-434.

Kaplow, L., 1996. "The Optimal Supply of Public Goods and the Distortionary Cost of Taxation." *National Tax Journal*, 49, 513–33.

_____, 2001, "A Framework for Assessing Estate and Gift Taxation," in *Rethinking Estate and Gift Taxation*, ed. by W. Gale, J. Hines, and J. Slemrod. Washington, D.C.: Brookings Institution Press.

_____, 2006a, "Optimal Control of Externalities in the Presence of Income Taxation," NBER working paper 12339, June.

_____, 2006b, "On the Undesirability of Commodity Taxation Even When Income Taxation Is Not Optimal," *Journal of Public Economics* 90(6/7): 1235-1250.

Kopczuk, W., and J. Lupton, 2007, "To Leave or Not to Leave: The Distribution of Bequest Motives," *Review of Economic Studies*, 74 (1): 207–235.

Kopczuk, W., AND J. Slemrod, 2001, "The Impact of the Estate Tax on Wealth Accumulation and Avoidance Behavior," in *Rethinking Estate and Gift Taxation*, ed. by W. Gale, J. R. Hines, and J. Slemrod. Washington, D.C.: Brookings Institution Press, 299–349

Korinek, A. and J. E. Stiglitz, 2008, "Political Economy in a Contestable Democracy: The Case of Dividend Taxation,", *2008 Meeting Papers*, Society for Economic Dynamics.

_____, 2009, "Dividend taxation and intertemporal tax arbitrage." *Journal Public Economics*, 93, 142–159

Mankiw, N. G., M. Weinzierl, D. Yagan, 2009, "Optimal Taxation in Theory and Practice," NBER Working Paper 15071, June.

Mattauch, L., O. Edenhofer, D. Klenert, and S. Bénard, 2016, "Distributional Effects of Public Investment when Wealth and Classes are Back," *Metroeconomica*, 67(3): 603-629.

Mattauch, L., D. Klenert, J. E. Stiglitz, and O. Edenhofer, 2017, "Overcoming wealth inequality: when can capital-tax financed public investment help?" paper presented to the 18th World Congress of the International Economic Association, Mexico City.

Mirrlees, J. A., 1971, "An Exploration in the Theory of Optimal Income Caxation," *Review of Economic Studies* 38(2): 175-208.

———, 1975, "Optimal Commodity Taxation in a Two-Class Economy," *Journal of Public Economics* 4(1): 27-33.

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 Econmics,* 71, 168-88

Nichols, A. L. and R. J. Zeckhauser, 1982, "Targeting Transfers through Restrictions on Recipients," *American Economic Review* 72: 372-377.

Ostry, M. J. D., M. A. Berg, and M. C. G. Tsangarides, 2014, "Redistribution, Inequality, and Growth." IMF Staff Discussion Notes, 14/2, available at: https://www.imf.org/external/pubs/ft/sdn/2014/sdn1402.pdf.

Pigou, A. C. , 1920, *The Economics of Welfare,* London (4th edition 1932).

———, 1947, *A study in public Finance*, Macmillan, London.

Pasinetti, L., 1962, "The Rate of Profit and Income Distribution in Relation to the Rate of Economic Growth," *Review of Economic Studies*, 29(4): 267-279.

Piketty, T., and E. Saez, 2012, "Optimal Labor Income Taxation," NBER Working Paper 18521, November.

———, 2013. "A theory of optimal inheritance taxation". *Econometrica*, 81(5): 1851-1886

Piketty, T., E. Saez, and S. Stantcheva, 2014: "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, 6, 230– 271

Ramsey, F., 1927, "A contribution to the theory of Taxation," *Economic Journal,* 37(145): 47-61

Roberts, K., 1982, "Long-Term Contracts", mimeo.

55

Rothschild, M. and J. E. Stiglitz, 1973, "Some Further Results on the Measurement of Inequality," *Journal of Economic Theory*, 6(2): 188-204.

Sadka, Efraim, (1976). "On Income Distribution, Incentive Effects and Optimal Income Taxation," Review of Economic Studies, 42, 261-268.

Saez, E., 2001. "Using elasticities to derive optimal income tax rates". *The Review of Economic Studies*, 68(1), 205-229.

_____, 2002a, "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics*, 83, 217-230.

_____, 2004, "Direct or Indirect tax instruments for redistribution: short run versus long-run," *Journal of Public Economics,* 88 pp. 503-18.

