

Taxation and Economic Efficiency

Alan J. Auerbach

University of California, Berkeley and NBER

James R. Hines Jr.

University of Michigan and NBER

February 2001

This paper has been prepared for a forthcoming volume of the Handbook of Public Economics, edited by Alan Auerbach and Martin Feldstein. We thank Charles Blackorby, Peter Diamond, Kenneth Judd, Louis Kaplow, Gareth Myles, Michel Strawczynski and Ronald Wendner for helpful comments on a previous draft.

Taxation and Economic Efficiency

ABSTRACT

This paper analyzes the distortions created by taxation and the features of tax systems that minimize such distortions (subject to achieving other government objectives). It starts with a review of the theory and practice of deadweight loss measurement, followed by characterizations of optimal commodity taxation and optimal linear and nonlinear income taxation. The framework is then extended to a variety of settings, initially consisting of optimal taxation in the presence of externalities or public goods. The optimal tax analysis is subsequently applied to situations in which product markets are imperfectly competitive. This is followed by consideration of the features of optimal intertemporal taxation. The purpose of the paper is not only to provide an up-to-date review and analysis of the optimal taxation literature, but also to identify important cross-cutting themes within that literature.

JEL Classification: H21.

Alan J. Auerbach
Department of Economics
549 Evans Hall
University of California
Berkeley, CA 94720-3880
auerbach@econ.berkeley.edu

James R. Hines Jr.
Office of Tax Policy Research
University of Michigan Business School
701 Tappan Street
Ann Arbor, MI 48109-1234
jrhines@umich.edu

Table of Contents

1. Introduction.....	1
1.1. Outline of the chapter.....	2
2. The theory of excess burden.....	2
2.1. Basic definitions.....	2
2.2. Variations in producer prices.....	7
2.3. Empirical issues in the measurement of excess burden.....	10
3. The design of optimal taxes.....	15
3.1. The Ramsey tax problem.....	15
3.2. Changing producer prices.....	21
3.3. The structure of optimal taxes.....	24
3.4. An example.....	25
3.5. The production efficiency theorem.....	26
3.7. Distributional considerations.....	27
4. Income taxation.....	30
4.1. Linear income taxation.....	30
4.2. Nonlinear income taxation: introduction.....	34
4.3. Nonlinear income taxation: graphical exposition.....	36
4.4. Nonlinear income taxation: mathematical derivation.....	39
5. Externalities, public goods, and the marginal cost of funds.....	46
5.1. The provision of public goods and the marginal cost of public funds.....	47
5.2. Externalities and the “double-dividend” hypothesis.....	51
5.3. Distributional considerations and the MCPF.....	54
6. Optimal taxation and imperfect competition.....	57
6.1. Optimal commodity taxation with Cournot competition.....	57
6.2. Specific and ad valorem taxation.....	63
6.3. Free entry.....	67
6.4. Differentiated products.....	72
7. Intertemporal taxation.....	75
7.1. Basic capital income taxation: introduction.....	76
7.2. The steady state.....	78
7.3. Interpreting the solution.....	79
7.4. Human capital accumulation and endogenous growth.....	82
7.5. Results from life-cycle models.....	90
8. Conclusions.....	94

1. Introduction

This chapter considers a subject at the very center of public finance analysis, the distortions introduced (and corrected) by taxation. Tax-induced reductions in economic efficiency are known as *deadweight losses* or the *excess burdens* of taxation, the latter signifying the added cost to taxpayers and society of raising revenue through taxes that distort economic decisions.

Taxes almost invariably have excess burdens because tax obligations are functions of individual behavior. The alternative, pure *lump-sum* taxes, are attractive from an efficiency perspective, but are of limited usefulness precisely because they do not vary with indicators of ability to pay, such as income or consumption, that are functions of taxpayer decisions. Thus, even though tax analysis often starts with the simple case of a representative household, it is household heterogeneity and the inability fully to observe individual differences that justify the restrictions commonly imposed on the set of tax instruments. Designing an *optimal tax* system means keeping tax distortions to a minimum, subject to restrictions introduced by the need to raise revenue and maintain an equitable tax burden.

The following sections discuss the theory and measurement of excess burden and the design of optimal tax systems. The analysis draws heavily on the chapters by Auerbach (1985) and Stiglitz (1987) in the original volumes of this Handbook, interweaving the most important results contained in these two chapters with the additional insights and areas of inquiry that have appeared since their publication. For more detailed analysis and a treatment of many other topics in this literature, the reader is referred to these original essays.

1.1. *Outline of the chapter*

The chapter begins with the basics and then turns to selected topics. Sections 2, 3, and 4 lay out the theory of excess burden, optimal commodity taxation, and optimal income taxation. Section 5 considers the provision of public goods and the correction of externalities, and how these problems interact with the manner in which revenues are raised. Section 6 discusses the impact on tax design of deviations from perfect competition, and Section 7 extends the theory of tax design to address issues that arise in intertemporal settings. Section 8 offers some brief conclusions regarding the evolution of the literature and promising directions for future research.

2. The theory of excess burden

2.1. *Basic definitions*

Excess burden (or deadweight loss) is well defined only in the context of a specific comparison, or conceptual experiment. If one simply seeks “the” excess burden of a particular tax policy, there are many equally plausible answers, so in order to obtain a unique meaning, it is necessary to be more specific. For example, the excess burden of a 10 percent tax on retail sales varies not only with the initial conditions of the tax system, but also with the direction of change, i.e., whether the tax is being added or removed.

To illustrate this ambiguity and its resolution, consider the simple case in which there are two goods, an untaxed numeraire good and a second good with a constant relative producer price of p_0 . In the absence of taxation, a population of identical consumers¹ demands quantity x_0 of the second good, as depicted by point 0 in Figure 2.1. The imposition of a tax per unit of $p_1 - p_0$ raises the consumer price of the taxed good to p_1 , with the producer price remaining at p_0 . Thus,

¹ We limit our discussion of excess burden to the case of identical consumers, thereby sidestepping issues of aggregation that arise in the case of heterogeneous consumers. See Auerbach (1985) for further discussion.

the quantity purchased falls to x_1 , and the government collects revenue equal to $(p_1 - p_0)x_1$, as represented in the figure by the shaded area labeled A.

What is the excess burden of this tax? If one were to use the Marshallian measure of the consumers' surplus generated by consumption in this market – the area under the demand curve, D, between $x=0$ and $x=x_0$ – it would appear that consumers lose an area equal to that of regions A+B, or B in excess of the revenue actually collected. By this approach, the roughly triangular area B – commonly known as a “Harberger” triangle in recognition of Arnold Harberger's influential empirical contributions – measures the excess burden of the tax.

Unfortunately (see Auerbach 1985), this particular measure of excess burden is not uniquely defined in a setting with more than one tax, due to the well-known problem of *path dependence* of consumers' surplus: the measure of excess burden is affected by the order in which one envisions the taxes being imposed. Path dependence is disconcerting, but more importantly reflects the imprecision of consumers' surplus-based measures of excess burden. There is no well-defined economic question to which the difference between the change in consumers' surplus and tax revenue is the answer. Thus, economists have sought alternative measures of excess burden that are not path-dependent and that answer meaningful questions.

Path dependence does not arise if excess burden is measured by Hicksian consumers' surplus, based on schedules that hold utility, rather than income, constant as prices vary. Because actual tax policy changes typically do not hold utility constant, it is therefore necessary to construct a measure based on a conceptual experiment in which utility is held constant. One intuitive experiment is to imagine that, as a tax is imposed, utility is held constant at its pre-tax level. Graphically, in Figure 2.2, this measure is based on the compensated demand curve $D(u_0)$, which by definition passes through the original, no-tax equilibrium point 0. If the tax is

imposed, and consumers are compensated to remain at original utility levels, then demand follows this schedule and the tax reduces consumption to point 1. At this point, revenue raised is the sum of areas A and C, rather than the actual level of revenue represented by area A, because compensation induces consumers to purchase more of the taxed good (if, as is assumed here, the good is normal) and hence pay more taxes. Excess burden is defined as the amount, in excess of this revenue, that the government must compensate consumers to maintain initial utility in the face of a tax-induced price change. The amount of compensation, which corresponds to the Hicksian measure of the *compensating variation* of the price change, may be calculated using the expenditure function as

$$(2.1) \quad E(\mathbf{p}_1, U_0) - E(\mathbf{p}_0, U_0) = \int_{\mathbf{p}_0}^{\mathbf{p}_1} \frac{dE(\mathbf{p}, U_0)}{d\mathbf{p}} d\mathbf{p} = \int_{\mathbf{p}_0}^{\mathbf{p}_1} x^c(\mathbf{p}, U_0) d\mathbf{p}$$

which is well-defined even for a vector of changing prices \mathbf{p} – the Hicksian variations are single-valued, regardless of the order of integration of the different price changes in (2.1). For each market, this measure equals the area between prices p_0 and p_1 to the left of the compensated demand curve $D^c(U_0)$. Thus, the deadweight loss equals area D in the figure – still approximately a “Harberger triangle”, but different than that defined by the ordinary demand curve in Figure 2.1.²

An alternative conceptual experiment is to begin with the tax already in place and then remove it, extracting from consumers in lump-sum fashion an amount that prevents them from changing their utility levels while the tax is removed. Because the initial tax is distortionary, it is

² Note that this definition is equally well-defined for the case of negative revenue, in which we would trace a path down the compensated demand curve from point 0. There, too, the tax system generates excess burden, in that the revenue lost exceeds the absolute value of the associated compensating variation. This serves as an important reminder that deadweight loss is the result of distortion, not of raising revenue *per se*.

necessary to extract more from consumers than the tax revenue, the difference representing the excess burden of the initial tax. Starting from point 1 in Figure 2.2, this experiment follows the compensated demand curve $D^c(U_1)$ down to point 0, where the price reaches its no-tax level but utility remains unchanged. Again using the expenditure function to calculate the amount the government extracts in this case – the Hicksian *equivalent variation*, based on the formula in (2.1) with U_1 in place of U_0 – the amount equals the area to the left of demand curve $D^c(U_1)$ between prices p_0 and p_1 . This exceeds the forgone revenue – in this case the actual revenue defined by area A – and again does so by a “triangle.”

Although these two measures are the most intuitive, they are actually just examples drawn from a class of measures based on arbitrary levels of utility, say U_i :

$$(2.2) \quad E(\mathbf{p}_1, U_i) - E(\mathbf{p}_0, U_i) - R(\mathbf{p}_0, \mathbf{p}_1, U_i)$$

where $R(\mathbf{p}_0, \mathbf{p}_1, U_i) \equiv (\mathbf{p}_1 - \mathbf{p}_0) \cdot x^c(\mathbf{p}_1, U_i)$ is the level of revenue collected with taxes in place and utility fixed at level U_i .

As Figure 2.3 shows, it is also possible to represent excess burden in a graph in commodity space. In the figure, the consumer’s indifference curve is tangent to the original budget line at point 0, which corresponds to point 0 in Figure 2.2. The tax rotates the consumer budget line as shown, leading to consumption at point 1 (corresponding to point 1 in Figure 2.2), at which tax revenue, measured in terms of the numeraire commodity, equals $R(p_0, p_1, U_1)$. The consumer could maintain utility level U_1 in the absence of taxes by consuming at point 0 (again, as labeled in Figure 2.2), where only $E(p_0, U_1)$ of expenditure would be required, which is less (as measured by the numeraire commodity) than the expenditure necessary to generate utility level

U_1 when consumption is distorted by taxes (as it is at point D). The difference is the equivalent variation measure of excess burden, based on expression (2.2) for utility level U_1 .

It is straightforward to generalize this class of measures to situations in which initial equilibria are not Pareto-optimal due to pre-existing taxes. The marginal excess burden of a tax change is the difference between the Hicksian variation associated with the price change and the *change* in tax revenue (which, in the absence of preexisting taxes, is simply tax revenue), at the chosen level of utility:

$$(2.3) \quad E(\mathbf{p}_2, U_i) - E(\mathbf{p}_1, U_i) - [R(\mathbf{p}_0, \mathbf{p}_2, U_i) - R(\mathbf{p}_0, \mathbf{p}_1, U_i)]$$

in which \mathbf{p}_2 is the price vector after the tax change. For a given reference utility level U_i , this definition has the important property that the marginal excess burden in moving from point 1 to point 2 equals the difference between the excess burden at point 2 and the excess burden at point 1, as defined in expression (2.2).

Figure 2.4 illustrates this measure for the case in which an initial tax in a single market that changed the consumer price from p_0 to p_1 is then increased, raising the price to p_2 . The figure illustrates the marginal excess burden of this tax increase, taking the reference utility level to be that obtained at point 1, the consumption point with the initial tax in place. The Hicksian variation of the additional price change equals the sum of areas A and B. The change in tax revenue (with utility held constant) equals the difference between final tax revenue (areas A+C) and tax revenue prior to the imposition of the second tax, (C+D), or a difference of A–D. That is, with a preexisting tax, it is necessary to net the revenue lost on forgone purchases against the revenue gained from a higher tax on remaining purchases. Thus, the marginal excess burden consists not only of the “triangle” B, but also the rectangle D. Marginal excess burden is no

longer just a second-order phenomenon (the triangle) that vanishes with a small tax increase, but instead is of first-order significance. The total excess burden (calculated at utility level U_1) of both taxes equals this marginal excess burden plus the excess burden of the initial tax, equal to area E.

2.2. *Variations in producer prices*

The analysis thus far adopts the simplifying assumption of fixed relative producer prices, but it is possible to extend the various measures of excess burden to the more general case in which producer prices vary. It is helpful to begin with a graphical exposition. Figure 2.5 repeats the experiment of Figure 2.3, but does so in a case in which the relative producer price of the taxed good – the inverse slope of the production possibilities frontier (*PPF*), shown in bold – varies with the output mix.

Starting again at an equilibrium in which a distortionary tax is used to raise revenue from the representative household, the household's consumption bundle is shown at point 1, which corresponds to point 1 in Figure 2.3. Production occurs at point 1^P in the figure, and the government raises revenue in the numeraire commodity equal to the horizontal distance between points 1 and 1^P . The consumer price p_1 exceeds the producer price q_1 by the tax per unit of output. The household's income (in units of the numeraire commodity) is y_1 , and its indifference curve is tangent to the consumer price line at point 1. Also passing through point 1 (but having a slope $-1/q_1$ and not tangent to the indifference curve) is a "private" production possibilities frontier – the original *PPF*, displaced to the left by the amount of the numeraire commodity corresponding to government consumption. Because the government is assumed to absorb only the numeraire commodity, this displacement is horizontal; otherwise, point 1 would not lie directly to the left of point 1^P . If, instead, the government devoted all tax revenues to purchases

of the taxed commodity, then point 1 would lie directly *below* point 1^P. It should be clear that (unlike in the experiment with fixed producer prices) the equilibrium is affected by how the government uses its revenue, since government purchases influence relative demand and hence relative producer prices of the two commodities.

Excess burden is the amount of additional revenue the government could collect without harming the consumer, were lump-sum taxes used instead of distortionary taxes. It is necessary to specify the form that this extra revenue takes. Here, all revenue takes the form of the numeraire commodity, shifting the “private” *PPF* horizontally to the left until tangent (at point 0⁰) with the indifference curve passing through point 1. Corresponding to consumption point 0⁰ is the production point 0^P. Excess burden is measured as the horizontal distance between this undistorted point 0⁰ and the corresponding point on the “private” *PPF* passing through point 1. Excess burden can be defined algebraically by noting that the horizontal distance between points 0⁰ and 0^P equals the sum of excess burden and tax revenue (the same revenue as that raised in the initial equilibrium, $R(q_1, p_1, U_1)$). Thus, letting y'_0 be the value of the household’s income from production at point 0⁰, excess burden equals

$$(2.4) \quad y'_0 - E(\mathbf{p}'_0, U_1) - R(\mathbf{q}_1, \mathbf{p}_1, U_1) = E(\mathbf{p}_1, U_1) - E(\mathbf{p}'_0, U_1) + y'_0 - y_1 - R(\mathbf{q}_1, \mathbf{p}_1, U_1)$$

with the last step in (2.4) following from the identity that $E(\mathbf{p}_1, U_1) = E(\mathbf{p}_1, U(\mathbf{p}_1, y_1)) \equiv y_1$. As in the case with fixed producer prices, the measure defined in (2.4) may be constructed for different reference utility levels.³ Also, differences in excess burden as measured by (2.4) correspond to changes in excess burdens due to additional taxes.

³ The expression for excess burden, and its graphical interpretation, becomes somewhat more complicated if the government absorbs both taxed and untaxed commodities. See Auerbach (1985) for further discussion.

Expression (2.4) collapses to (2.2) when producer prices do not change, for then income y is fixed and the net of tax price vector in the tax-distorted equilibrium, \mathbf{q}_1 , and the price vector in the undistorted equilibrium, \mathbf{p}'_0 , both are identical to the original price vector \mathbf{p}_0 . The extra term, $y'_0 - y_1$, is the change in income along the production possibilities frontier when moving from point 1^P to point 0^P . By the envelope theorem, the change in income equals $\int_{\mathbf{q}_1}^{\mathbf{p}'_0} \mathbf{x}(\mathbf{q}) d\mathbf{q}$, where $\mathbf{x}(\mathbf{q})$ is the quantity vector of goods produced at price vector \mathbf{q} . It is then possible to represent excess burden in a single market in price-quantity space, as does the diagram in Figure 2.6, in this case with an upward sloping supply curve for the taxed good, $x(q)$. The excess burden, according to expression (2.4), equals the sum of Hicksian consumers' surplus, areas A+B, plus the change in income, areas C+D (sometimes known as "producers' surplus") minus tax revenue, A+C, for a net excess burden of areas B+D.

For future reference, it is useful to present a very simple expression for the marginal excess burden of taxation. Totally differentiating the right side of (2.4) yields

$$\begin{aligned}
 dEB &= \frac{dE}{d\mathbf{p}} d\mathbf{p}_1 - \frac{dy}{d\mathbf{q}} d\mathbf{q}_1 - (\mathbf{p}_1 - \mathbf{q}_1)' \frac{d\mathbf{x}^c}{d\mathbf{p}} d\mathbf{p} - \mathbf{x}'(d\mathbf{p}_1 - d\mathbf{q}_1) \\
 (2.5) \quad &= \mathbf{x}^c(\mathbf{p}_1, U_1) d\mathbf{p}_1 - \mathbf{x}(\mathbf{q}_1) d\mathbf{q}_1 - (\mathbf{p}_1 - \mathbf{q}_1)' \frac{d\mathbf{x}^c}{d\mathbf{p}} d\mathbf{p}_1 - \mathbf{x}'(d\mathbf{p}_1 - d\mathbf{q}_1) = -\mathbf{t}' \frac{d\mathbf{x}^c}{d\mathbf{p}} d\mathbf{p}_1
 \end{aligned}$$

where the last step follows from the fact that $\mathbf{x}^c(\mathbf{p}_1, U_1) = \mathbf{x}(\mathbf{q}_1)$. That is, the change in excess burden equals the sum of the products of existing tax rates and changes in output. This result is extremely useful in searching for taxes that impose minimal excess burden. It is sometimes expressed as a first-order Taylor approximation for discrete changes, $-\mathbf{t} \Delta \mathbf{x}$, or a second-order approximation $-(\mathbf{t} \Delta \mathbf{x} + \frac{1}{2} \Delta \mathbf{t} \Delta \mathbf{x})$. The second-order approximation taken around the

undistorted point ($\mathbf{t} = 0$), with $\Delta \mathbf{t}$ set equal to the tax vector itself, approximates a measure of the total excess burden of the tax system (e.g. Harberger 1964a). From this approximation comes the common intuition that excess burden increases with the square of a tax. If one considers the second-order approximation for a single tax Δt_i and producer prices fixed, excess burden is $-\frac{1}{2} \Delta t_i (dx_i^c / dt_i) \Delta t_i$.

2.3. *Empirical issues in the measurement of excess burden*

While the theory of deadweight loss measurement has a long and colorful history that dates back to the nineteenth century contributions of Jules Dupuit (1844) and Fleeming Jenkin (1871/72), economists seldom measured actual deadweight losses prior to the pioneering work of Arnold Harberger in the 1950s and 1960s. In two influential papers published in 1964, Harberger (1964a) derived the approximation (2.5) used to measure deadweight loss and (1964b) applied the method to estimate deadweight losses due to income taxes in the United States. Harberger shortly thereafter (1966) produced estimates of the welfare cost of U.S. capital taxes. A generation of empirical studies by other scholars followed the publication of Harberger's subsequent survey article (1971).⁴

The empirical work that followed Harberger's efforts focussed on the use of simple deadweight loss formulas to estimate the welfare impact of a wide array of tax-induced distortions, including those to labor supply (Browning, 1975; Hausman, 1981a), saving (Feldstein, 1978), corporate taxation (Shoven, 1976), and the consumption of goods, such as housing and non-housing consumption items, that are taxed to differing degrees (King, 1983).⁵ In addition, some attention was devoted to refining the approximations used in applying

⁴ See Hines (1999) for an interpretive survey of this literature.

⁵ See the discussion in Auerbach (1985) and the more recent survey by Slesnick (1998).

estimated behavioral parameters to calculate deadweight losses. The variant of (2.5) used by Harberger, in which a form of uncompensated demand is used in place of compensated demand, approximates a compensated measure of welfare change (2.4). One question of interest to subsequent investigators is the practical difference between results obtained using Harberger-style approximations and those available from more exact measures. As Mohring (1971) and subsequent authors note, it is often the case that the same demand information necessary to calculate approximations to (2.5) can, if properly modified, be used to calculate Hicksian deadweight loss measures of the form (2.4). The extent to which these two methods generate different answers is, of course, an empirical question. Rosen (1978) finds that (2.4) and approximations to (2.5) track each other rather closely, but Hausman (1981b) offers some examples in which they differ considerably.

The generation of empirical work following Harberger calls attention to the importance of linking the strategy used to estimate demand and the ultimate goal of using the estimates to perform welfare analysis. Specifically, this entails estimating models that can be integrated to obtain expenditure functions from which expressions such as (2.4) can be derived.⁶ In the course of performing such estimation, it is of course desirable to make the model sufficiently flexible that its functional form imposes as few answers as possible. For this purpose it can be useful to employ algorithms that estimate expenditure functions numerically based on demand parameter estimates (Vartia, 1983).

A major practical difficulty in measuring the excess burden of a single tax, or of a system of taxes, is that excess burden is a function of demand interactions that are potentially very difficult to measure. For example, a tax on labor income is expected to affect hours worked, but

⁶ Examples of such estimation strategies include Deaton and Muellbauer (1980), Gallant (1981), and Jorgenson, Lau and Stoker (1982). Hausman and Newey (1995) offer a nonparametric alternative.

may also affect the accumulation of human capital, the intensity with which people work, the timing of retirement, and the extent to which compensation takes tax-favored (e.g., pensions, health insurance, and workplace amenities) in place of tax-disfavored (e.g., wage) form. In order to estimate the excess burden of a labor income tax, it is in principle necessary to estimate the effect of the tax on these and other decision margins. Analogous complications are associated with estimating the excess burdens of most other taxes. In practice, it can be very difficult to obtain reliable estimates of the impact of taxation on just one of these variables.

It is in reaction to the complicated nature of the problem of separately estimating the effect of taxation on all of a taxpayer's decision margins that a number of recent papers estimate variants of (2.5) in which the dependent variable is taxable income. The usefulness of this formulation is evident from considering the consumer's problem in maximizing

$$(2.6) \quad U(x_1, x_2, x_3, l),$$

in which x_1, x_2 , and x_3 are commodities taxed to differing degrees, and l is leisure. In order to illustrate the issues involved, we consider the case in which good 1 is an ordinary commodity that consumers purchase out of after-tax income, purchases of good 2 are fully deducted from taxable income, and purchases of good 3 are partially deductible for tax purposes. Given a labor endowment of \tilde{L} , a wage of w , and facing a (flat-rate, for purposes of simplicity) labor income tax rate of J , the consumer's budget constraint is

$$(2.7) \quad p_1 x_1 + p_2 x_2 (1 - \mathbf{t}) + p_3 x_3 (1 - \mathbf{at}) + w(1 - \mathbf{t})l \leq w(1 - \mathbf{t})\tilde{L},$$

in which \mathbf{a} denotes the degree to which purchases of x_3 are deductible for tax purposes. Feldstein (1999) notes that the budget constraint (here, 2.7) can be transformed to yield a variant of

$$(2.8) \quad \frac{p_1 x_1 + (1 - \alpha) p_3 x_3}{1 - t} \leq w(\tilde{L} - l) - p_2 x_2 - \alpha p_3 x_3.$$

The right side of (2.8) equals taxable income, since labor effort is given by $(\tilde{L} - l)$, purchases of commodity 2 are deductible from income, and a fraction α of purchases of commodity 3 is also deductible. In this environment, higher labor income tax rates create deadweight loss by discouraging consumption of good 1, and partially discouraging consumption of good 3, relative to consumption of leisure and of good 2. It is therefore possible to estimate deadweight loss by estimating the responsiveness of taxable income to changes in tax rates, since doing so traces the effect of changes in J on the numerator of the left side of (2.8).

Several empirical studies, including Lindsey (1987), Feldstein (1995), Auten and Carroll (1999), Goolsbee (2000), and Moffitt and Wilhelm (2000), consider the responsiveness of taxable income to tax rates, relying on major U.S. tax changes to provide variation in tax rates. The American tax reforms of 1981 and 1986 significantly reduced marginal tax rates, particularly those of high-income taxpayers, while tax reforms enacted in 1990 and 1993 had the opposite effect of raising tax rates on high-income taxpayers. The evidence indicates that taxable income is generally very responsive to tax changes, with estimated response elasticities that significantly exceed the typically very modest estimated effects of taxation on numbers of hours worked. Lindsey and Feldstein report elasticities of taxable income in excess of unity, while Auten and Carroll, Goolsbee, and Moffitt and Wilhelm provide a range of somewhat more modest estimates. All of these studies report that the taxable incomes of high-income taxpayers are far more responsive to tax rate changes than are the taxable incomes of the rest of the population.

There are two important considerations in interpreting this evidence. The first is that, in order to use the framework described by (2.7) as the basis of analysis, it is important to estimate the responsiveness to taxation of the present value of taxable income. Tax avoidance often takes the form of deferring a tax obligation from one period into another in order to reduce its present value. Consequently, the reaction of short-term taxable income to a tax change may exceed the reaction of the present value of taxable income, which Goolsbee (2000) finds occurred with executive compensation in response to the 1993 U.S. tax change. In addition to the difficulty of distinguishing empirically short-term from long-term reactions, there is the added complication that timing behavior depends on anticipated future tax policies that may not be known to the analyst.

