NBER WORKING PAPER SERIES

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

Joseph E. Stiglitz

Working Paper 23892
http://www.nber.org/papers/w23892

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2017

Paper written for special issue in memory of Tony Atkinson and dedicated to his memory. It develops ideas on which we worked jointly together, and I am greatly indebted to him. I am also indebted to INET and the Roosevelt Institute, and the supporters of its Inequality Project for financial support; to Matthieu Teachout and to Haaris Mateen for research assistance; to Debarati Ghosh for editorial assistance; and to Simcha Barkai, Martin Guzman and Linus Mattauch for comments on an earlier draft. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Pareto Efficient Taxation and Expenditures: Pre- and Re-distribution
Joseph E. Stiglitz
NBER Working Paper No. 23892
September 2017
JEL No. E2,H2,H41,H52,I24

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, 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 was 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; and with non-separability, capital taxation depends on the complementarity/substitutability of leisure during work with retirement consumption. ¸

The final section considers taxation when there are constraints on the imposition of intergenerational transfers (either political constraints or those derived from unobservability.) It 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

Joseph E. Stiglitz

Beginning with my supervision of Tony Atkinson in Cambridge in 1965-1966 while I was a junior research fellow, Tony and I enjoyed a close collaboration and friendship. One of our early results that received a great deal of attention was hat when there was separability in the utility function between consumption and leisure, 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.

Reflections on the Atkinson-Stiglitz Theorem

This and similar results from optimal tax theory were expanded upon and, taken seriously, seized upon as a basis of policy by those critical of capital taxation.[1] Those who did so typically did not 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 serious policy. Indeed, our analysis was motivated in part as a critique of Ramsey pricing (Ramsey, 1927), which had been used not only to justify high taxes on basic necessities like food (which

[1] See, for instance, Mankiw *et al* (2009), Atkeson *et al.* (1999), Lucas (1990). 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.

had a low elasticity of demand) but also patterns of pricing by monopolies like A T & T.[2] Ramsey had established his result 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 had shown that in the absence of an income tax, optimal commodity taxes looked markedly different from that suggested by Ramsey; and yet we could find a very simple Ramsey-like formula incorporating distributional effects. 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. The key 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.[3]

One 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. Ramsey (1927) and Diamond Mirrlees (1971) analysis 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. If there were two sets of commodities, one which could be taxed and the other which could not, the taxes on those that could be taxed will depend on those that could not, especially so if by taxing some

[2] See, e.g. Baumol and Bradford (1970) and Boiteux (1956). Elsewhere, David Sappington and I (1987) explained while there was some similarities between the two problems, there were also some critical differences.

[3] Subsequently (Stiglitz, 2009), I explored what optimal commodity taxes would look like in the presence of an optimal linear income tax. One doesn't obtain the clean result that Atkinson and I had obtained; but the structure of taxes still looked markedly different from that of Ramsey. See also Kaplow (2006)

commodity, one could tax rents, which should be taxed at 100% (Dasgupta and Stiglitz, 1971, 1972).

(b) 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. Even though there was a single “dimension” in which individuals differed (ability), in general, one could extract information about that difference 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.

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.

Should Capital be Taxed: Beyond the Atkinson-Stiglitz Theorem

And, of course, if individuals differed in other ways—as in fact they do-- there was no presumption that that special result would hold, even if the separability assumption held.

It was particularly clear to us that our theorem did not provide the basis of policy for capital taxation, and it was easy to generate more general models in which capital taxation was clearly desirable.

Thus, assume that there are two classes of individuals—a group of capitalists with large endowments of capital, and workers--with the capitalists sufficiently richer than the workers that the social marginal utility of a dollar to the capitalists is negligible. It is obvious that in such a situation the optimal utilitarian[4] tax on capitalists is close to a Rawlsian tax, maximizing the revenue that one could obtain from a tax on capital. It is also 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. We will explore such a model in the final section of the paper.

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 the income; and that there is disutility associated with the exertion of effort.[5] Then the optimal tax structure

---

[4] 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).

[5] 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.) In each of these cases, the desirability of a highly progressive capital tax is obvious. Here, we make the equally obvious point: even if there is an observable 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

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), similar results would be obtained; 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 depend on both wage and capital income, $T(W, C)$, with the tax function not in general being just a function of $W + C$, total income.

Of course, some individuals might 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 income generated by the assets—then there might be a trade-off, with higher tax rates generating more tax revenue but leading to less wealth accumulation. Even a Rawlsian would impose a tax that was less than 100%. It may be that the actual value of the wealth is not observable, but only the income generated by it. But there is no presumption that the optimal tax (whatever the social welfare function) on the income generated by that wealth would be

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).

taxed at zero. Quite the contrary: there is a presumption that it would be taxed, and possibly at high rates.

Taxing capital goods

Moreover, the result that one does not want to tax the returns to individuals was 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). The price of the asset would fall, to reflect the tax. All our theorem said was that (under the stipulated conditions) one should not tax the returns received by the individual.

Similarly, in a world in which there are imperfect risk markets, Domar and Musgrave (1944)[6] showed that a corporate income tax with appropriate loss offsets could encourage risk taking. A utilitarian government would again want to impose a corporate income tax. The price of shares would reflect the taxes imposed. At most, our analysis *suggested* that one should not tax the returns received by the individual (I say suggested, because we had no formal analysis of risk in our paper.)[7]

Extending Atkinson-Stiglitz towards a General Equilibrium Dynamic Model

[6] 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.)

[7] The design of the optimal tax system in a democracy where views about the appropriate social welfare function differs is a complicated matter, that goes beyond the scope of this paper. See Korinek and Stiglitz (2008)

The previous paragraphs have suggested a range of alternative models in which one could examine the structure of capital taxation. This paper is narrower. It extends and qualifies the Atkinson and Stiglitz results in several ways, but keeps with the assumption that individuals differ only in their abilities. Most importantly, it embeds the results in a simple general equilibrium model where relative wages are endogenous, affected by the supplies of different types of labor and different kinds of capital goods. Because we are interested in part in understanding the taxation of capital, we embed the analysis in an overlapping generations model.

