

ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

OPTIMAL TAXATION

Louis Kaplow

Discussion Paper No. 607

01/2008

Harvard Law School
Cambridge, MA 02138

This paper can be downloaded without charge from:

The Harvard John M. Olin Discussion Paper Series:
http://www.law.harvard.edu/programs/olin_center/

The Social Science Research Network Electronic Paper Collection:
http://papers.ssrn.com/abstract_id=1095899

Optimal Taxation

Louis Kaplow*

Abstract

Optimal taxation concerns how various forms of taxation should be designed to maximize social welfare. The task requires an integrated consideration of the revenue-raising and distributive objectives of taxation. The central instrument in developed economies is the labor income tax, the analysis of which was pioneered by Mirrlees (1971). Subsequently, Atkinson and Stiglitz (1976) showed how commodity taxes should be set in the presence of an optimal income tax, the results differing qualitatively from, and in important respects displacing, the teachings derived from Ramsey's (1927) seminal analysis of the pure commodity tax problem.

Forthcoming, *The New Palgrave Dictionary of Economics*, 2nd Edition

*Harvard University and National Bureau of Economic Research. I am grateful to the John M. Olin Center for Law, Economics, and Business at Harvard University for financial support. For more extensive treatments, see Kaplow (2007, 2008).

Optimal Taxation

Louis Kaplow

© Louis Kaplow. All rights reserved.

Optimal taxation concerns the question of how various forms of taxation should be designed in order to maximize a standard social welfare function subject to a revenue constraint. The task requires an integrated consideration of the revenue-raising and distributive objectives of taxation. The central instrument in developed economies is the labor income tax. Mirrlees (1971) pioneered the analysis of this challenging problem. Subsequently, Atkinson and Stiglitz (1976) showed how commodity taxes should be set in the presence of an optimal income tax. The results are qualitatively different from – and in important respects displace – prior teachings that originate in Ramsey’s (1927) analysis of the pure commodity tax problem. In addition to setting particular taxes optimally, it is also necessary to choose optimally among tax systems.

1. INCOME TAXATION

1.1. Model

The standard optimal income tax model involves a one-period setting in which individuals’ only choice variable is their degree of labor effort l . There is a single composite consumption good c . An individual’s utility is given by $u(c, l)$, where $u_c > 0$ and $u_l < 0$. An individual’s consumption is given by

$$(1) \quad c = wl - T(wl),$$

where w is the individual’s wage rate and T is the tax-transfer function.

The motivation for redistributive taxation is that individuals differ, in particular in their wages, that is, their earning abilities. The distribution of abilities will be denoted $F(w)$, with density $f(w)$. Individuals’ wage rates are taken to be exogenous. Their pre-tax earnings wl are the product of their wage rate and level of labor effort. More broadly, one can interpret labor effort as including not only hours of work but also intensity and not only productive effort but also investments in human capital.

Taxes and transfers, $T(wl)$, at any income level may be positive or negative. The (uniform) level of the transfer received by an individual earning no income, that is, $-T(0)$, is sometimes referred to as the grant g . Taxes may be interpreted broadly, to include sales taxes or value-added tax (VAT) payments in addition to income taxes. Transfers include those through the tax system in addition to welfare programs. The inclusion of transfers is important both

practically, since they are in fact significant, and conceptually, since otherwise redistribution would be limited to transfers between the rich and the middle class, once the poor were exempted from the tax system.

Taxes and transfers are taken to be a function of individuals' incomes, assumed to be observable, and it is this dependence of taxes on income that is the source of distortion. If taxes could instead depend directly on individuals' abilities, w , individualized lump-sum taxes would be feasible and redistribution could be accomplished without distorting labor supply. Ability, however, is assumed to be unobservable.

The government's problem is to choose $T(wl)$ to maximize social welfare, which can be stated as

$$(2) \int W(u(c(w), l(w))) f(w) dw,$$

where c and l are each expressed as functions of w to refer to the levels of consumption achieved and labor effort chosen by an individual of ability w . If W is linear, the welfare function is utilitarian, whereas if W is strictly concave, additional weight is given to inequality in utility levels (not just levels of marginal utilities).

This maximization is subject to a revenue constraint and to constraints regarding individuals' behavior. The former is

$$(3) \int T(wl(w)) f(w) dw = R,$$

where R is an exogenously given revenue requirement. Here, revenue is to be interpreted as expenditures on public goods that should be understood as implicit in individuals' utility functions; because these expenditures are taken to be fixed, they need not be modeled explicitly. Regarding the latter constraints, individuals are assumed to respond to the given tax schedule optimally, which determines the functions $c(w)$ and $l(w)$.