The second consideration is that tax changes that reduce one type of taxable income may have offsetting or reinforcing effects on other sources of taxable income. For example, increasing the personal income tax rate may encourage some high-income taxpayers to incorporate their personal businesses, thereby reducing total income earned by individuals through proprietorships while increasing corporate income. A simple calculation of the responsiveness of personal income to changes in personal income tax rates would then overstate the true effect of tax changes on total taxable income. Furthermore, individuals purchase commodities that are taxed to differing degrees, and tax collections from these sources are appropriately included in reactions to tax changes.⁷ Properly accounting for all of these reactions when performing welfare analysis is a daunting task, but one that is more likely than many of the available alternatives to provide useful answers.

⁷ Note that (2.7) would be unchanged if expenditures on commodity 3 were nondeductible, but purchases of commodity 3 were subject to an ad valorem tax at rate ($-J$). As a general matter, however, preexisting distortions due to taxes, imperfect competition, and other sources of divergence between price and marginal cost should be incorporated in measuring deadweight loss.

3. The design of optimal taxes

Taxes (other than lump-sum taxes) distort behavior, yet society needs to collect revenue to pursue various social objectives. The optimal taxation literature identifies tax systems that minimize the excess burden of taxation, subject to various restrictions on tax instruments and information available to the government, and under different assumptions about population heterogeneity and the functioning of private markets.

Historically, there are three strands in the development of the optimal taxation literature. One, initiated by the seminal work of Ramsey (1927) and carried on, perhaps most notably, by Diamond and Mirrlees (1971), concentrates on the design of commodity taxes. A second set of contributions, beginning with Mirrlees (1971), considers more general nonlinear income taxes and focuses on the role of such taxes in addressing distributional concerns. Finally, the work of Pigou (1947) and others analyzes the use of taxes to address two types of market failures: financing “public” goods not provided by the private sector, and correcting externalities associated with incomplete private sector markets.⁸ Although these three strands in the literature have converged, it is still useful to consider them separately in turn before discussing their interrelationship.

3.1. The Ramsey tax problem

The simplest version of the Ramsey tax problem abstracts from population heterogeneity and posits that the government must raise a fixed sum of tax revenue with proportional

⁸ One potentially important market failure not considered by this chapter is the incompleteness of markets in state-contingent claims that might otherwise be used to diversify risks. In such a setting, it is possible for taxation to improve welfare simply by reducing (after-tax) private returns – since the government can pool risks through its tax and spending actions. Diamond, Helms and Mirrlees (1980), Varian (1980), and Eaton and Rosen (1980) analyze the properties of optimal distortionary taxation in stochastic settings with missing state-contingent markets, while Sandmo (1985) provides a more general survey of the impact of taxation in settings characterized by risk.

commodity taxes, leaving to the side how such revenue is to be spent. With a population of identical individuals, typically analyzed as a single representative individual, the goal of optimal tax design is to minimize the excess burden associated with raising the needed revenue. We typically rationalize government's inability to use lump-sum taxes by saying that such taxes are inequitable, although this may seem a bit forced in a setting with identical individuals. It may help to think of this simple problem as a necessary building block, rather than as one that adequately models a realistic situation.

The representative consumer maximizes utility, $U(\mathbf{x})$, over a vector of commodities x_i ($i = 0, 1, \dots, N$), subject to the budget constraint $\mathbf{p}\mathbf{x} \leq y$, where \mathbf{p} is the corresponding vector of consumer prices and y is lump-sum income. To raise the required level of revenue, R , the government imposes a vector of taxes on the commodities, \mathbf{t} , driving a wedge between consumer prices and producer prices, \mathbf{q} . It is useful to assume initially that this vector of producer prices is fixed (perhaps by world prices), but as will be seen later, this is not a restrictive assumption in characterizing the optimum. With given producer prices, the government in setting tax rates is effectively choosing the consumer price vector, since $\mathbf{p} = \mathbf{q} + \mathbf{t}$. Thus, the government's optimal tax problem can be modeled as

$$(3.1) \quad \max_{\mathbf{p}} V(\mathbf{p}, y), \text{ subject to } (\mathbf{p} - \mathbf{q})'\mathbf{x} \geq R$$

where $V(\mathbf{p})$ is the household's indirect utility function.

To see the relationship between the optimal tax problem and the problem of excess burden, note that the problem in (3.1) is equivalent to

$$(3.2) \quad \min_{\mathbf{p}} y - E(\mathbf{q}, V(\mathbf{p}, y)) - R, \text{ subject to } (\mathbf{p} - \mathbf{q})'\mathbf{x} \geq R$$

because y and R are constants and $E(\mathbf{q}, V(\mathbf{p}, y))$ is monotonically increasing in $V(\mathbf{p}, y)$. But, as $y \equiv E(\mathbf{p}, V(\mathbf{p}, y))$, expression (3.2) amounts to minimizing the excess burden of taxation subject to the revenue constraint, in which excess burden is evaluated at the utility level $V(\mathbf{p}, y)$ that holds in the presence of taxation (that based on the Hicksian equivalent variation⁹).

Without further restrictions, the optimal tax problem is actually quite trivial, since excess burden can be avoided entirely simply by raising all prices by a uniform multiple. That is, let $\mathbf{p} = \mathbf{f}\mathbf{q}$, with $\mathbf{f} > 1$ chosen so that $(\mathbf{f} - 1)\mathbf{q}'\mathbf{x} = R$. Then excess burden is

$$\begin{aligned}
 & E(\mathbf{f}\mathbf{q}, V(\mathbf{f}\mathbf{q}, y)) - E(\mathbf{q}, V(\mathbf{f}\mathbf{q}, y)) - (\mathbf{f} - 1)\mathbf{q}'\mathbf{x}(\mathbf{f}\mathbf{q}, y) \\
 (3.3) \quad & = \mathbf{f}E(\mathbf{q}, V(\mathbf{f}\mathbf{q}, y)) - E(\mathbf{q}, V(\mathbf{f}\mathbf{q}, y)) - (\mathbf{f} - 1)\mathbf{q}'\mathbf{x}(\mathbf{f}\mathbf{q}, y) \\
 & = (\mathbf{f} - 1)\mathbf{q}'\mathbf{x}(\mathbf{f}\mathbf{q}, y) - (\mathbf{f} - 1)\mathbf{q}'\mathbf{x}(\mathbf{f}\mathbf{q}, y) = 0
 \end{aligned}$$

where the second step follows from the fact that the expenditure function is homogeneous of degree 1 with respect to prices, and the third step from the identity $E(\mathbf{q}, V(\mathbf{f}\mathbf{q}, y)) \equiv \mathbf{q}'\mathbf{x}(\mathbf{f}\mathbf{q}, V(\mathbf{f}\mathbf{q}, y)) = \mathbf{q}'\mathbf{x}(\mathbf{f}\mathbf{q}, y)$.

Raising revenue in this way entails no excess burden because it is equivalent to imposing a lump-sum tax; the household's budget constraint in the presence of uniform taxation is

$$(3.4) \quad \mathbf{f}\mathbf{q}'\mathbf{x} = y \quad \Rightarrow \quad \mathbf{q}'\mathbf{x} = y - (\mathbf{f} - 1)y / \mathbf{f}$$

⁹ This measure of excess burden based on the equivalent variation may be used more generally to compare any two tax systems, neither of which is necessarily optimal. This property has led some (e.g. Kay 1980) to prefer its use over measures based on other reference-utility levels.

Thus, it is necessary to impose taxes that create excess burden only if it is impossible to adjust the tax rates freely on all $N+1$ commodities, or else if exogenous income $y=0$, in which case uniform taxes raise no revenue.¹⁰

What does it mean for consumers to have no exogenous income? The interpretation of the condition that $y=0$ depends on the definition of commodities \mathbf{x} . Consider, for example, the simple case of three commodities, including two that the household purchases, x_1 and x_2 , and a third, labor, that the household supplies as a factor to the production process. It is customary to write the budget constraint for this problem as

$$(3.5) \quad p_1x_1 + p_2x_2 + wl = w\tilde{L}$$

where l is leisure consumed and \tilde{L} is the household's time endowment. Households divide their time between leisure and working at a wage of w per unit of working time. With the budget constraint written this way, it is clear that a uniform tax on consumption and leisure is equivalent to a lump-sum tax on the household's time endowment. It is standard to rule this out by specifying that leisure cannot be taxed, that the government restricted to taxing labor, $L = \tilde{L} - l$. With such a restriction, if leisure is taxed, the government must offer a matching subsidy to the time endowment, a requirement that eliminates the possibility of lump-sum taxation. That is

(3.5) can be rewritten as

$$(3.6) \quad p_1x_1 + p_2x_2 + w(l - \tilde{L}) = p_1x_1 + p_2x_2 - wL = 0$$

in which it is clear that uniform taxes on x_1 , x_2 , and L raise no revenue. This result may seem counterintuitive because the "tax" on the household's leisure purchases *raises* the price of labor,

¹⁰ Note that if $y < 0$, it is possible to raise revenue with uniform taxation by choosing $\mathcal{F} < 1$.

corresponding to what we normally think of as a wage *subsidy*. It is possible to raise revenue by lowering the wage while raising prices p_1 and p_2 , but this no longer leaves relative prices undistorted – it lowers the real wage in terms of each consumption good. Indeed, a labor income tax and a uniform tax on the two consumption goods are equivalent tax policies. With the budget constraint expressed as

$$(3.7) \quad p_1 x_1 + p_2 x_2 = wL,$$

it is clear that raising commodity prices is the same policy as reducing wages.

Thus, the need to use distortionary taxes results either from a restriction on the use of tax instruments (e.g., it is not possible to tax leisure, or the consumption of any other endowed commodity, separately from its endowment) or on the absence of exogenous income (if labor, rather than leisure, is the relevant commodity). Because it is standard to assume that the government cannot impose separate taxes on endowments in labor or other commodities,¹¹ it is easier to adopt the second interpretation, expressing commodities as flows between the household and production sectors and leaving only “pure” economic rent potentially on the right side of the budget constraint.

With no lump-sum income, two tax systems are equivalent if they differ by proportional taxes on all commodities. Without lump-sum income one is therefore free to normalize one of the taxes, say on good 0, to zero, and for convenience choose the same good as numeraire, i.e., $q_0 = p_0 = 1$. The maximization problem in (3.1), with the multiplier λ associated with the budget constraint, yields N first-order conditions:

¹¹ It is customary simply to assume that the government cannot tax an individual’s labor endowment because this endowment is not observable; equivalently, we assume that we can observe an individual’s labor income, but not the effort expended or leisure forgone in earning that income. Although there has been some work considering modifications of this assumption (e.g., Stern 1982), this issue has received relatively little attention in the literature.

$$(3.8) \quad -\mathbf{l}x_i + \mathbf{m}x_i + \sum_j t_j \frac{dx_j}{dp_i} = 0 \quad i = 1, \dots, N$$

in which $\mathbf{l} \equiv \mathcal{N}(\mathbf{p}, y)/\mathcal{Y}$ is the marginal utility of income. Making use of the Slutsky decomposition, (3.8) implies

$$(3.9) \quad \sum_j t_j S_{ji} = -\frac{(\mathbf{m} - \mathbf{a})}{\mathbf{m}} x_i \quad i = 1, \dots, N$$

where S_{ji} is the j^{th} element of the Slutsky matrix $S \equiv \frac{d\mathbf{x}^c}{d\mathbf{p}}$ and $\mathbf{a} = \mathbf{l} + \mathbf{m} \sum_j t_j \frac{dx_j}{dy}$ is the

“social” marginal utility of income that includes the value of the additional tax revenue raised when the household receives another unit of income.¹²

Although there is no independent condition for good 0, it may be shown (see Auerbach 1985) that the N first-order conditions in (3.9) imply a comparable condition for good 0, a result that should not be too surprising given that the choice of the good to bear the zero tax is arbitrary. Stacking these $N+1$ conditions yields

$$(3.10) \quad S\mathbf{t} = -\left(\frac{\mathbf{m} - \mathbf{a}}{\mathbf{m}}\right)\mathbf{x}$$

Premultiplying both sides of (3.10) by the tax vector \mathbf{t} , we obtain an equation in which the left side is a negative semi-definite quadratic form and the right side equals the product of the

¹² Samuelson (1951) uses the symmetry of the Slutsky matrix ($S_{ij} = S_{ji}$) to interpret (3.9) as implying that optimal taxes entail equiproportionate compensated reductions in demands for all commodities. While valid locally, this interpretation relies on constancy of the elements of the Slutsky matrix as tax rates change, a feature they do not generally exhibit.

constant term $-(\mathbf{m}\mathbf{a})/\mathbf{m}$ and tax revenue $\mathbf{t} \cdot \mathbf{x}$ ¹³. Thus, if revenue is positive, $\mathbf{m}^3 \mathbf{a}$ – the marginal social cost of raising additional revenue, \mathbf{m} is at least as large as the cost of raising revenue in lump-sum fashion, \mathbf{a} , i.e., marginal excess burden is nonnegative. This condition does not hold for arbitrary tax schedules, but starting from an optimal tax system for any given level of revenue means that there is no opportunity to reduce excess burden while raising taxes, for example by bringing up the tax rates on goods that initially are undertaxed.¹⁴ Note that this inequality relates \mathbf{m} to \mathbf{a} , not to \mathbf{l} , the private marginal utility of income. By the definition of \mathbf{a} , $\mathbf{m}^3 \mathbf{a} \geq \mathbf{m}^3 \mathbf{l}$ only if revenue is nondecreasing in income, i.e., if the tax base is a normal composite good. This distinction is important to keep in mind when considering the literature that seeks to identify the “marginal cost of funds.”

Before interpreting expression (3.10) further, it is useful to consider the more general case of variable producer prices.

3.2. *Changing producer prices*

Since the excess burden of a tax is a function of the extent to which the tax changes producer prices, it follows intuitively that allowing producer prices to vary alters the first-order conditions for the optimal tax schedule. Let the general production be characterized by

$$(3.11) \quad f(\mathbf{z}) \geq 0$$

where \mathbf{z} is the production vector and perfect competition insures that $q_i/q_j = f_i/f_j \forall i,j$. Without loss of generality, the units of the production function can be chosen such that $q_i = f_i$. If there are

¹³ Because the first element of the tax vector is zero, the relevant part of the Slutsky matrix is the submatrix formed by striking the first row and column of S . This submatrix and the associated quadratic form will generally be negative definite, as long as some of the omitted substitution terms are nonzero.

constant returns to scale, then $f(\mathbf{z})$ is homogeneous of degree zero in \mathbf{z} . Otherwise, there may be pure profits, $y = \mathbf{q} \cdot \mathbf{z} > 0$.

With changing producer prices, it is not appropriate to specify the constraint in the optimal tax problem as a scalar value of tax revenue to be collected, so it is necessary to posit that the government absorbs a vector \mathbf{R} of commodities. This implies that the consumption vector \mathbf{x} satisfies $f(\mathbf{x} + \mathbf{R}) \leq 0$, thereby incorporating both revenue and production constraints. The optimal tax problem, then, is to maximize the indirect utility function $V(\mathbf{p}, y)$ subject to this constraint, and not that given in (3.2). The associated Lagrangean expression is

$$(3.12) \quad V(\mathbf{p}, y) - \mathbf{m} f(\mathbf{x} + \mathbf{R})$$

and the government's problem is still that of choosing the consumer price vector \mathbf{p} , rather than the tax vector \mathbf{t} , even though the relationship between changes in the two vectors is more complicated than when producer prices are fixed.¹⁵ The resulting first-order conditions are (using the normalized form of production function)

$$(3.13) \quad -\mathbf{I}x_i + \mathbf{I} \frac{dy}{dp_i} + \mathbf{m} \left[-\sum_j q_j \frac{dx_j}{dp_i} \right] = 0 \quad i = 1, \dots, N$$

Differentiating the household's budget constraint $\mathbf{p} \cdot \mathbf{x} = y$ with respect to p_i yields

$$(3.14) \quad x_i + \sum_j p_j \frac{dx_j}{dp_i} - \frac{dy}{dp_i} = 0 \quad i = 1, \dots, N$$

¹⁴ Note that marginal excess burden is nonpositive when revenue is initially negative, because raising revenue means reducing the level of distortions caused by subsidies.

and adding the left side of this equation to the expression inside the brackets in (3.13) yields

$$(3.15) \quad -\mathbf{I}x_i + \mathbf{I} \frac{dy}{dp_i} + \mathbf{m} \left[x_i + \sum_j t_j \frac{dx_j}{dp_i} - \frac{dy}{dp_i} \right] = 0 \quad i = 1, \dots, N$$

Since producer prices, and hence profits, change with \mathbf{p} , the derivative dx_j/dp_i in (3.15) includes the indirect effect of p_i on profits through changes in production:

$$(3.16) \quad \frac{dx_j}{dp_i} = \frac{\mathcal{J}x_j}{\mathcal{J}p_i} + \frac{dx_j}{dy} \cdot \frac{dy}{dp_i}$$

Using this and the Slutsky decomposition, (3.15) can be rewritten as

$$(3.17) \quad -\sum_j t_j S_{ji} = \frac{(\mathbf{m} - \mathbf{a})}{\mathbf{m}} \left(x_i - \frac{dy}{dp_i} \right) \quad i = 1, \dots, N$$

which differs from expression (3.9), the first-order condition in the case of fixed producer prices, by the term dy/dp_i on the right side. Thus, if there are constant returns to scale ($y \equiv 0$), the first-order conditions are identical (Diamond and Mirrlees 1971). The same is true if the government imposes a pure profits tax, so that the after-tax value of y accruing to households is uniformly zero (Stiglitz and Dasgupta 1971).

From expression (2.5), the left side of (3.17) equals the marginal excess burden associated with an increase in p_i . The second term on the right side of (3.17) is the net

¹⁵ As discussed in Auerbach (1985), $d\mathbf{p}/dt = [I-HS]^{-1}$, where H is the Hessian of $f(\mathbf{x})$, so there is a one-to-one relationship between changes in t and changes in \mathbf{p} as long as $[I-HS]$ is of full rank.

compensation required to maintain the individual's utility as p_i rises¹⁶ which, by definition, exceeds the marginal revenue raised by the marginal excess burden induced by the price change. Thus, (3.17) says that the excess burden of a marginal increase in any tax must be proportional to the sum of marginal revenue plus marginal excess burden, or:

$$(3.18) \quad \frac{dEB}{dp_i} = \frac{(\mathbf{m} - \mathbf{a})}{\mathbf{m}} \left(\frac{dR}{dp_i} + \frac{dEB}{dp_i} \right) \quad i = 1, \dots, N$$

It follows that the marginal excess burden per dollar of revenue raised, $(\mathbf{m} - \mathbf{a})/\mathbf{a}$, is also constant,

$$(3.19) \quad \frac{dEB}{dp_i} = \frac{(\mathbf{m} - \mathbf{a})}{\mathbf{a}} \frac{dR}{dp_i} \quad i = 1, \dots, N$$

which is an intuitive condition for minimizing the total excess burden induced by raising a given amount of revenue from alternative sources.

3.3. *The structure of optimal taxes*

The optimal tax rules just derived generally do not imply that the government should impose taxes at uniform rates, even in the simple case in which producer prices are fixed. For example, consider the three-good case, in which the two first-order conditions (3.9) yield

$$(3.20) \quad \frac{t_1}{t_2} = \frac{-S_{22}x_1 + S_{12}x_2}{-S_{11}x_2 + S_{21}x_1}$$

¹⁶ This term equals $-\frac{dV(\mathbf{p}, y)/dp_i}{dV(\mathbf{p}, y)/dy}$; according to Roy's identity, this equals the net increase in income required to maintain the household's utility level as p_i increases.

which, using the fact that $\sum_j p_j S_{ij} = 0$, and defining $\mathbf{q} \equiv t_i/p_i$ as the tax rate on good i , may be rewritten as

$$(3.21) \quad \frac{\mathbf{q}_1}{\mathbf{q}_2} = \frac{\mathbf{e}_{20} + \mathbf{e}_{21} + \mathbf{e}_{12}}{\mathbf{e}_{10} + \mathbf{e}_{21} + \mathbf{e}_{12}}$$

where \mathbf{e}_{ij} is the compensated cross-price elasticity of demand for good i with respect to the price of good j .

This expression indicates that two goods should be taxed at equal rates (i.e., $\mathbf{q}_1 = \mathbf{q}_2$) if and only if the goods are equally complementary with respect to the untaxed good 0. The intuition sometimes offered for this result comes from the case in which the untaxed good 0 is labor, making it desirable to tax more heavily the good that is more complementary with leisure because it is impossible to tax leisure directly. But since expression (3.20) would also apply if a consumption good were chosen to bear the zero tax, it may be more accurate to say that complements to untaxed goods are taxed more heavily to achieve reductions in the untaxed goods without taxing them directly.

In the special case of zero cross-elasticities among all taxed goods, the first-order conditions (3.9) yield the “inverse elasticity rule” that $\mathbf{q} \mu 1/\mathbf{e}_i$, since in this case each good’s demand responds only to its own tax, so achieving a reduction of equal proportion means keeping $\mathbf{q} \mathbf{e}_i$ constant.

3.4. *An example*

Suppose that household preferences over goods and leisure are described by the Stone-Geary utility function,

$$(3.22) \quad U(x_1, x_2, l) = (x_1 - a_1)^{b_1} (x_2 - a_2)^{b_2} l^{1-b_1-b_2}$$

For this utility function, the cross elasticity ϵ_{0i} equals $(1-b_1-b_2)(1-a_i/x_i)$, so optimal taxes fall more heavily on the consumption good whose “basic need” a_i represents a larger portion of total consumption x_i . In terms of underlying preferences, it can be shown that this is equivalent to taxing more heavily the good with the higher value of $p_i a_i / b_i$, the good for which expenditures on basic needs are a greater fraction of the good’s discretionary budget share, b_i . In the special case where $a_1 = a_2 = 0$, the Stone-Geary utility function collapses to the Cobb-Douglas function, and uniform taxes are optimal. The Cobb-Douglas utility function is separable into goods and leisure (or, to be more exact, into the taxed and untaxed commodities) and homogenous in goods – it can be written in the form $U(\mathbf{f}(\mathbf{x}), l)$, where $\mathbf{f}(\cdot)$ is a homogeneous function. This homothetic separability is a sufficient condition for uniform taxation (Atkinson and Stiglitz 1972). Separability alone does not suffice – as the general Stone-Geary example illustrates.

3.5. *The production efficiency theorem*

All of the tax instruments considered so far are proportional taxes on transactions between the household sector and the production sector. Production itself is assumed to face no distortions, and perfect competition ensures that the economy achieves a point on the production frontier. However, the government has access to policies that distort production while raising revenue, either through explicit taxes or through government production schemes that allocate inputs and outputs on the basis of criteria possibly different than those used by the private sector. One might think that such policy instruments would favorably augment the government’s options, but this may well not be so.

Consider the case in which there is a second production sector, say controlled directly by the government, with production function $g(\mathbf{s})$ and production vector \mathbf{s} , with the production set defined by $g(\mathbf{s}) \leq 0$. Distortions between the two sectors occur implicitly through the government's choice of the vector \mathbf{s} , with each sector, but not necessarily the two sectors in combination, assumed to be on its own production frontier. Further assume that production in both sectors is subject to constant returns to scale.

Because private production now equals the difference between purchases $\mathbf{x} + \mathbf{R}$ and government production \mathbf{s} , the government's problem is to maximize $V(\mathbf{p}, y)$ subject to $f(\mathbf{x} + \mathbf{R} - \mathbf{s}) \leq 0$ and $g(\mathbf{s}) \leq 0$. Forming the Lagrangean as before, with the multiplier \mathbf{z} associated with the second sector's production, we obtain the same first-order conditions as before with respect to \mathbf{p} , and the conditions that $\mathbf{p}f_i - \mathbf{z}g_i = 0 \forall i$ with respect to the vector \mathbf{s} . This implies that all marginal rates of substitution in production should be equal, $f_i/f_j = g_i/g_j$, i.e., production should not be distorted. This result does not hold if there are pure profits received by the household, and this helps provide insight into why it *does* hold when no such profits are received. In this special case, all household decisions are based on the relative price vector \mathbf{p} . It is possible to bring about any configuration of this vector that is consistent with the revenue constraint, without resorting to production distortions. Thus, production distortions can serve only to reproduce what can already be achieved, but with the additional social cost of lost production. Of course, if the government is not free to adjust all relative prices directly, it may find production distortions useful, and political realities may often dictate such an indirect policy.

3.7. *Distributional considerations*

The rules derived thus far apply to the case of identical individuals, but heterogeneity with respect to taste and ability is an important consideration. Taking account of individual

differences in a population of H individuals means replacing the indirect utility function of the representative individual, $V(\mathbf{p}, y)$, with a social welfare function, $W(V^1(\mathbf{p}, y^1), \dots, V^H(\mathbf{p}, y^H))$. With either fixed producer prices or constant returns to scale, there is no lump-sum income y^h and social welfare is still simply a function of the price vector \mathbf{p} . This has the immediate implication that the production efficiency theorem just derived still holds, because there is no scope for improving social welfare once the price vector is established through the optimal tax vector \mathbf{t} . However, the shape of the social welfare function influences the choice of \mathbf{t} itself.

The first-order conditions corresponding to maximizing this social welfare function subject to the revenue constraint in (3.1) are analogous to those in (3.8):

$$(3.23) \quad -\sum_h W_h \mathbf{I}^h x_i^h + \mathbf{m} \left[x_i + \sum_j t_j \sum_h \frac{dx_j^h}{dp_i} \right] = 0 \quad i = 1, \dots, N$$

where W_h is the partial derivative of W with respect to the utility of individual h , \mathbf{I}^h is individual h 's marginal utility of income, and x_i^h is individual h 's consumption of good i . Again defining $\mathbf{a}^h \equiv W_h \mathbf{I}^h + \mathbf{m} \sum_j t_j \frac{dx_j^h}{dy^h}$ as individual h 's social marginal utility of income, (3.23) can be expressed in more compact form (Diamond 1975) as

$$(3.24) \quad \sum_j t_j S_{ji} = -\frac{(\mathbf{m} - \tilde{\mathbf{a}})}{\mathbf{m}} x_i \quad i = 1, \dots, N$$

where $S_{ji} = \sum_h S_{ji}^h$ is an aggregation of comparable terms from individual Slutsky matrices and

$$(3.25) \quad \tilde{\mathbf{a}}_i \equiv \sum_h \left(\frac{x_i^h}{x_i} \right) \mathbf{a}^h$$

is the social marginal utility of income taken from households via a tax on good i . It is higher, the greater the share of the tax burden borne by individuals with a high social marginal utility of income, which is typically thought to be those of lower income.

Equation (3.24) is easy to understand by reference to (3.18), which still holds in this case, for $\tilde{\mathbf{a}}$ in place of \mathbf{a} . Now, the marginal excess burden, rather than being equal for each source of funds, should be reduced for those commodities for which the associated loss in real income is costly ($\tilde{\mathbf{a}}$ is high). Because the ultimate objective is to equalize \mathbf{m} across sources of revenue, those with higher distributional costs should have lower efficiency costs.