There is a second key set of results, concerning the intertemporal discount rate and the optimal rate of investment for public goods which our general equilibrium intertemporal framework allows us to ask. There has been some controversy over whether the discount rate should reflect the intertemporal marginal rate of substitution or the marginal rate of transformation. Anything other than the latter would suggest a deviation from production efficiency, counter to the spirit of Diamond and Mirrlees (1971). Consistent with the spirit of our analysis of optimal taxation, it turns out the distributive effects of public expenditures needs to be taken into account; in the absence of those effects, in model of overlapping generations, the marginal rate of transformation is equal to the social marginal rate of substitution. Ramsey argued persuasively that there should be no discounting, implying a near zero rate of return for public investment[8] But public investments which improve the distribution of income should be

[8] This ignores the return to risk. Ramsey argued that intergenerational equity required no discounting, but that leads to problems for the maximization of intertemporal welfare. We ignore these issues by having discounting, but possibly at a very low rate.

pushed beyond that point (i.e. to a negative rate of return), and conversely for those that worsen it.[9]

The basic insight in both capital taxation and public investment is one which has become at the center of more recent policy discussions, including that of Atkinson (2015) and Stiglitz *et al* (2015): Tax and expenditure policies can affect the market distribution of income, and by creating a more equalitarian market distribution, there is smaller burden imposed on costly *redistribution.* (Changing the market distribution of income has come to be called *pre-distribution.* See Hacker and Pierson (2011)). Formally, the result comes from observing that taxes and expenditures which change relative wages affect the all-important self-selection constraints which are the source of the distortion in taxation.

Structure of the paper

Section I presents the basic model, reformulating the Atkinson Stiglitz model for two types of individuals and extending the analysis to Pareto efficient taxation. In this framework, we show that, under the conventional assumption of time separability of utility functions, whether capital should be taxed depends not on the properties highlighted by Ramsey (the elasticity of "demand" for consumption later in life) but simply on whether consumption while one is working is complementary to or a substitute for leisure. Section II then shows the results hold with an overlapping generations model. Section III demonstrates that both in taxation and

[9] This paper ignores risk, but it should be obvious that public investments that act as insurance policies, delivering high returns in states of nature where the marginal value of income is high should also have a negative discount rate. (Arrow *et al* , 1995).

public investment, the effects on the market distribution of income have to be taken into account. This implies that even with separability, it may be desirable to tax the returns to capital, and that the long run return to some kinds of public capital might be lower than the social discount rate. The fourth section looks at inherited wealth, and argues that if there is a group of rich capitalists, there is a presumption in favor of a progressive capital income tax, if inequality is large enough. I suggest that this is the real argument behind the taxation of capital.

The Basic Model

We begin our discussion with the simplest possible model, in which there are only two types of individuals, differing in ability, but having the same utility function. The i-th 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) $C_i = w_i L_i$

where C_i = the i-th individual's consumption and L_i = number of hours worked by the i-th individual. Neither w_i nor L_i is observable, but the product, the i-th individual's income, is:

(2) $Y_i = w_i L_i$

The i-th individual receives utility from consuming goods and disutility from work:

(3) $U^i = U^i(C_i, L_i)$

where $\frac{\partial U^i}{\partial C_i} > 0, \frac{\partial U^i}{\partial L_i} < 0$, and U is quasi-concave. Assume the government imposes a tax as a function of income:

(4) $T_i = T(Y_i)$

The individual's consumption now is his income minus his tax payment

(5) $C_i = Y_i - T(Y_i)$

The individual maximizes his utility subject to his budget constraint:

(6) $$\max_{\{C_i, Y_i\}} U^1(C_1) \quad s.t \quad C_i \leq w_i L_i - T(w_i L_i)$$

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

(7) $$\frac{dU^i/dL_i}{dU^i/dC_i} = -w_i(1 - T')$$

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.

(8) $$\frac{1}{w_i}\frac{dU^i/dL_i}{dU^i/dC_i} + 1 = T'$$

T' is the marginal tax rate.

The problem of the government concerned with Pareto efficiency[10] is to maximize the utility of, say, individuals of type 2, subject to (a) individuals of type 1 having at least a given 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.[11]

(9) $$\max_{\{C_1, C_2, Y_1, Y_2\}} V^2(C_2, Y_2)$$

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

[10] The notion of Pareto efficient tax structure is a slight generalization of the "optimum" tax analysis of Mirrlees and Diamond and Mirrlees. We identify properties of the tax structure which hold regardless of the social welfare function. See, e.g. Stiglitz (1982a, 1987) and Brito *et al* (1990).

[11] 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.

(11) $V^2(C_2, Y_2) \geq V^2(C_1, Y_1)$

(12) $V^1(C_1, Y_1) \geq V^1(C_2, Y_2)$

And the revenue constraints:

(13) $R = (Y_1 - C_1)N_1 + (Y_2 - C_2)N_2 \geq \bar{R}$

(where $\bar{R}$ is government revenue, R is the revenue requirement, and N_i the number of individuals of type i). Consumption may be interpreted as a vector, if we write the budget constraint as:

(13') $R = N_1Y_1 + N_2Y_2 - \sum_{i=1}^{2} \sum_j \left(C_{ij}N_i\right) \geq \bar{R}$

Where C_{ij} is the i-th individual's consumption of the j-th consumption good, and accordingly, to get total consumption, we sum over all j, and where we have chosen our units so a unit of each commodity costs the same to produce. 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 the Lagrangian is

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

Then the first order conditions are:

(15a) $\frac{d\Lambda}{dC_{1j}} = \mu \frac{dV^1}{dC_{1j}} - \lambda_2 \frac{dV^2}{dC_{1j}} + \lambda_1 \frac{dV^1}{dC_{1j}} - \gamma N_1 = 0$

(15b) $\frac{d\Lambda}{dC_{2j}} = \frac{dV^2}{dC_{2j}} + \lambda_2 \frac{dV^2}{dC_{2j}} - \lambda_1 \frac{dV^1}{dC_{2j}} - \gamma N_2 = 0$

(15c) $\frac{d\Lambda}{dY_1} = \mu \frac{dV^1}{dY_1} - \lambda_2 \frac{dV^2}{dY_1} + \lambda_1 \frac{dV^1}{dY_1} + \gamma N_1 = 0$

(15d) $\frac{d\Lambda}{dY_2} = \frac{dV^2}{dY_2} + \lambda_2 \frac{dV^2}{dY_2} - \lambda_1 \frac{dV^1}{dY_2} + \gamma N_2 = 0$

Some preliminary remarks

Before proceeding with the analysis, two preliminary remarks are in order. First, 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.

