THE THEORY OF EXCESS BURDEN AND OPTIMAL TAXATION

ALAN J. AUERBACH*

University of Pennsylvania, Philadelphia, PA
National Bureau of Economic Research, Cambridge, MA

1. Introduction

The theory of excess burden and optimal commodity taxation is one of the oldest subjects of study in public finance, dating back to Dupuit (1844), and yet is also closely associated with the rapid analytical development of the field which commenced in the early 1970s. Perhaps more than in most areas of economics, there has been a tendency to overlook contributions made in earlier decades. As a result, much of the "new" public economics of the last decade may be viewed, in part, as a restatement and extension, perhaps in less arcane language and terminology, of previously proven propositions.

Probably the most celebrated example of such "rediscovery" is that of Ramsey's (1927) derivation of optimal commodity tax formulae, now referred to as the Ramsey rule. The lapse here is even harder to understand in that Ramsey's results were succinctly described in Pigou's classic public finance text (1947) and rederived by Boiteux (1956). The deadweight loss "triangles" made popular by the work of Harberger (1964) were considered by Hotelling (1938), and appear implicitly in Dupuit (1844):

> "It follows that when the change in consumption brought about by a tax is known, it is possible to find an upper limit to the amount of the utility lost by multiplying the change in consumption by half the tax."[1]

Indeed, the generalization of such excess burden formulae by Boiteux (1951) and Debreu (1951, 1954) has until recently[2] been almost entirely ignored in the subsequent literature. Even the "Laffer curve", popular for a time among non-economists, might more appropriately be called the "Dupuit curve":

[1] Dupuit (1844).
[2] See, for example, Diewert (1981).

"If a tax is gradually increased from zero up to a point where it becomes prohibitive, its yield is at first nil, then increases by small stages until it reaches a maximum, after which it gradually declines until it becomes zero again. It follows that when the state requires to raise a given sum by means of taxation, there are always two rates of tax which would fulfill the requirement, one above and one below that which would yield the maximum. There may be a very great difference between the amounts of utility lost through these taxes which yield the same revenue."[3]

The purpose of this chapter is to present the chronological development of the concept of excess burden and the related study of optimal tax theory. A main objective is to uncover the interrelationships among various apparently distinct results, so as to bring out the basic structure of the entire problem.

1.1. Outline of the chapter

Any discussion of welfare economics inevitably begins with the problem of welfare measurement, which in the present context involves a treatment of Marshall's consumers' surplus and its relationship to Hicks' (1942) notions of compensating and equivalent variations. These are discussed in Section 2, where special attention is paid to the distinction between the measurement of the welfare effects of price changes and the distortionary impact of tax changes. Section 3 develops the various measures of excess burden, focusing on issues of approximation, informational requirements and aggregation over individuals, and the effects of a more general technology than the commonly supposed one with fixed producer prices. Section 4 reviews some of the empirical attempts to estimate various deadweight losses. Section 5 presents and interprets the basic rules for optimal commodity taxation, including a discussion of the role of profits taxation and the desirability of production efficiency. The analysis in Section 6 concerns the relative desirability of direct and indirect taxation and the structure of individual preferences. Section 7 presents some applications of optimal tax theory to questions such as the provision of public goods, correction of externalities, and the allocation of risk. Finally, in Section 8, we explore the issue of tax reform, as distinct from *de novo* tax design. This literature dates back to Corlett and Hague (1953–54), and asks whether specified local movements away from an initial suboptimal equilibrium will improve social welfare. In general, movement of prices in the direction of their optimal levels does not guarantee such an improvement.

---

[3] Dupuit, *op. cit.*, p. 278. For this particular rediscovery, I am indebted to the historical analysis of Atkinson and Stern (1980).