Mirrlees's (1971) original exposition has been followed by subsequent elaborations, much of which is synthesized and extended in Atkinson and Stiglitz (1980), Stiglitz (1987), Tuomala (1990), and Salanié (2003). Because the problem is formidable, the present discussion will be confined to stating basic results, such as are embodied in first-order conditions and produced by simulations.

1.2. Linear income tax

Substantial illumination with greatly reduced complexity is provided by first examining a linear income tax,

$$(4) T(wl) = twl - g,$$

where t is the (constant, income-independent) marginal tax rate and g , as previously noted, is the uniform per-capita grant. Because of the presence of the grant g , a linear income tax can be highly redistributive (consider setting t at 100% and g equal to mean income net of any per capita revenue requirement – in the absence of incentive constraints) or not at all redistributive (t may be 0% and g equal to the negative of the per capita revenue requirement). Foreshadowing discussion of the nonlinear income tax, the degree of redistribution is more directly related to the levels of t and g than to the shape (deviation from linearity) of the tax schedule.

To derive the optimal linear income tax, the government's maximization problem can be written in Lagrangian form as choosing t and g to maximize

$$(5) \int \left[W(u((1-t)wl(w) + g, l(w))) + \lambda(twl(w) - g - R) \right] f(w) dw,$$

where λ is the shadow price of revenue, referring to the constraint (3), and expression (4) is substituted into expression (1) so that consumption is expressed in terms of the specific linear tax system under consideration. Following Atkinson and Stiglitz (1980) and Stiglitz (1987), the first-order condition for the optimal tax rate can usefully be expressed as

$$(6) \frac{t}{1-t} = - \frac{\text{cov}(\alpha(w), y(w))}{\int y(w)\varepsilon(w)f(w)dw},$$

where $y(w) = wl(w)$, income earned by individuals of ability w ; $\varepsilon(w)$ is the compensated elasticity of labor effort of individuals of ability w ; and $\alpha(w)$ is the net social marginal valuation of income, evaluated in dollars, of individuals of ability w :

$$(7) \alpha(w) = \frac{W'u_c(w)}{\lambda} + tw \left(\frac{\partial l(w)}{\partial g} \right).$$

The numerator of the first term on the right side of expression (7) indicates how much additional (lump-sum) income to an individual of ability w contributes to social welfare – u_c indicates how much utility rises per dollar and W' indicates the extent to which social welfare increases per unit of utility – and this product is converted to a dollar value by dividing by the shadow price of government revenue. The second term takes into account the income effect, namely that giving additional lump-sum income to an individual of ability w will reduce labor effort ($\partial l(w)/\partial g < 0$), which in turn reduces government tax collections by tw per unit reduction in $l(w)$.

Expression (6) indicates how various factors affect the optimal level of a linear income tax. Beginning with the numerator on the right side, a higher (in magnitude) covariance between α and y favors a higher tax rate. In the present setting, $\alpha(w)$ will (under assumptions ordinarily postulated) be falling with income. Note that a larger covariance does not involve a closer (negative) correlation but rather a higher dispersion (standard deviation) of α and of y . The dispersion of α will tend to be greater the more concave (egalitarian) is the welfare function W

and the more concave is utility as a function of consumption (i.e., the greater the rate at which marginal utility falls with income). Income, y , will have a higher dispersion (again, under standard assumptions) when the distribution of underlying abilities is more unequal. In sum, more egalitarian social preferences, more rapidly declining marginal utility of consumption, and higher underlying inequality each contribute to a higher optimal tax rate.

The denominator indicates that a higher compensated labor supply elasticity favors a lower tax rate. The other terms in the integrand indicate that, ceteris paribus, the labor supply elasticity matters more with regard to high-income individuals and at ability levels where there are more individuals (typically the middle of the income distribution) because of the greater sacrifice in revenue.

The foregoing exposition is incomplete in not emphasizing the various respects in which income effects are relevant (they influence α and also λ) and in ignoring that the values on the right side of expression (6) are endogenous. Especially for the latter reason, the literature has relied heavily on simulations.