Secondly, the standard case of *optimal utilitarian tax policy*, where the government maximizes $\sum N_i V^i$ implies that $\mu = 1$.[12]

Basic conditions of Pareto efficient taxation

From (15a-d) we obtain

(16a) $\frac{dV^2/dC_{2j}}{dV^2/dC_{2k}} = 1,$ (16a') $\frac{dV^2/dC_{2j}}{dV^2/dY_2} = 1$

(16b) $\frac{dV^1/dC_{1j}}{dV^1/dC_{1k}} = \frac{N_1\gamma + \lambda_2 dV^2/dC_{1j}}{N_1\gamma + \lambda_2 dV^2/dC_{1k}}$

Equations (16a) and (16a') yield the familiar result that *There should be no distortionary taxation on the individual with the highest ability.* Note that this result goes beyond that of Mirrlees [1971] which said that the marginal income tax rate should be zero. (16a) says that there should be no commodity taxation as well.[13]

Taxation of lower ability individuals

The interpretation of (16b) is, however, somewhat more subtle. For simplicity, denote

$$\alpha_{jk}^{i} = MRS_{ijk}(C_1, Y_1)$$

i.e., i's marginal rate of substitution between j and k at the bundle $\{C_1, Y_1\}$. We know that high ability individual's MRS at $\{C_2, Y_2\}$ is 1 (from (16a) and that leisure for these individuals is higher

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

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

at Y_1 than at Y_2. Individual 1 and 2 differ only in their abilities, so where α_{jk}^1 > = or < 1 simply depends on how leisure affects the MRS. In the case of separability, α_{jk}^2= 1.

Let

$N_1\gamma = a$

$\lambda_2 \frac{\partial V^2}{\partial C_{1k}} = b$

Then (16b) can be rewritten

$$\alpha_{jk}^1 = \frac{a + b\,\alpha_{jk}^2}{a + b}$$

It immediately follows that α_{jk}^2 > = < 1 as α_{jk}^1 > = < α_{jk}^2. In the case of a separable utility function, there is no effect, so that α_{jk}^1 = α_{jk}^2 = 1, that is *There should be no commodity taxation if leisure and consumption are separable.* Note that in this case, we have established that there should not be commodity taxation on either the more or less able individuals. *This result does not depend on any of the properties upon which Ramsey, and the subsequent Diamond- Mirrlees and Boiteux literature focused*. 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--again a property that has nothing to do with Ramsey's focus on deadweight loss.

The intuition

Ramsey analyzed the impact of commodity taxes in creating distortions, beginning from a situation in which the economy was efficient. But the fundamental issue is that redistributive income taxes create large distortions, as we have seen: low income individuals are induced to curtail their work; the self-selection constraints which come into play if there is to be redistribution in an environment in which governments cannot tell who is able and who is not, create first order distortions. Commodity taxation can, under certain conditions, reduce the force of those distortions.

Assume that only more able individuals use pencils in their work, and that pencil consumption is observable. If taxpayers didn't realize this, then the government could use pencil consumption as a basis of taxation—recognizing that pencil consumption is a signal of ability. This is true even if taxpayers realize this. Assume the government provides a subsidy on pencil consumption,[14] up to some level (below that which corresponds to the efficient level of pencil consumption by high ability individuals.) Then the high ability's indifference curve becomes flatter--he needs less of an increase in consumption to compensate him for an increase in income (decrease in leisure). The shift up in the indifference curve weakens the self-selection constraint (makes it less binding), so that the utility of the less able can be increased.

In our model, both the more and less able individuals have the same preferences, so matters are somewhat more subtle. Assume that the marginal rate of substitution between pencils and food changes systematically with the amount of leisure. Assume we can impose non-linear consumption taxes and subsidies, and that we tax pencils and use the proceeds to subsidize food, at levels of consumption that exceed the levels of consumption of the low ability. The low ability pay no commodity taxes. In the case of non-separability, a small revenue neutral tax/subsidy which leaves utility at the point selected unaffected (the dead weight loss

[14] Offset by an increase in the implicit lump sum tax on upper income individuals; in this example, only high-income individuals buy pencils, and the pencil tax is only an infra-marginal tax.

depends on the square of the tax/subsidy) can still affect the marginal rate of substitution between "total consumption" and leisure, and as such, it can have a *first order* effect on the low ability individual's utility (the intersection between the high ability individual's indifference curve and the budget constraint augmented by the transfer) just as in the earlier case where only the high ability individual consumes a particular commodity.

Indeed, this result is more general: even if we are restricted to linear commodity taxes, so long as consumption patterns differ between the high and low ability individuals at each level of after tax income (which they will in the case of non-separability) there exists combinations of commodity taxes and subsidies which leave the upper income individual at the same level of utility, but weaken the force of the self-selection constraint, and thereby increase the welfare of the low ability individual.

Finally, note that dividing (15a) by (15c) we obtain

(17) $$\frac{dV^1/dC_{1j}}{dV^1/dY_1} = \frac{N_1\gamma + \lambda_2 dV^2/dC_{1j}}{N_1\gamma + \lambda_2 dV^2/dY_1}$$

Implying that *in general, the labor supply of the less able is distorted.*

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 most of the simplified models explored in the 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,

(18) $U = u_1(C_1, L) + u_2(C_2)$

where there is time separability, but not separability between consumption and leisure. Then the marginal rate of substitution between second period consumption and first period consumption is

(19) $$\frac{\partial u_1/\partial C_1}{\partial u_2/\partial C_2}$$

Whether there should be a tax or a subsidy depends on whether an increase in leisure increases or decreases the marginal utility of consumption in the first period, i.e. on whether leisure and consumption are complements or substitutes.[15] 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). [16]

II. Extension to Many Periods with Overlapping Generations

[15] This result is strongly reminiscent of that of Collette and Hague (1953), though the reasoning is very different.
[16] It should be clear from the structure of the analysis that if individuals supplied labor at different dates, so $U^i = U(C_1, C_2, \dots C_T, L_1, L_2, \dots L_T)$, but the utility function remained separable between consumption and leisure (work) $U^i = u^i (C_1, C_2, \dots C_T) + v^i (L_1, L_2, \dots L_T)$ 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.