2. Measures of surplus and excess burden

2.1. Consumers' surplus and the Hicksian variations

We begin with Marshall's (1920, p. 811) diagram, in Figure 2.1, depicting consumers' and producers' surplus. The consumers' surplus is defined, somewhat vaguely, to be the amount that consumers would pay in excess of the amount they are paying, $p_0 x_0$, for the amount they are purchasing, x_0. Interpreting the demand curve as an expression of willingness to pay, we obtain area A as such a measure by integrating the vertical gap between the demand curve and p_0 over x. Similarly, interpreting producers' surplus as the level of profits received in supplying the quantity sold, and assuming that competitive supply causes the marginal social cost to coincide with the supply schedule S, we obtain the area B. The sum $A + B$ is maximized when price equals marginal cost, and changes in each measure following from a price change are easily calculated. For example, if the price rises from p_0 to p_1, the change in consumers' surplus is the area of a
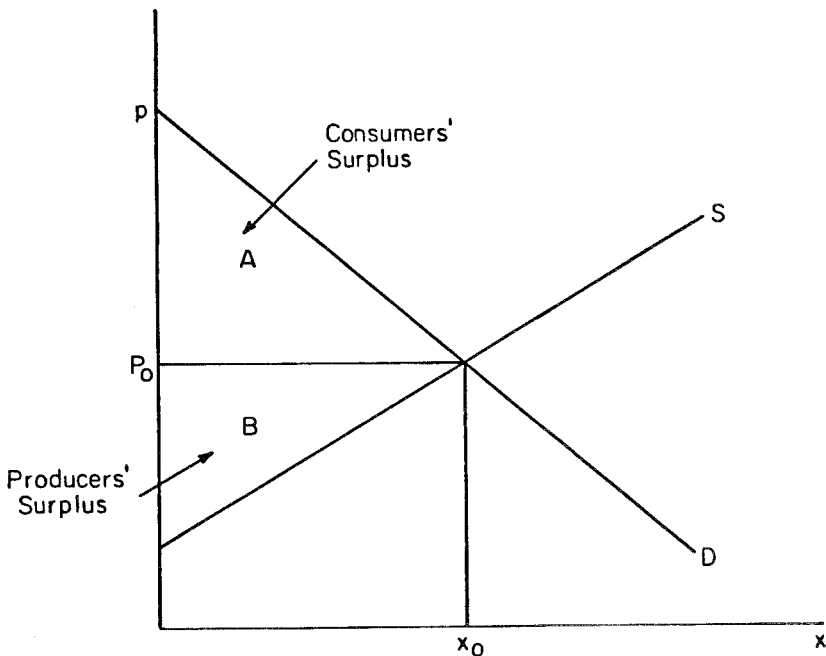

Figure 2.1. Consumers' and producers' surplus.

trapezoid which equals

$$\Delta S = - \int_{p_0}^{p_1} x(p) \, dp, \tag{2.1}$$

where $x(\cdot)$ is the demand function with respect to the good's own price, holding other prices fixed.[4]

The basic problem with consumers' surplus as a welfare measure is that it does not come directly from underlying consumer preferences. As a result, it has the serious flaw of path-dependence: if more than one price changes, the order in which the trapezoids in (2.1) are calculated matters. That is, if we let x^i and p^i be the quantity demanded and price in the ith market, the sum of individual changes in consumers' surplus, ΔS^i, i.e., the line integral

$$\Delta S = \sum_i \Delta S = - \int_{p_0}^{p_1} \sum_i x^i \, dp^i, \tag{2.2}$$

takes on different values according to the path of integration from the initial price vector p_0 to the ultimate price vector p_1. To see this, consider a simple example with two markets. If we change the price in market 1 first, the change in surplus is

$$\Delta S_1 = - \int_{p_0^1}^{p^1} x^1 (p^1, p_0^2) \, dp^1 - \int_{p_0^2}^{p_1^2} x^2 (p_1^1, p^2) \, dp^2, \tag{2.3a}$$

while if we change the price in market 2 first, we obtain

$$\Delta S_2 = - \int_{p_0^1}^{p_1^1} x^1 (p^1, p_1^2) \, dp^1 - \int_{p_0^2}^{p_1^2} x^2 (p_0^1, p^2)(dp^2). \tag{2.3b}$$

Subtracting ΔS_1 from ΔS_2, we obtain

$$\Delta S_2 - \Delta S_1 = - \int_{p_0^1}^{p_1^1} \left[x^1 (p^1, p_1^2) - x^1 (p^1, p_0^2) \right] dp^1$$

$$+ \int_{p_0^2}^{p_1^2} \left[x^2 (p_1^1, p^2) - x^2 (p_0^1, p^2) \right] dp^2. \tag{2.4}$$