To illustrate this trade-off between equity and efficiency in the choice of tax structure, consider again the three-good case in which two consumption goods are taxed. Now, the ratio of the tax rates on the two goods should satisfy

$$(3.26) \quad \frac{\mathbf{q}_1}{\mathbf{q}_2} = \frac{\mathbf{p}_1 \mathbf{e}_{20} + \mathbf{p}_2 \mathbf{e}_{12}}{\mathbf{p}_2 \mathbf{e}_{10} + \mathbf{p}_1 \mathbf{e}_{21} + \mathbf{p}_2 \mathbf{e}_{12}}$$

where $\mathbf{p}_i \equiv (\mathbf{m} \tilde{\mathbf{a}}_i) / \mathbf{m}$. Here, $\mathbf{q}_1 > \mathbf{q}_2$ if and only if $\mathbf{e}_{10} / \mathbf{e}_{20} < \mathbf{p}_1 / \mathbf{p}_2$. If the good most complementary with leisure is also the good with the greater social valuation $\tilde{\mathbf{a}}$, it is not clear which good will be taxed more heavily – the answer depends in part on the strength of distributional preferences.

If preferences satisfy the restriction of homothetic separability mentioned above in section 3.4, it will still be true that commodity taxes should be uniform (as long as preferences over consumption are the same across individuals). When preferences take this form, Engel curves (relating consumption to income) are linear and pass through the origin. Thus, there will be no variation in the relative budget shares of different goods among individuals of different

abilities, and hence nothing to be gained from a distributional perspective by imposing differential taxation; this leaves the optimality of uniform taxation undisturbed.

An instance in which distributional preferences necessarily work in the opposite direction of minimizing excess burden is that in which the social welfare is the sum of individual utilities and individuals have identical Stone-Geary utility functions of the type considered in the example above, differing only with respect to ability (as measured by the wages received per unit of labor supplied). To see this, note first that the ordinary demand functions $x_i(\mathbf{p}, y)$ are linear in income. Thus, the change in tax revenue generated when a household changes its consumption in response to receiving a dollar of income is constant across households. This implies that differences in $\tilde{\mathbf{a}}$ arise only from differences in consumption patterns of households with differing social marginal utilities of income ($W_h \mathbf{I}_h = \mathbf{I}_h$). Next, note that the derivative of good- i consumption with respect to household utility is $dx_i^c(\mathbf{p}, U) / dU = (x_i - a_i) / U$, so that the elasticity of x_i with respect to U is $(x_i - a_i) / x_i$. Thus, the good with the higher elasticity of consumption with respect to utility – the good more concentrated among higher-utility individuals and hence with the lower value of $\tilde{\mathbf{a}}$ – is the good with the lower value of a_i relative to x_i and therefore has a higher demand cross-elasticity with respect to leisure. Thus, the good that is desirable to tax more heavily for distributional reasons is also the good that is desirable to tax less heavily for efficiency reasons.

4. Income taxation

4.1. Linear income taxation

In analyzing taxes on a representative individual, it was convenient to side step the question of why the government might not be able to use lump-sum taxes. With population

heterogeneity now an explicit aspect of the analysis, it is appropriate to revisit this question. In practice, governments include *uniform* lump-sum taxes among their tax instruments. Indeed, the use of lump-sum taxes permits the introduction of the most rudimentary of progressive income taxes, the *linear income tax*. For example, in the three-good case considered earlier, with the household's budget constraint given by (3.7) and suitably modified by introducing a lump-sum tax and choosing one of the consumption goods (good 1) as the untaxed numeraire commodity, the household faces the budget constraint:

$$(4.1) \quad q_1 x_1 + \frac{q_2}{1 - \mathbf{q}} x_2 = -T + \frac{w}{(1 - \mathbf{q})} L = wL - (T + \mathbf{t}wL)$$

where $\mathbf{t} = -\mathbf{q}/(1 - \mathbf{q})$ is the household's marginal income tax rate. As (4.1) shows, the government has the option of using differential commodity taxation to supplement the linear income tax schedule. This leads immediately to two questions. First, when will the government wish to use the commodity tax \mathbf{q} or, for the case of several commodities 1, ..., N , the commodity taxes $\mathbf{q}_1, \dots, \mathbf{q}_N$? Second, under what conditions will the income tax be progressive, with average tax rates rising with income (e.g., with $T < 0$)?

In answer to the first question, a sufficient condition for the optimality of uniform commodity taxes or, equivalently, taxes only on labor income, is that preferences are weakly separable into goods and leisure, and that commodities have linear Engel curves with identical slopes across households (Deaton 1979).¹⁷ Such preferences include the case of homothetic separability, for which Engel curves pass through the origin. It is noteworthy that this condition is the same as that required for exact aggregation of consumers, and that for an aggregate

¹⁷ An example is the Stone-Geary utility function considered above.

measure of excess burden to be independent of the distribution of resources across consumers. Note also that a weaker condition suffices with a nonlinear income tax schedule, the design of which is discussed below. In that case, it is possible to dispense with the requirement that Engel curves be linear, since weak separability of goods and leisure suffices (Atkinson and Stiglitz 1976).

If the government taxes only labor income, then equation (3.25) implies (because purchases of labor are negative) that

$$(4.2) \quad t_0 S_{00} = -\frac{(\mathbf{m} - \tilde{\mathbf{a}}_0)}{\mathbf{m}}(-L)$$

where L and S_{00} are aggregate measures, with labor measured in efficiency units so that it is possible to aggregate over individuals of different abilities. The availability of lump-sum taxes adds a marginal condition that $\mathbf{m} = \bar{\mathbf{a}}$, the unweighted average value of \mathbf{a} across individuals: since the government can use positive or negative lump-sum taxes at the margin, the marginal cost of funds must equal the cost of raising funds with lump-sum taxes. Substituting this condition into (4.2) and rearranging terms yields

$$(4.3) \quad \frac{(-t_0) p_0 (-S_{00})}{p_0 L} = -\frac{(\tilde{\mathbf{a}}_0 - \bar{\mathbf{a}})}{\bar{\mathbf{a}}}$$

which, for a household labor price of $p_0 = w(1 - \mathbf{t})$ and $t_0 = -\mathbf{t}w$ (recall that in this notation a positive value of t_0 raises the after-tax wage rate) may be expressed (Dixit and Sandmo 1977) as

$$(4.4) \quad \frac{\mathbf{t}}{(1 - \mathbf{t})} = -\frac{(\tilde{\mathbf{a}}_0 - \bar{\mathbf{a}})}{\bar{\mathbf{a}}\bar{\mathbf{e}}} = -\frac{cov\left(\frac{L^h}{L}, \frac{\mathbf{a}^h}{\bar{\mathbf{a}}}\right)}{\bar{\mathbf{e}}}$$

where $\bar{\epsilon} \equiv w(1 - \tau)(-S_{00}) / L$ is the aggregate compensated labor supply elasticity (which must be positive), L^h is household h 's labor supply, and \bar{L} is the average value of L^h across households. Since labor is expressed in efficiency units (at the common wage w), higher ability translates, for a given fraction of time worked, into higher labor supply. Expression (4.4) says that the marginal tax rate on labor income is positive if and only if the marginal social valuation of income falls as labor supply (in efficiency units) rises, a condition that is met by utilitarian social welfare functions together with labor supply schedules that are increasing in ability.

The value of marginal tax rate, and whether it is sufficiently high to make the linear income tax progressive ($T < 0$), depends on the weight of the social welfare function's redistributive component – how fast a^h declines as L^h rises. Properties of the marginal tax rate also depend on the amount of tax revenue required. To understand why, consider the case in which the government's revenue requirement is zero. Then it is possible to obtain a Pareto optimum by setting the marginal income tax rate, and the lump-sum tax T , to zero. Since the social marginal utility of income differs across individuals, and since there is no first-order excess burden from the introduction of a small tax, it must then be optimal to introduce some distortion (i.e., a positive marginal tax rate) to redistribute income from those with high incomes and low social marginal utility of income to those with lower incomes and higher social marginal utility of income. Thus, the linear income tax is progressive at zero net revenue. As the government's revenue requirement rises, holding T constant, the marginal excess burden of raising revenue also rises, and so too does the cost of redistribution. As Stiglitz (1987) notes, there exists a point at which maximum revenue is collected via marginal tax rates (i.e., the marginal excess burden per dollar of revenue is infinite), at which point the government *must*

rely on lump-sum taxes for additional revenue. Greater reliance on lump-sum taxes obviously reduces the progressivity of the tax schedule. Indeed, simulations confirm that the lump-sum transfer falls as revenue rises (Stern 1976), and that it becomes negative for sufficiently high revenue requirements (Slemrod et al. 1994).

4.2. *Nonlinear income taxation: introduction*

In practice, governments use income tax systems with multiple marginal tax rates. Although the linear income tax just considered can have progressive average tax burdens, its redistributive potential is limited by the fact that the average tax burden must approach the marginal tax rate asymptotically and can rise no higher. Historically, many in government have felt that only a schedule of rising marginal tax rates could deliver the appropriate degree of progressivity toward the top of the income distribution, and have implemented income tax systems with top marginal tax rates in some instances exceeding 90 percent.¹⁸

Governments certainly can impose income tax systems more complicated than the linear income tax, but what should these systems look like? As in the case of the linear income tax, the issue involves balancing efficiency and equity, with the surprising conclusion that high and rising marginal tax rates may well not be appropriate even when the government has a strong redistributive motive.

At first, it might seem that the ability to choose an arbitrary income tax function $T(\cdot)$ offers the government the opportunity to impose individual-specific lump-sum taxes, for the function could be chosen to pass through values of tax burdens appropriate to individuals at each

¹⁸ For example, just prior to the Kennedy-Johnson tax cut of 1964, the top marginal federal income tax rate in the United States was 91 percent.

level of income. However, as is rapidly apparent, the endogeneity of income strongly limits the government's ability to impose differential lump-sum taxation.

To begin, suppose that there is a single consumption good, that labor supply is the only source of income, and that individuals have common preferences $U(c,l)$ over consumption and leisure, differing only in their abilities, as measured by wage rates w . Imagine that the government needs to raise a certain amount of revenue, R , using an income tax, and that it is desirable to assign a lump-sum income tax burden T_i to individual i . With the consumption good as numeraire, the problem may be expressed as

$$(4.5) \quad \max_{\mathbf{T}} W(V^1(w^1, -T^1), V^2(w^2, -T^2), \dots, V^H(w^H, -T^H)) \text{ subject to } \sum_h T^h \geq R$$

If \mathbf{m} is the Lagrange multiplier associated with the revenue constraint, then the H first-order conditions are simply that $W_h \mathbf{I}^h = \mathbf{m}$ —that the marginal social utility of income is the same across all individuals.

What does this condition imply for tax burdens? For the utilitarian social welfare function $W(U^1, \dots, U^H) = \sum_h U^h$, it implies that the marginal utility of income \mathbf{I}^h is constant across individuals, which (from the first-order conditions for utility maximization) implies that the marginal utility of consumption is constant across households, but that the marginal utility of leisure is proportional to w^h . Equating the marginal social cost of *income* across individuals, the government in effect forces high-wage individuals to work until they reach the point that leisure is very valuable to them. In the process, this tax system makes high-wage individuals *worse off* than low-wage individuals, a paradoxical outcome that is guaranteed if leisure is a normal good.

For example, suppose the common utility function takes the quasi-linear form $U(c,l) = c - v(1-l)$, with $v > 0$ and $v > 0$. Then, with optimal household-specific taxation, all households

have the same level of consumption, and leisure declines monotonically with the wage rate. The lowest wage household obtains the highest level of utility, which illustrates quite clearly the problem to be faced in attempting to implement such a tax system. Aside from the political implausibility of the outcome, this scheme could be implemented only if government knew each household's ability level and assigned taxes accordingly. Otherwise, all other households would have incentives simply to masquerade as the household with the lowest ability by supplying the amount of labor necessary to produce that household's income level, thereby leaving themselves better off than the lowest-ability household (because they forgo less leisure to reach this level of income), rather than worse off. But this, in turn, leaves the government with a uniform lump-sum tax and too little revenue. While the government could respond by increasing the lump-sum tax, it is clear from the previous discussion of the linear income tax that this policy alone is not likely to be optimal. Rather, the government seeks to impose a tax system more progressive than the lump-sum tax, while still accounting for the absence of information about individual types and the endogeneity of household income. A linear income tax is but one such tax system.

4.3. *Nonlinear income taxation: graphical exposition*

Much of the intuition behind the design of the optimal nonlinear income tax emerges from consideration of an income tax imposed on an economy composed of two individuals, one (H) of high ability and one (L) of low ability.¹⁹ Because the government observes only income, $Y=w(1-l)$, rather than labor supply and wage rates separately, it is useful to express each individual's preferences over consumption and leisure (or labor) in terms of preferences over consumption and income, as depicted in Figure 4.1. On the left side of the figure is an

¹⁹ We follow the mnemonic notation in the literature in denoting the two ability classes as H and L for the following graphical exposition, but remind the reader that the variable L represents labor supply in all other parts of the chapter.

indifference curve over consumption and leisure, based on the utility function $U(c,l)$. On the right are two corresponding indifference curves for the same level of utility but different wage rates, based on the same utility function, $U(c,1-y/w)$. The curve corresponding to the higher wage rate is flatter because a given change in labor translates into a greater change in income. This suggests that when indifference curves of two individuals do cross, as at point A, the indifference curve of the higher ability individual is flatter.

Figure 4.2 illustrates the outcome of attempting to impose the previously-discussed lump-sum tax “solution”, with consumption equal to c_0 for both high- and low-ability individuals and the higher ability type on a lower indifference curve, as indicated by the relative consumption at zero income (at which ability differences are irrelevant). Rather than accept the bundle (c_0, y^H) , the high-ability household would prefer to earn income y^L and receive the same level of consumption. The problem with this plan is that it violates the *self-selection constraint* that each household prefer its government-designated bundle among the available options. In this instance, the high-ability household prefers the bundle designated for the low-ability household. It is typically the self-selection constraint of the high-ability person with which the government must be concerned.

As Figure 4.3 illustrates, the self-selection constraints limit the scope for redistribution through differential lump-sum taxation. For the sake of exposition, assume that the required level of revenue, R , equals zero. With no redistribution, each household’s budget constraint has unit slope (since a dollar of income produces a dollar of consumption) and passes through the origin. The high-ability and low-ability households choose points H and L, respectively. Each household strictly prefers its own bundle, so neither self-selection constraint is binding. As a result, it is possible to impose a lump-sum tax on H and provide an equal lump-sum transfer to L

until reaching the point that H 's self-selection constraint binds, which occurs at points H_1 and L_1 . The government cannot do more with lump-sum taxation *without* violating H 's self-selection constraint, but it *can* do more.

Slopes of the indifference curves of individuals H and L differ at point L_1 . Because this point is an optimum for L (since L 's indifference curve is tangent to the budget line) but not for H , a slight movement in any direction along the budget line has no first-order effect on the utility of L , but does have a first-order effect on the utility of H . Moving toward the origin along the budget line makes H worse off, because H already is working inefficiently “too little” at point L_1 – H 's indifference curve is flatter than the budget line. This suggests a way to relax H 's self-selection constraint and achieve more redistribution, as illustrated in Figure 4.4. By shifting individual L from point L_1 to point L_2 , the government imposes on L only a “second-order” excess burden (since L is initially at an undistorted point) but raises “first-order” tax revenue by being able to shift individual H down to point H_2 . This tax revenue equals the distance CD in Figure 4.4. The extra revenue extracted from H (net of the amount – distance AB in Figure 4.4 – needed to compensate for the small distortion to L 's choice) can then be allocated between L and H , with H receiving just enough to keep the self-selection constraint satisfied. The final result is that L is better off than at L_1 and H is worse off than at H_1 .

The limits that govern this redistribution are the government's success in carrying it out (which reduces disparities in the social valuation of marginal incomes received by different households) and by marginal excess burdens that rise as one moves further away from the initial point L_1 . L 's bundle can be thought of as being implemented via a marginal tax rate on L 's income that produces a budget line with slope less than one. This offers the insight that it is optimal to impose a positive marginal tax rate on individual L not to raise revenue from L , but to

raise revenue from those with incomes higher than L 's – in this case, individual H . A corollary is that, as there is no one of higher ability than H in this example, it is not optimal to impose a marginal tax rate on H 's income. Doing so would distort H 's behavior and reduce the revenue the government could extract from H without violating H 's self-selection constraint. These lessons are useful in considering the case in which there is a continuum of agents.

4.4. *Nonlinear income taxation: mathematical derivation*

The mathematics of optimal income taxation with a continuum of agents is not straightforward, because it is not possible to rule out such phenomena as nondifferentiability of the tax function $T(\cdot)$. These phenomena are not simply “anomalies.” As discussed in Stiglitz (1987), nondifferentiability arises in cases in which it is optimal to pool individuals with different skill levels at a single point in (c,y) space. To understand why, consider the case in which there are many individuals of type H (as considered above) and an equal number of individuals of type L . The optimal tax policy is obviously identical to that with one individual of each type. Then introduce an additional individual at some intermediate wage rate between L and H . If this individual, say M , is offered an allocation that H prefers to L 's bundle, then H 's self-selection constraint is violated. It is possible to maintain individuals of type H at their initial allocations only by reducing the attractiveness of M 's bundle. This, in itself, distorts the choice of M 's bundle, but if there are many more individuals of types H and L than of type M , society gains from doing so until M 's bundle approaches that of L .

In spite of the importance of this complication, it is useful for intuition to derive results for cases in which such problems do not arise. Our approach closely follows that in Atkinson and Stiglitz (1980). For further discussion of the more general mathematical issues, see Mirrlees (1976, 1986).

Continuing to assume, for simplicity, that overall revenue $R = 0$, the government seeks to maximize some general social welfare function of individual utilities, subject to the constraint that total consumption equal total before-tax income. Letting $f(w)$ be the fraction of the population endowed with wage rate/skill level w , the government's objective is

$$(4.6) \quad \max \int_w G(U(w))f(w)dw \quad \text{subject to} \quad \int_w (c(w) - y(w))f(w)dw \leq 0$$

where $c(w)$ and $y(w)$ are the levels of consumption and income chosen by each individual at wage rate w and $U(w)$ is the utility of that individual based on these values, $U(c(w), 1 - y(w)/w)$.

The optimization problem is further constrained by the requirement that wage- w individuals voluntarily choose the bundle $(c(w), y(w))$ – the self-selection constraint discussed above. The requirement that the bundle $(c(w), y(w))$ is individually rational for people of wage w means that utility $U(c(w), 1 - y(w)/w)$ achieves a maximum at $w = w$. This may be expressed in terms of the first-order condition,

$$(4.7) \quad \frac{\partial U}{\partial c} \frac{dc}{dw'} + \frac{\partial U}{\partial y} \frac{dy}{dw'} = 0$$

that indicates that the individual cannot increase utility through a local change in labor supply. This then implies, for common preferences, that the change in utility as the wage rate rises is simply the derivative of the utility function with respect to w , holding c and y fixed:

$$(4.8) \quad \frac{dU}{dw} = \frac{\partial U}{\partial w} = U_2 \frac{y}{w^2} = U_2 \frac{L}{w}$$

Thus, the optimal tax problem is that expressed in (4.6), subject to the additional constraint given in (4.8). While it is expressed as one of choosing the bundle (c, y) , it can equally well be viewed

as a choice of the utility level u and the level of labor supply L , as $u=U(c,1-L)$ and $y=wL$. To solve the problem expressed this way, it is helpful to form the Hamiltonian:

$$(4.9) \quad H = [G(u) - \mathbf{m}(c(L,u) - y(L,u))]f(w) - \mathbf{h}U_2(L,u) \cdot \frac{L}{w}$$

with control variable L , state variable u , Lagrange multiplier \mathbf{m} and costate variable \mathbf{h} . The first-order conditions are

$$(4.10) \quad (a) \quad \frac{\mathcal{H}}{\mathcal{L}} = 0 \quad (b) \quad \frac{\mathcal{H}}{\mathcal{u}} = -\frac{d\mathbf{h}}{dw}$$

Condition (4.10a), as applied to (4.9), implies that

$$(4.11) \quad -\mathbf{m} \left[\frac{\mathcal{K}}{\mathcal{L}} \Big|_u - \frac{\mathcal{Y}}{\mathcal{L}} \Big|_u \right] f(w) - \mathbf{h} \left[\frac{\mathcal{U}_2}{\mathcal{L}} \Big|_u \cdot \frac{L}{w} + \frac{U_2}{w} \right] = 0$$

Note that $y=wL \Rightarrow \mathcal{Y}/\mathcal{L}|_u = w$ and $du/dL|_u = 0 = U_1 \mathcal{K}/\mathcal{L}|_u - U_2 \Rightarrow \mathcal{K}/\mathcal{L}|_u = U_2/U_1$.

Further, individual utility maximization ensures that $U_2/U_1 = w(1 - T')$. Thus, (4.11) can be

rewritten as

$$(4.12) \quad \frac{T'}{1 - T'} = \left(\frac{U_1 \mathbf{h}}{\mathbf{m}} \right) \frac{\mathbf{y}}{wf(w)}$$

where $\mathbf{y} \equiv \frac{\mathcal{U}_2}{\mathcal{L}} \Big|_u \cdot \frac{L}{U_2} + 1$. This expression says that the optimal marginal tax rate is increasing

in $(U_1 \mathbf{h}/\mathbf{m})$ and \mathbf{y} and decreasing in $wf(w)$. The last of these effects is straightforward: the more

effective labor supply that is subject to the marginal tax rate at w , the greater is the excess burden associated with that tax rate.

To interpret the other two terms in (4.12) and their effects, consider the special case of quasilinear preferences, $U(c,l) = c - v(1-l) = c - v(L)$, where $v(\cdot)$ is convex. For this case, it may be shown that $\mathbf{y} = 1 + 1/\mathbf{e}$, where \mathbf{e} is the compensated labor supply elasticity at w . Thus, a higher labor supply elasticity leads to a lower value of \mathbf{y} , which by (4.12) leads to a lower marginal tax rate. This is sensible, as a higher labor supply elasticity is also associated with greater excess burden per dollar of revenue raised. A similar effect appears in (4.4) for the case of the linear income tax, but here it is the labor supply elasticity at the particular wage rate w , rather than the aggregate labor supply elasticity, that is important because the government is free to choose different marginal tax rates for different levels of income.

Finally, consider the remaining term in (4.12), $(U_1 \mathbf{h}/\mathbf{m})$. From the first-order condition (4.10b),

$$(4.13) \quad -\frac{d\mathbf{h}}{dw} = \frac{\mathcal{H}}{\mathcal{U}} = \left[G' - \mathbf{m} \left(\frac{\mathcal{C}}{\mathcal{U}} \Big|_L - \frac{\mathcal{Y}}{\mathcal{U}} \Big|_L \right) \right] f(w) - \mathbf{h} U_{21} \frac{L}{w} \frac{\mathcal{C}}{\mathcal{U}} \Big|_L$$

As $\mathcal{Y}/\mathcal{U} \Big|_L = 0$ and because $du/du \Big|_L = 1 \Rightarrow dc/du \Big|_L = 1/U_1$, (4.13) can be rewritten as

$$(4.14) \quad \frac{U_1}{\mathbf{m}} \frac{d\mathbf{h}}{dw} = - \left[\frac{G' U_1}{\mathbf{m}} - 1 \right] f(w) + \frac{\mathbf{h} U_{21} L}{w \mathbf{m}}.$$

To interpret this further, it is again helpful to impose the simplifying assumption of quasilinear preferences, thereby implying that U_1 is constant (here normalized to 1) and $U_{21} = 0$. Then, integrating both sides of (4.14) and imposing the transversality condition $(\mathbf{h} \mathcal{R})$ as $w \mathcal{R} \mathbf{Y}$ yields

$$(4.15) \quad \frac{U_1 \mathbf{h}}{\mathbf{m}} = \int_w^\infty \left(1 - \frac{G'(\tilde{w})U_1}{\mathbf{m}} \right) f(\tilde{w}) d\tilde{w} = [1 - F(w)] - \int_w^\infty \frac{G'(\tilde{w})U_1}{\mathbf{m}} f(\tilde{w}) d\tilde{w}$$

where $F(\cdot)$ is the cumulative density function based on $f(\cdot)$.²⁰

This expression equals the social value, scaled by the marginal cost of funds \mathbf{m} of raising a dollar through marginal taxation at wage level w . This value has two components. The first term is the amount of revenue raised, equal to the taxes collected from all those who pay the extra tax – those with wages rates at least as high as w . The second term is the value, again in revenue units, of the social welfare lost by these individuals in paying the extra tax. Each of these terms declines with w , because we collect less revenue and impose less burden by raising taxes on fewer people, but it is the difference between the terms that matters. What pattern does this difference follow? The difference must be positive if marginal tax rates are positive, and the difference converges to zero as $w \rightarrow \infty$. If G declines with w , then the second term in (4.15) – the social cost of an increase in the marginal tax rate at w – converges to zero more rapidly than does the first term. Hence, there may be a range of w over which the difference between the two terms increases. The intuition is that high marginal tax rates at high levels of income are very inefficient because they produce so little revenue, while high marginal tax rates at low levels of income are inequitable because they impose burdens on those with very high social marginal utilities of income G . The best compromise may be to raise marginal tax rates at middle income levels, where tax obligations are not imposed on those for whom the burden of higher taxes is most socially costly but where higher tax rates still raise considerable revenue.

²⁰ In recent work, Saez (2000a) derives an analytical expression extending (4.15) to the case of more general preferences. While he offers an intuitive discussion of this expression, it is necessarily more complicated than the basic intuition presented here.

As should be clear from this discussion, the exact pattern that the term in (4.15) follows as w rises depends on the social welfare function and the shape of the wage distribution. Even if this term does indicate higher marginal tax rates somewhere in the middle of the wage distribution, this is precisely where one of the other terms in (4.12), $wf(w)$, is also likely to be greatest, which has the effect of reducing T .

Thus, it is possible to say very little about the general shape of the optimal marginal tax rate schedule, although since the seminal work of Mirrlees (1971) there has been a general tendency to find that optimal marginal tax rates should either fall throughout most of the income distribution or else have an “inverted-u” shape, reflecting the effect of the term in (4.15) (see, e.g., Kanbur and Tuomala 1994). This conclusion is in some sense predetermined by findings that, under certain circumstances, the optimal marginal tax rate equals zero at both the top and bottom of the income distribution.

The rationale for a zero top marginal tax rate appeared already, in the graphical presentation of the two-person case. For the general case with a bounded distribution of wage rates, the result (see Phelps 1973, Sadka 1976 and Seade 1977) follows directly from the fact that the term in (4.15) approaches zero as the wage w approaches its upper support, \bar{w} . As to why the marginal rate might be zero at the bottom of the wage distribution (see Seade 1977), consider the value of expression (4.15) at the lower support of the wage distribution, say \underline{w} . As $F(\underline{w})=0$, the expression indicates that $T = (1-T) \mu [1 - \bar{a}/m]$, where \bar{a} is the average social marginal utility of income over the entire distribution.²¹ But, as discussed in the case of the linear income tax, $\bar{a}=m$ when there is a uniform lump-sum tax available, so T must equal zero. The intuition for this result follows the algebra. At the very bottom of the income distribution, an increase in the

marginal tax rate has the same revenue and distributional effects as a uniform lump-sum tax – it raises revenue from the entire population. But it also distorts the behavior of the lowest income individuals, which a lump-sum tax does not. Thus, a lump-sum tax dominates any positive marginal tax on lowest-wage individuals.

