Journal of Public Economics 30 (1986) 137–143. North-Holland

THEORY OF OPTIMAL TAXATION

P.A. SAMUELSON

M.I.T., Cambridge, MA 02139, USA

1.

Assuming the constant-returns-to-scale conditions most suitable for viable perfect competition, let us represent production of goods and services by $(x_1,\dots,x_n)$, where factors of production or inputs can be regarded as negative outputs. Let the government require for public purposes amounts of the respective goods $(G_1,\dots,G_n)$ so that what is left over for private purchases is $(X_1=x_1-G_1,\dots,X_n=x_n-G_n)$. Let $(P_1,\dots,P_n)$ represent the prices paid for goods by private consumers, and let there be specific excises on goods and services of $(T_1,\dots,T_n)$ so that what producers receive in prices is $(p_1=P_1-T_1,\dots,p_n=P_n-T_n)$. Note that subsidies can be treated as algebraically negative taxes, and in such cases prices paid by consumers, P_i, are *less* than prices received by producing firms, p_i; note too that if an x refers to an input and so is negative, then its P represents what consumers *receive* for their service, its p what producers *pay* for these services, and its T must be *negative* if the government is to be collecting revenue rather than paying it out.

In a community of more than one person, it can be shown that the redistributional effects of small taxes are much more important than their distortive effects. Hence to concentrate on the latter, we assume either (1) a single Robinson Crusoe, or (2) a mass of representative citizens all affected in the same direction by the tax program, or (3) that the appropriate distribution of income has already been achieved so as to realize and maintain an optimum according to some prescribed norms of interpersonal equity. Any of these three assumptions enables us to treat the problem as a single person one.

Our single consumer spends (and saves, if we wish to include variables over time) that which he earns from sale of his owned factors of production: hence, he is subject to the algebraic net budget equation $\sum P_iX_j=M=0$; and with all price ratios given, his real quantities demanded and supplied are determinate functions of the form

$$X_i=X^i(P_1,\dots,P_n;0),\quad i=1,2,\dots,n, \tag{1}$$

where $M=0$ represents the absence of any lump-sum taxes or subsidies.

Under strong conditions of constant-returns-to-scale, we can summarize all relevant producer relations by an optimum production possibility schedule relating all the final outputs and inputs; namely

$$F(x_1,\ldots,x_n)=0, \tag{2}$$

where scale does not matter, but only proportions; and where the partial derivatives of this function are marginal costs or productivities, proportional to producers' competitive prices $(p_1,\ldots,p_n)$.

The basic problem of optimal taxation can now be simply stated: given the government requirements for goods and services, the consumer's demand–supply reactions, and competitive production relations, what is the optimal pattern of taxes on goods and services so as to leave the consumer at the highest feasible level of preference?[1]

It is easy to see mathematically and intuitively that a lump-sum tax is ideal:[2] it does lead to a loss to the consumer, but all of that loss is unavoidable; there is no further deadweight loss. However, the problem gains in analytic interest and practical relevance if we assume that the real resources must be raised *by taxing goods and services* and that we seek to minimize the consumer's deadweight loss as far as feasible.

2.

The solution to the problem can be briefly sketched. The demand functions of (1) can be thought of as having arisen from an ordinal utility function $U=U(X_1,\ldots,X_n)$, and substituting (1), we get ordinal utility depending upon the P's alone, or $U=V(P_1,\ldots,P_n;0)$. Our problem can be considered that of picking P's so as to maximize

$$U=V(P_1,\ldots,P_n;0) \tag{3}$$

subject to

$$F[X_1+G_1,\ldots,X_n+G_n]=0$$

$$=F[X^1(P_1,\ldots,P_n;0)+G_1,\ldots,X^n(P_1,\ldots,P_n;0)+G_n]$$

and G's given.

[1]Aspects of the correct answer have been hinted at by Ramsey (1927) and Joseph (1939). Neither in its local or marginal aspects nor in its larger aspects does the problem benefit from consumer's surplus techniques: it is interesting that Marshall, who stumbled on to some interesting cases for taxation and subsidy that have no relevance to the problem as stated here, was led to false answers by consumer's surplus; where the later Pigouvian analysis adopted the conclusions of Marshall on increasing cost industries, etc. that analysis had finally to give way to the criticisms of Young, Knight and Robertson, leaving a valid core dependent on external economies and diminished rather than increased surplus.

[2]The incidence of this tax will almost certainly shift relative P's; but such shifts are optimal provided that $P_i=p_i$ in every case.

Using the usual technique of the Lagrangean multiplier, remembering the definition of the variables and the proportionality of the producer p's to the partial derivatives or marginal productivities of the production possibility function, we finally end up with the following formulation,[3] valid for large or small G's and taxes:

An optimal pattern of taxes is one which, if it were imposed but compensated for by giving the consumer enough lump-sum income to keep him on the same level of satisfaction, would then result in an equal percentage change in all goods and services.