---

[4] Note that, by integrating (2.1) by parts, we obtain the formula for ΔS based on the difference between the two levels of surplus themselves, i.e.,

$$\Delta S = \int_{x(p_0)}^{x(p_1)} p(x) \, dx - [p_1 x_1 - p_0 x_0] = \int_0^{x(p_1)} p(x) \, dx - p_1 x_1 - \left[\int_0^{x(p_0)} p(x) \, dx - p_0 x_0 \right].$$

For this term to equal zero, it must generally be zero over all subintervals between p_0 and p_1. In particular, for small changes in p^1 and p^2, with $p_1^2 = p_0^2 + d p_0^2$ and $p_1^1 = p_0^1 + d p^1$, (2.4) becomes

$$\Delta S_2 - \Delta S_1 = \frac{\partial x^2\left(p_0^1, p_0^2\right)}{\partial p^1} \, d p^1 \, d p^2 - \frac{\partial x^1\left(p_0^1, p_0^2\right)}{\partial p^2} \, d p^1 \, d p^2, \qquad (2.5)$$

which equals zero only if the cross-price derivatives $\partial x^1/\partial p^2$ and $\partial x^2/\partial p^1$ are equal.[5] Such symmetry holds for *compensated* demands: the Slutsky matrix *is* symmetric [Hicks (1946)]. However, ordinary demand derivatives also possess income effects that are not generally equal.

The path-dependence problem does not arise from surplus measures based on compensated commodity demands, for which the symmetry property holds. Here, however, we face a different question: since utility does change with the change in prices, which utility level should be used as a reference level for the compensated demand functions? Two natural candidates are the levels of utility prevailing before and after the price changes. Following Hicks (1942), we define the *compensating* variation of a price change to be that amount of income the consumer must receive to leave utility unaffected by the price change, and the *equivalent* variation as the amount of income the consumer would forego to avoid the price change. By definition, the compensating variation of a price change from p_0 to p_1 equals the equivalent variation of a change from p_1 to p_0. Using the expenditure function, defined by the minimization of expenditure at given prices to satisfy a given level of utility:

$$E(p, \overline{U}) = \min(p \cdot x) \quad \text{subject to} \quad U(x) \geq \overline{U}, \qquad (2.6a)$$

we may express concisely the equivalent and compensating variations as $E(p, \overline{U}) - E(p_0, \overline{U})$, where $\overline{U}$ is the pre-change utility level in the case of the compensating variation, and the post-change utility level in the case of the equivalent variation. Letting y be the consumer's actual income,[6] we can express these two measures as functions of prices and income alone through use of the indirect utility function, $V(p, y)$, defined by

$$V(p, y) = \max U(x) \quad \text{subject to} \quad p \cdot x \geq y. \qquad (2.6b)$$

Substituting (2.6b) into (2.6a), we obtain for the compensating variation of a price

---

[5] See Hotelling (1938) for the original statement of this result.

[6] y should be thought of as a comprehensive "full income" measure not affected by individual decisions regarding, for example, labor supply. This is discussed further in Section 5 below.

change from p_0 to p_1,

$$CV(p_0, p_1) = E(p_1, V(p_0, y)) - E(p_0, V(p_0, y))$$

$$= E(p_1, V(p_0, y)) - y, \qquad (2.7a)$$

and for the corresponding equivalent variation,

$$EV(p_0, p_1) = E(p_1, V(p_1 y)) - E(p_0, V(p_1, y))$$

$$= y - E(p_0, V(p_1, y)), \qquad (2.7b)$$

[where we use the identity $y = E(p, V(p, y))$].