The most-reported optimal linear income taxation simulations are those of Stern (1976). For his preferred case – an elasticity of substitution between consumption and labor of 0.4, a government revenue requirement of 20% of national income, and a social marginal valuation of income that decreases roughly with the square of income – he finds that the optimal tax rate is 54% and that individuals' lump-sum grant equals 34% of average income. To illustrate the benefits of redistribution, he finds that a scheme that uses a lower tax rate, just high enough to finance government programs (that is, with a grant of zero), produces a level of social welfare that is lower by an amount equivalent to approximately 5% of national income. If there is very little weight on equality, the optimal tax rate is only 25%, whereas if there is extreme weight on equality, the optimal tax rate is 87%. Returning to his central case, an extremely low labor supply elasticity implies an optimal tax rate of 79%, and an elasticity as high as had been used in some earlier literature implies an optimal tax rate of 35%. In the absence of the need to finance government expenditures, the optimal tax rate is 48%, and if government expenditures are twice as high, the optimal tax rate is 60%.

1.3. Nonlinear income tax

Mirrlees (1971) and subsequent investigators employ control-theoretic techniques to address the more general formulation of the optimal nonlinear income taxation problem, which requires choosing an entire tax schedule $T(wl)$ rather than a single tax rate. In this maximization, the constraints regarding individuals' maximizing behavior entail that no individual of any type w will prefer the choice specified for any other type w° . This approach is related to the use of the revelation principle in work on mechanism design, and in similar spirit many researchers following Stiglitz (1982) and others analyze a simpler, discrete variant of the problem, often involving two types, in which the binding incentive constraint is usually that the high-ability type not have an incentive to mimic the low-ability type in order to pay less tax.

The analysis for the continuous case can be summarized in a first-order condition for the optimal marginal income tax rate at any income level y^* , where w^* and l^* correspond to the

ability level and degree of labor effort supplied by the type of individual who would earn y^* . Making the simplifying assumptions that utility is separable between consumption and labor effort and that marginal utility u_c is constant, the condition can be expressed as

$$(8) \frac{T'(w^*l^*)}{1 - T'(w^*l^*)} = \frac{1 - F(w^*)}{\xi^* w^* f(w^*)} \frac{\int_{w^*}^{\infty} \left(1 - \frac{W'(u(w))u_c}{\lambda}\right) f(w)dw}{1 - F(w^*)},$$

where $\xi^* = 1/(1+l^*u_{ll}/u_l)$ – which, when marginal utility is constant as assumed here, equals $\varepsilon/(1+\varepsilon)$, where ε again is the elasticity of labor supply. For derivations of related expressions, see, for example, Auerbach and Hines (2002), Atkinson and Stiglitz (1980), Dahan and Strawczynski (2000), Diamond (1998), Saez (2001), and Stiglitz (1987). Note that this formulation (like those in recent literature) includes $1-F(w^*)$ in both the numerator and the denominator on the right side. The motivation is that, in the first term, $(1-F(w^*))f(w^*)$ is purely a property of the distribution of w , and, in the second term, because the numerator is an integral from w^* to ∞ , the term as a whole gives an average value for the expression in parentheses in the integrand. Both aspects aid intuition, as will be seen in the discussion to follow.

Expression (8), being a first-order condition, should be interpreted by reference to an adjustment that slightly raises the marginal tax rate at income level y^* (say, in a small interval from y^* to $y^*+\delta$), leaving all other marginal tax rates unaltered. There are two effects of such a change. First, individuals at that income level face a higher marginal rate, which will distort their labor effort, a cost. Second, all individuals above income level y^* will pay more tax, but these individuals face no new marginal distortion. That is, the higher marginal rate at y^* is inframarginal for them. Since those thus giving up income are an above-average-income slice of the population (it is the part of the population with income above y^*), there tends to be a redistributive gain.

The right side of expression (8) can readily be interpreted in terms of this perturbation (although it should be kept in mind that this interpretation omits, inter alia, income effects and the endogeneity of variables). Begin with the first term. Revenue is collected from all individuals with incomes above y^* , which is to say all ability types above w^* ; hence the $1-F(w^*)$ in the numerator. This factor favors marginal tax rates that fall with income: As there are fewer individuals who face the inframarginal tax, the core benefit of a higher marginal rate declines. In the extreme, if there is a highest known type in the income distribution, the optimal marginal rate at the top would be zero because $1-F$ would be zero: A higher rate collects no revenue but distorts the behavior of the top individual. However, when there is no highest type, known with certainty in advance, this result is inapplicable. Furthermore, with a known highest type, simulations suggest that zero is not a good approximation of the optimal marginal tax rate even quite close to the top of the income distribution, so the zero-rate-at-the-top result is of little practical importance.