We now modify the above analysis by assuming that the wage rate and interest rate are determined endogenously. They depend on capital accumulation (the saving decision) and on the labor-supply decision, both of which are affected by tax policy. 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 i-th type in the t-th generation is denoted by

(20) $$V_t^i(C_{it,t}, C_{it,t+1}, L_{i,t})$$

where $C_{it,t}$ is the t[th] generation's consumption during the first (working) period, and $C_{it,t+1}$is the t[th] generation's consumption during the second (retirement) period of this life. We reformulate our model to let the government take { C_{it}, L_{it} } as control variables, and as before, we will embed the self-selection constraints into the maximization problem. We now write the constraint as

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

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

(22) $$\hat{L}_{2t} = \frac{w_{1t}L_{1t}}{w_{2t}} = vL_{1t}$$

where ν is the ratio of the productivity of the low ability individual to the high ability individual, assumed now fixed.

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

(23) $$Q_t = F\ (K_t, E_t)$$

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

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

where we have assumed that capital wears out each period. Forming the Lagrangian as before[17], but with separate constraints for each period (including for the self-selection constraints (21) and the resource constraint (24) which replaces the budget constraint), and differentiating with respect to the consumption vector, we obtain

(25a) $$\frac{dV^i/dC_{it,t}}{dV^i/dC_{it,t+1}} = \frac{\gamma_t}{\gamma_{t+1}}$$

and

(25b) $$\frac{dV^1/dC_{it,t}}{dV^1/dC_{it,t+1}} = \frac{N_1\gamma_t + \lambda_2 dV^2/dC_{1jt,t}}{N_1\gamma_t + \lambda_2 dV^2/dC_{1jt,t+1}}$$

where γ_t is the Lagrange multiplier associated with the t-th period government budget constraint (24).

By differentiating the Lagrangian with respect to K_t, we obtain

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

It thus follows that the high ability individual's marginal rate of substitution between current and future consumption should equal the marginal rate of transformation: *there should be no interest income tax*.

---

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

In steady state, $\frac{\gamma_t}{\gamma_{t+1}} = (1 + F_K(K^*, E^*)) = r^*$, and $\frac{\lambda_{2,t}}{\lambda_{2,t+1}} = r^*$, implying that the RHS of (25b) is a constant.

All of the results obtained in the partial equilibrium static model remain valid in the steady state of the overlapping generations model, where both wages and interest rates are endogenously determined; out of steady state the marginal rate of substitution between consumption in time t and time t + 1 for the low ability individual may vary with the shadow price on the self-selection constraint (relative to that on the resource constraint.)

Note that, in the optimal tax-cum-growth version of this problem, with an additively separable social welfare function, where the welfare of future generations are discounted at the rate δ ($\mu_{i,t} = \delta^t$) , i.e. the government maximize $\sum_t \delta^t (N_1 V_t^1 + N_2 V_t^2)$, in steady state $\frac{\gamma_t}{\gamma_{t+1}} = 1/\delta$, so (25) and (26) imply that the economy with distortionary taxation still converges to the modified golden rule, where the rate of interest equals the pure rate of *social* time preference. With separabiltiy, individual's MRS is equal to that rate, but not necessarily otherwise.

Proposition 1A (i) At every date, upper income groups should face no commodity or marginal income taxation. (ii) At every date, with separability, lower income groups should face no commodity taxation, but without separability they should. (iii) At ever date, there will be distortionary income taxes imposed on the lower income groups.

But within the life-cycle model, we can establish two propositions for the dynamic equilibrium:

Proposition 1B. With (and, in general, only with) separability, there should be no interest income tax or subsidy.

The high ability individual's marginal rate of substitution between current and future consumption should equal the marginal rate of transformation; as before, the low ability individual should have a tax or a subsidy on interest depending on how the (high ability individual's) MRS between consumption in the two dates is affected by leisure. The results obtained in the partial equilibrium static model remain valid in the overlapping generations model, where both wage levels (but not relative wages) and interest rates remain endogenously determined. In the optimal tax-cum-growth we have also established

Proposition 1C. *In steady state of the optimal tax problem, with, and, in general, only with separability, the rate of interest (both the producer and consumer rate of interest) equals the pure rate of time preference.*

III.A Public and Private Investment

It is easy to extend the analysis to incorporate public and private investment. Let K_g represent public investment, K_p private investment, and let the production function now be

$$Q = F(K_g, K_p, E)$$

It immediately follows that

$$F_{K_g} = F_{K_p}$$

Proposition 2A. *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.

Public Consumption

The analysis so far assumes that public expenditures only affect consumers indirectly, through the impact on aggregate output. But public investments may yield services directly to consumers, in which case we need to rewrite the utility functions

$$V_t^i = V^i(C_{t,t}, C_{t,t+1}, L, \boldsymbol{K}_{g,t}, \boldsymbol{K}_{g,t+1})$$

where it is important to note that here $\boldsymbol{K}_{g,t}$ is a vector of capital goods that generate public consumption[18]. Now, upon taking the derivative of the Lagrangian with respect to public investment we obtain two sets of *added* terms, the direct impact on utility, and the indirect impacts on the self-selection constraints:

$$\sum_i \mu_{i,t} V^i_{t,\boldsymbol{K}_{g,t}} + \sum_i \mu_{i,t-1} V^i_{t-1,\boldsymbol{K}_{g,t}} + \lambda_{2,t}\left(V^2_{t,\boldsymbol{K}_{g,t}} - \hat{V}^2_{t,\boldsymbol{K}_{g,t}}\right) + \lambda_{2,t-1}\left(V^2_{t-1,\boldsymbol{K}_{g,t}} - \hat{V}^2_{t-1,\boldsymbol{K}_{g,t}}\right)$$

where $\hat{V}^2 = V^2(C_{1t,t}, C_{1t,t+1}, \hat{L}_{1,t}, \boldsymbol{K}_{g,t}, \boldsymbol{K}_{g,t+1})$, i.e., the high-ability individual's utility evaluated at the low-ability individual's "bundle." Again, with separability, all self-selection terms are zero.

[18] That is the consumption of the public good $G_t = K_{g,t+1}$, the investment made for it at t-1, and aggregate output at time t, Q_t, can be used either for consumption or investment in public or private goods. The public good here is a pure public good, affecting everyone's utility; but the analysis applies equally to the case of publicly provided private goods (Atkinson and Stiglitz, 1980). Similar results hold if the vector $\boldsymbol{K}_{g,t}$ yields consumption directly.

Proposition 2B. With separability, Pareto efficient public expenditures require the return on public investment (including the return in terms of consumption services rendered by the public capital good) to be equal to the return on private investments.