These measures may be depicted graphically. By the envelope theorem, the derivative of the expenditure function with respect to an individual price p^i is
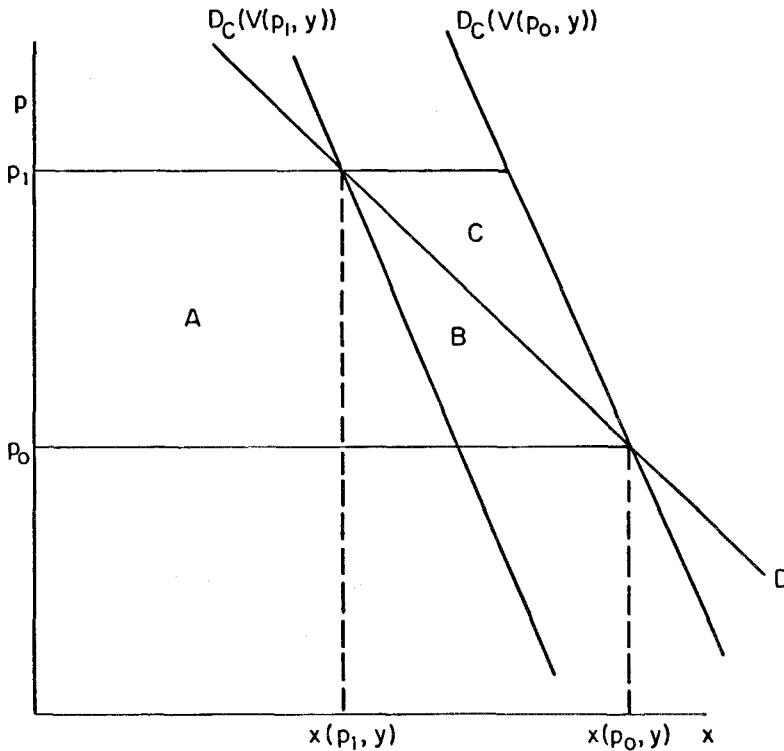

Figure 2.2. Compensating and·equivalent variations.

simply the Hicksian or compensated demand $x_C^i(p, \overline{U})$. Thus, either of the Hicksian variations may be expressed (for the appropriate value of $\overline{U}$) as

$$E(p_1, \overline{U}) - E(p_0, \overline{U}) = \int_{p_0}^{p_1} \frac{dE}{dp}(p, \overline{U}) \, dp = \int_{p^0}^{p^1} x_C(p, \overline{U}) \, dp. \qquad (2.8)$$

Since the cross-price derivatives are symmetric for compensated demands, these measures are path-independent. For the case of a single price change, they may be easily compared to the simple change in consumers' surplus, which is then well-defined. This is shown in Figure 2.2, where $D_C(\overline{U})$ is the compensated demand curve corresponding to the compensated demands $x_C(p, \overline{U})$, drawn more steeply than the ordinary demand curve D under the assumption of normality. The ordinary consumers' surplus changes by the area $A + B$ with an increase in price from p_0 to p_1. The compensating variation of the change equals the area $A + B + C$, while the equivalent variation equals the area A. The bracketing of the Marshallian measure by the two Hicksian measures was emphasized by Hicks (1942) and Willig (1976) in their attempts at rehabilitation of consumers' surplus as a welfare measure. However, their argument becomes weaker when more than one price changes, for then consumers' surplus is not even single-valued. Moreover, for estimating the excess burden of a tax, it is not the entire loss to the consumer in which we are interested but rather the loss in excess of revenue collected. It turns out that in such a case, the felicitous outcome with respect to the relative sizes of the three measures no longer holds.

2.2. Definitions of excess burden

The deadweight loss from a tax system is that amount that is lost in excess of what the government collects. Unfortunately, while this definition makes intuitive sense, it is too vague to permit a single interpretation.

We begin again with the simple Marshallian approach, which is adequate for purposes of illustrating the concept of excess burden in a single market. We can see the effects of a tax t in Figure 2.3. By raising the consumer price from p_0 to $p_1 + t$, the tax reduces consumers' surplus by the area $A + B$. Producers' surplus is reduced by $C + D$, by the drop in producer price to p_1, but tax revenues amount only to $A + C$, yielding a social loss of $B + D$, or approximately $\frac{1}{2}t(x_0 - x_1) = -\frac{1}{2}t\Delta x$, as suggested by Dupuit.