However, neither of these results is robust to reasonable changes in assumptions. As its derivation suggests, the result regarding the marginal tax rate at the bottom requires that the entire population works. Otherwise, the marginal tax rate applied to the lowest-wage *worker* does not collect tax revenue from all individuals, and the logic just given breaks down.²² At the top of the wage distribution, optimal marginal tax rates need not approach zero, even in the limit, if the wage distribution is unbounded, nor is the “inverted-u” shape of the marginal tax rate distribution robust, as demonstrated by Diamond (1998) for the case of a Pareto distribution of wages and quasilinear preferences.²³

Even for bounded wage distributions where optimal marginal tax rates must eventually decline, marginal tax rates may rise over most of the income distribution, although numerical simulations of the more restricted optimal two-bracket linear tax system (Slemrod et al. 1994) find that the second/top marginal rate is lower than the first. This has quite interesting implications for the recent debate about the equity effects of the *flat tax* (Hall and Rabushka

²¹ As there are no income effects on labor supply for the quasilinear utility function, it is possible to ignore the indirect effect of income on revenue.

²² A different departure from this logic occurs if individuals at the bottom end of the income distribution make discrete choices of whether or not to work, as analyzed by Saez (2000b). In this case, the optimal marginal tax rate on the lowest income is negative, since the tax system thereby induces greater labor force participation and higher incomes.

²³ Diamond finds the optimal marginal tax rate schedule to be u-shaped in the example he analyzes. As clarified by Dahan and Strawczynski (2000), though, Diamond’s result of a rising marginal tax rate at the top depends on the joint assumptions of an unbounded ability distribution *and* quasilinear preferences. The result need not hold, even for the Pareto distribution of abilities, if one adopts a more general utility function. For another variation in assumptions, Stiglitz (1982) notes that if the effort of high-skilled workers is an imperfect substitute for that of low-skilled workers, it may be optimal to *subsidize* income at the top of the wage distribution to increase skilled labor effort and thereby raise the wages of the less skilled.

1995), a close relative of the linear income tax under which tax liabilities are constrained to be nonnegative. Although some (e.g. Bradford 1986) have suggested modifying the flat tax to permit additional, higher marginal tax rate brackets on higher wage individuals, these simulation results suggest that adding an additional bracket should occasion lower, not higher marginal tax rates at higher wage levels.

5. Externalities, public goods, and the marginal cost of funds

The analysis to this point ignores the use to which public funds may be put, other than redistribution to other taxpayers. In reality, of course, a major reason for raising revenue is to finance public expenditures, and it is important to consider how this affects the conclusions. In turn, it is interesting to ask how the use of distortionary taxation influences the optimality conditions of Samuelson (1954) regarding the provision of public goods. At the same time, it is convenient to consider how the distortionary nature of taxation alters the prescriptions concerning the use of Pigouvian taxation to correct externalities.

Basic results relating the provision of public goods and the correction of externalities to the use of distortionary taxes may be found, respectively, in Atkinson and Stern (1974) and Sandmo (1975). Auerbach (1985) presents and interprets these results in some detail, so we will offer only a brief derivation here. Both models assume that the government is limited to the use of indirect proportional taxes, and avoid any discussion of distribution by assuming that individuals are identical, i.e., that the population consists of H copies of the representative individual. In this context, it is natural to assume that the government seeks to maximize the utility of each representative individual or, equivalently, the sum of individual utilities.

5.1. *The provision of public goods and the marginal cost of public funds*

Consider first the case in which the government wishes to provide a public good, G , using all its tax revenue. Individuals choose consumption \mathbf{x} treating G as given, so their utility function may be written in semi-indirect form as $V(\mathbf{p}, y; G)$, with $\partial V / \partial G = \partial U / \partial G |_{x(\mathbf{p}, y; G)}$. For simplicity, the economy's production function $f(\mathbf{X}, G)$ (where $\mathbf{X} = H\mathbf{x}$) is taken to obey constant returns, so that there are no pure profits and $y = 0$. This set-up gives rise to the Lagrangean (compare to 3.12):

$$(5.1) \quad HV(\mathbf{p}; G) - \mathbf{m}f(\mathbf{X}, G)$$

with first-order conditions with respect to each price and the level of public goods, G . The first-order conditions with respect to price are identical to those derived above in section 3 for the case of $y = 0$, in (3.15), repeated here for convenience:

$$(5.2) \quad -\mathbf{I}X_i + \mathbf{m} \left[X_i + \sum_j t_j \frac{dX_j}{dp_i} \right] = 0 \quad i = 1, \dots, N$$

except that X_i is now the sum of individual purchases of good i , equivalently the product of H and the purchase of the representative consumer. The first-order condition with respect to the public good is

$$(5.3) \quad H \frac{\partial V}{\partial G} - \mathbf{m} \left[f_G + \sum_i f_i \frac{\partial X_i}{\partial G} \right] = 0.$$

The utility function implies that $\partial V / \partial G = U_G^h$, in which U_i^h is individual h 's marginal utility of good i . The economy's production constraint and private production efficiency impose the

condition that $f_i \propto q_i$, while the consumer's budget constraint implies that $\mathbf{p}' \partial \mathbf{X} / \partial G = 0$. Taking good 0 to be the untaxed numeraire commodity, and \mathbf{I} to be the marginal utility of income, it follows that $U_0^h = \mathbf{I} p_0 = \mathbf{I} f_0$, and (5.3) implies

$$(5.4) \quad \sum_h \frac{U_G^h}{U_0^h} = \left(\frac{\mathbf{m}}{\mathbf{I}} \right) \left(\frac{f_G}{f_0} - \frac{dR}{dG} \right)$$

where R is tax revenue, $\mathbf{t} \cdot \mathbf{X}$ and the variable \mathbf{m} is the shadow cost of the government's revenue constraint (measured in units of utility). The ratio $(\mathbf{m} \mathbf{I})$, which measures the shadow price of revenue units of the numeraire, is often referred to as the *marginal cost of public funds (MCPF)*, because it measures the cost of each unit of public funds, taking account of the deadweight loss from the additional taxes associated with those funds.

Expression (5.4) deviates in two respects from the Samuleson rule of equating the marginal rate of transformation, f_G/f_0 , and the sum of the marginal rates of substitution, $\sum_h U_G^h / U_0^h$. First, it indicates that the implicit cost of public goods is reduced to the extent that public spending increases spending on taxed commodities, i.e., $dR/dG > 0$ – a point noted by Diamond and Mirrlees (1971). Second, it requires that one adjust the relative price of public goods, f_G/f_0 , for the *MCPF*, consistent with intuition provided by Pigou (1947). However, as noted by Atkinson and Stern, the *MCPF* as defined need not exceed 1. Recall from section 3 that optimal taxes ensure that $\mathbf{m} \succ \mathbf{a}$, where $\mathbf{a} = \mathbf{I} + \mathbf{m} dR/dy$ is the “social” marginal utility of income – the value to society of giving an individual an extra unit of income, taking account of the revenue provided by induced spending on taxed goods. However, if dR/dy is negative, then it is possible that the *MCPF* is equal to or even less than 1.

A simple example illustrating this possibility is provided by Ballard and Fullerton (1992). Consider the case in which the utility function is weakly separable into private and public goods, so that $dR/dG = 0$. Suppose that there are just two private goods, leisure and consumption, so that there is just one independent tax instrument, and normalize this tax instrument so that only the tax on labor income is positive. The first-order condition with respect to the price of labor – the wage rate w – is, from (5.2),

$$(5.5) \quad -L + m L - t dL/dw = 0$$

where L is the aggregate supply of labor and t is the tax per unit of labor supplied.²⁴ Defining h_{Lw} as the uncompensated labor supply elasticity and q as the tax rate t/w , (5.5) may be rewritten:

$$(5.6) \quad \frac{m}{L} = \frac{1}{1 - q h_{Lw}}$$

from which it is obvious that the *MCPF* exceeds 1 if and only if the uncompensated labor supply elasticity is positive. For the “benchmark” case of Cobb-Douglas preferences, the uncompensated labor supply elasticity is zero, and the *MCPF* = 1.²⁵ Given that a zero uncompensated labor supply elasticity lies within the range of existing estimates, this result is not simply a theoretical curiosity, and suggests that we may well err in automatically assuming that the existence of distortionary taxation raises the *MCPF* significantly.²⁶

²⁴ The term $t dL/dw$ enters in expression (5.5) with a minus sign because the tax is subtracted from the wage.

²⁵ Ballard and Fullerton argue based on an informal survey that this outcome was generally a surprise to a group of public finance economists.

²⁶ More generally, if the utility function is not separable, one may show that the Samuelson rule holds whenever the supply of labor is unaffected by the increase in spending on the public good – whenever the combined impact on L of the increase in G and the decrease in w equals zero. In this case, the marginal cost of funds as defined in (5.3) is not equal to 1, but its deviation from 1 is offset by the dR/dG term.

The reason that this assumption has the potential to go wrong is that the deadweight loss of a tax system and the *MCPF* are two entirely separate concepts. Deadweight loss is a measure of the potential gain from replacing distortionary taxes with an efficient lump-sum alternative, and marginal deadweight loss is simply the change in this magnitude as tax revenue changes. By contrast, the *MCPF* reflects the welfare cost, in units of a numeraire commodity, of raising tax revenue for exhaustive government expenditure.

While this result seems simple and straightforward, much has been written on the topic of how the *MCPF* should be defined. Without reviewing this extensive literature (see, for example, the survey by Håkonsen 1998), we note that the disagreements relate largely to terminology and questions of normalization. As an illustration (see Schöb, 1997), consider the same example (one public good, labor, and one other private good), but normalize the proportional taxes so that the tax on labor is zero. The first-order condition with respect to the price, p , of the taxed commodity, instead of (5.5), would be

$$(5.7) \quad -lX + m(X + t dX/dp) = 0$$

where X is the aggregate purchase of the commodity and t is the tax per unit of that commodity.

Defining h_{xp} as the uncompensated own-price demand elasticity and q as t/p , (5.7) can be rewritten as

$$(5.8) \quad \frac{m}{l} = \frac{1}{1 + qh_{xp}}$$

which says that the *MCPF* should exceed 1 if and only if $h_{xp} < 0$ – i.e., X is not a Giffen good.

Since this is a much weaker condition than that $h_{Lw} > 0$, it is easy to see how one might become

confused, given that these conditions supposedly reflect the same underlying experiment.

Indeed, when $h_{Lw} = 0$, $h_{xp} = -1$, so $ml = 1/(1-q)$. This apparent paradox is resolved by noting that the normalization does not affect the underlying outcome, but does change the units of (ml). In the first instance, the *MCPF* is defined in units of the commodity; in the second, it is measured in terms of units of labor.

The impact of this difference may be understood using the standard approach of cost-benefit analysis (e.g., Harberger 1972), that weights the costs of funds according to sources. When the labor supply elasticity is zero, an increase in the tax on labor has no impact on the amount of labor supplied. Thus, the extra taxes that finance additional spending on the public good are absorbed fully through reduced consumption. Hence, the marginal cost of funds equals the marginal value of a unit of the commodity. Therefore, if the commodity is chosen as the numeraire, the marginal cost of funds equals 1. If labor is chosen as the numeraire, the marginal cost of funds still equals 1 unit of the commodity, but this equals $1/(1-q)$ units of labor, due to the tax wedge between labor and private consumption. The equilibrium is the same regardless of normalization, but the *MCPF* is different. This discussion also highlights that the *MCPF* reflects only the presence of a distortion on one particular margin – between the public good and the numeraire. This distortion can be positive, negative or zero, independent of the presence of deadweight loss due to taxation.

5.2. Externalities and the “double-dividend” hypothesis

A similar logic applies to the analysis of externalities, as in Sandmo (1975). Suppose that, rather than there being a public good, there is an externality, E , that enters into each person’s utility function and which cannot be avoided, so that the representative individual’s indirect utility function may be written $V(\mathbf{p};E)$. Suppose also, for simplicity, that the externality

is the product of aggregate consumption of a single good, say the good with the highest index, N . Then, the Lagrangean,

$$(5.9) \quad HV(\mathbf{p}; X_N) - \mathbf{m}(\mathbf{X})$$

implies the following N first-order conditions with respect to the prices of goods $1, \dots, N$ (compare to 3.8):

$$(5.10) \quad -\mathbf{l}x_i + \mathbf{m} \left[x_i + \sum_j t_j^* \frac{dx_j}{dp_i} \right] = 0 \quad i = 1, \dots, N$$

where

$$t_j^* = t_j \quad j = 1, \dots, N$$

$$t_N^* = t_N + \frac{HV_E}{\mathbf{m}} = t_N + \frac{HV_E / \mathbf{l}}{\mathbf{m} / \mathbf{l}}$$

Expression (5.10) is the standard optimal tax solution, except that it calls for the tax on the externality-producing good, t_N , to equal the sum of the “optimal” tax that ignores the externality, t_N^* , plus a term that reflects the cost of the externality. This second term equals the corrective Pigouvian tax – the social cost per unit of consumption of the good, measured in terms of the numeraire commodity – divided by the *MCPF*, \mathbf{m} / \mathbf{l} .

Thus, in a result analogous to that just presented for the provision of public goods, the presence of distortionary taxation leads to “undercorrection” of the externality if and only if the *MCPF* exceeds 1. As before, though, one must exercise care in interpreting this result. Suppose, following the previous example, that the externality enters the utility function in a separable manner, and that preferences over direct consumption of goods and leisure are Cobb-Douglas.

Also assume that there are just two consumption goods, a “clean” good and a “dirty” good that causes the externality. Absent the externality (and if various regularity conditions are satisfied), the optimal tax structure calls for equal taxes on the two consumption goods, i.e., $t_1^* = t_2^*$. This can be achieved either through a tax on wages alone or through uniform taxes on the two consumption goods. In the first case, letting the clean good be numeraire, it is clear that $m\mathbf{l} = 1$, so the Pigouvian tax should be implemented without adjustment. In the second case, letting labor be numeraire, $m\mathbf{l} = 1/(1-q) > 1$, so it is necessary to “undercorrect” for the externality.

It is tempting to conclude in the latter case that one “undercorrects” because the corrective tax is piled on top of the preexisting consumption tax, while in the former case no initial preexisting consumption tax exists. However, the two equilibria are identical, with the same distortions present on all margins.²⁷ Thus, the intuition is misleading. While there is no initial consumption tax when only labor is taxed, there is still a distortion of the labor-leisure choice. Taxing the dirty consumption good exacerbates the distortion between that good and labor, just as if the initial tax were on the two consumption goods instead. The fact that it is overall distortions that matter, and not the levels of individual taxes, also exposes a serious interpretive difficulty in what is known as the “double-dividend” hypothesis. This hypothesis, as discussed in much more detail in the chapter in this Handbook by Bovenberg and Goulder, states that corrective taxes have an added benefit in the presence of other distortionary taxes – the revenue that allows a reduction in the other tax rates and their associated deadweight loss. Corrective taxes do not merely raise revenue and correct externalities, but also exacerbate

²⁷ For example, let q be the producer price of the dirty good, and t^p the Pigouvian tax based on the standard formula. When the clean good is the untaxed numeraire and labor is taxed, the net wage rate relative to the price of the dirty good is $w(1-q)/(q+t^p)$. When labor is untaxed, each consumption good faces a tax that raises its price by the factor $q/(1-q)$, and the dirty good also faces the corrective tax of $t^p/(m\mathbf{l}) = t^p/(1-q)$, so the net wage relative to the price of the dirty good is $w[q/(1-q)+t^p/(1-q)] = w(1-q)/(q+t^p)$.

existing distortions. Taxing consumption and using the proceeds to reduce taxes on labor has no net impact on the consumption-leisure choice in this instance.

5.3. *Distributional considerations and the MCPF*

With a heterogeneous population, the provision of public goods and the correction of externalities take on added complications. Even in the absence of distortionary taxation, the optimal rules then reflect the social valuations of utilities of different individuals. In addition, the costs and benefits of public goods, externalities, and the taxes that address them all have distributional consequences. For example, the government might wish to expand provision of public goods that have favorable distributional consequences; Sandmo (1998) offers a detailed analysis of the general problem. Also see Slemrod and Yitzhaki (2001), who illustrate how one can decompose both the costs and benefits of public expenditure projects in terms of efficiency and distributional consequences. However, it is also useful to consider circumstances in which the problem becomes much simpler, which is the case when the government has sufficient flexibility in its choice of tax instruments.

There is a close analogy here to the standard optimal income tax problem, under which it may not be necessary to tax luxury goods more heavily for purposes of distribution if the government can use a nonlinear income tax (as in Atkinson and Stiglitz 1976). Indeed, the analysis yields a parallel result, namely that distributional considerations should not enter into the provision of public goods or the correction of externalities when there is a nonlinear income tax and preferences are weakly separable into goods and leisure. This result is described by Kaplow (1996), building on previous work of Hylland and Zeckhauser (1979).

Kaplow's observation is that the Samuelson rule for public goods provision is unaffected by the presence of distortionary taxation when preferences are separable and the government

uses a nonlinear income tax. The argument has two pieces. First, following the intuition given above for the proportional tax case, there will be no change in labor supply, so that all of the expenditures on the public good come through reductions in the untaxed numeraire commodity. Hence, there is no tax wedge at the margin between public and private goods. Second, because of the availability of the nonlinear income tax, the distributional consequences of an increase in public goods spending can be offset, so that distributional weights will also be absent from the decision.

To expand on the reasoning Kaplow provides for his result, we present a detailed proof here. Suppose that households vary with respect to wage rates, w , but that each household's preferences take the form $U(v(c,g),1-L)$, where c is private good consumption, g is the level of the public good, and L is labor supplied. Public goods are financed using a nonlinear tax on labor income $T(wL; g)$, where T_1 is the household's marginal tax rate. Consider an experiment in which g is increased, with taxes raised on each individual so that net utility is unchanged. (Continuing to spend and tax in this way will eventually lead to an optimal level of public goods provision, if the government persists to the point that marginal revenue from additional spending is zero.) The claim is that this policy results in no change in labor supply.

The household's initial optimum labor supply decision implies that

$$(5.11) \quad \frac{\partial U}{\partial L} = U_1 v_1 (w - T_1 w) - U_2 = 0$$

and that (5.11) holds as g changes:

$$(5.12) \quad v_1 \frac{dU_1}{dG} + U_1 \left(v_{11} \frac{dc}{dg} + v_{12} \right) w(1 - T_1) - U_1 v_1 w \left(T_{11} w \frac{dL}{dg} + T_{12} \right) - U_{21} \frac{dv}{dg} - U_{22} \frac{dL}{dg} = 0$$

The claim is that (5.12) holds with both U and L constant. Note that if U and L remain constant, so must v , and hence U_1 . Thus, the claim implies that

$$(5.13) \quad v_{11} \frac{dc}{dg} + v_{12} = \frac{v_1}{1-T_1} T_{12}$$

or, using $dv/dg = v_1 dc/dg + v_2 = 0 \Rightarrow dc/dg = -(v_2/v_1)$,

$$(5.14) \quad \frac{\partial(v_2/v_1)}{\partial c} = \frac{1}{1-T_1} T_{12}$$

By the assumption that L is fixed, $dc/dg = dT/dg$ and $dT/dg = T_2$. Thus, $v_2/v_1 = T_2$. Moreover, this equality does not hold simply at a particular point, but rather at all points in the income distribution. That is, the functions $v_2/v_1(c,g)$ and $T_2(wL;g)$ are equal for any value of $c = wL - T(wL;g)$. Thus,

$$(5.15) \quad \frac{\partial(v_2/v_1)}{\partial c} = T_{21} \frac{dwL}{dc} \Big|_g = T_{21} \frac{1}{1-T_1}$$

Because $T_{12} = T_{21}$, (5.14) holds, consistent with the initial claim.

Just as in the case previously considered in section 5.2, a parallel analysis applies to externalities, with the implication that, under the maintained assumptions regarding preferences and the use of the nonlinear income tax, no adjustment to the standard Pigouvian tax formula is warranted.

While these results do depend on two key assumptions, those concerning the separability of individual preferences and the flexibility of the income tax, they are still quite important because they identify the source of deviations from the basic rules of Samuleson and Pigou. As discussed

in this Handbook's chapter by Kaplow and Shavell, they also have additional implications regarding the extent to which government policies should be influenced by distributional issues.

6. Optimal taxation and imperfect competition.

The analysis to this point concerns the optimal design of tax policies in economies with perfectly competitive industries. Since some economic situations are characterized by imperfect competition, it is useful to consider the implications of differing degrees of market competition for optimal tax design. One of the difficulties of summarizing the implications of imperfect competition for optimal taxation stems from the multiplicity of imperfectly competitive market structures. Nevertheless, it is possible to identify common welfare implications by considering a range of tax instruments and market situations. Our analysis follows closely that of Auerbach and Hines (2001).

6.1. Optimal commodity taxation with Cournot competition.

It is useful to start with the behavior of a firm that acts as a Cournot competitor in an industry with a fixed number (n) of firms. The government imposes a specific tax on output at rate t , so firm i 's profit is given by

$$(6.1) \quad Px_i - tx_i - C(x_i),$$

in which P is the market price of the firm's output, x_i the quantity it produces, and $C(x_i)$ the cost of producing output level x_i . In this partial-equilibrium setting, it is appropriate to take P to be a univariate function of industry output, denoted X .

The firm's first-order condition for profit maximization is

$$(6.2) \quad P + x_i \frac{dP}{dX} (1 + \mathbf{q}) - t = C'(x_i),$$

in which $\mathbf{2}$ is firm i 's conjectural variation, corresponding to $(\frac{dX}{dx_i} - 1)$. Differing market structures correspond to differing values of $\mathbf{2}$. In a Cournot-Nash setting, in which firm i believes that its quantity decisions do not affect the quantities produced by its competitors, then $\mathbf{2} = 0$. In a perfectly competitive setting, $\mathbf{2} = -1$. Various Stackelberg possibilities correspond to values of $\mathbf{2}$ that can differ from these, and indeed, need not lie in the $[-1, 0]$ interval.

It is useful to consider the pricing implications of (6.2). Differentiating both sides of (6.2) with respect to t , taking $\mathbf{2}$ to be unaffected by t , and limiting consideration to symmetric equilibria (so that $x_i = \frac{X}{n}$, $C(x_i) = C(X/n)$, and, since $\frac{dX}{dt} = \frac{dP/dt}{dP/dX}$, it follows that

$$\frac{dx_i}{dt} = \frac{dP/dt}{n dP/dX}, \text{ then}$$

$$(6.3) \quad \frac{dP}{dt} = \left\{ 1 + \frac{1 + \mathbf{q}}{n} (1 + \mathbf{h}) - \frac{C''(X/n)}{n dP/dX} \right\}^{-1},$$

in which $\mathbf{h} \equiv \frac{d^2 P}{dX^2} \frac{X}{dP/dX}$ is the elasticity of the inverse demand function for X . From (6.3), it is

clear that $\frac{dP}{dt}$ can exceed unity, a possibility that is consistent with the firm's second-order

condition for profit maximization and with other conditions (discussed by Seade, 1980a, 1980b) that correspond to industry stability.

Equations (6.2) and (6.3) identify the potential welfare impact of taxation in the presence of imperfect competition. From (6.2), the combination of imperfect competition ($\mathbf{2} > -1$) and a

downward-sloping inverse demand function ($\frac{dP}{dX} < 0$) implies that firms choose output levels at which price exceeds marginal cost. Hence there is deadweight loss in the absence of taxation, and, in this simple partial equilibrium setting, tax policies that stimulate additional output reduce deadweight loss, while those that reduce output make bad situations worse. In some circumstances the imposition of a tax may reduce industry output sufficiently that after-tax profits actually rise.

Tax policy can be used to reduce or eliminate the allocative inefficiency due to imperfect competition, though other policy instruments (such as antitrust enforcement) are also typically available and may be more cost-effective at correcting the problem.²⁸ Taking alternative remedies to be unavailable, the optimal policy, if the government has access to lump-sum taxation, is to guarantee marginal cost pricing by setting $t = \frac{X}{n} \frac{dP}{dX} (1 + \mathbf{q})$.²⁹ Since $\frac{dP}{dX} < 0$, this corrective method entails subsidizing the output of the imperfectly competitive industry, so in realistic situations in which tax revenue is obtained through distortionary instruments, it follows that the optimal policy may not fully eliminate the problems due to imperfect competition.

In order to explore this issue further, consider the setup of section 3.1, in which all commodities are produced at constant cost. There are $N+1$ commodities, of which the first $M+1$, indexed $0, \dots, M$, are produced by perfectly competitive firms, and the remaining commodities, $M+1, \dots, N$, are produced in imperfectly competitive markets, each of whose pricing satisfies

²⁸ One possibility, explored by Katz and Rosen (1985), is that tax authorities design corrective policies on the basis of imperfect understanding of the extent of competition in oligopolistic industries.

²⁹ Such a corrective subsidy was proposed by Robinson (1933, pp. 163-165), who attributes it to her husband and presents it as an “ingenious but impractical scheme.”

(6.2).³⁰ Denoting the (constant) per-unit production cost of commodity i by q_i , it follows that $p_i = q_i + t_i, \forall i = 0, \dots, M$. As in section 3, we assume that the tax on the numeraire commodity, good 0, equals 0. Firms in the imperfectly competitive industries generate profits, and someone in the economy receives these profits as income.³¹ Taking consumers in the economy to be identical, it follows that the utility of the representative consumer can be represented by

$$(6.4) \quad V(\mathbf{p}, B),$$

in which \mathbf{p} is the vector of $N+1$ commodity prices, and B represents profits earned by the imperfectly competitive firms. Commodity demands are then functions of (\mathbf{p}, B) , but to simplify the calculations that follow, we consider the case in which firms ignore the indirect impact of their pricing decisions on demand through induced changes in profits. In industry $j > M$, the representative firm's first-order condition for profit maximization is

$$(6.5) \quad p_j - t_j - q_j = -\frac{X_j (1 + \mathbf{q}_j)}{n_j \partial X_j / \partial p_j},$$

where n_j and \mathbf{q} are defined for industry j in the usual way. Thus, the price-cost margin imposed

by imperfect competition is $m_j = -\frac{X_j(1 + \mathbf{q}_j)}{n_j(\partial X_j / \partial p_j)}$ in industry j .

The optimal taxation problem consists of maximizing (6.4) with respect to the specific taxes \mathbf{t} subject to these mark-up conditions, the revenue constraint,

³⁰ We follow much of the literature in assuming that preferences and technology support a unique stable market equilibrium, which, as Roberts and Sonnenschein (1977) note, need not exist in the presence of imperfect competition.