Continuing with the first term, raising the marginal rate at a particular point distorts only the behavior of the marginal type, which explains the $f(w^*)$ in the denominator. For standard distributions, this factor is rising initially and then falling, which favors falling marginal rates at

the bottom of the income distribution and rising rates at the top. The denominator also contains weights of ξ^* , indicating the extent of the distortion, and w^* , indicating how much productivity is lost per unit of reduction in labor effort. The elasticity is often taken to be constant, although some empirical evidence on the elasticity of taxable income supports a rising elasticity due to the greater ability of higher-income individuals to avoid taxes. This consideration may favor marginal rates that fall with income. Finally, w^* is rising, which also favors falling marginal rates: The greater the wage (ability level), the greater the revenue loss from a given decline in labor effort.

The second term applies a social weighting to the revenue that is collected. The expression in parentheses in the integrand in the numerator is the difference between the marginal dollar that is raised and the dollar equivalent of the loss in welfare that occurs on account of individuals above w^* paying more tax. As in the interpretation of expression (7), u_c is the marginal utility of consumption to such individuals, W' indicates the impact of this change in utility on social welfare, and division by λ , the shadow price on the revenue constraint, converts this welfare measure into dollars. This integral is divided by $1-F(w^*)$, which as noted makes the second term an average for the affected population.

This term tends to favor marginal rates that rise with income. The greater is w , the lower is W' (unless the welfare function is utilitarian, in which case this is constant) and the lower would be the marginal utility of income u_c (had we not abstracted from this effect in the simplifying assumptions). Hence, at a higher w^* , the average value of the term subtracted in the integrand is smaller, making the entire term larger. Note further that if social welfare or utility is reasonably concave, $W'u_c$ will approach zero at high levels of income, at which point this term will be nearly constant in w^* . That is, the term favors rising marginal tax rates when income is low or moderate, but has little effect on the pattern of marginal tax rates near the top of the income distribution.

Because of difficulties in determining the shape of the optimal income tax schedule by mere inspection of the first-order condition (8), analysts beginning with Mirrlees (1971) have used simulations to help join the theoretical analysis with empirical estimates of labor supply elasticities and of the distribution of skills or income in order to provide further illumination. Tuomala (1990) offers a useful survey and set of calculations. In all the cases he reports, marginal tax rates fall as income increases, except at very low levels of income. Mirrlees's (1971) original calculations had displayed a similar tendency, but subsequent researchers had questioned the extent to which this result may have depended on the social preferences he stipulated or the arguably high labor supply response he assumed. Later work, however, suggests that a greater social preference for equality or a lower labor supply response tends to increase the level of optimal marginal tax rates but does not generally result in a substantially different shape. This phenomenon is also illustrated by Slemrod et al. (1994), who examine the optimal two-bracket income tax. In all of their simulations, the optimal upper-bracket marginal rate is lower than the lower-bracket rate; indeed, this gap widens as the social preference for equality increases because of the additional value of raising the lower-bracket rate in generating funds to increase the grant, which is of greatest relative benefit to the lowest-income individuals.

Subsequent work further explores the circumstances in which optimal marginal tax rates

might rise with income. Kanbur and Tuomala (1994) find that when inequality in individuals' abilities (wages) is significantly greater than previously assumed (but at levels they suggest to be empirically plausible), optimal marginal tax rates do increase with income over a substantial range, although for upper-income individuals optimal marginal rates still fall with income. Diamond (1998) examines a Pareto distribution of skills (instead of the commonly used lognormal distribution), under which the $(1-F)/f$ component of expression (8) rises more rapidly at the top of the distribution, and finds that optimal marginal tax rates are rising at the top. However, Dahan and Strawczynski's (2000) simulations indicate that Diamond's result was driven in large part by his additional assumption that preferences were quasi-linear, thus removing income effects. (Nevertheless, their diagrams do suggest that, consistent with Diamond's claim, moving from a lognormal to a Pareto distribution favors higher rates – still falling, but notably less rapidly – at the top of the income distribution.) Saez (2001), using income distribution data in the United States from 1992 and 1993, finds that the shape of the distribution of $(1-F)/wf$ is such that optimal rates should fall substantially well into the middle of the income distribution, to an income of approximately \$75,000, rise until approximately \$200,000, and then be essentially flat thereafter.