Mathematically, let the Slutsky compensated change or substitution terms $S_{ij}(P_1,\dots,P_n;0)=\partial X^i/\partial P_j+X^j\,\partial X^i/\partial M=S_{ji}$, where M stands for a change in lump-sum tax or net income. Then the basic condition for an optimal tax pattern is

$$\sum_{j=1}^{n}\frac{S_{ij}(P_1,\dots,P_n;0)\,T_j}{X^i}=\sum_{j=1}^{n}s_{ij}t_j=-K, \tag{4}$$

where K can be evaluated when we know the size of the government program and where $s_{ij}=S_{ij}P_j/X_i$ is the dimensionless elasticity form of the Slutsky terms, and where $t_j=T_j/P_j$ is the percentage or ad valorem value of the specific tax.

If the tax program is 'small', the above literary statement stands. If large, the optimal pattern must be interpreted as giving taxes proportional to rates of price change which would, if compensated, give rise (at the new margin) to equal percentage rates of change in every good and service.

3.

The formula (4) is basic for tax policy, In the present state of our econometric knowledge it consists substantially of 'empty boxes'. Nonetheless, every legislator and treasury official is either acting implicitly in terms of his best intuitive guess as to the terms involved or else he is acting without thought of maximizing feasible consumer well-being.

The more sophisticated textbook discussions of public finance implicitly make certain partial equilibrium assumptions about the elasticities of tobacco, liquor, etc. often treating each of these categories as if it had zero cross-

[3]Ramsey (1927) came close to this conclusion, but by virtue of his assumptions about zero income elasticity of all but one good, his conclusion referred to uncompensated changes. Mrs. Joseph's (1939) important graphical analysis established the important point that it is neither price elasticity nor income elasticity upon which avoidable deadweight loss depends, but rather the elasticity of substitution along the indifference contours. See Hicks (1947, ch. X) for further discussion.

elasticities with other goods. One of the advantages of a formulation in terms of S_{ij} substitution terms is the fact that everyone will instantly recognize that changing one good alone *must* change some other good or goods if we are to remain at the same level of indifference; and hence, there is less danger of rushing to premature policy prescriptions based on questionable independence assumptions.

There is a deceptive simplicity about (4) that must be warned against. It is the final summary of a long chain of mathematical argument, here omitted. As soon as the tax program becomes at all large the formula cannot be applied until we have knowledge about the perplexing problem of tax incidence, since the S terms depend upon the unknown P's that will result from any selected pattern of T's. Furthermore, (4) turns out to involve explicitly only consumption substitution terms and not to involve explicitly production substitution terms which determine the rate of increase of relative marginal costs. This is really remarkable. Some years ago I worked out (4) for the special case of constant marginal costs, where the production problems are very simple. Not until I saw a brilliant analysis by Marcel Boiteux[4] of small-scale changes was I able to imagine that a final result could depend upon consumption terms alone; and encouraged by his result, I was able to develop formula (4), which in the case of small taxes is essentially his formula.

Finally, note that (4) involves a set of simultaneous equations that must be inverted. Few people can intuitively envisage the results of such a process; most likely they will get their answers by partial equilibrium analysis, treating all cross-terms as zero and concentrating on 'own' terms, which may lead to misleading results.[5]

An interesting question for further careful analysis is the following: How do we form intuitive notions about the relevant elasticities, and how do we make best approximate guesses based upon such notions? I suspect that in the absence of knowledge, many economists implicitly treat all the s_{ii}'s as approximately equal, with all cross-terms of diffused small magnitude. An assessment of the validity and relevance of such an assumption is urgently needed.

[4]Boiteux [1951, eq. (17)]. Note that for finite G's, the solution does depend on the P's, which in turn depend on the form of F and its substitution terms; but the dependence is indirect, and for small government programs of little significance.

[5]Economists often confuse two different questions: (1) How much should each good be taxed, which the most heavily and which the least heavily? (2) If we can only tax one good, which should it be? The latter question can be rigorously answered – at least in the small tax case – by considering the own-terms alone. It can be shown that deadweight loss can be beautifully measured, free of all cardinal utility or illegitimate consumer's surplus assumptions, by $\sum(\dot{X}_i+\partial X_i/\partial M)\dot{T}_i$, where $\sum X_i\dot{T}_i=1$, and where dots indicate differentiation of the variables with respect to the size of the government's tax and expenditure programs. We can manipulate the deadweight loss into the form $\sum\sum S_{ij}\dot{T}_i\dot{T}_j$, which is to be minimized subject to the above equality. If, for reason of feasibility, certain $\dot{T}$'s must be zero or preassigned, we can continue to minimize with respect to the remaining $\dot{T}$'s at our discretion.

4.

This brief discussion is not the place to interpret significance for tax policy. But a few tentative conclusions may be ventured.

(1) Since we have been able to treat factors of production simply as algebraically negative goods, in principle there can be no basic differences between the two categories; hence the alleged differences between direct and indirect taxes can at best be a quantitative one, a difference in degree rather than of kind.