A key aspect of this measure is that it is greater than zero whether the tax is positive or negative. The case of a subsidy at rate s is depicted in Figure 2.4. Here, there is an increase in consumption to x_1, and consumers' surplus and producers' surplus both rise by the areas $H + I$ and $F + G$, respectively. But the
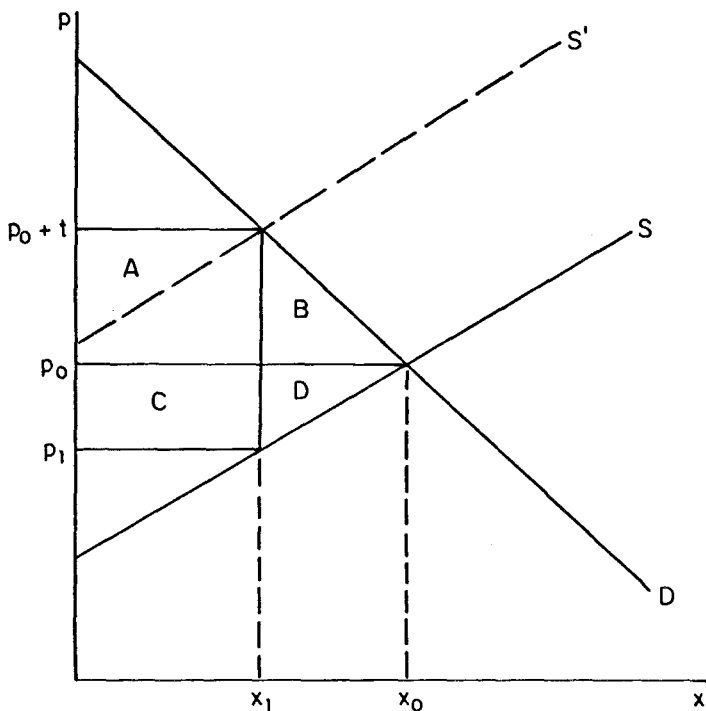
Figure 2.3. Excess burden of a tax.

amount of the subsidy exceeds those gains by the area J, equal to $\frac{1}{2}s\Delta x$ or, again, $-\frac{1}{2}t\Delta x$ for $t = -s$ being the algebraic value of the tax. The loss comes from the distortion of a Pareto optimal allocation, not simply the reduction in output.

For the case where a tax already exists, we may ask what additional excess burden would be caused by a tax increase. In this case, we subtract the *change* in government revenue from the change in producers' and consumers' surplus, since revenue is positive at the initial point. The resulting measure is shown in Figure 2.5.

By raising the consumer price from $p_1 + t_1$ to $p_2 + t_2$, the tax causes a loss in consumers' surplus of $A + B$. Producers' surplus declines by $C + D$, and, as before, the government collects additional revenue on the purchases x_2 equal to $(t_2 - t_1) \cdot x_2$, or area $A + C$. However, the government loses the revenue it was collecting on the purchases in excess of x_2, equal to area E. Thus, the welfare loss of the tax increase equals the trapezoidal area $B + E + D$, or approximately $-(t\Delta x + \frac{1}{2}\Delta t\Delta x)$. Thus, even if Δt is very small, the additional excess burden
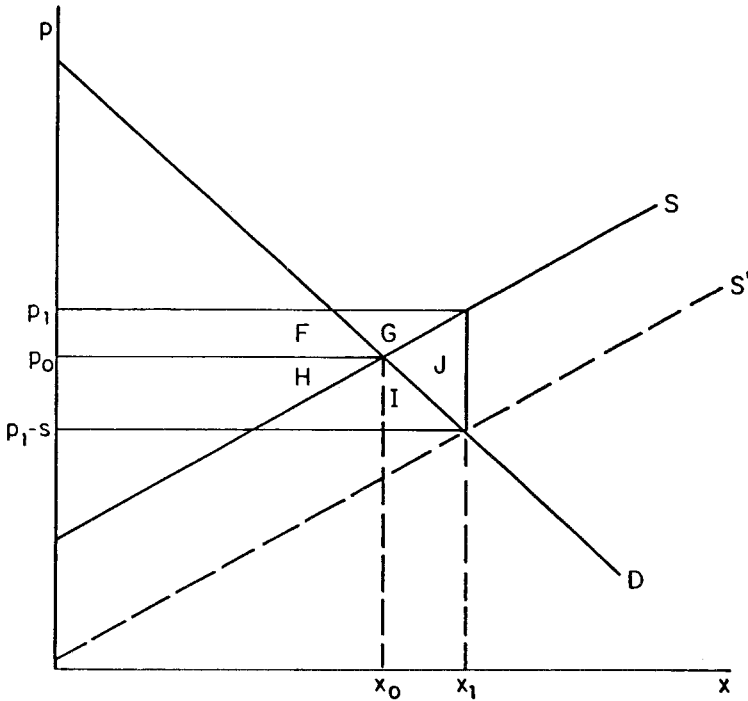
Figure 2.4. Excess burden of a subsidy.

need not be, unlike in the case where no tax exists initially: there is now a first-order welfare loss resulting from marginal tax changes.