An additional result from the simulations is that, at the optimum, a nontrivial fraction of the population does not work, and this fraction is larger when social preferences favor greater redistribution and when the labor supply elasticity is higher. This outcome should hardly be surprising because, as the analysis of expression (8) and the simulations suggest, high marginal rates tend to be optimal at the bottom of the income distribution, along with a sizable grant. Relatedly, little productivity and thus little tax revenue is sacrificed when those with very low abilities are induced not to work (whereas substantial revenue is raised from the rest of the population, for whom marginal tax rates on the first dollars of income are inframarginal).

1.4. Extensions

Given the central importance of income taxation to the revenue and distributive objectives of government, further exploration of various aspects of the problem should be a high research priority. A number of features have received some, although generally quite limited, attention. For broader discussions and further references, see Atkinson and Stiglitz (1980), Stiglitz (1987), Tuomala (1990), Salanié (2003), and Kaplow (2008).

A critical assumption in optimal income tax analysis is that earning ability is unobservable so that income, a signal of ability, is taxed instead, which is the source of distortion. Hence, it is worth considering the possibilities for basing taxation more directly on ability. To some degree, hours may be observable, and ability (wages) can thus be inferred. But in many occupations (notably, self-employment) hours are difficult to observe, and both hours and wages are manipulable, such as by extending reported hours and lowering the reported wage. Another approach would be to measure proxies of earning ability, such as through testing. Unfortunately, skills measurable by testing explain only some of the variance in earning ability, and if taxes were to be based on test results or other ability measures, individuals would adjust their performance and thereby distort the measurement. A third technique – one sometimes employed – is to adjust taxes and transfers for observable personal attributes, such as physical disability, age, or family composition.

In general, tax and transfer schedules could be made a function of various imperfect signals of ability (or of other pertinent differences, such as in utility functions). For each value of the signal, there would in essence be a different tax schedule, governed by the first-order condition (8); each of these tax schedules would, however, be linked in a common optimization by the shadow price λ . One might view models like those of Akerlof (1978), in which he assumes that a subset of the lowest-ability group can be identified perfectly (“tagged”), and Stern (1982), in which he examines the usefulness of a noisy signal of ability in a two-type model, as special cases of this more general formulation.

There exist myriad additional complications. One is that income may be a noisy signal of ability, whether because of variations in occupations (for a given ability, one job may pay more to compensate for specific disamenities) or in preferences (an individual may earn more not because of greater ability but rather due to a higher marginal utility of consumption or a lower marginal disutility of labor effort). Another possibility is that individuals may have preferences concerning redistribution itself, perhaps due to altruism or envy. Other topics that have been explored include liquidity constraints, general equilibrium effects of taxation on the distribution of pre-tax wages, uncertainty, interactions with nontax distortions, and human capital.

2. COMMODITY TAXATION

2.1. Commodity Taxation with Income Taxation

To examine optimal commodity taxation with labor income taxation, the foregoing model can be modified as follows. In place of consumption c , individuals choose commodity vectors \mathbf{x} and, as before, labor effort l to maximize the utility function $u(\mathbf{x}, l)$. On the left side of individuals’ budget constraints (1), c is replaced by $\mathbf{p}\mathbf{x}$, where \mathbf{p} is the consumer price vector equal to $\mathbf{p} + \boldsymbol{\tau}$: the sum of a producer price vector (taken to be constant and equal to production costs) and a vector of commodity taxes (which, if negative, are subsidies).

Atkinson and Stiglitz (1976) demonstrate that, when the income tax is set optimally, commodity taxes should be undifferentiated, i.e., $\boldsymbol{\tau} = 0$, when utility is weakly separable in labor (on which more in a moment). Alternatively, other levels of $\boldsymbol{\tau}$ are similarly optimal as long as the ratio of any two consumer prices equals the ratio of producer prices, with the difference in consumer price level being offset by an adjustment to the income tax schedule. (For example, if all commodity taxes are ten percent rather than zero, the income tax schedule may be reduced so that, at all levels of pre-tax income wl , disposable income is ten percent higher.) Subsequent work extends this uniformity result to examine cases in which the income tax need not be optimal and to assess various partial reforms, one result being that any proportionate reduction in nonuniform commodity taxes can generate a Pareto improvement. See Kaplow (2006) and also Konishi (1995) and Laroque (2005).