³¹ In the competitive context, assuming a zero tax rate on one commodity restricts the government effectively from imposing a tax on pure profits through a uniform tax on all commodities. Here, though, before-tax profits would

$$(6.6) \quad \sum_{j=1}^N t_j X_j = R$$

and the household's budget constraint,

$$(6.7) \quad \sum_{j=M+1}^N (p_j - t_j - q_j) X_j = \mathbf{p}.$$

Combining the revenue constraint (6.6) and the budget constraint (6.7), we may recast the problem as one of maximizing (6.4) with respect to consumer prices \mathbf{p} , subject to the constraint,

$$(6.8) \quad \sum_{j=1}^N (p_j - q_j) X_j \geq R + \mathbf{p},$$

where profits are given by³²

$$(6.9) \quad \mathbf{p} = - \sum_{j=M+1}^N \frac{X_j}{n_j} \frac{(1 + q_j)}{\partial X_j / \partial p_j} X_j.$$

With \mathbf{m} defined as the multiplier of the constraint given in (6.8), the first-order conditions for this problem are:

respond to such uniform taxation, leaving the government's problem unchanged. We show this below, after presenting an expression for equilibrium profits.

³² Examination of expression (6.9) clarifies that taxing all goods uniformly would not reduce real profits. Taxing all goods at the same rate would raise prices by a factor \mathbf{I} , so it is necessary to verify that (6.9) continues to hold if profits, \mathbf{p} simultaneously increased by \mathbf{I} (and were therefore unchanged in real terms). Multiplying prices and profits by \mathbf{I} has no effect on X_j , since consumer demands are homogeneous of degree zero in income and prices. But this magnification of prices and income multiplies $\mathcal{K}_j / \mathcal{P}_j$ by the factor $1/\mathbf{I}$, as a unit change in price represents only $1/\mathbf{I}$ as large a proportional change as before. Thus, the right-hand side of (6.9) equals its original value, multiplied by \mathbf{I} . As left-hand side of (6.9) also equals its original value (\mathbf{p} multiplied by \mathbf{I}), the expression still holds.

$$(6.10) \quad -\mathbf{I}X_i + \mathbf{I} \frac{d\mathbf{p}}{dp_i} + \mathbf{m} \left[x_i + \sum_j (p_j - q_j) \frac{\partial X_j}{\partial p_i} + \sum_j (p_j - q_j) \frac{\partial X_j}{\partial y} \frac{d\mathbf{p}}{dp_i} - 1 \right] = 0 \quad i = 1, \dots, N$$

where, as before, \mathbf{I} is the marginal utility of income. Once again defining $\mathbf{a} = \mathbf{I} + \mathbf{m} \sum_{j=1}^N t_j \frac{\partial X_j}{\partial \mathbf{p}}$

to be the “social” marginal utility of income, we may rewrite (6.10) as

$$(6.11) \quad -\mathbf{I}X_i - \mathbf{m} \left[X_i + \sum_{j=1}^N t_j^* \frac{\partial X_j}{\partial p_i} - \left(\frac{\mathbf{m} - \mathbf{a}}{\mathbf{m}} \right) \frac{d\mathbf{p}}{dp_i} \right] = 0$$

in which

$$\begin{aligned} t_j^* &= t_j & j \leq M \\ t_j^* &= p_j - q_j & j > M \end{aligned}$$

is the total wedge in market j , equal to $t_j + m_j$ in noncompetitive industries.

Equation (6.11) is analogous to (5.10), and carries precisely the interpretation offered by Sandmo for the optimal tax conditions in the presence of externalities. Intuitively, the “externality” in the case of imperfect competition is the outcome of the oligopolistic output selection, resulting in the extra mark-up m_j . The definition of t_j^* takes into account the need to correct this pre-existing distortion. Were this the only term on the right side of (6.11), then it would be optimal fully to correct for the extra distortions in noncompetitive industries and then impose the standard optimal taxes. Presumably, the net result in industry would be an incomplete offset of oligopolistic mark-ups, the optimal tax component normally being positive. The second term in brackets in (6.11) accounts for the existence of profits, taking the form laid

out in expression (3.17) above and explained in that context.³³ In this instance, tax-induced price changes affect the profitability of the imperfectly competitive industry, the difference (: – ”) capturing the welfare effect of increasing industry profits by one unit. To the extent that a higher price of a commodity directly or indirectly augments oligopoly profits, this must be included in computing the price change’s overall welfare effect. Doing so has the effect of making the price increase less attractive as a policy tool.

6.2. *Specific and ad valorem taxation.*

In competitive markets the distinction between specific and ad valorem taxation arises only from minor tax enforcement considerations. In imperfectly competitive markets these two tax instruments are no longer equivalent, since the imposition of an ad valorem tax makes the tax rate per unit of sales a function of a good’s price, which is partly under the control of individual firms. As a result, ad valorem and specific taxes that raise equal tax revenue will typically differ in their implications for economic efficiency, ad valorem taxation being associated with much less deadweight loss.³⁴ Intuitively, ad valorem taxation removes a fraction (equal to the ad valorem tax rate) of a firm’s incentive to restrict its output level in order to raise prices.

The welfare superiority of ad valorem taxation is evident in the simple partial equilibrium setting considered initially above. Now, the government is assumed to have access both to an ad valorem tax and to a specific tax, and tax revenues are assumed costly to obtain (for reasons omitted from the model). In this setting the firm’s profits equal

$$(6.1) \quad (1 - \mathbf{t})Px_i - tx_i - C(x_i)$$

³³ Auerbach and Hines (2001) present a longer, alternative derivation of (6.11) that includes explicit expressions for the terms $d\mathbf{p}/dp_i$.

³⁴ Suits and Musgrave (1953) provide a classic analysis of this comparison; their treatment is greatly expanded and elaborated by Deliapalla and Keen (1992).

in which J is the ad valorem tax rate. Assuming the n -firm outcome to be symmetric, the first-order condition for profit maximization becomes

$$(6.2) \quad (1 - \mathbf{t}) \left[P + \frac{X}{n} \frac{dP}{dX} (1 + \mathbf{q}) \right] - t = C' \left(\frac{X}{n} \right),$$

and its pricing implications are

$$(6.12) \quad \frac{dP}{dt} = \left\{ (1 - \mathbf{t}) \left[1 + \frac{1 + \mathbf{q}}{n} (1 + \mathbf{h}) \right] - \frac{C''(x/n)}{n dP/dX} \right\}^{-1}$$

$$(6.13) \quad \frac{dP}{d\mathbf{t}} = \left[P + \frac{X}{n} \frac{dP}{dX} (1 + \mathbf{q}) \right] \frac{dP}{dt}.$$

Since a unit change in J raises more tax revenue than does a unit change in t , it is unsurprising that $\frac{dP}{d\mathbf{t}} > \frac{dP}{dt}$. Much more revealing is the effect of these tax instruments normalized by dollar

of marginal tax revenue. Since total tax revenue is given by $Rev = \mathbf{t}PX + tX$, it follows that

$$(6.14a) \quad \frac{dRev}{dt} = X \left(1 + \mathbf{t} \frac{dP}{d\mathbf{t}} \right) + (t + \mathbf{t}P) \frac{\partial X}{\partial P} \frac{dP}{dt}$$

$$(6.14b) \quad \frac{dRev}{d\mathbf{t}} = PX \left(1 + \frac{\mathbf{t}}{P} \frac{dP}{d\mathbf{t}} \right) + (t + \mathbf{t}P) \frac{\partial X}{\partial P} \frac{dP}{d\mathbf{t}}.$$

In this simple partial equilibrium model, the change in deadweight loss associated with one of these tax changes is equal to the product of the induced change in X and the difference between marginal cost and price. Consequently,

$$\frac{d(DWL)/dt}{d(DWL)/d\mathbf{t}} = \frac{-\left(\frac{\partial X}{\partial P}\right) \left(\frac{dP}{dt}\right) \left(P - C\left(\frac{X}{n}\right)\right)}{-\left(\frac{\partial X}{\partial P}\right) \left(\frac{dP}{d\mathbf{t}}\right) \left(P - C\left(\frac{X}{n}\right)\right)} = \frac{dP/dt}{dP/d\mathbf{t}},$$

which, together with (6.14a) and (6.14b), implies that

$$(6.15) \quad \frac{\left[\frac{d(DWL)/dt}{d(DWL)/d\mathbf{t}}\right]}{\left[\frac{dRev/dt}{dRev/d\mathbf{t}}\right]} = \frac{X\left(\frac{P}{dP/d\mathbf{t}} + \mathbf{t}\right) + (t + \mathbf{t}P)\frac{\partial X}{\partial P}}{X\left(\frac{1}{dP/d\mathbf{t}} + \mathbf{t}\right) + (t + \mathbf{t}P)\frac{\partial X}{\partial P}}.$$

From (6.13), $\frac{dP}{d\mathbf{t}} < P \frac{dP}{dt}$, so if tax revenue is an increasing function of tax rates, then the right

side of (6.15) is greater than unity. Hence revenue-equal substitution of ad valorem for specific taxation reduces deadweight loss at any (t, J) combination.³⁵ Of course, such substitution works at the expense of firm profitability, and would, if used excessively, drive profits negative and supply presumably to zero. But assuming the firm profitability constraint not to bind, the optimal tax configuration entails ad valorem rather than specific taxation.

The preceding comparison of ad valorem and specific taxation compares their effectiveness per dollar of foregone revenue, but does not address the question of the optimal rate of ad valorem taxation when the government is unable or unwilling to provide specific subsidies. While this problem is typically thought (e.g., Myles, 1989) to entail a very different solution than

³⁵ Consequently, if the government is able to impose negative specific taxes (specific subsidies), then it can completely eliminate the distortion due to imperfect competition through a judicious combination of ad valorem tax and specific subsidy, as noted by Myles (1996). The effectiveness of this corrective method is limited by any constraints on ad valorem tax rates, such as a restriction that they be nonnegative.

that for specific taxation, properly framed it becomes clear that the solution has the same character regardless of the type of available tax instrument.

Following the analysis of specific taxes, we seek to maximize the indirect utility function in (6.4) subject to the revenue constraint,

$$(6.16) \quad \sum_{j=1}^N \mathbf{t}_j p_j X_j \geq R,$$

the definition of profits,

$$(6.17) \quad \sum_{j=M+1}^N (p_j(1-\mathbf{t}_j) - q_j) X_j = \mathbf{p},$$

and the characterization of producer behavior in noncompetitive industries,

$$(6.18) \quad p_j(1-\mathbf{t}_j) - q_j = -(1-\mathbf{t}_j) \frac{X_j}{n_j} \frac{(1+\mathbf{q}_j)}{\partial X_j / \partial p_j} \quad j > M.$$

As before, we combine household and government budget constraints to express economy's resource constraint as

$$(6.19) \quad \sum_{j=1}^N (p_j - q_j) X_j \geq R + \mathbf{p},$$

and analyze the problem as one of maximizing (6.4) with respect to \mathbf{p} , subject to this constraint,

where profits are given by

$$(6.20) \quad \mathbf{p} = \sum_{j=M+1}^N \left[\frac{q_j}{p_j - \mathbf{f}_j} \right] \mathbf{f}_j X_j, \quad \text{where } \mathbf{f}_j = -\frac{X_j}{n_j} \frac{(1+\mathbf{q}_j)}{\partial X_j / \partial p_j}.$$

Note that expression (6.20) differs from (6.9) by the term multiplying $f_j X_j$ on the right-hand side of (6.20), which equals $(1-t_j)$. Otherwise, the problem is identical to that for specific taxes, and the first-order conditions given in (6.11) still hold, for t_i inserted in place of t_i/p_i . The resulting equilibrium will generally be different, of course, because profits, and hence the terms dp/dp_i , will be different.

Auerbach and Hines (2001) provide some numerical simulations confirming that, in cases for which a noncompetitive industry's tax is positive under specific taxation, it should be higher in the case of ad valorem taxation. They also extend the analysis to the case in which the government is uncertain about the degree of noncompetitive behavior, as represented by the parameter q . This uncertainty tends to reduce the extent of the desired corrective subsidy, for the subsidy tends to be most effective precisely when it is least needed, i.e., when q is small.

6.3. *Free entry.*

The standard Cournot model takes as its point of departure an industry with a fixed number of firms. The ability of firms to enter and leave an industry changes the optimal tax problem, and introduces some interesting features of the solution (such as the possibility of welfare-improving positive tax rates even if the government has access to nondistortionary sources of revenue). In spite of these differences, many of the main implications of the preceding analysis, including the welfare superiority of ad valorem to specific taxation, persist in a model with free entry.

Consider an industry consisting of identical firms that behave according to (6.2). In this model, entry and exit are free, but new entrants do not necessarily select output levels that

minimize cost, since they behave in a manner that is cognizant of the effect of output on price.³⁶

The government imposes ad valorem and specific taxes, so the zero-profit condition for industry entry (assuming, for convenience, that it is possible to have fractional numbers of firms) is

$$(6.21) \quad (1 - t) \left(\frac{X}{n} \right) P - \left(\frac{X}{n} \right) t - C \left(\frac{X}{n} \right) = 0$$

Assuming that the government has access to lump-sum tax instruments, the social total cost (TC)

of producing industry output is given simply by its resource cost, or $TC = nC \left(\frac{X}{n} \right)$. For a small

change in a tax instrument, τ (either an ad valorem or a specific tax), it follows that

$$(6.22) \quad \frac{dTC}{d\mathbf{x}} = C' \left(\frac{X}{n} \right) \frac{dX}{d\mathbf{x}} + \left[C \left(\frac{X}{n} \right) - \frac{X}{n} C' \left(\frac{X}{n} \right) \right] \frac{dn}{d\mathbf{x}}.$$

The value to consumers for which the tax change is responsible is given by $P \frac{dX}{d\mathbf{x}}$.

Consequently, the change in the difference between consumer value and social cost, say Δ , is

$$(6.23) \quad \frac{d\Delta}{d\mathbf{x}} = \left[P - C' \left(\frac{X}{n} \right) \right] \frac{dX}{d\mathbf{x}} - \left[\frac{C \left(\frac{X}{n} \right)}{\left(\frac{X}{n} \right)} - C' \left(\frac{X}{n} \right) \right] \frac{X}{n} \frac{dn}{d\mathbf{x}}.$$

Equation (6.24) succinctly captures the two competing considerations in changing a tax rate that applies to imperfectly competitive industries. The first term is the product of the induced change in output and the difference between price and marginal cost of production for

³⁶ New entrants are assumed to exhibit the same oligopolistic behavior (as reflected in 2) as do other firms in the industry; see Mankiw and Whinston (1986) for an analysis of the welfare effects of entry in such a setting.

firms in the industry. If the number of firms in the industry were fixed, then this would be the only expression on the right side of (6.23), and it would carry the previous implication that, with the availability of lump-sum tax instruments, efficient taxation consists of equating price and marginal cost. The difficulty, of course, is that it is not the only term on the right side of (6.23). In this model it is necessary to subsidize an industry in order to equate price and marginal cost, and government subsidies encourage inefficient entry of new firms.

The welfare effect of tax policy on entry is captured by the second term on the right side of (6.23). This term is the product of the amount of output produced by new entrants and the difference between average and marginal costs for each firm in the industry. Subtracting (6.2) from (6.21) implies that

$$(6.24) \quad \frac{C\left(\frac{X}{n}\right)}{X/n} - C'\left(\frac{X}{n}\right) = -(1-t)\frac{X}{n} \frac{dP}{dX} (1+q) > 0,$$

which simply follows from the fact that price exceeds marginal cost. Hence average cost exceeds marginal cost, and new entry is inefficient, since marginal output is less expensively produced by existing firms than by new entrants.³⁷

The effect of introducing taxes can be identified by differentiating the identity that

$$X \equiv n\left(\frac{X}{n}\right), \text{ which yields}$$

³⁷ This equilibrium condition requires the production technology to exhibit decreasing average costs over some range of output.

$$(6.25) \quad \frac{dX}{d\mathbf{x}} = \left(\frac{X}{n}\right) \frac{dn}{d\mathbf{x}} + n \frac{d\left(\frac{X}{n}\right)}{d\mathbf{x}}.$$

Together, (6.21), (6.23) and (6.25) imply

$$(6.26) \quad \frac{d\Lambda}{d\mathbf{x}} = [P\mathbf{t} + t] \frac{dX}{d\mathbf{x}} + \left(\frac{N}{X}\right) \left[nC\left(\frac{X}{n}\right) - XC\left(\frac{X}{n}\right) \right] \frac{d\left(\frac{X}{n}\right)}{d\mathbf{x}}.$$

Starting from $t = J = 0$, it follows from (6.26) and (6.24) that $\frac{d\Lambda}{d\mathbf{x}} > 0$ if $\frac{d\left(\frac{X}{n}\right)}{d\mathbf{x}} > 0$, regardless of the effect of taxation on entry and exit. The intuition behind this result is that, while greater output by existing firms promotes efficiency (since price exceeds marginal cost), in the absence of taxation, price equals average cost and there is no welfare impact of marginal entry.

Recall from (6.24) that average cost exceeds marginal cost in equilibrium, and hence is a declining function of a firm's output. Therefore, increases in output per firm will reduce average cost and increase welfare. From the zero-profit condition (6.21), average cost is

$$(6.27) \quad AC\left(\frac{X}{n}\right) = \frac{C\left(\frac{X}{n}\right)}{X/n} = (1 - \mathbf{t})P - t.$$

Hence output per firm rises, and therefore welfare rises, in response to the introduction of taxes that reduce the right side of (6.27).

Equation (6.2) describes the firm's first-order condition for profit maximization. By (6.27), average output per firm (X/n) can be expressed as a decreasing function of $[P(1-J) - t]$, while the market demand curve allows us to express total output, X , as a function of P .

Appropriately differentiating both sides of (6.2) with respect to t , evaluating the resulting expression at $J = t = 0$, and collecting terms yields

$$(6.28) \quad \frac{dP}{dt} = \frac{1 + \frac{d(X/n)}{d(P-t)} \frac{dP}{dX} (1 + \mathbf{q}) - C'' \left(\frac{X}{n} \right) \frac{d(X/n)}{d(P-t)}}{1 + \frac{d(X/n)}{d(P-t)} \frac{dP}{dX} (1 + \mathbf{q}) - C'' \left(\frac{X}{n} \right) \frac{d(X/n)}{d(P-t)} + \frac{\mathbf{h}(1 + \mathbf{q})}{n}}.$$

where O is the elasticity of the inverse demand function, as defined above at (6.3). Since the conditions for industry stability imply that both the numerator and the denominator of the expression on the right side of (6.28) are positive,³⁸ it follows that $\left(\frac{dP}{dt} - 1 \right)$ has the same sign as $-O$. Hence a positive value of O implies that the introduction of a (positive) specific tax increases the market price by less than the amount of the tax, expanding per-firm output and thereby improving welfare.³⁹ The reason is that the reduced industry output due to a higher tax rate reduces $\frac{dP}{dX}$, which is a factor in the oligopolistic markup by which price is elevated above marginal cost. While the same consideration applies in other settings, the existence of free entry and exit is critical to the welfare result due to the induced attenuation of the effect of taxes on price.

Ad valorem taxation continues to be more attractive than specific taxation in industries with free entry and exit. Starting from $J = t = 0$, the introduction of an ad valorem tax reduces

³⁸ Seade (1980a) demonstrates that stability requires $C''(X/n) > (1 + \mathbf{q}) dP/dX$, and since $d(X/n)/d(P-t) < 0$, it follows that the numerator of (6.28) is positive. Seade (1980b) also adopts $\mathbf{h} + n/(1 + \mathbf{q}) > 0$ as a stability condition, noting (1980a) that it is a sufficient condition for a firm's marginal revenue to fall when other firms expand output, and that this condition implies that new entry is associated with greater industry output. Together, these stability conditions guarantee that the denominators of (6.28) and (6.29) are positive.

³⁹ See Besley (1989) and Delipalla and Keen (1992) for additional results and interpretation.

the right side of (6.27) if $\frac{dP}{dt} < P$. Appropriately differentiating both sides of (6.2) with respect

to J yields

$$(6.29) \quad \frac{dP}{dt} = P \left\{ \frac{1 + \frac{d(X/n)}{d[P(1-t)]} \frac{dP}{dX} (1+q) - C'' \left(\frac{X}{n} \right) \frac{d(X/n)}{d[P(1-t)]} + \left(\frac{X}{n} \right) \frac{dP}{dX} \frac{(1+q)}{P}}{1 + \frac{d(X/n)}{d[P(1-t)]} \frac{dP}{dX} (1+q) - C'' \left(\frac{X}{n} \right) \frac{d(X/n)}{d[P(1-t)]} + \frac{h(1+q)}{n}} \right\}.$$

Since the stability conditions imply that the denominator of the right side of (6.29) is positive, it

follows that $\left(\frac{dP}{dt} - P \right)$ has the same sign as $\left(\frac{dP}{dX} \frac{X}{P} - h \right)$. Hence the introduction of an ad

valorem tax improves welfare not only if O is positive, but also if O is negative but smaller in absolute value than the elasticity of the inverse demand function. This condition for welfare improvement is weaker than that for the introduction of specific taxes, thereby reflecting the relatively more potent effect of ad valorem taxes in reducing an imperfectly competitive firm's return from restricting output in order to elevate price.

6.4. Differentiated products.

Certain types of oligopolistic situations take the form of competition among firms selling products that are imperfect substitutes. Firms take actions that affect product attributes as well as output levels, and these actions are potentially affected by tax policies. Since there are many forms of competition between sellers of differentiated products, it can be difficult to draw general welfare conclusions concerning the impact of taxation in such settings; it is, however, possible to identify the major considerations on which the results turn.

Consider an industry of n firms selling products that differ along a univariate quality scale, indexed by α , so that firm i sells products of quality α_i , in which α_i represents a profit-maximizing choice made by the firm. Firm i produces output x_i at quality level α_i , with idiosyncratic costs given by $c_i(x_i, \alpha_i)$. The representative consumer's preferences are then responsible for the inverse demand function $\mathbf{p}(\mathbf{x}, \boldsymbol{\alpha})$, and the government imposes an ad valorem tax at a uniform rate on all sales in the industry.

Production takes place in two stages. First, firms select values of α_i , taking as fixed the elements of the $\boldsymbol{\alpha}$ vector other than α_i (interesting generalizations are possible by incorporating strategic interaction in the choice of $\boldsymbol{\alpha}$). Second, firms choose output levels x_i contingent on $\boldsymbol{\alpha}$ and taking the output of other firms as fixed. Of course, first stage choices of $\boldsymbol{\alpha}$ are made in anticipation of induced pricing and output effects in the second stage. Conditional on $\boldsymbol{\alpha}$, firm i 's optimal choice of x_i in the second stage must satisfy

$$(6.30) \quad (1 - \mathbf{t}) \left(p_i(x, \mathbf{n}) + \frac{\partial p_i(x, \mathbf{n})}{\partial x_i} x_i \right) = \frac{\partial c_i(x_i, \mathbf{n}_i)}{\partial x_i}.$$

Denoting the vector of values of x_i that solve (6.30) by $\mathbf{x}^*(\boldsymbol{\alpha})$, the first-order condition for the optimal choice of α_i is

$$(6.31) \quad (1 - \mathbf{t}) \left[\frac{\partial p_i(x^*(\mathbf{n}), \mathbf{n})}{\partial \mathbf{n}_i} + \sum_{j \neq i} \frac{\partial p_i(x^*(\mathbf{n}), \mathbf{n})}{\partial x_j} \frac{\partial x_j(\mathbf{n})}{\partial \mathbf{n}_i} \right]_{x_i} = \frac{\partial c_i(x_i, \mathbf{n}_i)}{\partial \mathbf{n}_i}.$$

Oligopolistic situations offer differing interpretations of the context and welfare interpretations of (6.30) and (6.31). From (6.30), it is clear that, conditional on $\boldsymbol{\alpha}$, imperfect competition leads to too little production, in the sense that prices exceed marginal costs. From

this observation it is tempting to conclude that (as before) the optimal tax policy is one that subsidizes the output of imperfectly competitive firms. The endogeneity of α has the potential to reverse this reasoning, however, since there is no presumption, from the general form of (6.31), that quality choices are optimal in the absence of taxation.

Quality choice may be suboptimal for many reasons. The first is that firms select quality levels based on their impact on marginal demand and not on the valuation of inframarginal output by the same firm. A second reason is that one firm's return to quality may come at the expense of other firms, and such pecuniary externalities affect welfare in situations in which prices differ from marginal costs. And a third reason is that quality choice in the first stage affects the output decisions of other firms in the second stage, a strategic consideration that creates inefficiencies whenever demand for one commodity is affected by the prices of others.

The examples analyzed in the literature generally share the feature that the introduction of (positive) ad valorem taxation can improve welfare.⁴⁰ Equation (6.31) identifies the strategic consideration responsible for this effect, since, if commodities i and j are substitutes in demand

$\left(\frac{\partial p_i}{\partial x_j} < 0 \right)$, and strategic substitutes in supply $\left(\frac{\partial x_j}{\partial n_i} < 0 \right)$, then, in the absence of taxation,

quality is oversupplied in the sense that $\frac{\partial p_i}{\partial n_i} < \frac{\partial c_i}{\partial n_i}$. Ad valorem taxation typically reduces

quality levels, thereby quite possibly improving welfare even though it serves further to distort the output level choice reflected in (6.30). This implication is very similar to the result (from the previous section) that ad valorem taxation is desirable in a model with free entry and exit, and indeed, these cases share many similarities. Firms described by (6.30) and (6.31) select output

⁴⁰ See, for example, Kay and Keen (1983) and Cremer and Thisse (1994). Besley and Suzumura (1992) analyze a two-stage game of strategic investment in cost-reducing technology with similar features. Kay and Keen (1991) consider the nature of preferences that determine the effect of taxation on product quality.

levels at which prices exceed marginal costs, but also select quality levels at which marginal costs exceed non-strategic returns. One can think of (6.31) as characterizing excessive “entry” along the quality dimension, and therefore positive ad valorem taxation as being desirable to the extent that it stimulates output per unit of effective quality. Hence, there is potentially a salutary role of taxes in reducing quality, particularly if oligopolistic competition is aggressive in non-price dimensions.

7. Intertemporal taxation

This section considers optimal taxation in intertemporal settings, generally resuming the assumption of perfect competition. Due in part to interest generated by the “consumption tax” advocacy of Fisher and Fisher (1942), Kaldor (1955), and others, one intertemporal issue in particular has received extensive attention: the optimal tax rate on capital income. One of the notable developments of modern optimal tax theory is the finding that, in a simplified second-best setting with identical individuals and in which the government can tax both capital income and labor income, welfare maximization implies zero taxes on capital income in the steady state. This finding reflects, of course, the highly distortionary nature of capital income taxes over long periods of time, but is nevertheless surprising in view of the standard Ramsey intuition that the deadweight loss is zero for the first dollar collected by any tax – and therefore, in the absence of spillovers between markets, all optimal tax rates are strictly positive. Where this intuition fails in the intertemporal context is that it does not account for just how extremely distortionary capital taxation can be even at very low rates of tax – specifically, that low tax rates correspond to distortionary intertemporal tax wedges that grow over time.