If we wish to consider the effects of several taxes at once, we must use more sophisticated measures based on the Hicksian variations. For the remainder of this subsection, we focus on the case of a single consumer facing fixed producer prices. These restrictions are relaxed in Section 3.

Using the equivalent variation, Mohring (1971) suggests that the excess burden of taxation is the amount in excess of taxes being collected that the consumer would give up in exchange for the removal of all taxes; that is, how much more could be collected from the consumer (and thrown away) than is currently being collected, with no loss in utility, if the collection method were lump sum taxation. In the terminology used above, we may write this measure as

$$EB_E = E(p_1, V(p_1, y)) - E(p_0, V(p_1, y)) - R(p_1, y)$$

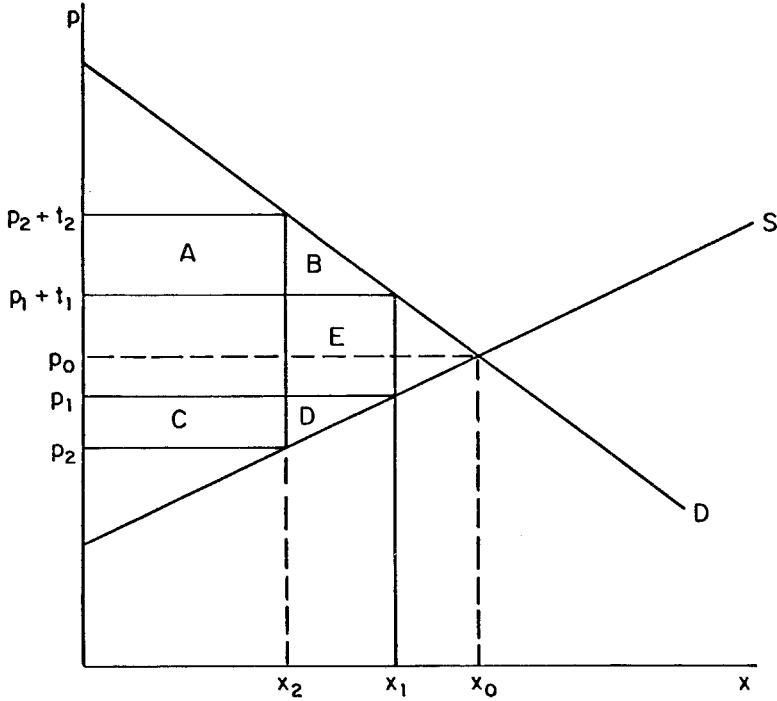$$= y - E(p_0, V(p_1, y)) - (p_1 - p_0) \cdot x(p_1, y), \qquad (2.9)$$

Figure 2.5. Excess burden with a pre-existing tax.

where $R(p_1, y)$ is the tax revenue collected when prices are at p_1 and the consumer's income equals y.

Alternatively, Diamond and McFadden (1974) suggest the use of the compensating variation by defining excess burden to be that amount, in addition to revenues collected, that the government must supply to the consumer to allow him to maintain the initial utility level. That is, how much must come from "outside" the system to compensate for the tax distortion. To avoid double-counting, we include in the government's revenue the additional amount it collects because the individual is compensated and (for a normal good) demands more of the taxed commodity. Thus, the Diamond–McFadden measure may be written

$$EB_C = E(p_1, V(p_0, y)) - E(p_0, V(p_0, y)) - R(p_1, E(p_1, V(p_0, y)))$$

$$= E(p_1, V(p_0, y)) - y - (p_1 - p_0) \cdot x(p_1, E(p_1, V(p_0, y)))$$

$$= E(p_1, V(p_0, y)) - y - (p_1 - p_0) \cdot x_C(p_1, V(p_0, y)), \qquad (2.10)$$

[where the last step uses the identity $x(p, E(p, \overline{U})) = x_C(p, \overline{U})$]. As with EB_E, EB_C must be non-negative.