Sah, R. and J. E. Stiglitz, 1992, *Peasants versus City-Dwellers: Taxation and the Burden of Economic Development*, Oxford and New York: Oxford University Press, 1992

Samuelson, P., 1948, "International Trade and the Equalization of Factor Prices," *Economic Journal* 58:163-184.

Samuelson, P. A., and F. Modigliani, 1966, "The Pasinetti Paradox in Neoclassical and More General Models," *Review of Economic Studies*, 33(4): 269-301.

Sandmo, A. 1975, "Optimal Taxation in the Presence of Externalities," *The Swedish Journal of Economics,* 77(1): 86-98

Sappington, D. and J. E. Stiglitz, 1987, "Information and Regulation," in *Public Regulation*, E. Bailey (ed.), London: MIT Press, 1987, 3-43

Seade, J. K., 1977, "On the Shape of Optimal Tax Schedules," *Journal of Public Economics*, 7: 203-236.

Shell, K., M. Sidrauski, and J. E. Stiglitz, 1969, "Capital Gains, Income and Savings,", *Review of Economic Studies*, 36, 15-26.

Stiglitz, J.E., 1966, "The Distribution of Income and Wealth Among Individuals," Presented at the December 1966 meetings of the Econometric Society, San Francisco , December 29. (MIT, mimeo)

_____,1967, "A Two-Sector Two Class Model of Economic Growth," *The Review of Economic Studies*, 34(2): 73 227-238

_____, 1969a, "Distribution of Income and Wealth Among Individuals," *Econometrica*, 37(3), July 1969, pp. 382-397. (shortened version of a paper presented at the December 1966 meetings of the Econometric Society, San Francisco.)

_____, 1969b, "The Effects of Income, Wealth and Capital Gains Taxation on Risk-Taking," *Quarterly Journal of Economics*, 83(2), May, 263-283.

_____,1975 "The Theory of Screening, Education and the Distribution of Income," *American Economic Review*, 65(3), June, pp. 283-300.

_____, 1976a, "Simple Formulae for the Measurement of Inequality and the Optimal Linear Income Tax," IMSSS Technical Report No. 215, Stanford University, August.

_____, 1976b, "Notes on Estate Taxes, Redistribution, and the Concept of Balanced Growth Path Incidence," *Journal of Political Economy*, 86(2), part 2, pp. 137-150.

_____, 1978, Equality, Taxation and Inheritance, *Personal Income Distribution*, W. Krelle and A.F. Shorrocks (eds.), North-Holland Publishing Company, 271-303. (Proceedings of IEA Conference, Noordwijk aan Zee, Netherlands, April 1977.) "Notes on Estate Taxes, Redistribution, and the Concept of Balanced Growth Path Incidence," *Journal of Political Economy*, 86(2), part 2, pp. 137-150.

_____, 1982a, "Self-Selection and Pareto Efficient Taxation," *Journal of Public Economics*, 17, 1982a, 213-240.

_____, 1982b, "Utilitarianism and Horizontal Equity: The Case for Random Taxation," *Journal of Public Economics*, 18, 1-33

_____, 1985a,"Inequality and Capital Taxation," IMSSS Technical Report 457, Stanford University, July.

57

_____, 1985b, "The General Theory of Tax Avoidance," National Tax Journal, 38(3), September 1985, pp. 325-338.

_____, 1987, "Pareto Efficient and Optimal Taxation and the New New Welfare Economics," in *Handbook on Public Economics*, A. Auerbach and M. Feldstein (eds.), North Holland: Elsevier Science Publishers, 991-1042

_____, 1998, "Pareto Efficient Taxation and Expenditure Policies, With Applications to the Taxation of Capital, Public Investment, and Externalities," presented at conference in honor of Agnar Sandmo, Bergen, January 1998

_____, 2009a, "Simple Formulae for Optimal Income Taxation and the Measurement of Inequality," in *Arguments for a Better World: Essays in Honor of Amartya Sen, Volume I, Ethics, Welfare, and Measurement*, K. Basu and R. Kanbur, eds., Oxford, UK: Oxford University Press, 2009, pp. 535-566. (Revision of IMSSS Technical Report No. 215, Stanford University, August 1976)

_____2009b, *The Selected Works of Joseph E. Stiglitz, Volume I: Information and Economic Analysis*, Oxford: Oxford University Press.

_____, 2012, *The Price of Inequality: How Today's Divided Society Endangers Our Future*, New York: W.W. Norton.

_____, 2015, "New Theoretical Perspectives on the Distribution of Income and Wealth Among Individuals: Parts I-IV" NBER Working Papers *21189-21192*, May.