The main findings concerning optimal capital taxation are reported by Chamley (1986) and Judd (1985). Subsequent research by Jones Manuelli, and Rossi (1993, 1997), Milesi-

Ferretti and Roubini (1998), and others extends its logic to the intertemporal taxation of factors other than capital. In particular, to the extent that wages represent returns to the accumulation of human capital, labor income taxes have capital components and are likewise optimally zero in the steady state. Indeed, the logic of optimal intertemporal taxation is such that there are plausible circumstances in which all taxes may be zero in the steady state. Of course, governments that attempt to implement such optimal taxes would need to amass considerable unspent tax revenue in years prior to the steady state in order to maintain intertemporal budget balance. Before considering these implications, however, it is useful to review the source of the basic intertemporal results concerning capital taxation alone.

7.1. *Basic capital income taxation: introduction*

The logic of the result that capital is untaxed in the steady state is apparent from working through a simplified version of Chamley's problem. Consider the case of an economy consisting of identical consumers who maximize the present discounted value of utility over infinite horizons:

$$(7.1) \quad \sum_{t=0}^{\infty} \mathbf{b}^t u(C_t, L_t)$$

in which \mathbf{b} is the one-period discount factor ($\mathbf{b} = (1+\mathbf{d})^{-1}$, \mathbf{d} being an individual's subjective discount rate), taken to be constant for all individuals in all periods. $u(C_t, L_t)$ is a consumer's contemporaneous utility in year t , an increasing function of consumption (C_t) and a decreasing function of labor supplied (L_t).

Consumers have initial wealth of K_0 and earn labor income in period zero equal to $w_0 L_0$, in which w_0 is the after-tax wage rate in period zero. Labor income is received at the start of

each period, and consumption also takes place at the start of each period, so any capital income is earned while a period elapses. A consumer therefore dissaves $(C_0 - w_0L_0)$ in the initial period, and has the lifetime budget constraint

$$(7.2) \quad \sum_{t=1}^{\infty} (C_t - w_t L_t) \prod_{s=1}^t (1 + r_{s-1})^{-1} \leq K_0 - (C_0 - w_0 L_0)$$

in which r_t is the (after-tax) return earned by capital during period t .

Assuming that the constraint (7.2) is binding (and that the solution entails interior optima), the first-order conditions that characterize the maximum of (7.1) are

$$(7.3) \quad w_t \frac{\partial u}{\partial C_t} = - \frac{\partial u}{\partial L_t}$$

$$(7.4) \quad \frac{\partial u}{\partial C_t} = \frac{\partial u}{\partial C_{t+1}} (1 + r_t) \mathbf{b}.$$

Equation (7.4) in turn implies

$$(7.5) \quad \frac{\partial u}{\partial C_0} = \frac{\partial u}{\partial C_n} \mathbf{b}^n \prod_{i=0}^{n-1} (1 + r_i)$$

Combining the budget constraint, (7.2), and the first-order conditions, (7.3) and (7.5), yields

$$(7.6) \quad \sum_{t=0}^{\infty} \mathbf{b}^t \left[\frac{\partial u}{\partial C_t} C_t - \frac{\partial u}{\partial L_t} L_t \right] \leq K_0 \frac{\partial u}{\partial C_0}.$$

As the economy consists of identical individuals, we consider the most notationally simple case of one such individual. The period-by-period resource constraint for such an economy is

$$(7.7) \quad C_t + G_t + K_{t+1} \leq F_t(K_t, L_t) + K_t, \forall t$$

in which G_t is government consumption in period t , and $F(K_t, L_t)$ is the economy's production function. The path of government consumption is taken to be exogenous and (for simplicity) capital is assumed not to depreciate. Inequality (7.7) expresses the idea that the sum of private and public consumption, plus net capital accumulation, cannot exceed the output of the economy.

7.2. *The steady state*

The most straightforward way to evaluate the properties of optimal taxation is to consider the first-order conditions that correspond to maximizing (7.1) subject to (7.6) and (7.7), taking C_t , L_t and K_t to be control variables. (It is noteworthy that (7.7) actually represents a separate constraint for each period.) The first-order condition corresponding to an interior choice of C_t is

$$(7.8) \quad \mathbf{b}' \left\{ \frac{\partial u}{\partial C_t} - \mathbf{I} \left[\frac{\partial u}{\partial C_t} + C_t \frac{\partial^2 u}{\partial C_t^2} - \frac{\partial^2 u}{\partial L_t \partial C_t} L_t \right] \right\} = \mathbf{m}_t,$$

in which \mathbf{I} is the Lagrange multiplier corresponding to the constraint (7.6), and \mathbf{m}_t is the Lagrange multiplier corresponding to condition (7.7) in period t . The first-order condition corresponding to an interior choice of K_t is

$$(7.9) \quad \mathbf{m}_t \left(1 + \frac{\partial F}{\partial K_t} \right) = \mathbf{m}_{t-1}.$$

Consider an economy that ultimately settles into a long-run steady state in which economic variables, specifically C_t and L_t , are unchanging. Since the term in braces on the left side of equation (7.8) is unchanging in this steady state, it follows that $\mathbf{m} = \mathbf{b}\mathbf{m}_{-1}$. Imposing this equality on (7.9) yields

$$(7.10) \quad \mathbf{b} \left(1 + \frac{\partial F}{\partial K_t} \right) = 1.$$

Equation (7.4), one of the consumer's first-order conditions, implies that, if $C_t = C_{t+1}$ and

$L_t = L_{t+1}$, then $\mathbf{b}(1 + r_t) = 1$. Consequently, (7.10) implies that $r_t = \frac{\partial F}{\partial K_t}$ in the steady state.

Recall that r_t is the after-tax return received by savers during period t . In a competitive market,

$\frac{\partial F}{\partial K_t}$ is the pre-tax return to investors. The equality of r_t and $\frac{\partial F}{\partial K_t}$ therefore implies that savings

are untaxed.

7.3. *Interpreting the solution*

The finding that capital income should be untaxed in the steady state contradicts the naïve intuition that, since taxes on labor income distort labor-leisure choices in the steady state, a minor reduction in labor taxes financed by a very small tax on capital income would improve the welfare of the representative individual. Where this intuition fails is that even very low-rate taxes on capital income generate first-order consumption distortions over long horizons. The reason is that a capital income tax at a very low rate creates a small distortion between consumption in periods t and $(t+1)$, but a large distortion between consumption in period t and consumption in period $(t+n)$, for large n .

It does not by any means follow from the steady-state properties of the optimal program that capital income taxes are always zero. Indeed, Chamley (1986) offers an example in which consumers have utility functions that are additively separable in consumption and leisure and iso-elastic in consumption, for which the optimal dynamic tax configuration is one in which the government imposes a capital income tax at a 100 percent rate for an initial period and 0 thereafter.⁴¹ Chamley offers the intuition that high initial rates of capital tax serve to tax away the value of initial capital, thereby acting in part as a lump-sum tax and in part as a very distortionary tax on capital accumulation during the regime of 100 percent tax rates.

This intuitive interpretation of the optimal tax pattern is correct but incomplete, since even if the government possessed an additional tax instrument, permitting it to extract up to 100 percent of the value of initial capital from the private sector, it might still wish to use standard capital income taxes to raise additional revenue in the short run. The reason is that capital income taxes in early years distort the choice between present and future consumption, but leave the margins among consumption at different future dates unaffected; nonzero capital income taxes in later years also distort the pattern of future consumption. If one thinks of consumption at different dates as separate commodities, then the Ramsey analysis suggests that optimal policy entails equal (revenue-adjusted) marginal distortions to consumption in each period. Because consumption taxes are not included in the government's instrument set, this outcome is approximated by the use of capital income taxes in early years but not in later years.

Analytically, the equations (7.8) and (7.9) that characterize the optimal path would be formally unchanged even if the government had access to an additional instrument that extracts the value of initial capital. Of course, these conditions would then imply a different tax rate path, but its

⁴¹ Chamley constrains the government not to impose capital income taxes at greater than a 100 percent rate in order to rule out nondistortionary lump-sum initial capital levies as a method of government finance.

general feature that capital income tax rates fall over time would persist, and therefore not reflect the desire to tax the value of initial capital.

The time-varying nature of optimal capital taxation makes such a policy time-inconsistent, in that whatever profile of future taxes that is optimal as of year t would not be optimal as of year $t+1$, and optimizing governments might therefore be tempted not to follow through on previously announced tax plans. Private agents, anticipating such behavior by governments, could not then be expected to respond to announced tax plans in the same way that they would if the government could commit reliably to the taxes that it announces. This is just one of many examples of the time inconsistency of optimal plans, a feature that takes on special significance in an economy in which private agents hold capital, the value of which governments might find attractive to seize through their tax policies. While there are attempts to identify optimal time-consistent capital tax policies by somehow constraining government behavior, all such efforts confront the fundamental problem that the mere existence of capital, together with the distortionary nature of income taxation, creates incentives for benevolent governments to behave in a time-inconsistent fashion.⁴² The analysis of this section follows the majority of the literature in considering government policies under the assumption that it is possible to make credible commitments.

⁴² There is an entirely separate, but relevant, issue that arises concerning the benevolence of governments over time. The optimal tax path is one that accumulates enormous government revenues in the early years in order to finance expenditures in later years (in which capital income tax rates will be zero). Given the implausibility of actual governments bestowing upon their successors such hard-won budget surpluses in order to finance efficient taxation in the future, it is worth bearing in mind that optimal taxation is a useful ideal if not a reality. In practice, the opposite pattern – in which governments run sizable deficits partly to constrain the fiscal choices of future governments (as in Persson and Svensson, 1989) – is much more common.

7.4. *Human capital accumulation and endogenous growth*

The model described by (7.1) – (7.10) carries implications for the taxation of labor income, but these are very difficult to characterize succinctly (other than to say that labor income taxes are positive and unchanging in the steady state). The treatment of labor as a factor of production is somewhat stylized, in that all labor is homogeneous and represents forgone leisure opportunities (with which individuals are endowed). The economy described by (7.1) – (7.10) grows via capital accumulation (and shrinks during periods of capital decumulation). As shown by Lucas (1990), Laitner (1995) and others, the qualitative features of optimal taxation are unaffected by introducing exogenous technical progress that generates economic growth and causes the economy to settle into a balanced growth path in the long run. Judd (1999) obtains the similar result that the long-run average optimal capital income tax rate is likewise zero for economies that do not converge to steady states. Extensions to economies with production subject to stochastic shocks, such as those by Zhu (1992) and Chari, Christiano, and Kehoe (1994), produce the result that the optimal tax on capital income is generally very low or zero.

The impact of fiscal policies in settings in which economies grow endogenously is the subject of a closely related literature. There is more than one potential source of endogenous growth, perhaps the most obvious being the accumulation of human capital, along with others that include social increasing returns to scale due to the productivity-enhancing effects of infrastructure and other public goods.⁴³ These models share in common the characteristic that the endogeneity of the growth rate arises from some positive externality. As in traditional public finance analysis, the presence of externalities means that an equilibrium without distortionary

⁴³ See, for example, Lucas (1990), King and Rebelo (1990), Rebelo (1991), Trostel (1993), and Stokey and Rebelo (1995). The sources of endogenous growth analyzed by Eaton (1981) and Hamilton (1987) differ from these in reflecting the saving and portfolio preferences of consumers, and need not entail any productive externalities.

taxes will generally not be Pareto-optimal. Thus, optimal tax design must take the presence of such externalities into account, as discussed in Section 5.2 above.

In endogenous growth models, the accumulation of human capital generates externalities through intergenerational transmission of acquired skills. However, one may consider the accumulation of human capital and its associated externality separately, and it is useful to do so in understanding the effects on optimal tax results. Human capital accumulation itself (without any intergenerational transmission of skills) is easily incorporated in the model (7.1) - (7.10), as labor income then represents the return to past forgone consumption and leisure (assuming that both goods and time contribute to the accumulation of human capital), as well as contemporaneous forgone leisure. Since labor income taxes then effectively tax intertemporal labor/leisure choices in much the same way that capital income taxes effectively tax intertemporal consumption choices, it is not surprising that the optimal dynamic tax path is one in which labor income taxes, as well as capital income taxes, are zero in the steady state (as in Jones, Manuelli, and Rossi (1993, 1997) and Milesi-Ferretti and Roubini (1998)).

To show this more formally, consider the case in which consumers have three uses for their time: they can work, for which they receive a wage, they can accumulate human capital, which increases future wages, and they can consume leisure. Denote by E_t the amount of time that the consumer devotes to human capital accumulation in period t . In the simple case in which utility is a function only of consumption and leisure, so that the disutility of time working equals the disutility of devoting the same amount of time to human capital accumulation, the consumer's maximand becomes

$$(7.1) \quad \sum_{t=0}^{\infty} \mathbf{b}^t u(C_t, L_t + E_t).$$

Let H_t denote the consumer's period- t stock of human capital; purely for simplicity assume that human capital does not depreciate. Accumulation of human capital occurs by devoting time and valuable goods and services (e.g., educational resources) to producing additional human capital. Let $M(E, B)$ denote the (time-invariant) human capital production function, in which B represents the value of goods and services devoted to human capital. The accumulation of human capital is therefore constrained by the relationship:

$$(7.11) \quad H_{t+1} \leq M(E_t, B_t) + H_t, \forall t.$$

The ability of consumers to allocate some of the economy's output to the accumulation of human capital requires a modification in the economy's resource constraint, as well as a slightly different specification of the production function, so that (7.7) becomes

$$(7.7 \text{)} \quad C_t + B_t + G_t + K_{t+1} \leq F_t(K_t, L_t, H_t) + K_t, \forall t.$$

The existence of human capital does not change (7.6), the consumer's intertemporal budget constraint.

The introduction of human capital adds a new state variable (H_t) to the optimal tax problem, as well as two new choice variables (E_t and B_t), a new constraint (7.11), and requires the modification of the objective function and one of the previous constraints. Once again, the most straightforward way to describe the properties of the optimal solution is to maximize (7.1) subject to (7.6), (7.7), and (7.11), taking $C_t, L_t, K_t, B_t,$ and H_t to be control variables. Equations (7.8) and (7.9) continue to hold, and so, therefore, does (7.10) and its implication that the return to saving is untaxed in the steady state.

The first-order condition corresponding to an interior choice of H_t is

$$(7.12) \quad \mathbf{m}_t \frac{\partial F}{\partial H_t} + \mathbf{y}_t = \mathbf{y}_{t-1},$$

in which \mathbf{R}_t is the Lagrange multiplier on the constraint (7.11) in period t . The first-order condition corresponding to an interior choice of B_t is

$$(7.13) \quad \mathbf{y}_t \frac{\partial M}{\partial B_t} = \mathbf{m}_t.$$

Since (7.8) continues to characterize the optimal solution, it follows that a steady state in which C , L , E and B are unchanging implies that $\mathbf{m}_t = \mathbf{b}\mathbf{m}_{t-1}$. From (7.13), it then follows that, in the steady state in which $\frac{\partial M}{\partial B_t}$ is unchanging, it must be the case that $\mathbf{y}_t = \mathbf{b}\mathbf{y}_{t-1}$. Together,

(7.12), (7.13), and $\mathbf{y}_t = \mathbf{b}\mathbf{y}_{t-1}$ imply

$$(7.14) \quad 1 + \frac{\partial M}{\partial B_t} \frac{\partial F}{\partial H_t} = \frac{1}{\mathbf{b}}.$$

From the steady state condition $1/\mathcal{S} = (1+r)$ it follows that

$$(7.15) \quad \frac{\partial M}{\partial B_t} \frac{\partial F}{\partial H_t} = r.$$

Equation (7.15) characterizes the steady state economy under optimal taxation, so it is instructive to compare (7.15) to the consumer's first-order conditions. An individual who defers consumption invests either in physical capital or in human capital. (7.4) describes the (interior) first-order condition for investing in physical capital; the analogous first-order condition for investing in human capital is

$$(7.16) \quad \frac{\partial u}{\partial C_t} = \frac{\partial u}{\partial C_{t+1}} \left(1 + \frac{\partial M}{\partial B_t} \frac{\partial w}{\partial H_t} \right) \mathbf{b},$$

in which w is the after-tax wage. The term $\frac{\partial w}{\partial H_t}$ in (7.16) therefore equals the single-period after-tax private return from accumulating an additional unit of human capital.

Equations (7.16) and (7.4) together imply that

$$\frac{\partial M}{\partial B_t} \frac{\partial w}{\partial H_t} = r_t,$$

which, together with (7.15), implies that

$$(7.17) \quad \frac{\partial w}{\partial H_t} = \frac{\partial F}{\partial H_t}.$$

The left side of (7.17) is the amount of additional after-tax income received by a worker who accumulates one more unit of human capital; the right side of (7.17) is the marginal product of this additional unit of human capital. Assuming that there are no productivity spillovers, so that the productivity gains from additional human capital are embodied in the effective labor supply of workers who possess the human capital, factor market competition guarantees that the right side of (7.17) equals the effect of human capital accumulation on pretax wages. Since the left side of (7.17) is the effect of human capital accumulation on after-tax wages, it follows that labor income must be untaxed in the steady state.

Note that this result depends on (7.16), which applies only if human capital accumulation requires inputs of goods – forgone consumption – as well as leisure. If this is not the case – if

human capital is accumulated simply through forgone leisure – then the results that follow will not hold. In particular, the tax on labor income will no longer distort the accumulation of human capital, because the entire cost of investment will be tax deductible. It follows, then, that if goods inputs are deductible, the human capital decision will remain undistorted by labor income taxes, in which case there is no requirement that labor income taxes equal zero in the steady state. As shown by Milesi-Ferretti and Roubini (1998), governments with a sufficient number of tax instruments can effectively decouple the taxation of human capital accumulation from the taxation of the return to forgone leisure.

The analysis of human capital accumulation is really a subset of a broader range of issues in which tax instruments are restricted in one way or another. In other settings, Jones, Manuelli and Rossi (1993, 1997) observe that restrictions on the range of tax instruments available to the government, or the presence of public goods in the aggregate production function, change the nature of even steady state taxation in a way that can make it optimal for the government to impose taxes on capital income. For example, there might be two types of labor in the economy, with properties (such as differing labor supply elasticities) that would make it optimal to tax the incomes they generate at different rates. If the government is constrained to select a single labor income tax rate, then the optimal tax rate on capital income might differ from zero in the steady state in order to compensate for the government's inability to tailor its labor income taxes. Judd (1997) analyzes the implications of restrictions on the ability of the government to control monopolistic and other noncompetitive market behavior, in which case tax policy may function as a different kind of second-best corrective mechanism; his work identifies circumstances under which the optimal tax on capital income may then be negative in the steady state. Coleman (2000) comes to a similar conclusion in a setting in which the government can impose separate

consumption and labor income taxes, and there are restrictions on the range of available tax instruments. Aiyagari (1995) considers the implications of market incompleteness that leaves individuals incapable of diversifying idiosyncratic risks. The resulting demand for precautionary saving leads to a positive optimal tax rate on capital income, even in the steady state.

Correia (1996) notes that many of these considerations stem from the existence of an important productive factor that the government is unable (for some reason) to tax or to subsidize. Depending on the application, this factor might represent inframarginal profits from decreasing returns to scale activity, the returns to monopolistic rents, positive or negative productivity spillovers, labor or capital of specific types, or the value of goods devoted to human capital accumulation. The effect of such a factor on optimal capital taxation is instructive. Consider the case in which consumers provide an additional productive service, denoted A_t , for which they experience disutility and which the government is unable to tax. The consumer's utility becomes

$$(7.1) \quad \sum_{t=0}^{\infty} \mathbf{b}^t u(C_t, L_t, A_t)$$

which the government maximizes subject to the conditions:

$$(7.6) \quad \sum_{t=0}^{\infty} \mathbf{b}^t \left[\frac{\partial u}{\partial C_t} C_t - \frac{\partial u}{\partial L_t} L_t - \frac{\partial u}{\partial A_t} A_t \right] \leq K_0 \frac{\partial u}{\partial C_0}$$

and

$$(7.7) \quad C_t + G_t + K_{t+1} \leq F_t(K_t, L_t, A_t) + K_t.$$

Greater levels of activity A generate pretax returns of $\frac{\partial F}{\partial A}$. The inability of the government to tax the return to A therefore imposes the additional constraint:

$$(7.18) \quad \frac{\partial F}{\partial A_t} \leq \frac{\partial U / \partial A_t}{\partial U / \partial C_t}$$

The first order condition corresponding to an interior choice of C_t is

$$(7.19) \quad \mathbf{b}' \left\{ \frac{\partial u}{\partial C_t} - \mathbf{l} \left[\frac{\partial u}{\partial C_t} + C_t \frac{\partial^2 u}{\partial C_t^2} - \frac{\partial^2 u}{\partial L_t \partial C_t} L_t \right] \right\} - \mathbf{q}_t \frac{\left\{ \frac{\partial^2 U}{\partial A_t \partial C_t} \frac{\partial U}{\partial C_t} - \frac{\partial^2 U}{\partial C_t^2} \frac{\partial U}{\partial A_t} \right\}}{\left(\frac{\partial U}{\partial C_t} \right)^2} = \mathbf{m}_t,$$

in which \mathbf{q} is the Lagrange multiplier corresponding to the constraint (7.18). The first-order condition corresponding to an interior choice of K_t is

$$(7.20) \quad \mathbf{m}_t \left(1 + \frac{\partial F}{\partial K_t} + \mathbf{q}_t \frac{\partial^2 F}{\partial A_t \partial K_t} \right) = \mathbf{m}_{t-1}$$

Taking the Lagrange multiplier \mathbf{q} to grow at rate \mathbf{b} in the steady state, these conditions together imply that, in the steady state,

$$(7.21) \quad r_t = \frac{\partial F}{\partial K_t} + \mathbf{q}_t \frac{\partial^2 F}{\partial A_t \partial K_t}.$$

Equation (7.21) is inconsistent with zero capital taxation whenever two conditions hold simultaneously: that constraint (7.18) binds, and that changes in K affect the marginal productivity of A .

In the case of ordinary human capital accumulation, the government does not seek to tax A (which can be interpreted as past labor effort used to accumulate human capital), so $\mathbf{q} = 0$ and physical capital is untaxed as well. In the case of economies with public goods or other types of productive externalities, or those in which heterogeneous inputs must receive identical tax treatment, a government that cannot use corrective taxation to induce efficient decentralized behavior will change its other taxes to accommodate the missing market.⁴⁴ As a result, steady state tax rates on capital will be greater than, equal to, or less than zero according to the nature of the externality (positive or negative) and the complementarity or substitutability of the untaxed factor with capital – a standard implication along the lines of Corlett and Hague (1953) in a static setting.

7.5. *Results from life-cycle models*

Though undoubtedly a powerful and illuminating result, the convergence of the optimal capital income tax to zero rests on the implausible assumption that agents live forever or behave in an equivalent manner with respect to their heirs. Without infinite lifetimes, no such result holds, although intuition suggests that long but finite lifetimes still would place strong bounds on the size of the optimal capital income tax. However, with finite lifetimes also comes the complication of heterogeneity with respect to age cohort, which tax policy optimization must take into account. Thus, there is more to learn from consideration of finite-lifetime, overlapping generation (*OG*) models than that capital income taxes should be low, if not zero, in the long run.

⁴⁴ Auerbach (1979) offers a similar analysis of the optimal taxation of heterogeneous capital goods in the presence of other constraints. Coleman's (2000) analysis of optimal consumption and labor income taxes takes the path of future government spending to be fixed in nominal terms, which implies that, in the steady state, the combination of a consumption tax and a labor subsidy relaxes the government's revenue requirement by reducing real government spending. Coleman finds that, if the labor income tax is constrained to be non-negative, then the optimal steady state labor income tax rate is zero and the tax on income from capital (which is a substitute for labor) is negative.

The Diamond (1965) model, in which each generation lives for two periods, consuming in both and working in the first, provided the basis for the initial research on optimal taxation in OG models. In this model, without bequests, the lifetime budget constraint for the representative household born in period t may be written:

$$(7.22) \quad C_t^1 + \left(\frac{1}{1+r_{t+1}} \right) C_{t+1}^2 = w_t L_t$$

where C^1 is consumption when young, C^2 is consumption when old, L is labor supply when young, and subscripts indicate periods in which activity occurs.

As is clear from this expression, endowing the government with two instruments, proportional taxes on labor income (which affect w) and capital income (which affects r), is equivalent to allowing the government to tax first- and second-period consumption, at possibly different rates. A zero-tax on capital income – a labor income tax – would result in uniform taxation of consumption in the two periods.

Using this model, papers by Diamond (1973), Pestieau (1974), Auerbach (1979), and Atkinson and Sandmo (1980) characterized optimal steady-state taxes under different assumptions about instruments available to the government. Two general results from this literature are that (1) with government debt available to redistribute resources across generations, the marginal product of capital should converge to the intertemporal discount rate embodied in the government's social welfare function; and (2) in this equilibrium, optimal taxes on labor and capital facing individual cohorts should follow the standard three-good analysis of static optimal tax theory, with a zero tax on capital income being optimal only for a certain class of preferences. Result (1) confirms that Cass's (1965) "modified Golden rule" result holds even in the presence of distortionary taxation. It is analogous to the Chamley-Judd result discussed

above. However, as result (2) confirms, this does not imply that capital income taxes converge to zero. The marginal product of capital is being equated to the *government's* discount rate (for comparing the consumption of different cohorts at different points in time), not the discount rate used by individual households in comparing their own first- and second-period consumption.

These results, like those derived for the infinite-lifetime case, tell us little about the nature of optimal tax schedules in transition; nor are they useful in determining how the long-run optimum might differ if the government faced constraints on its short-run policy. For example, if the optimal path for capital income taxes were one of high taxes declining to zero (as in Chamley's analysis), but the government's decision whether or not to abolish capital income taxes had to be made on a once-and-for-all basis, would it still improve economic efficiency to abolish capital income taxes? As transition constraints are a major concern of actual tax policy decisions, understanding the linkage between transition and long-run policy is important.

Analyzing the efficiency (and incidence) effects of tax policies in transition has been a major objective of the literature utilizing dynamic computable general equilibrium (*CGE*) models based on more realistic characterizations of life-cycle behavior. Auerbach, Kotlikoff and Skinner (1983) and Auerbach and Kotlikoff (1987) developed a 55-generation *OG* model with endogenous labor supply and retirement, in which agents alive during the transition from one steady state to another have perfect foresight about future factor prices and tax rates. Their central simulations consider the impact of switching immediately from a uniform tax on labor and capital income to a tax on labor income or a consumption tax. While such taxes appear equivalent in terms of the lifetime budget constraint represented in (7.22), as well as in the 55-period version of this budget constraint, they are *not* the same with respect to transition generations, who begin the transition with previously accumulated life-cycle wealth. For these

transition generations, a consumption tax is equivalent to a tax on labor income *plus* a tax on existing wealth – a capital levy. This can be seen by considering an amended version of (7.22) that has some measure of existing assets, A_t , on the right side. Thus, the transition to a consumption tax is more attractive than a transition to a labor income tax from the standpoint of economic efficiency.