However,

Proposition 2C: Without separability, public investments need to take into account the impact on the self-selection constraints.

By curtailing investment in public capital which is a complement to leisure,[19] the {consumption, leisure} package of the less able becomes less attractive to the more able; the self-selection constraint is thus relaxed, and welfare is thus enhanced.

III.B Atmospheric Externalities

Still another generalization that is easy to incorporate into the model is atmospheric externalities[20], 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:

$$V^i = V^i(C, L, B_t, B_{t+1}) \quad \text{where } V^i_{B_t} < 0 \textit{ and } V^i_{B_{t+1}} < 0$$

and

$$Q = F(K, L, B) \qquad \text{where } F_B < 0$$

Then we have

---

[19] In the sense that $\frac{dV_{K_g}}{dL} < 0$

[20] An atmospheric externality is an externality the value of which depends on aggregate consumption. In the notation below C_t is aggregate consumption of all commodities at date *t.*

$$\frac{d\Lambda}{dC_{it,t}} = [\ldots] + \left\{ \underbrace{\sum \mu_{it} \frac{dV_t^i}{dB_t} + \sum \mu_{it-1} \frac{dV_{t-1}^i}{dB_t}}_{A} + \underbrace{\gamma_t \frac{dF}{dB_t}}_{B} + \underbrace{\lambda_{2t}\left(V_{tB}^2 - \hat{V}_{tB}^2\right)}_{C} + \underbrace{\lambda_{2t}\left(V_{t-1B}^2 - \hat{V}_{t-1B}^2\right)}_{D} \right\} \frac{dB}{dC_{it,t}}$$

where […] represents the derivative of the Lagrangian in the absence of externalities. There are four additional terms: {A} represents the direct consumption externality; {B} represents the production externality; {C} represents the impact on the self-selection constraint of those who are young at the time the externality occurs; and {D} represents the impact on the self-selection constraint of those who are old at the time the externality occurs.

In the case of a separable utility function, where

$$V^i = V(C, L) - Z(B),$$

the third and fourth terms disappear, and there are only the direct externality effects.

Proposition 3A. If the impact of an atmospheric externality is separable from private consumption and leisure, then a straightforward Pigouivan tax/subsidy is optimal. In other words, the complicated formulae underlying the literature on corrective taxation that focus on interactions between the corrective taxes and other taxes are not relevant.

On the other hand, if the utility functions are not separable, simple Pigouvian corrective taxation is not appropriate, but the deviations do not depend on indirect revenue effects, but on the impacts on self-selection constraints. For instance, 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.

Proposition 3B. With non-separable utility functions, Pigouvian corrective taxation has to be modified. The deviations from the simple Pigouvian formula do not depend on the second order effects on government revenue but on the first order effects on the self-selection restraints.

IIIC. Distributive effects of tax policy

The analysis so far has done three things: it 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; 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, the first set of results are not robust*; in more general models, the consequences of tax and expenditure policies for the before tax distribution of income cannot be ignored*. In this section, we extend the life cycle model of the previous section to include distributive effects of taxation.

First, we assume that the two types of labor are not perfect substitutes, i.e., the aggregate production is of the form

(28) $$Q_t = F(K_t, L_{1t}, L_{2t})$$

with $$w_{i,t} = F_{L_{i,t}}$$

Now, relative wages depend on government policies. If we assume constant returns to scale

(29) $$\nu_t = \frac{w_{1t}}{w_{2,t}} = \Phi\left(\frac{L_{1t}}{K_t}, \frac{L_{2t}}{K_t}\right)$$

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 (i.e. (25a) and (25b)), but now there is an extra term in each equation to reflect the effect on relative wages on the self-selection constraint. Because $\partial V_t^2/\partial L_{2t} < 0$, anything that increases ν, i.e. the wage of the low skilled worker relative to the high skilled worker, loosens the self-selection constraint and increases welfare. Thus, in the case of K,

(30) $$\gamma_{t+1}(1+F_K) = \gamma_t + \lambda_{t+1}^2 L_{1t} \frac{dV^2}{dL_{it}}\left(C_{it+1,t+1}, C_{it+1,t+2}, \nu L_{it+1}\right)\frac{d\nu}{dK}$$

It thus immediately follows that:

Proposition 3A: If the level of capital affects relative wages, the marginal rate of substitution is not equal to the marginal rate of transformation. It is optimal to have an interest income tax or subsidy.

Increases in capital change wage inequality. We impose a tax on interest income, if an increase in K increases wage inequality. The reason we do so is related to how the self-selection constraints are affected, not to how the income distribution is affected directly.

Whether there should be a tax or 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.[21] Thus, if an increase in capital accumulation increases the wage

[21] Conditions under which either result obtains may easily be derived, simply by differentiating

differential (if capital is a complement to high skill labor and a substitute for low skilled workers) the optimal level of $1 + F_k$ (assuming separability in consumption) consumers' discount rate.

Moreover in the optimal growth-cum-distribution interpretation of the model, the economy does not converge to the standard modified golden rule; though the consumer rate of discount converges to δ, the producer rate is greater or less than δ as increasing K decreases or increases v, the relative wage.

III.D. More General Models

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

(31) $$\mathfrak{J}(C_1, C_2, L_1, L_2, K_{t+1}, K_t) = 0$$

and $v = \frac{\mathfrak{J}_{L_2}}{\mathfrak{J}_{L_1}}$

(31) 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 ((16a) and (16b)), which we have shown hold in the more general overlapping generations model, has to be changed; it follows that

Proposition 3B. 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*

$\frac{F_{L_1}\left(\frac{N_1 L_1}{K}, \frac{N_2 L_2}{K}\right)}{F_{L_2}\left(\frac{N_1 L_1}{K}, \frac{N_2 L_2}{K}\right)}$ with respect to K

rate of substitution, (ii) the rate of return in the public sector should not equal that in the private one, and (iii) even asymptotically, it will not equal the pure rate of time preference. 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 consume different commodity bundles; if the elderly on average consume more skilled labor-intensive services (e.g. expensive medical procedures), then we want to encourage individuals to consume more later in life, because that drives up wages, 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. (As we explain in the next section, we do not believe that this model captures accurately the most important aspects of capital taxation.)

Similarly, the optimal Pigouvian tax formulae have to be corrected for induced wage effects: the externality may have an effect on relative wages, and thus on the self-selection constraints. The required modifications in the optimal formulae are straightforward.