(2) Since price ratios, rather than absolute prices, alone count, the optimal tax pattern is not unique – unless we have specified some one good whose price is to be the *numeraire*, remaining untaxed and unchanged. This reinforces point (1), since a tax on strawberries is just like a tax on the labor of picking strawberries. More generally, a factor tax pattern that raised all factor prices in the same proportion while leaving all goods prices unchanged – such a pattern could be exactly replaced by a commodity tax pattern that changed all goods prices in the opposite direction and left factor prices unchanged.

(3) At least since the time of Hume, economists have recongnized that taxes upon an unchangeable supply of land had minimum distortion effects. (It may be noted that a zero tax on rent, coupled with a proportionate tax on *all other* goods and services, would have the same effect for the reason given above.) Often the total supply of labor is believed to be rather inelastic to wage changes, so that a similar argument is adduced in favor of personal income taxes generally. To the extent that the price-inelasticity of labor is due to a cancellation of income and substitution effects, the alleged absence of deadweight loss is false. It must be re-emphasized that it is not price-elasticity that is crucial, nor income elasticity, but rather substitution terms that determine deadweight loss.

(4) Taxes which people can and will avoid by changing their behavior give an illusion of being unburdensome, precisely because they can be avoided. The truth is just the opposite: per dollar of revenue collected, such 'voluntary taxes' do the maximum harm and are to be avoided; they occasion the greatest distortions and do so without achieving the purpose of releasing to the government real resources.

(5) Equity and feasibility aside, the best tax is a lump-sum tax. We approach as near as we can to this ideal to the degree that we are able to tax activities that go on the same regardlesss of the tax: e.g., a tax on land rent, on smoking, etc. Indeed, this same type of argument would lead ultimately to taxes on objects which the consumer cannot affect: e.g., his age, sex, serial number, or potentialities.

(6) Taxes that can be evaded by smugglers, bootleggers, and others are to be avoided for the reasons given above, if no other. The resulting distortions are for no good revenue purposes.

(7) Taxes on items that will be in zero net demand or supply are bad, since they collect no revenue and still have distorting effects. Similarly reducing the tax base by exempting home produced and consumed goods is, ceteris paribus, a step in the wrong direction, as are other exemptions of imputed incomes.

(8) In contrast to an expenditure tax, a tax on personal income is regarded by Irving Fisher and Pigou as involving 'double taxation'.[6] Such a tax is often regarded as being discriminatory against saving. Actually, the real problem must be formulated as follows: After we have introduced goods and services of all time periods into our analysis, does it show that income taxes give rise to more deadweight loss than do expenditure taxes? No dogmatic answer is possible. Certainly one can imagine a pattern of s_{ij} terms such that income taxes are to be preferred to consumption taxes. One can also imagine an opposite case. I am not sure whether there is any a priori presumption that one case is likely to be more optimal than the other; but if there is, formula (4) should be a good tool for deciding the issue.

(9) Arguments based on ignorance are notoriously dangerous. Nonetheless, if we neglect costs of tax collection, it may be that we can find a presumption against the customary practice of governments, like the United Kingdom, which derive most of their consumption revenue from a few standard commodities. (The fact that most of us vaguely disapprove of tobacco and alcohol complicates the issue, since we thereby discard the consumer's own indifference curves in favor of our moral preferences, and consequently do not regret the fact that, in consuming less of these goods, the consumer is moving to a lower indifference curve than would be obtainable. I ignore such complications.) If there can be found some presumption toward equal s coefficients, then the same ad valorem tax rates might be considered optimal. Other things being equal, this would provide a case for as wide a selection of commodity taxation as possible.[7]

[6]Pigou (1947, chs. X, IX).

[7]It may be that some economists will grant that, in the absence of detailed econometric knowledge, we should treat the s's as equal for two goods with the same algebraic PX, but that where the goods differ in expenditure, some systematic difference in presumption concerning the s's must be made. I, myself, have a distrust of arguments based on ignorance; but I do occasionally find myself deviating from the straight and narrow path. See Lerner (1946, pp. 29–32) for a somewhat similar attempt to deduce that equal incomes will maximize the *probable* total of utility. If we don't know which individuals are the better marginal utility or pleasure machines, Lerner argues that we run the least risk if we treat them as equal. It seems to me that he proves only that an equal distribution is better than a random movement around it; and that exactly the same argument will prove that *any given unequal* distribution of income is better than a random movement around *it*.

References

Boiteux, M., 1951, Le 'revenu distribuable' et les pertes economiques, Econometrica 19, April, 112–133.

Hicks, U., 1947, Public finance (Cambridge Handbook Series), (Pittman, London), ch. X.

Joseph, M.F.W., 1939, The excess burden of indirect taxation, Review of Economic Studies 6, June, 226–231.

Lerner, A.P., 1946, Economics of control (Macmillan, New York) 29–32.

Pigou, A.C., 1947, A study in public finance, 3rd edition (Macmillan).

Ramsey, F.P., 1927, A contribution to the theory of taxation, Economic Journal 37, March, 47–61.