Determining the efficiency differences between these two reforms is complicated by the fact that the reforms also have different intergenerational incidence, the consumption tax harming initial generations at the expense of future generations, the labor income tax doing the reverse. As a result, the steady-state welfare gain overstates the efficiency gain in the case of the consumption tax, for it reflects not only efficiency gains but also transfers from transition generations. By the same logic, the steady-state welfare gain understates the efficiency gain in the case of the labor income tax. To separate incidence from efficiency effects, the authors construct a hypothetical “lump-sum redistribution authority” that makes balanced-budget lump-sum taxes and transfers among generations to ensure that all transition generations are kept at the pre-reform utility level and all post-transition generations enjoy an equal increase in lifetime utility, an increase that can be viewed as a measure of the policy’s efficiency gain (or loss, if negative). With this adjustment, and for base case parameter assumptions, the transition to a consumption tax is predicted to increase economic efficiency, while the transition to a labor income tax would reduce economic efficiency.

The key lesson of these simulations is that tax systems that appear to be equivalent from the perspective of a representative individual may differ significantly in an economy with different age cohorts. A corollary is that adopting a consumption tax but simultaneously providing transition relief for those harmed by the tax in transition will offset not only adverse

distributional effects, but also the efficiency benefits of the capital levy. Auerbach (1996) illustrates this result in an analysis of a range of consumption-type tax reform proposals that vary in the extent to which they provide transition relief. The putative efficiency advantage of the consumption tax relies, of course, on the ability of the government to use the implicit capital levy “just once” and raises the question of dynamic inconsistency discussed above.

Just as it is possible to extend the representative-agent, infinite-horizon model to include human capital accumulation, this has been one direction in which dynamic *CGE* models have been extended in recent years, most notably by Heckman, Lochner and Taber (1998).

8. Conclusions

The analysis of excess burden and optimal taxation is one of the oldest subjects in applied economics, yet research continues to offer important new insights that build on the original work of Dupuit, Jenkin, Marshall, Pigou, Ramsey, Hotelling, and others. Fundamentally, it remains true that departures from marginal cost pricing are associated with excess burden, that the magnitude of excess burden is roughly proportional to the square of any such departure, and that efficient tax systems are ones that minimize excess burden subject to achieving other objectives. The contribution of modern analysis is to identify new and important reasons for prices and marginal costs to differ, to assess their practical magnitudes, and to consider their implications for taxation.

One of the major developments of the last fifty years is the widespread application of rigorous empirical methods to analyze the efficiency of the tax system. Empirical work not only assists the formation and analysis of economic policy, but also plays a critical role in distinguishing important from less-important theoretical considerations, thereby contributing to further theoretical development. Properly executed, empirical analysis is not only consistent

with the welfare theory that underlies normative public finance, but also takes the theory further by testing its implications and offering reliable measurement of parameters that are critical to the assessment of tax systems.

Recognition of the importance of population heterogeneity and of the potential complications of evaluating policy reforms with pre-existing distortions has motivated much of the recent normative work in public finance. The new learning serves generally to highlight the value of Ramsey's insights by demonstrating their application to a variety of settings, including those with population heterogeneity and a wide range of available tax instruments. Mirrlees differs from Ramsey in focussing on the role of informational asymmetries between governments and taxpayers as a determinant of the shape of optimal tax schedules; nevertheless, Ramsey-like conditions characterize optimal tax policy even in this setting.

The efficiency of the tax system is a topic of enduring importance and continuing investigation. Economic analysis has much to offer on the topic of efficiency, and indeed, is occasionally criticized for offering too much. The other chapters in this Handbook offer what is perhaps an illustration of this proposition by examining both positive and normative aspects of taxation in a wide variety of settings.

References

- Aiyagari, S. Rao, 1995, Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting, *Journal of Political Economy* 106, 1158-1175.
- Atkinson, Anthony B. and Agnar Sandmo, 1980, Welfare implications of the taxation of savings, *Economic Journal* 90, 529-549.
- Atkinson, Anthony B. and Nicholas H. Stern, 1974, Pigou, taxation and public goods, *Review of Economic Studies* 41, 119-128.
- Atkinson, Anthony B. and Joseph E. Stiglitz, 1972, The structure of indirect taxation and economic efficiency, *Journal of Public Economics* 1, 97-119.
- Atkinson, Anthony B. and Joseph E. Stiglitz, 1976, The design of tax structure: Direct versus indirect taxation, *Journal of Public Economics* 6, 55-75.
- Atkinson, Anthony B. and Joseph E. Stiglitz, 1980, *Lectures in public economics* (McGraw-Hill, New York).
- Auerbach, Alan J., 1979, The optimal taxation of heterogeneous capital, *Quarterly Journal of Economics* 93, 589-612.
- Auerbach, Alan J., 1985, The theory of excess burden and optimal taxation, in: Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, vol. 1 (North-Holland, Amsterdam), 61-127.
- Auerbach, Alan J., 1996, Tax reform, capital allocation, efficiency and growth, in: Henry J. Aaron and William G. Gale, eds., *Economic effects of fundamental tax reform* (Brookings, Washington, D.C.), 29-81.
- Auerbach, Alan J., and James R. Hines, Jr., 2001, Perfect taxation and imperfect competition, NBER Working Paper No. 8138.
- Auerbach, Alan J. and Laurence Kotlikoff, 1987, *Dynamic fiscal policy* (Cambridge University Press, Cambridge, UK).
- Auerbach, Alan J., Laurence J. Kotlikoff and Jonathan Skinner, 1983. The efficiency gains from dynamic tax reform, *International Economic Review* 24, 81-100.
- Auten, Gerald and Robert Carroll, 1999, The effect of income taxes on household income, *Review of Economics and Statistics* 81, 681-693.
- Ballard, Charles L. and Don Fullerton, 1992, Distortionary taxes and the provision of public goods, *Journal of Economic Perspectives* 6, 117-131.

- Besley, Timothy, 1989, Commodity taxation and imperfect competition: A note on the effects of entry, *Journal of Public Economics* 40, 359-366.
- Besley, Timothy and Kotaro Suzumura, 1992, Taxation and welfare in an oligopoly with strategic commitment, *International Economic Review* 33, 413-431.
- Bradford, David, 1986, *Untangling the income tax* (Harvard University Press, Cambridge, MA).
- Browning, Edgar K., 1975, Labor supply distortions of social security, *Southern Economic Journal* 42, 243-252.
- Cass, David, 1965, Optimum growth in an aggregative model of capital accumulation, *Review of Economic Studies* 32, 233-240
- Chamley, Christophe, 1986, Optimal taxation of capital income in general equilibrium with infinite lives, *Econometrica* 54, 607-622.
- Chari, V. V., Lawrence J. Christiano and Patrick J. Kehoe, 1994, Optimal fiscal policy in a business cycle model, *Journal of Political Economy* 102, 617-652.
- Coleman, Wilbur John, II, 2000, Welfare and optimum dynamic taxation of consumption and income, *Journal of Public Economics* 76, 1-39.
- Corlett, W.J. and D.C. Hague, 1953, Complementarity and the excess burden of taxation, *Review of Economics Studies* 10, 295-337.
- Correia, Isabel H., 1996, Should capital income be taxed in the steady state? *Journal of Public Economics* 60, 147-151.
- Cremer, Helmuth and Jacques-François Thisse, 1994, Commodity taxation in a differentiated oligopoly, *International Economic Review* 35, 613-633.
- Dahan, Momi, and Michel Strawczynski, 2000, Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates: Comment, *American Economic Review* 90, 681-686.
- Deaton, Angus S., 1979, Optimally uniform commodity taxes, *Economic Letters* 2, 357-361.
- Deaton, Angus S. and John Muellbauer, 1980, An almost ideal demand system, *American Economic Review* 70, 312-326.
- Delipalla, Sofia and Michael Keen, 1992, The comparison between ad valorem and specific taxation under imperfect competition, *Journal of Public Economics* 49, 351-366.

- Diamond, Peter A., 1965, National debt in a neoclassical growth model, *American Economic Review* 55, 1126-1150.
- Diamond, Peter A., 1973, Taxation and public production in a growth setting, in: James A. Mirrlees and Nicholas H. Stern, eds., *Models of economic growth* (Macmillan, New York), 215-235.
- Diamond, Peter A., 1975, A many-person Ramsey rule, *Journal of Public Economics* 4, 335-342.
- Diamond, Peter A., 1998, Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates, *American Economic Review* 88, 83-95.
- Diamond, Peter A., L. Jay Helms, and James A. Mirrlees, 1980, Optimal taxation in a stochastic economy: A Cobb-Douglas example, *Journal of Public Economics* 14, 1-29.
- Diamond, Peter A. and James A. Mirrlees, 1971, Optimal taxation and public production I: Production efficiency and II: Tax rules, *American Economic Review* 61, 8-27 and 261-278.
- Dixit, Avinash K. and Agnar Sandmo, 1977, Some simplified formulae for optimal income taxation, *Scandinavian Journal of Economics* 79, 417-423.
- Dupuit, Arsène Jules Étienne Juvénal, 1844, De la mesure de l'utilité des travaux publics, *Annales des ponts et chaussées*, Second series, 8. Translated by R.H. Barback as On the measurement of the utility of public works, *International Economic Papers*, 1952, 2, 83-110; reprinted in: Kenneth J. Arrow and Tibor Scitovsky, eds., *Readings in welfare economics* (Richard D. Irwin, Homewood, IL, 1969), 255-283.
- Eaton, Jonathan, 1981, Fiscal policy, inflation and the accumulation of risky capital, *Review of Economic Studies* 48, 435-445.
- Eaton, Jonathan and Harvey S. Rosen, 1980, Labor supply, uncertainty, and efficient taxation, *Journal of Public Economics* 14, 365-374.
- Feldstein, Martin, 1978, The welfare cost of capital income taxation, *Journal of Political Economy* 86, S29-S51.
- Feldstein, Martin, 1995, The effect of marginal tax rates on taxable income: A panel study of the 1986 Tax Reform Act, *Journal of Political Economy* 103, 551-572.
- Feldstein, Martin, 1999, Tax avoidance and the deadweight loss of the income tax, *Review of Economics and Statistics* 81, 674-680.
- Fisher, Irving and Herbert W. Fisher, 1942, *Constructive income taxation, a proposal for reform* (Harper and Brothers, New York and London).

- Gallant, Ronald A., 1981, On the bias in flexible functional forms and an essentially unbiased form: The Fourier functional form, *Journal of Econometrics* 15, 211-245.
- Goolsbee, Austan, 2000, What happens when you tax the rich? Evidence from executive compensation, *Journal of Political Economy* 108, 352-378.
- Håkonsen, Lars, 1998, An investigation into alternative representations of the marginal cost of public funds, *International Tax and Public Finance* 5, 329-343.
- Hall, Robert E. and Alvin Rabushka, 1995, *The flat tax*, 2nd ed. (Hoover Institution Press, Stanford).
- Hamilton, Jonathan H., 1987, Taxation, savings, and portfolio choice in a continuous time model, *Public Finance* 42, 264-282.
- Harberger, Arnold C., 1964a, The measurement of waste, *American Economic Review* 54, 58-76.
- Harberger, Arnold C., 1964b, Taxation, resource allocation, and welfare, in: John F. Due, ed., *The role of direct and indirect taxes in the federal revenue system* (Princeton University Press, Princeton, NJ), 25-70.
- Harberger, Arnold C., 1966, Efficiency effects of taxes on income from capital, in: Marian Krzyzaniak, ed., *Effects of corporation income tax* (Wayne State University Press, Detroit), 107-117.
- Harberger, Arnold C., 1971, Three basic postulates for applied welfare economics: An interpretive essay, *Journal of Economic Literature* 9, 785-797.
- Harberger, Arnold C., 1972, The opportunity costs of public investment financed by borrowing, in: R. Layard, ed., *Cost-benefit analysis* (Penguin, Harmondsworth, UK), 303-10.
- Hausman, Jerry A., 1981a, Labor supply, in: Henry J. Aaron and Joseph A. Pechman, eds., *How taxes affect economic behavior* (Brookings, Washington), 27-72.
- Hausman, Jerry A., 1981b, Exact consumer's surplus and deadweight loss, *American Economic Review* 71, 662-676.
- Hausman, Jerry A. and Whitney Newey, 1995, Nonparametric estimation of exact consumer surplus and deadweight loss, *Econometrica* 63, 1445-1476.
- Heckman, James J., Lance Lochner, and Christopher Taber, 1998, Tax policy and human capital formation, *American Economic Review* 88, 293-97.
- Hines, James R., Jr., 1999, Three sides of Harberger triangles, *Journal of Economic Perspectives* 13, 167-188.

- Hylland, Aanund and Richard Zeckhauser, 1979, Distributional objectives should affect taxes but not program choice or design, *Scandinavian Journal of Economics* 81, 264-284.
- Jenkin, Henry Charles Fleeming, 1871-1872, On the principles which regulate the incidence of taxes, in: *Proceedings of the Royal Society of Edinburgh*, 618-631; reprinted in: Sidney Colvin and J.A. Ewing eds., *Papers, literary, scientific, &c. by the late Fleeming Jenkin*, vol. 2 (Longmans Green, London, 1887), 107-121.
- Jones, Larry E., Rodolfo E. Manuelli and Peter E. Rossi, 1993, Optimal taxation in models of endogenous growth, *Journal of Political Economy* 101, 485-517.
- Jones, Larry E., Rodolfo E. Manuelli and Peter E. Rossi, 1997, On the optimal taxation of capital income, *Journal of Economic Theory* 73, 93-117.
- Jorgenson, Dale W., Lawrence J. Lau, and Thomas M. Stoker, 1982, The transcendental logarithmic model of aggregate consumer behavior, in: Robert L. Basmann and G. Rhodes eds., *Advances in econometrics*, vol. 1 (JAI Press, Greenwich, CT), 97-238.
- Judd, Kenneth L., 1985, Redistributive taxation in a simple perfect foresight model, *Journal of Public Economics* 28, 59-83.
- Judd, Kenneth L., 1997, The optimal tax rate for capital income is negative, National Bureau of Economic Research Working Paper No. 6004.
- Judd, Kenneth L., 1999, Optimal taxation and spending in general competitive growth models, *Journal of Public Economics* 71, 1-27.
- Kaldor, Nicholas, 1955, *An expenditure tax* (Allen and Unwin, London).
- Kaplow Louis, 1996, The optimal supply of public goods and the distortionary cost of taxation, *National Tax Journal* 49, 513-33.
- Katz, Michael L. and Harvey S. Rosen, 1985, Tax analysis in an oligopoly model, *Public Finance Quarterly* 13, 3-20.
- Kay, John A., 1980, The deadweight loss from a tax system, *Journal of Public Economics* 13, 111-120.
- Kay, John A. and Michael Keen, 1983, How should commodities be taxed? Market structure, product heterogeneity and the optimal structure of commodity taxes, *European Economic Review* 23, 339-358.
- Kay, John A. and Michael Keen, 1991, Product quality under specific and ad valorem taxation, *Public Finance Quarterly* 19, 238-246.

- King, Mervyn A., 1983, Welfare analysis of tax reforms using household data, *Journal of Public Economics* 21, 183-214.
- King, Robert G. and Sergio Rebelo, 1990, Public policy and economic growth: Developing neoclassical implications, *Journal of Political Economy* 98, S126-S150.
- Laitner, John, 1995, Quantitative evaluations of efficient tax policies for Lucas' supply side models, *Oxford Economic Papers* 47, 471-92.
- Lindsey, Lawrence B., 1987, Individual taxpayer response to tax cuts: 1982-1984: With implications for the revenue maximizing tax rate, *Journal of Public Economics* 33, 173-206.
- Lucas, Robert E., Jr., 1990, Supply-side economics: An analytical review, *Oxford Economic Papers* 42, 293-317.
- Mankiw, N. Gregory and Michael D. Whinston, 1986, Free entry and social inefficiency, *Rand Journal of Economics* 17, 48-58.
- Milesi-Ferretti, Gian-Maria and Nouriel Roubini, 1998, On the taxation of human and physical capital in models of endogenous growth, *Journal of Public Economics* 70, 237-254.
- Mirrlees, James A., 1971, An exploration in the theory of optimum income taxation, *Review of Economics Studies* 38, 175-208.
- Mirrlees, James A., 1976, Optimal tax theory: A synthesis, *Journal of Public Economics* 6, 327-358.
- Mirrlees, James A., 1986, The theory of optimal taxation, in: Kenneth J. Arrow and Michael D. Intriligator, eds., *Handbook of Mathematical Economics*, vol. 3 (North-Holland, Amsterdam), 1197-1249.
- Moffitt, Robert and Mark Wilhelm, 2000, Taxation and the labor supply decisions of the affluent, in: Joel B. Slemrod ed., *Does Atlas shrug? The economic consequences of taxing the rich* (Harvard University Press, Cambridge, MA), 193-234.
- Mohring, Herbert, 1971, Alternative welfare gain and loss measures, *Western Economic Journal* 9, 349-368.
- Myles, Gareth D., 1987, Tax design in the presence of imperfect competition: An example, *Journal of Public Economics* 34, 367-378.
- Myles, Gareth D., 1989, Ramsey tax rules for economies with imperfect competition, *Journal of public Economics* 38, 95-115.
- Myles, Gareth D., 1995, *Public economics* (Cambridge, UK, Cambridge University Press).

- Myles, Gareth D., 1996, Imperfect competition and the optimal combination of ad valorem and specific taxation, *International Tax and Public Finance* 3, 29-44.
- Persson, Torsten, and Lars E. O. Svensson, 1989, Why a stubborn conservative would run a deficit: Policy with time-inconsistent preferences, *Quarterly Journal of Economics* 104, 325-345.
- Pestieau, Pierre M., 1974, Optimal taxation and discount rate for public investment in a growth setting, *Journal of Public Economics* 3, 217-235.
- Phelps, Edmund S., 1973, The taxation of wage income for economic justice, *Quarterly Journal of Economics* 87, 331-354.
- Pigou, A.C., 1947, *A study in public finance*, 3rd ed. (Macmillan, London).
- Ramsey, F.P., 1927, A contribution to the theory of taxation, *Economic Journal* 37, 47-61.
- Rebelo, Sergio, 1991, Long-run policy analysis and long-run growth, *Journal of Political Economy* 99, 500-521.
- Roberts, John and Hugo Sonnenschein, 1977, On the foundations of the theory of monopolistic competition, *Econometrica* 45, 101-113.
- Robinson, Joan, 1933, *The economics of imperfect competition* (London, Macmillan).
- Rosen, Harvey S., 1978, The measurement of excess burden with explicit utility functions, *Journal of Political Economy* 86, S121-S135.
- Sadka, Efraim, 1976, On income distribution, incentive effects and optimal income taxation, *Review of Economic Studies* 43, 261-268.
- Saez, Emmanuel, 2000a, Using elasticities to derive optimal income tax rates, NBER Working Paper No. 7628.
- Saez, Emmanuel, 2000b, Optimal income transfer programs: intensive versus extensive labor supply responses, NBER Working Paper No. 7708.
- Samuelson, Paul A., approx. 1951, Theory of optimal taxation, unpublished memorandum for the U.S. Treasury; published in: *Journal of Public Economics* 30, 1986, 137-43.
- Samuelson, Paul A., 1954, The pure theory of public expenditure, *Review of Economics and Statistics* 36, 387-389.
- Sandmo, Agnar, 1975, Optimal taxation in the presence of externalities, *Swedish Journal of Economics* 77, 86-98.

- Sandmo, Agnar, 1985, The effects of taxation on savings and risk taking, in: Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, vol. 1 (North-Holland, Amsterdam), 265-311.
- Sandmo, Agnar, 1998, Redistribution and the marginal cost of public funds, *Journal Public Economics* 70, 365-382.
- Schöb, Ronnie, 1997, Environmental taxes and pre-existing distortions: The normalization trap, *International Tax and Public Finance* 4, 167-176.
- Seade, Jesus, 1977, On the shape of optimal tax schedules, *Journal of Public Economics* 7, 203-237.
- Seade, Jesus, 1980a, On the effects of entry, *Econometrica* 48, 479-489.
- Seade, Jesus, 1980b, The stability of Cournot revisited, *Journal of Economic Theory* 23, 15-26.
- Shoven, John B., 1976, The incidence and efficiency effects of taxes on income from capital, *Journal of Political Economy* 84, 1261-1283.
- Slemrod, Joel and Shlomo Yitzhaki, 2001, Integrating expenditure and tax decisions: the marginal cost of funds and the marginal benefit of projects, *National Tax Journal*, forthcoming.
- Slemrod, Joel, Shlomo Yitzhaki, Joram Mayshar, and Michael Lundholm, 1994, The optimal two-bracket linear income tax, *Journal of Public Economics* 53, 269-90.
- Slesnick, Daniel T., 1998, Empirical approaches to the measurement of welfare, *Journal of Economic Literature* 36, 2108-2165.
- Stern, Nicholas H., 1976, On the specification of models of optimum income taxation, *Journal of Public Economics* 6, 123-162.
- Stern, Nicholas H., 1982, Optimum taxation with errors in administration, *Journal of Public Economics* 17, 181-211.
- Stiglitz, Joseph E., 1982, Utilitarianism and horizontal equity: The case for random taxation, *Journal of Public Economics* 18, 1-33.
- Stiglitz, Joseph E., 1987, Pareto efficient and optimal taxation and the new new welfare economics, in: Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, vol. 2 (North-Holland, Amsterdam), 991-1042.
- Stiglitz, Joseph E. and Partha S. Dasgupta, 1971, Differential taxation, public goods and economic efficiency, *Review of Economic Studies* 38, 151-174.

- Stokey, Nancy L. and Sergio Rebelo, 1995, Growth effects of flat-rate taxes, *Journal of Political Economy* 103, 519-50.
- Suits, Donald B. and Richard A. Musgrave, 1953, Ad valorem and unit taxes compared, *Quarterly Journal of Economics* 67, 598-604.
- Trostel, Philip A., 1993, The effect of taxation on human capital, *Journal of Political Economy* 101, 327-350.
- Varian, Hal R., 1980, Redistributive taxation as social insurance, *Journal of Public Economics* 14, 49-68.
- Vartia, Yrjö, 1983, Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions, *Econometrica* 51, 79-98.
- Zhu, Xiaodong, 1992, Optimal fiscal policy in a stochastic growth model, *Journal of Economic Theory* 58, 250-289.