IV. Inheritances

The most critical assumption underlying Atkinson and Stiglitz 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. Much of the argument for capital taxation is that taxation of these 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. Here we formulate two different models. The first explains why a straightforward extension of the two group model of Atkinson and Stiglitz is of limited help in thinking about these issues. The second presents an alternative which we believe captures more of the underlying rationale for capital and inheritance taxation.

A. Overlapping generations model with a bequest motive

Assume that each generation cares only about its descendants, and for simplicity, only about the direct utility (i.e. not about the descendant's utility from its descendant's descendants.[22]) Then the utility of the t^th generation (for an individual of any type; we assume types are perfectly inherited) is $W^t = V^t + \delta V^{t+1}$. We assume that the individual gets directly pleasure out of giving, and thus write $V^t(\boldsymbol{C}_t^t, \boldsymbol{C}_t^{t+1}, L^t, \boldsymbol{b}^t)$, where the individuals life-time budget constraint obviously depends on $\mathbf{b}^{t-1} - \mathbf{b}^t$ the difference between what he inherits and what he bequeaths. The final term represents his direct pleasure from giving. The individual chooses b^t so that (in the absence of taxation) $-V_I^t + V_b^t + \delta V_I^{t+1} = 0$, or in steady state, where $b^t = b^{t-1}$, and $(1 + r) = 1/\delta$, $V_b^t(\boldsymbol{C}_t^{t*}, \boldsymbol{C}_t^{t+1}, L^t, \boldsymbol{b}^*) = (1 - \delta)V_I^t$. Bequests go to the point where the marginal benefit of giving is $(1 - \delta)V_I^t$, and individuals pass on just the amount that they receive.

A Benthamite social welfare function maximizes $\sum W^t \delta_s^t$. (There is no reason that the social discount rate should be the same as that of the individual.) If the government could directly control b, and maximized steady state utility, V, it would recognize that in steady state, inheritances equal bequests, and set $V_b^t(\boldsymbol{C}_t^{t*}, \boldsymbol{C}_t^{t+1}, L^t, \boldsymbol{b}^*) = 0$ i.e. at a higher level than the individual does. If it maximized the steady state value of $(1 + \delta)\, V$, i.e. utility as perceived by

[22] It is straightforward to extend the analysis to other cases.

each generation, taking account of the utility it gets from the well-being of future generations, then, of course, the solution is the same.

More generally, bequests, in a Benthamite social welfare function, give utility twice, both to those giving and those receiving. Moreover, the size of the bequest may differ from the size of what is received—the difference being the inheritance tax. Hence, denoting the bequests received by β, $\boldsymbol{b}$ and β are chosen so that

(32a) $V_b^t(\boldsymbol{C}_t^{t*}, \boldsymbol{C}_t^{t+1}, L^t, \boldsymbol{b}^*) - V_I^t + \gamma^t = 0$

and

(32b) $(\delta + \delta_s)V_I^{t+1}(\boldsymbol{C}_t^{t*}, \boldsymbol{C}_t^{t+1}, L^t, \boldsymbol{b}^*) - \gamma^{t+1} = 0$

In steady state

(33) $V_b^*(\boldsymbol{C}_t^{t*}, \boldsymbol{C}_t^{t+1}, L^t, \boldsymbol{b}^*) - V_I^*\left[1 - \frac{\delta+\delta_s}{\delta_s}\right] = 0,$

So, for instance, if $\delta = \delta_s$, $V_b^* = -V_I^*$, i.e. the level of bequests is greater than in the "free market" solution—and greater than the level that maximizes steady state utility V.

Notice that in our world with two groups, low ability and high ability, the values of $\boldsymbol{b}$ and β will differ between the two groups if the government could costlessly distinguish between them. But the whole focus of this paper is that the government cannot distinguish between the two, except through the self-selection constraints. Thus, bequests and inheritances need to become part of the "package" confronting the two groups among which they must choose. The main distinction between bequests and other expenditures is that noted, from a social welfare perspective, they increase social welfare doubly, because of the effect they have on both the giver and the receiver.

B. Inheritances as a source of inequality

Since one of the major differences across individuals is what they receive from their parents in the form of either financial or human capital, a full analysis of taxation and inequality has to come to terms with the drivers of financial and human capital bequests—the determinants of the intergenerational transmission of advantage and advantage.[23] While it is natural that economists begin with a discussion of *choices and tastes* (some individuals or dynasties might give greater weight to their descendants than others),[24] almost surely more relevant, at least in the extremes of inequality, are differences in "lifetime wealth," the sum of what one inherits and what one "earns": those with very high lifetime wealth tend to give large bequests. And there are two sources of high levels of lifetime wealth: children who have the good fortune of having parents who bequeath substantial amounts of that wealth to their children, and those who happen to have high earnings themselves. High life time earnings, in turn, are related to inherited ability, human capital, effort, and a large dose of lucks; those include the inventors and entrepreneurs who earn outsized incomes, often through the exploitation of market power. Gates and Rockefeller are examples. In many emerging markets (and earlier periods in the US), it is those individuals who were particularly effective in receiving favors from the state, e.g. in the form of land grants. 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, there is still a third source of differences in inherited wealth

[23] See Stiglitz (1969a, 2015, 2017) and Becker and Tomes (1979, 1986) for a more extensive discussion.
[24] See, for instance Piketty and Saez (2013).

(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.[25]

In none of these cases is there any substantial evidence that taxing at a high marginal rate such income would have significant adverse effects on their economic activity. Indeed, there is some positive evidence and theory that it might reduce incentives for rent seeking and exploitation, thereby increasing economic efficiency.[26]

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) and would consume more.[27] These concerns limit the taxation of bequests, implying that even with bequests taxes, there will be individual with substantial incomes from inherited capital. Thus, taxation on capital (at high incomes) can be viewed as an indirect (and imperfect) way of taxing bequests.[28] Other recipients of high capital income are those who have been very lucky in their investments, including some who have put a great effort in this work. It is their elasticity of effort with which public policy should perhaps be most concerned; but there is little

[25] This theory supports a Pareto tail distribution. See Stiglitz (1978) and Flemming (1979).
[26] See Stiglitz (2017) and Saez, Piketty, and Stantcheva (2014).
[27] 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.
[28] See Cremer and Pestieau (2006) and Cremer, Pesteiau and Rochet (2003) Farhi and Werning (2010), Piketty and Saez (2013), and Stiglitz (2017).

evidence that curtailing some of the "unexpected high end" returns would result in significant reductions in entrepreneurial efforts. The final group of those with high capital income are those upon whom the rest of this paper has focused, lifetime savers with high wages.