The intuition behind the uniformity result is that, despite the second-best setting (due to the inherently distortionary character of a redistributive labor income tax), there is nothing to be gained – except distortion of consumption – by differentiating commodity taxes when the utility

function is weakly separable in labor. When that assumption is relaxed, one has the qualification – due originally to Corlett and Hague (1953) in a Ramsey setting – that complements to leisure (labor) should be taxed (subsidized). For example, taxing beach attendance or the purchase of novels may make leisure less attractive, encouraging labor effort and thereby reducing the distortion due to the income tax. Other qualifications, including with regard to preferences that depend on ability, other preference heterogeneity, and administrative and enforcement concerns, are cataloged in Kaplow (2008).

2.2. The Ramsey Problem: Commodity Taxation Alone

The foregoing analysis is usefully contrasted with that of Ramsey (1927), who considered how to set commodity taxes on a population of identical individuals to meet a revenue requirement. The familiar result is that commodity taxes should be inversely proportional to the elasticity of demand, with refinements for demand interdependencies. Introducing nonidentical individuals leads to modifications reflecting distributive concerns that entail higher taxes than otherwise on luxuries and lower taxes on necessities. See generally Atkinson and Stiglitz (1976, 1980), Auerbach and Hines (2002), Salanié (2003), and Stiglitz (1987).

As initially emphasized in Atkinson and Stiglitz (1976) and elaborated in Stiglitz (1987) and Kaplow (2008), however, neither prescription is apt if there is also an income tax. In the original Ramsey model in which all individuals are identical and thus there are no distributive concerns, the optimal tax obviously would be a uniform lump-sum extraction (a limiting case of an income tax), which, it should be noted, neither requires information about individuals' types nor is distributively objectionable in this setting. When differences in earning ability are admitted, the optimal tax is a nonlinear income tax, and in typical cases the lump-sum component involves a uniform lump-sum subsidy. Nevertheless, optimal commodity taxation still is not guided either by the familiar inverse-elasticity rule or by the general preference for harsher treatment of luxuries than of necessities. As noted, in the basic case optimal differentiation is nil regardless of the demand elasticity or how demand changes with income, and qualifications such as that favoring taxation (subsidization) of leisure complements (substitutes) are largely unrelated to the level of the own-elasticity of demand for a commodity or its income elasticity.

2.3. Applications

Optimal commodity taxation is, in an important sense, a building block for the analysis of many other important problems. For example, Atkinson and Stiglitz (1976) explain how the analysis of optimal capital taxation can be assimilated into the framework, for it involves nonuniform taxation of consumption in different time periods, which may be interpreted in terms of the model simply as differently indexed commodities. Hence, in the basic case, the optimal tax on capital is zero.

Furthermore, as discussed by Kaplow (2004, 2008), other types of government policy may be analyzed in a similar fashion. Allowing for externalities, the no-differential-tax prescription may be interpreted as requiring that consumer price ratios equal not producer price ratios but instead ratios of full social costs; hence, first-best Pigouvian taxes and subsidies (i.e.,

set equal to marginal external effects) are optimal despite second-best concerns about distortionary income taxation and distributive effects. For public goods, the analogy to differential taxation is a departure from the pure Samuelson rule, so in the basic case, that cost-benefit test also does not require modification on account of income tax distortions and distributive concerns. Likewise, deviations from marginal cost pricing of public production is counter-indicated.

By contrast, much prior and ongoing work examines these problems and others in a Ramsey-like setting. As Stiglitz (1987) observes, this course may be appropriate for developing economies in which income taxation is largely infeasible, but not for developed economies with an income tax.

3. OPTIMAL TAX SYSTEMS

Most optimal taxation analysis simply assumes that certain tax instruments are available and others are not. Mirrlees (1971), Atkinson and Stiglitz (1980), and Slemrod (1990), however, emphasize the importance of motivating the presumed set of available instruments by administrative and enforcement concerns that indicate what actually is feasible. Ideally, these concerns would not be stipulated but rather would be made endogenous. Often, feasibility is a matter of degree, and one must choose among various imperfect systems, the quality of each being determined by policy choices regarding administration and enforcement and also by how the instrument is used.