Figure 2.1. The Measurement of Excess Burden

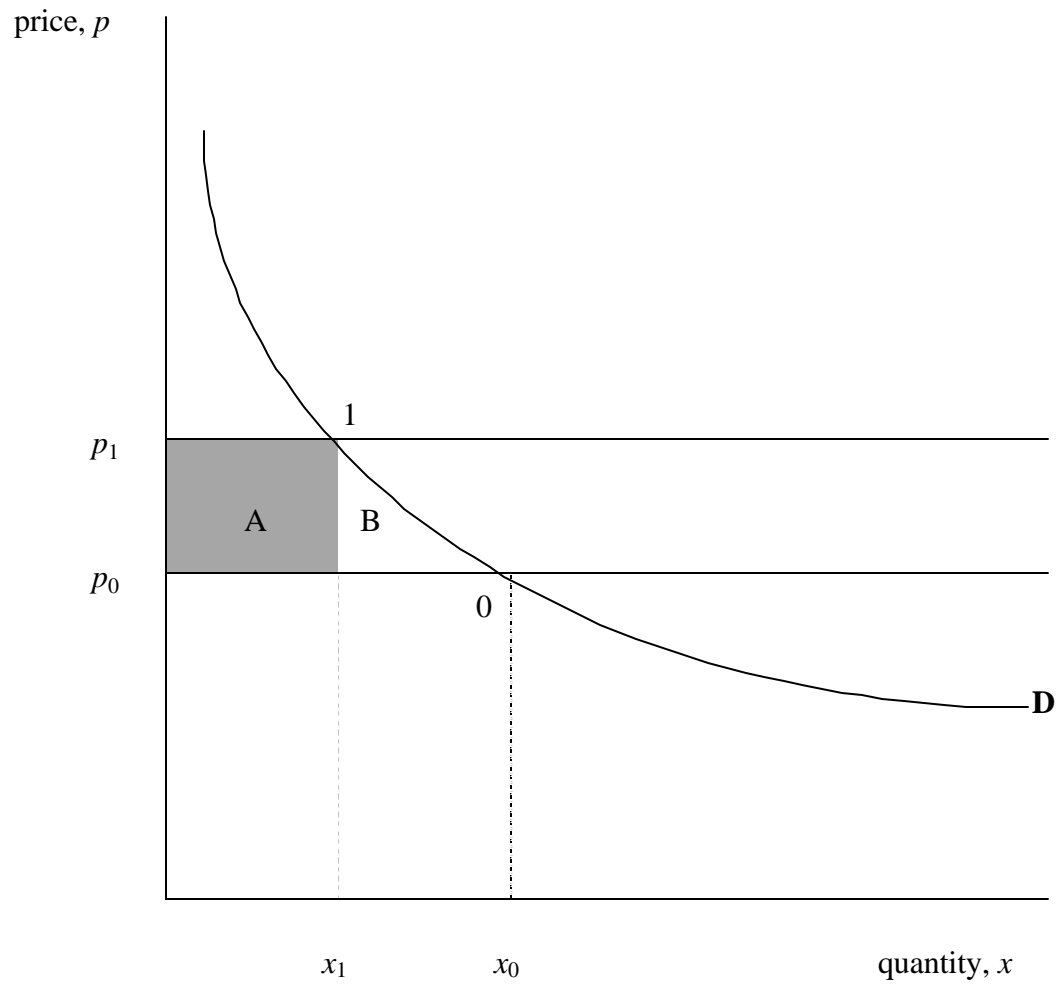


Figure 2.2. Using Hicksian Variations to Measure Excess Burden

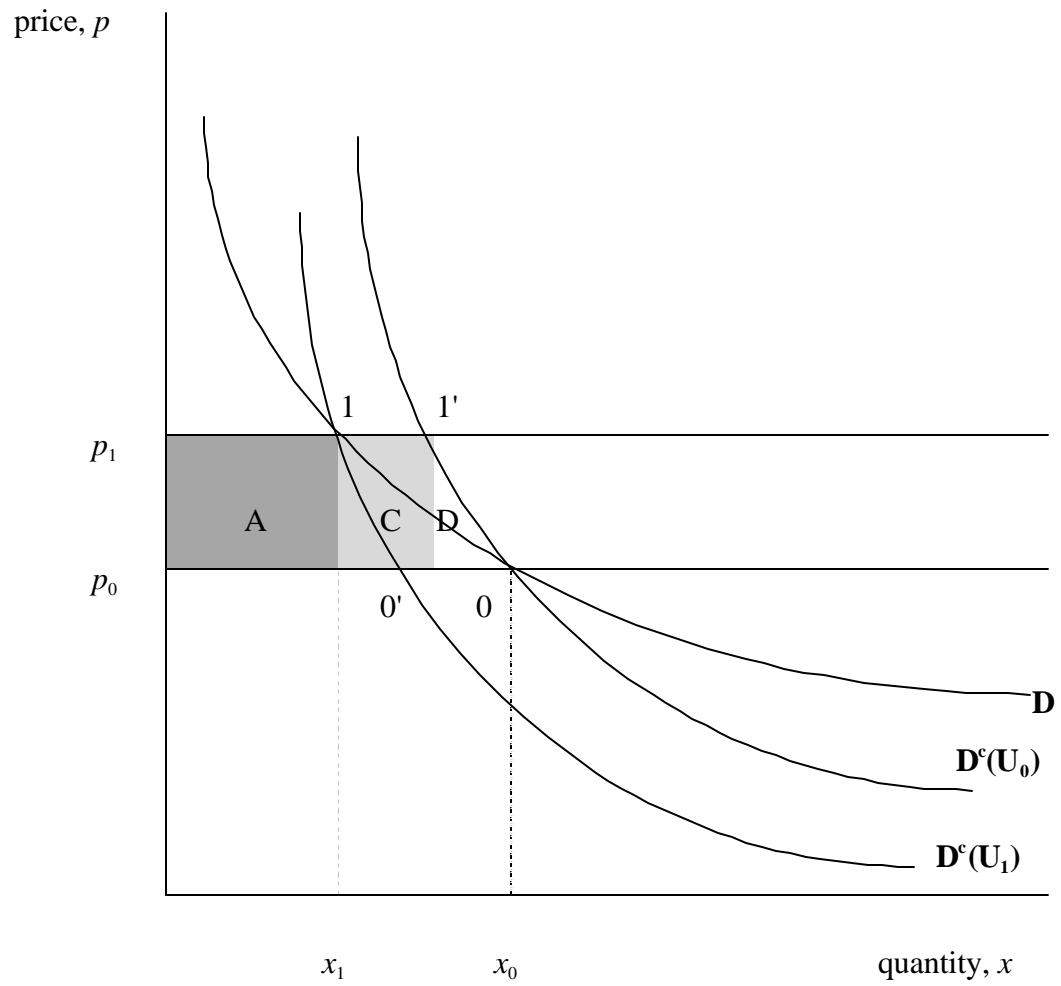


Figure 2.3. Excess Burden: An Alternative Graphical Representation

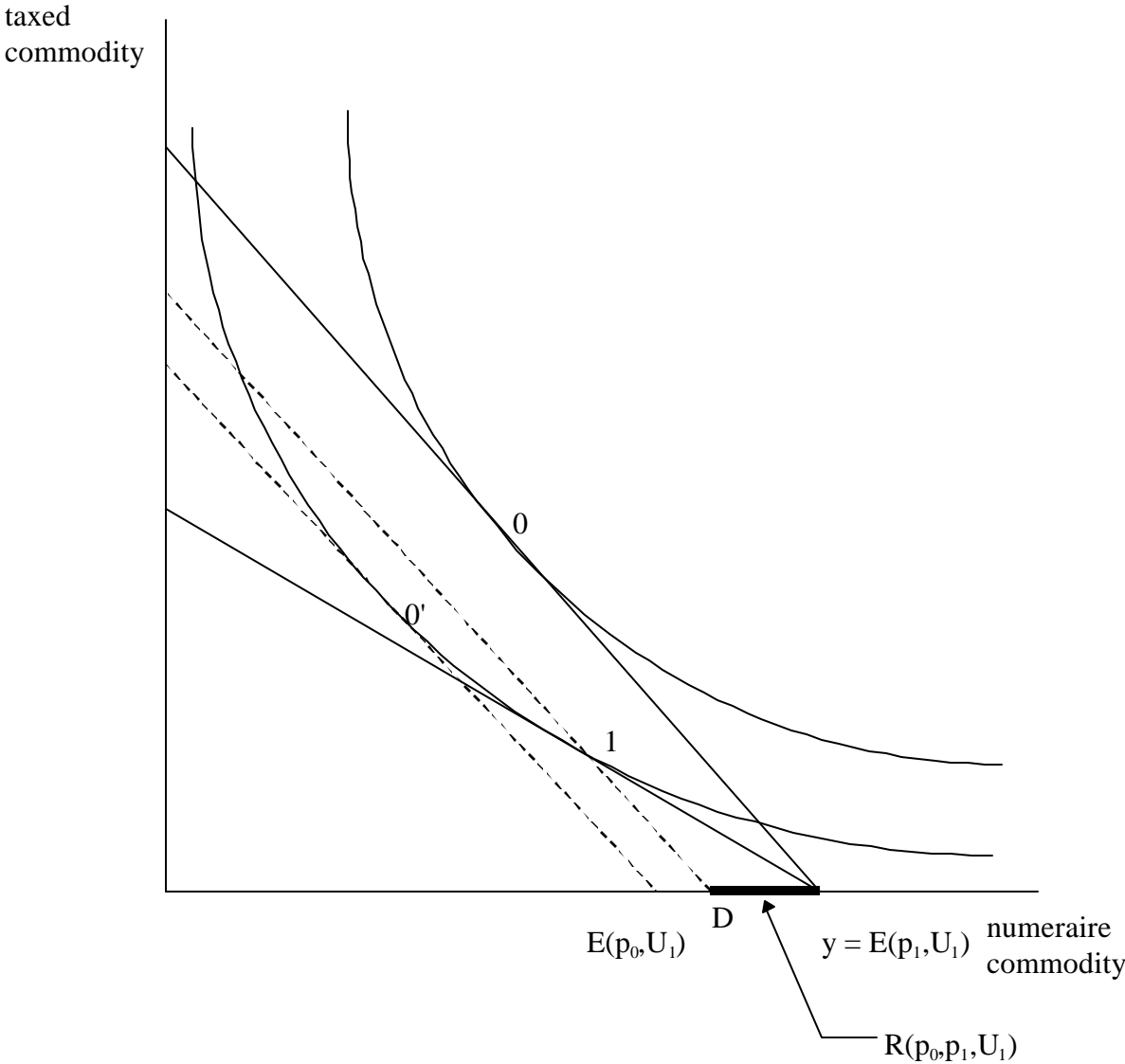


Figure 2.4. Marginal Excess Burden of a Pre-Existing Tax

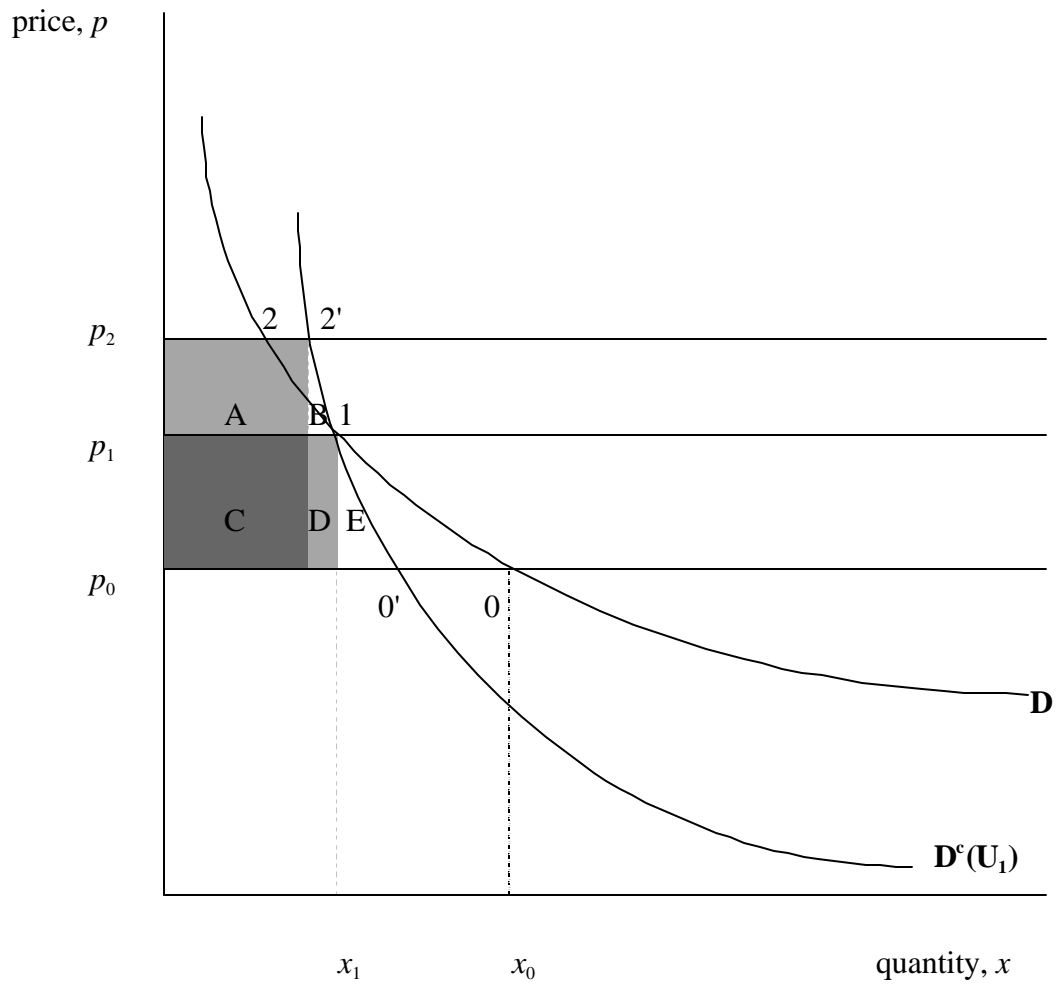


Figure 2.5. Excess Burden with Varying Producer Prices

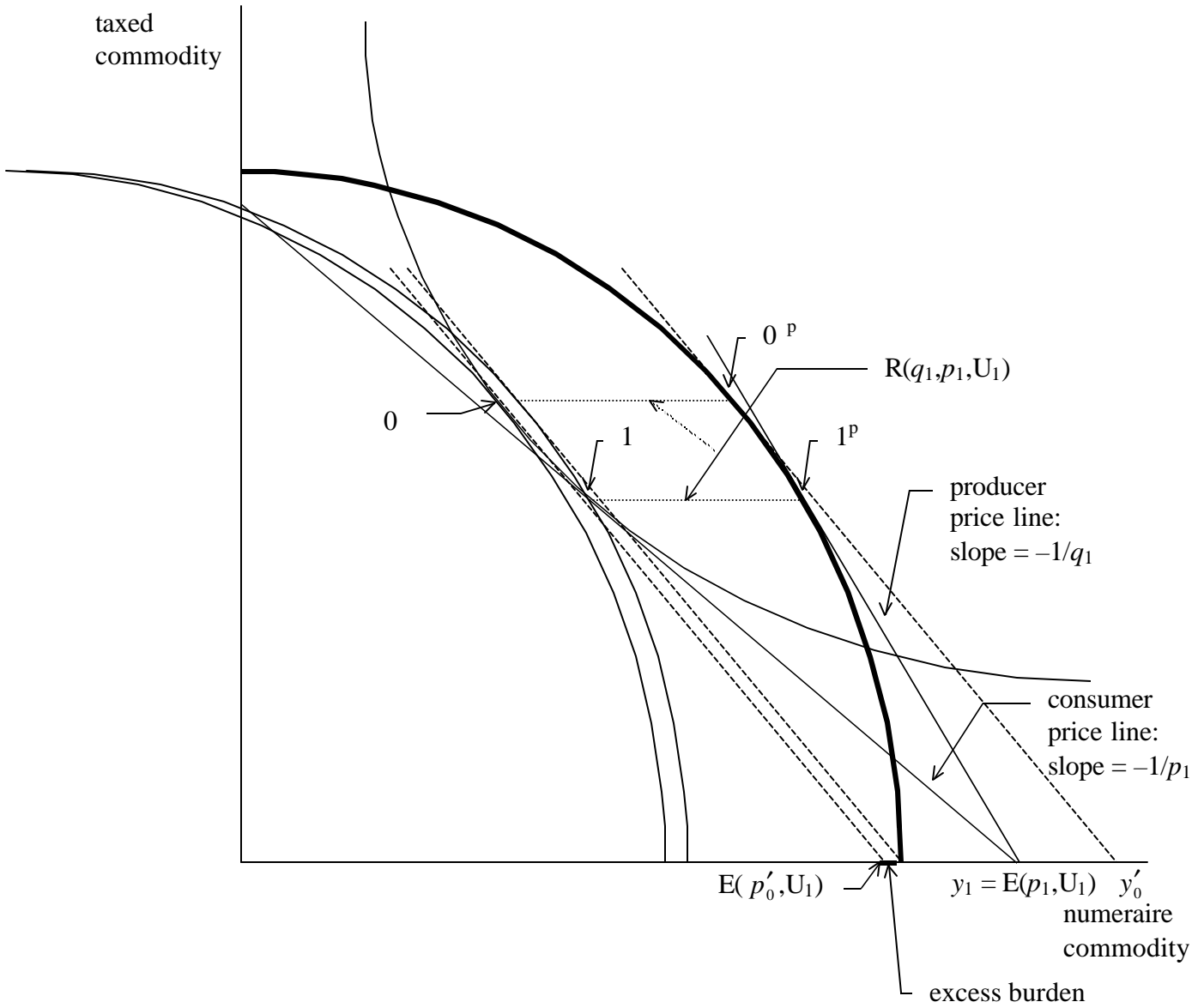


Figure 2.6. Excess Burden with an Upward Sloping Supply Curve

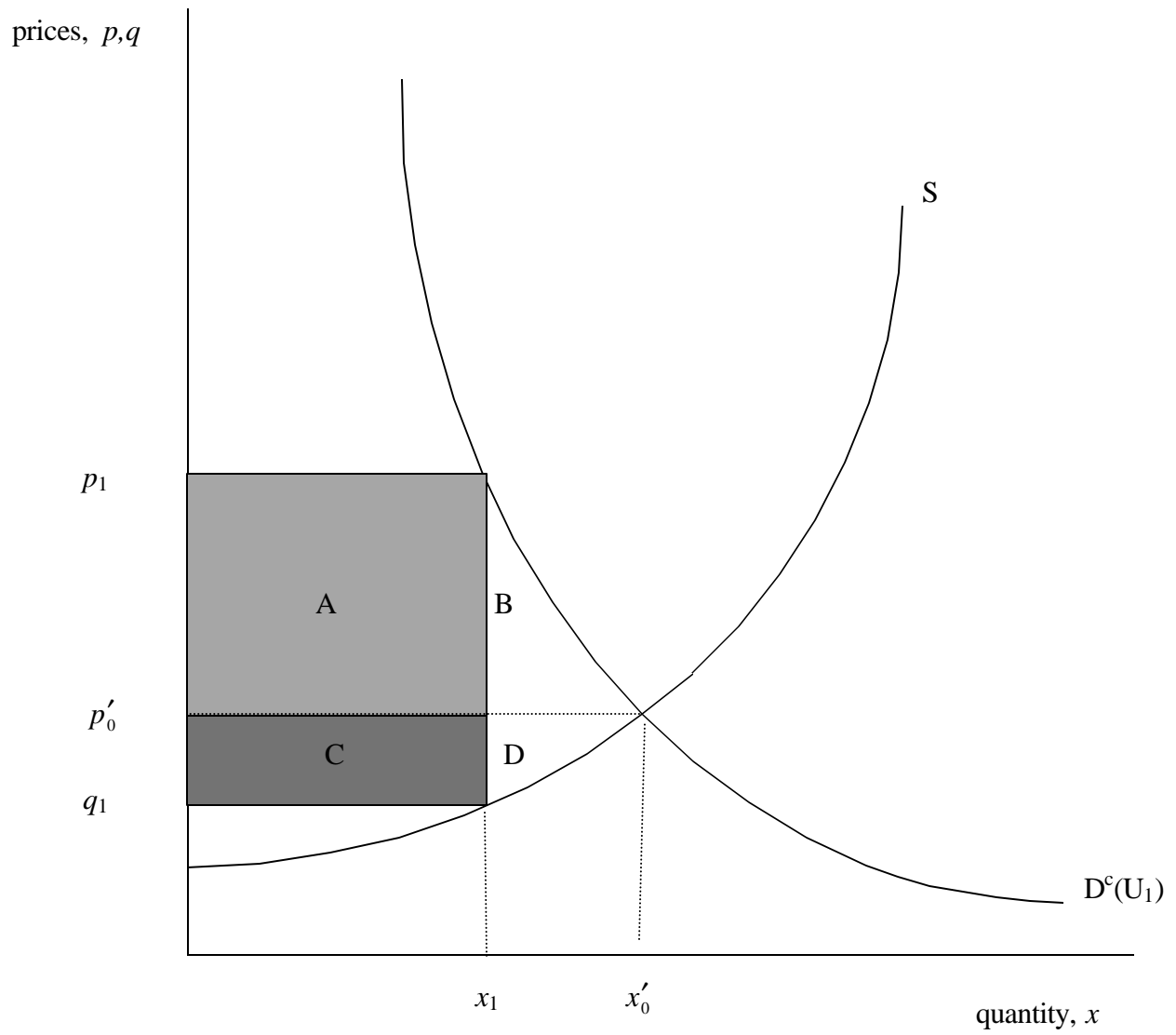


Figure 4.1. Indifference Curves over Consumption and Income

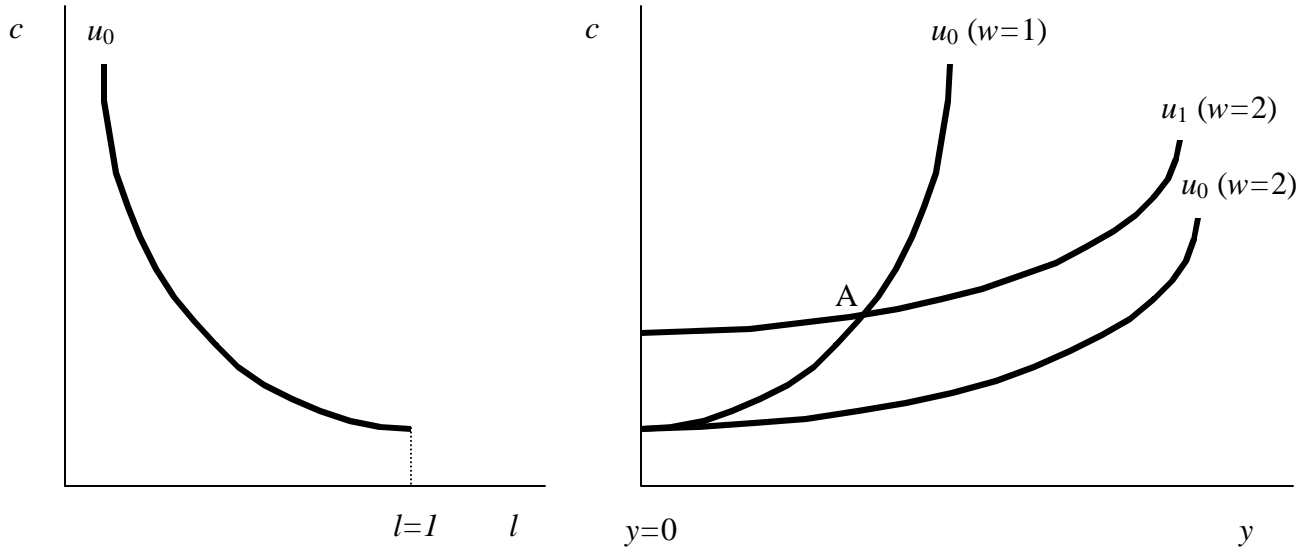


Figure 4.2. Violation of the Self-Selection Constraint

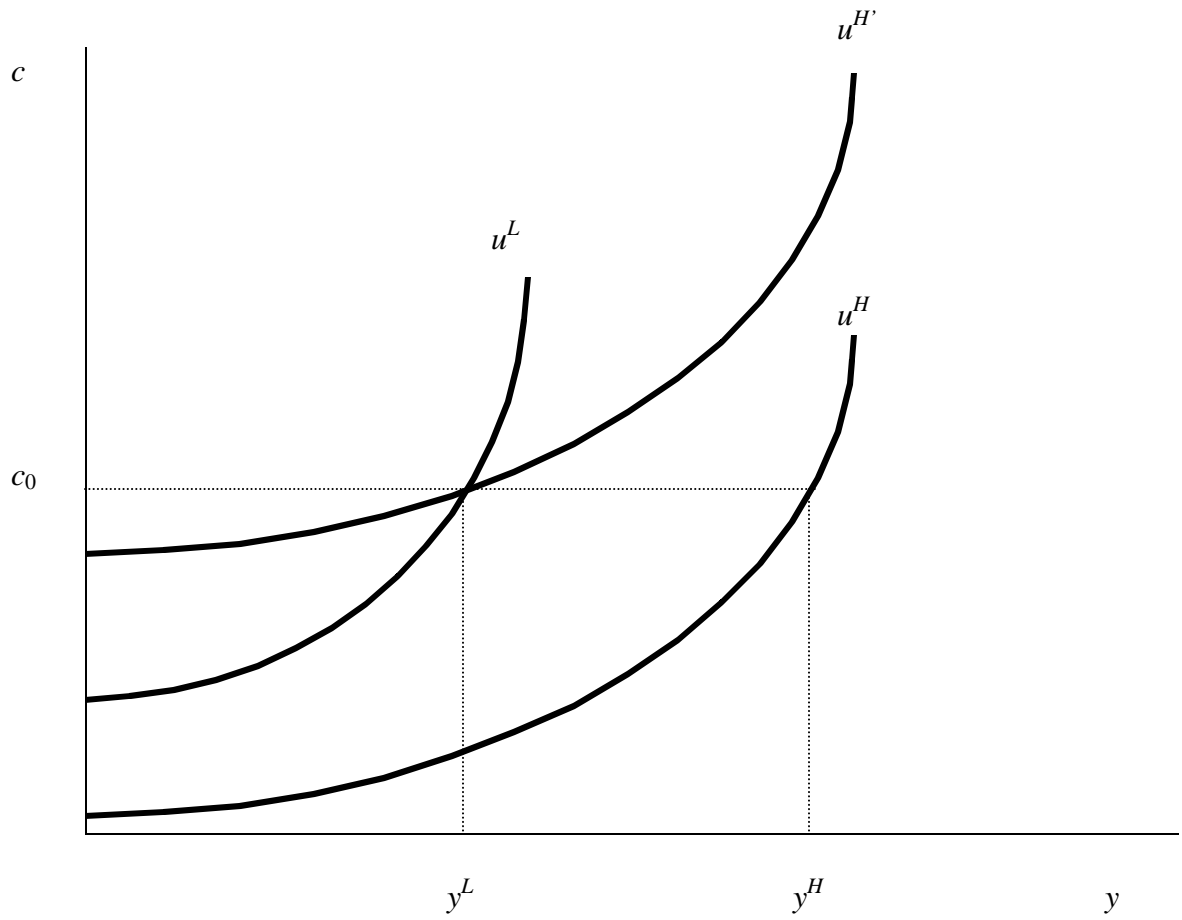


Figure 4.3. The Scope for Lump-Sum Taxation

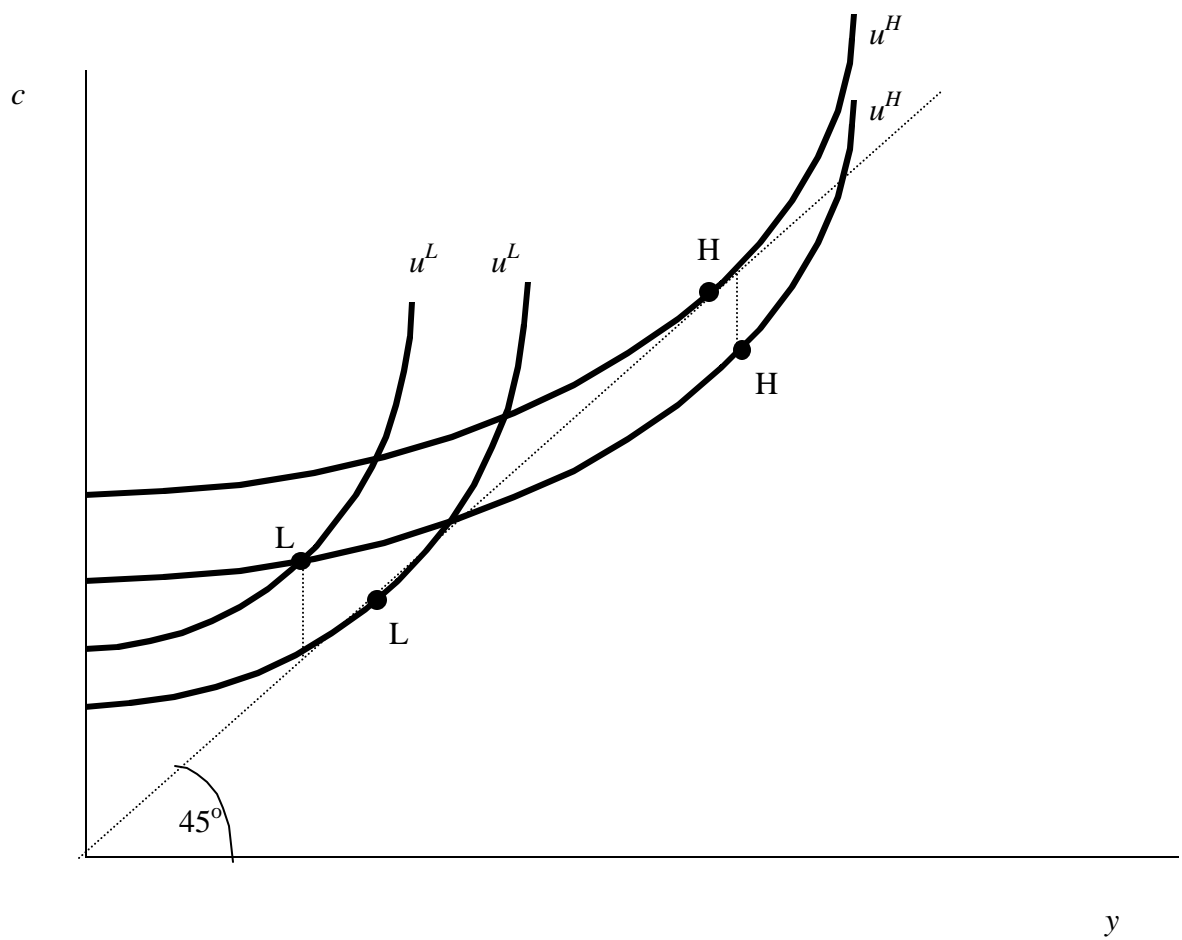


Figure 4.4. Using Distortionary Income Taxation