_____, 2016, "New Theoretical Perspectives on the Distribution of Income and Wealth among Individuals," in *Inequality and Growth: Patterns and Policy, Volume I: Concepts and Analysis*, Kaushik Basu and Joseph E. Stiglitz (eds.), IEA Conference Volume No. 156-I, Houndmills, UK and New York: Palgrave Macmillan.

_____, 2017, "Wealth and Income Inequality in the Twenty-First Century", Paper prepared for presentation at the 18th World Congress of the International Economic Association, Mexico City, June 19-23.

_____, with N. Abernathy, A. Hersh, S. Holmberg and M. Konczal, 2015, *Rewriting the Rules of the American Economy*, New York: W.W. Norton.

Stiglitz, J. E. and B. Greenwald, 2014, *Creating a Learning Society: A New Approach to Growth, Development, and Social* Progress, New York: Columbia University Press. Reader's Edition published 2015.

Stiglitz, J. E. and J. Yun, 2013, "Optimality and Equilibrium in a Competitive Insurance Market under Adverse Selection and Moral Hazard," NBER Working Paper 19317.

Tuomala, M. , 1990, *Optimal Income Tax and Redistribution*, Oxford: Clarendon Press.
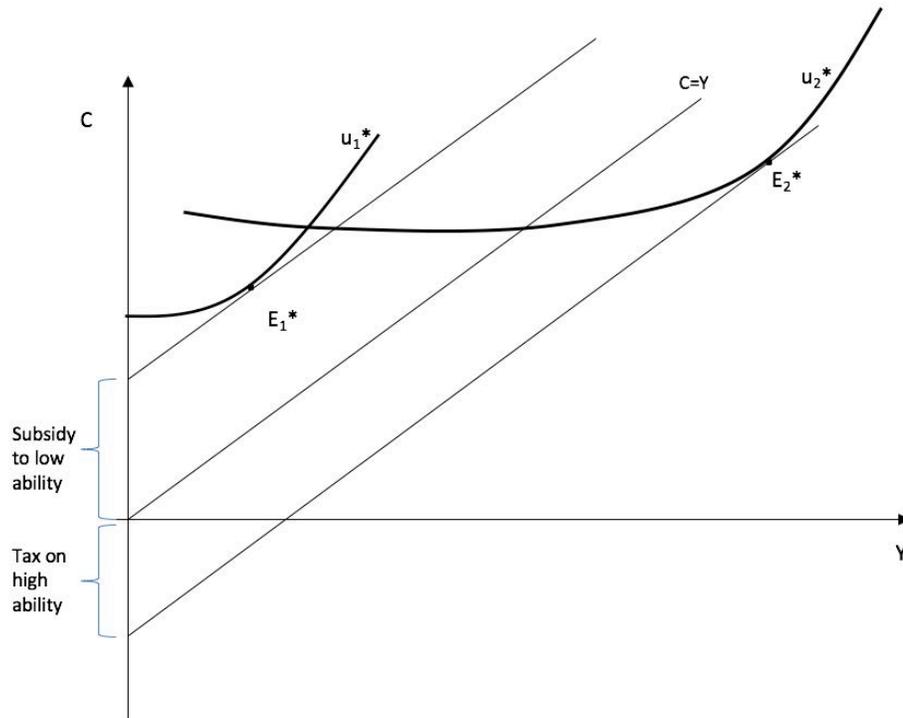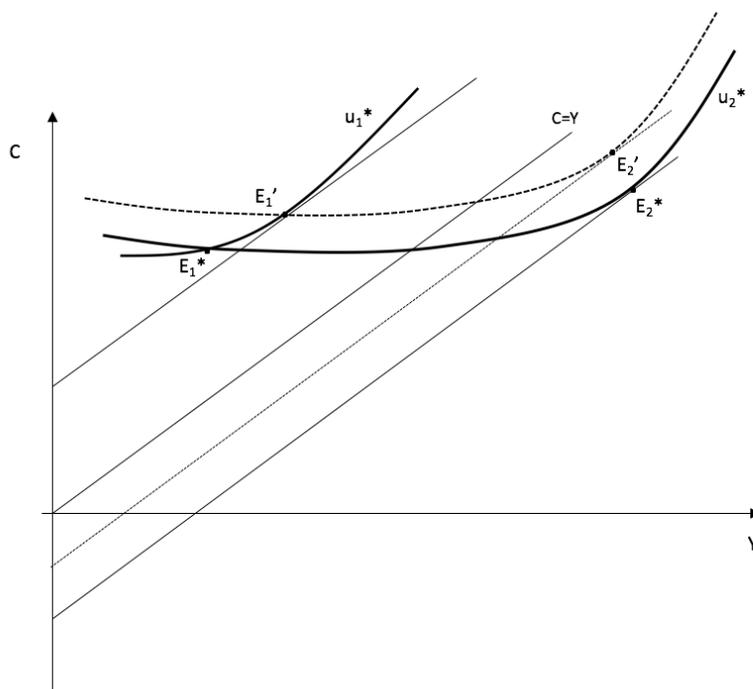
Figure 1(a)