With one exception, 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.[29] The one possible exception to this perspective is that explored by Bevan and Stiglitz (1979) and Bevan (1979), where individuals themselves think about their children and, when they have good luck, 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. It might seem that so long as the family's intertemporal discount rate is the same as the social welfare function[30], 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. 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

[29] 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 each 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.

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

returns on capital. A utilitarian would wish to redistribute income between the lucky individual and those who had a bad roll of the dice. If the government can't perfectly tell whether the high wages or high returns to capital are the result of inherited abilities or human capital provided by parents (or connections, or any of the multiple other advantages that can be transmitted across generations) or effort and hard work, then the redistributive taxes would have to take into account the self-selection constraints described in previous sections of the paper. In effect, the model in section 2 was a special case of this model, where there was perfect inheritability of talents. The more general case is that where there is regression towards the mean, with imperfect inheritability: A government looking for a Pareto Efficient Tax system would engage in precisely the kind of analysis described earlier in this paper[31], with both an income and a capital tax.

C. A Two Class Model

A simplification 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 group as workers. 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, generalizing Mattauch *et*

[31] Though in the earlier analysis individuals differed only in their earnings ability, here, they may differ also in the amount of inherited wealth and the ability to generate earnings out of capital. In the formal model presented below, however, we assume all workers are identical, all capitalists are identical, and everyone receives the same (before tax) return on capital.

al (2016) and Mattauch *et al* (2017).[32] All workers and capitalists are the same (except different capitalists may inherit more money)[33]. Capitalists maximize a standard intertemporal utility function with discount δ. We normalize by assuming workers have a (fixed) population of unity, capitalists of N, so capitalists' consumption is N C_c. We simplify the analysis by assuming the government simply wishes to maximize workers' steady state consumption.[34] We assume, for simplicity, that the labor supply is fixed (at unity as well), and all revenues from taxation are spent on education. Then output, Q, is a constant returns to scale function of capital, K, and effective labor:

(34) $Q = F\left(K, \phi\left(K_g\right)\right)$

where $\phi\left(K_g\right)$ describes the increased productivity from public education K_g on labor, i.e. $\phi\left(K_g\right)$ is the effective labor supply, with $\phi'\left(K_g\right) \geq 0$ and $\phi''\left(K_g\right) \leq 0$. Investments in education increase productivity, but there are diminishing returns. Because of constant returns to scale in capital and effective labor, we can write

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

where $k = \frac{K}{\phi}$, the capital-"effective labor" ratio. We assume a tax rate of τ on the net return to capital of capitalists but no taxation on the stock or returns of capital of workers. Current tax

[32] Their models used logarithmic utility functions for both capitalists and workers. Mattauch *et al* 2016 did not analyze as here investments in education. Mattauch *et al* 2017 focused on the issue when the tax rate on capital was sufficiently high that all the capital was held by workers.

[33] In the formal model presented below, we do not discuss inequalities among capitalists or among workers. With stochastic wages and returns to capital, these can be analyzed along the lines of Stiglitz (2015). Thus we focus here on differences between groups, rather than differences within each group. As discussed in the previous section and in footnote 5, Pareto efficient taxation within each group can be analyzed in standard models such as those presented in earlier sections of this paper.

[34] 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.

law in many countries differentiates between the return to capital of workers and capitalists, e.g. through "IRA" accounts where the returns to capital are not taxed. In our model, only workers would be allowed to make contributions to such accounts.[35] This implies that in long run equilibrium k is given by

(36) $(1-\tau)(f'-\eta)=\delta$

where η is the rate of depreciation. The after tax net return to capital is equal to the pure rate of time discount. (36) determines the equilibrium value of k simply as a function of τ (given η and δ).[36]

(37) $k=\psi(\tau)$

Capitalists' capital, K_c is the difference between the total capital stock, K, and workers' capital, K_w, so tax revenue is $\tau\,(f'-\eta)(K-K_w)$, all of which is spent on public education. Hence

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

using (36), giving K_g as a function of just τ and K_w. Assuming workers have homothetic indifference curves[37], worker savings are given by $s_w(r)\,w$, where r is the (before tax) rate of return on capital—the return received by life cycle savers:

(39a) $r = (f'-\eta)=\frac{\delta}{1-\tau}$

and w is the wage, where

(39b) $w=\phi(K_g)\,g(k)=\phi(K_g)\,(f-kf')$

[35] Again, similar results obtain if one cannot 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.
[36] 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.
[37] If they don't, then $K_w = S(w,r)$, and the analysis proceeds much as below.

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

(40) $K_w = s_w(r)\,\phi(K_g)\,g(k) = s_w\left(\frac{\delta}{1-\tau}\right)\phi(K_g)\,g(\psi(\tau)) = H(\tau,\ K_g)$

using (37): K_w is just a function of τ and K_g. For any value of τ, we can solve (38) and (40) for the long run equilibrium values of $\{K_g, K_w\}$. More directly, substituting (40) into (38), using the definition of $k = \frac{K}{\phi}$, we obtain

(41a) $K_g = \frac{\tau\delta}{1-\tau}\,\phi(K_g)\left(\psi(\tau) - s_w\left(\frac{\delta}{1-\tau}\right)g(\psi(\tau))\right) = \frac{\tau\delta}{1-\tau}\,\phi(K_g)k\,\xi_K.$

where ξ_K = capitalists' share of the capital stock $(1 - K_w/K)$. We can also solve for workers steady state utility,

(41b) $V^w = V^w(w, r)$

which depends just on $\{w, r\}$, which from (37) (39b), (40) and (41a) depends just on τ. Differentiating logarithmically (39b), using (37) and (41a), it is easy to show that taxing capital to invest in education increases wages *if* ϕ' is large enough. The direct effect of the tax is to lower k, the effective capital labor ratio, and thus lower the return per unit of effective labor. At the same time, it increases the return on capital of workers[38], so that at the optimum, wages are actually declining:

Proposition 4: Provided $g(k*)\phi'(0) > 1$, it is always desirable to tax the return to capital of capitalists,

[38] We make use of the fact that (in what is call Roy's formula) $\partial V^w/\partial r = V_I K_w > 0$ and $d w/d\tau_{|\tau=0} = -\delta[\phi k - g K_c \phi']$, so $dV_w/d\tau = -\delta K_c[1 - g\phi']$.

where k^* is the capital labor ratio prevailing with $\tau = 0$. (This will always be the case when productivity with no education is very low, with initial investments in education yielding very high returns.)