To illustrate these tradeoffs, note that a nonlinear income tax may be a more fine-tuned redistributive instrument than a linear income tax but is subject to additional types of manipulations that are costly to regulate. Likewise, if nonuniform commodity taxation is employed, there exist incentives to reclassify commodities. More comprehensive tax bases may avoid unnecessary distortions but be more costly to administer. The extent of evasion under any system may depend on the level of tax rates and on what other taxes are in place.

Greater attention to the choice among tax systems seems warranted. Whether or not to have a 20% VAT, relying far less on income taxes, is probably a more important decision than how to set commodity tax differentials in light of subtle qualifications to the uniformity result. System choices are likely to be particularly important for developing countries, where fewer options are feasible and the available instruments are changing over time and in ways that are influenced by other government policies.

BIBLIOGRAPHY

- Akerlof, G.A. 1978. The Economics of “Tagging” as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning. *American Economic Review* 68, 8-19.
- Atkinson, A.B. and Stiglitz, J.E. 1976. The Design of Tax Structure: Direct Versus Indirect Taxation. *Journal of Public Economics* 6, 55-75.
- Atkinson, A.B. and Stiglitz, J.E. 1980. *Lectures on Public Economics*. New York: McGraw-Hill Book Company.
- Auerbach, A.J. and Hines, J.R. 2002. Taxation and Economic Efficiency. In *Handbook of Public Economics*, vol. 3, edited by A.J. Auerbach and M. Feldstein. Amsterdam: North-Holland. 1347-1421.
- Corlett, W.J. and Hague, D.C. 1953. Complementarity and the Excess Burden of Taxation. *Review of Economic Studies* 21, 21-30.
- Dahan, M. and Strawczynski, M. 2000. Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates: Comment. *American Economic Review* 90, 681-686.
- Diamond, P.A. 1998. Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review* 88, 83-95.
- Kanbur, R. and Tuomala, M. 1994. Inherent Inequality and the Optimal Graduation of Marginal Tax Rates. *Scandinavian Journal of Economics* 96, 275-282.
- Kaplow, L. 2004. On the (Ir)Relevance of Distribution and Labor Supply Distortion to Government Policy. *Journal of Economic Perspectives* 18(4), 59-75.
- Kaplow, L. 2006. On the Undesirability of Commodity Taxation Even When Income Taxation Is Not Optimal. *Journal of Public Economics* 90, 1235-1250.
- Kaplow, L. 2007. Taxation. In *Handbook of Law and Economics*, vol. 1, edited by A.M. Polinsky and S. Shavell. Amsterdam: Elsevier 647-755.
- Kaplow, L. 2008 (forthcoming). *The Theory of Taxation and Public Economics*. Princeton: Princeton University Press.
- Konishi, H. 1995. A Pareto-Improving Commodity Tax Reform Under a Smooth Nonlinear Income Tax. *Journal of Public Economics* 56, 413-446.
- Laroque, G. 2005. Indirect Taxation Is Harmful under Separability and Taste Homogeneity: A Simple Proof. *Economics Letters* 87, 141-144.
- Mirrlees, J.A.. 1971. An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies* 68, 175-208.
- Ramsey, F.P. 1927. A Contribution to the Theory of Taxation. *Economic Journal* 37, 41-61.
- Saez, E. 2001. Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies* 68, 205-229.
- Salanié, B. 2003. *The Economics of Taxation*. Cambridge, Mass.: MIT Press.
- Slemrod, J. 1990. Optimal Taxation and Optimal Tax Systems. *Journal of Economic Perspectives* 4(1), 157-178.
- Slemrod, J., Yitzhaki, S., Mayshar, J. and Lundholm, M. 1994. The Optimal Two-Bracket Linear Income Tax. *Journal of Public Economics* 53, 269-290.
- Stern, N.H. 1976. On the Specification of Models of Optimum Income Taxation. *Journal of Public Economics* 6, 123-162.
- Stern, N.H. 1982. Optimum Taxation with Errors in Administration. *Journal of Public Economics* 17, 181-211.

- Stiglitz, J.E. 1982. Self-Selection and Pareto Efficient Taxation. *Journal of Public Economics* 17, 213-240.
- Stiglitz, J.E. 1987. Pareto Efficient and Optimal Taxation and the New New Welfare Economics. In *Handbook of Public Economics*, vol. 2, edited by A.J. Auerbach and M. Feldstein. Amsterdam: North-Holland. 991-1041.
- Tuomala, M. 1990. *Optimal Income Tax and Redistribution*. Oxford: Clarendon Press.