The required revenue $\bar{R}$ can be raised in a Pareto efficient way which gives individual 1 utility $\overline{U^1}$ by a non-linear income tax (dotted line) which confronts both types with zero marginal tax rates. At E_2*, high ability individuals do not want to mimic low ability individuals.

C

$u_1{}^*$ C=Y $u_2{}^*$

$E_1{}'$

$E_2{}'$

$E_1{}^*$ $E_2{}^*$

Y

T

Figure 1(b)

The more normal case: at maximal revenue associated with $\overline{U^1}$, point $E_1{}'$, we can identify the maximal revenue we can extract from high ability, by drawing type 2's indifference curve through $E_1{}'$. The maximal revenue corresponds to point $E_2{}'$. It is insufficient to raise required revenue $\overline{R}$. We thus have to impose a higher tax on 2, lowering his utility, which result in his mimicking 1 if we had left 1 at $E_1{}'$. We force 1 to work less, moving him to $E_1{}^*$, raising less from the low ability types, necessitating an even higher tax on the high ability. At the (Pareto) efficient tax, type 2 is indifferent between his optimal point ($E_2{}^*$) and the low ability's optimal point $E_1{}^*$. The Pareto efficient allocation can be implemented through a tax which is linear with slope equal to that of type 1 at $E_1{}^*$ (i.e. effectively a demi-grant plus a *positive* tax rate), then at some income level above $Y_1{}^*$ but below $Y_2{}^*$, a high (infinite) tax rate, and then above a critical threshold somewhat below $Y_2{}^*$, a zero marginal tax rate. There are many tax schedules that can implement the Pareto efficient allocation. This is not so when there are a continuum of types, for then we have to be sensitive to marginal tax rates at every income level.
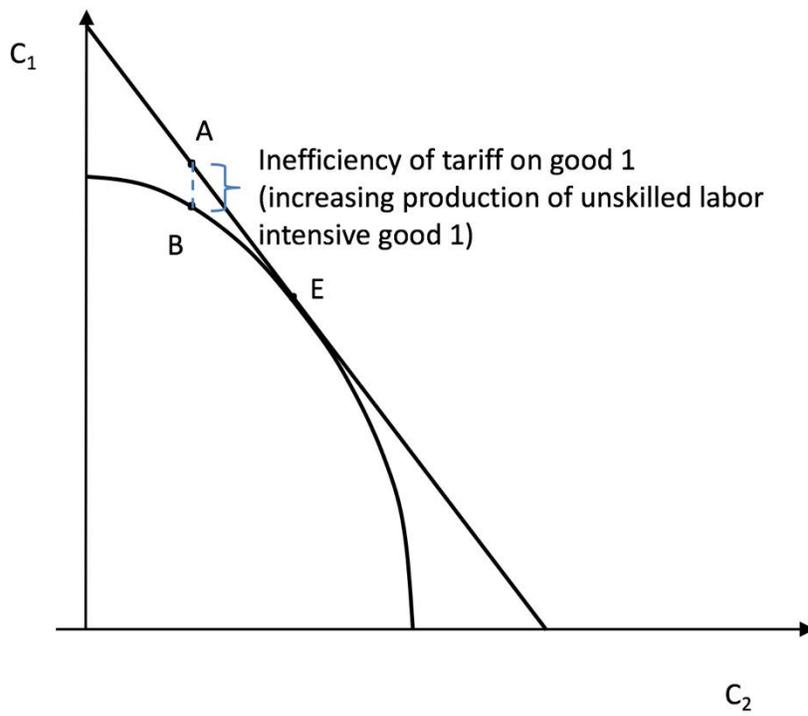
Figure 2

By producing at B rather than at E, the economy can consume less than it otherwise could have. The effective loss in output in terms of good 1 is measured by AB.