We can get a general formula for the optimal tax simply by differentiating V_w with respect to τ. (Recall that we have an inequality averse social welfare function, with a gap between workers and capitalists sufficiently large that we place essentially no weight on the welfare of the capitalists. The only impediment to imposing a higher tax on capitalists is the adverse effect on wages). We can show that that if τ is not too large (ignoring, in other words, terms of order τ^2 or higher)

(42) $\tau \approx (s\delta + \zeta)/(S + \zeta)$

provided that at the optimum $S > s\delta$, where $S = \frac{S_K}{1-S_K} = (f' - \eta)k/(f - kf')$ is the ratio of the net income of capital to that of labor, and $\zeta = \frac{d \ln \phi}{d \ln K_g}$, the elasticity of the productivity of labor with respect to investments in education. (In 42, all variables are evaluated at the optimum, e.g. $s_w(r^*) = s_w(\delta/(1 - \tau^*))$.) Not surprisingly, the higher the return to education, the higher the tax rate.[39]

[39] The result follows from observing that $(dV_w/d\tau)/V_l = -\delta K_c [1 - g\phi']/(1-\tau)^2 + O(\tau^2)$. (We used this result in the previous footnote to show that when $\tau = 0$, $g(k*)\phi'(0) > 1$ is necessary and sufficient for $\tau > 0$) . We can rewrite $g\phi'$, first making use of the definition of the elasticity of ϕ, ζ, then of (41a), then using (39a) and the definition of S_K:
$g\phi' = g\zeta\phi/K_g = g\zeta(1-\tau)/[\delta\tau(k - sg)] = \zeta(1 - \tau)/[\tau[(1-\tau)S_K/(1 - S_K) - s\delta]]$.
Substituting, we obtain

$$dV_w/d\tau = -\delta K_c / [\tau[(1-\tau)S_K/(1 - S_K) - s\delta]]\,[\tau[(1-\tau)S_K/(1 - S_K) - s\delta] - \zeta(1 - \tau)] + O(\tau^2)$$

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

(43) $W = \phi(f - kf') + \tau\,(K - K_W) - K_g$ - t

where output is a general function of labor, public capital goods allocated to augmenting labor and capital productivity and substituting for private capital:

(44) Q = F(K, K_g, L).

If public capital augments the productivity of private capital, so Q = F(H(K_g/K)K, L)[40], then in equilibrium the effective capital labor ratio is determined so that

(45) (Hf'(k) – η) (1 – τ) = δ,

so that the tax spent on private capital productivity augmenting investment leads to a further decrease in k, hurting workers even more.[41]

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) may increase, and so will their welfare.[42] With the return to capitalists' capital below the critical threshold, then eventually consume their wealth. The economy is supported then by public

[40] 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.)
[41] A further question is, should t, the tax on workers, used to finance additional human capital, be positive? Maximizing steady state utility, noting that the tax leaves unchanged r, V is maximized when $\phi' g = 1$, so it is optimal to set t = 0 if (and only if) at the optimal value of τ*, $\phi' g \leq 1$. The above analysis (previous footnote) made clear that at the optimum ϕ'g ≈ 1. If it is desirable to impose a labor tax, it will be small. Under plausible conditions, it can be shown to be zero.
[42] That is, if Q = F(K + K_g, L), then the expenditures on public goods drives down the after tax rate of interest (f' – η)(1 – τ) < δ, in which case capitalist start to "eat" their capital.

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: under quite general conditions[43], 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 in human capital, the optimal tax on capitalists' capital may be very high.

V. 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—a general principle of considerable importance. When there was an optimal 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 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 –

[43] 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, rather than a tax on their net income.

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.[44]

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.

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. 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. 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.

[44] 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).

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.[45] 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 to be adverse.[46] Indeed, there is increasing evidence that inequality itself may affect productivity.[47] 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.

[45] 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).

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

[47] 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.

References

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.

Arrow, K., W.R. Cline, K-G. Maler, M. Munasinghe, R. Squitieri, and J. E. Stiglitz, 1995, "Intertemporal Equity and Discounting," in *Global Climate Change: Economic and Policy Issues,* M. Munasinghe (ed.), World Bank Environment Paper 12, Washington, D.C., pp. 1-32. Reprinted in an abbreviated format as "Intertemporal Equity, Discounting, and Economic Efficiency," *Climate Change 1995: Economic and Social Dimensions of Climate Change,* J. Bruce, H. Lee, and E. Haites (eds.), Cambridge: Cambridge University Press, 1996, 125-144.

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.

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, July-August, pp. 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.

Baumol, W. J, J. C. Panzar and R. D. Willig, 1982, "Contestable Markets and the Theory of Industry Structure", *San Diego, CA : Harcourt Brace Jovanovich*.

Bevan, D., 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 a l'equilibre 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.

Bovenberg, A. L. and L. Goulder, 1996, "Optimal Environmental Taxation in the Presence of Other Taxes: General-Equilibrium Analyses," *American Economic Review,* v. 86, no. 4 (sept), pp. 985-1000.

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.

————, 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.

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.

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

—— 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.

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.

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.

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 Assosciation 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*, Volume 14, Issue 4, December, pp. 419-434.

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

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

Lucas, R., 1990, "Supply-Side Economics: An Analytical Review," *Oxford Economic Papers*, New Series, 42(2): 293-316.

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.

——, 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.

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.

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, 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– 271Ramsey, F., 1927, “A contribution to the theory of Taxation,” *Economic Journal,* 37(145): 47-61

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

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

Sappington, D. and J. E. Stiglitz, 1987, “Information and Regulation,” Public Regulation, E. Bailey (ed.), London: MIT Press, 1987, pp. 3-43

Stiglitz, J.E., 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.

————, 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.)

————, 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

————, 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

————, 2009, “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.

————, 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 J. Yun, 2013, “Optimality and Equilibrium in a Competitive Insurance Market under Adverse Selection and Moral Hazard,” NBER Working Paper 19317, August.