For a single price change, these two measures of excess burden may be graphically compared to the Marshallian measure shown in Figure 2.3. The three measures together are shown in Figure 2.6. To obtain the equivalent variation measure or the consumers' surplus measure of excess burden, we subtract the revenue actually collected at $x(p, y)$ from the respective measures shown in Figure 2.2. For the compensating variation measures, we subtract the revenue that would be collected if utility were kept at $V(p_0, y)$. This yields the areas A, $A + B$, and C for the three respective measures. Note that the two Hicksian
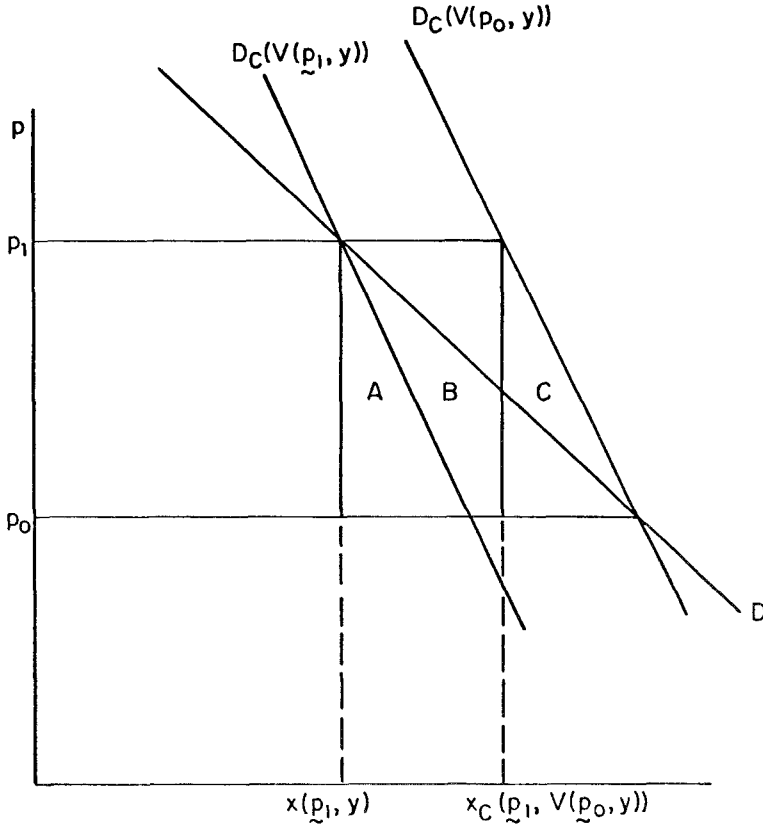

Figure 2.6. A comparison of excess burden measures.

measures no longer bracket the Marshallian one.[7] If the taxed good is normal, the latter is necessarily larger than each of the former, and the discrepancy may be quite large.

Other logical measures of excess burden involving the equivalent and compensating variations may be conceived.[8] In addition, it is easy to adapt the two measures already derived to the case where the initial equilibrium is not Pareto optimal but is already distorted by taxes. The equivalent variation measure of additional excess burden would then be the amount, in excess of additional tax revenues, that the consumer would pay to avoid the latest price increase from p_1 to p_2,

$$EB_E = E(p_2, V(p_2, y)) - E(p_1, V(p_2, y))$$

$$- [R(p_2, y) - R(p_1, E(p_1, V(p_2, y)))]$$

$$= y - E(p_1, V(p_2, y)) - (p_2 - p_0) \cdot x(p_2, y)$$

$$+ (p_1 - p_0) \cdot x_C(p_1, V(p_2, y))$$

$$= y - E(p_1, V(p_2, y)) - (p_2 - p_1) \cdot x(p_2, y)$$

$$+ (p_1 - p_0) \cdot (x_C(p_1, V(p_2, y)) - x(p_2, y)). \qquad (2.11)$$

Comparing (2.11) with (2.9), we find that (2.11) contains an additional expression representing the reduction in tax revenues as demand declines with the new rise in price, with utility held constant at $V(p_2, y)$. This additional term corresponds to that found for the basic consumers' surplus measure in Figure 2.6. Likewise, the compensating variation measure would be the amount in excess of the change in revenues that would be required to maintain the initial utility level, or

$$EB_C = E(p_2, V(p_1, y)) - E(p_1, V(p_1, y))$$

$$- [R(p_2, E(p_2, V(p_1, y))) - R(p_1, y)]$$

$$= E(p_2, V(p_1, y)) - y - (p_2 - p_0) \cdot x_C(p_2, V(p_1, y))$$

$$+ (p_1 - p_0) \cdot x(p_1, y)$$

$$= E(p_2, V(p_1, y)) - y - (p_2 - p_1) \cdot x_C(p_2, V(p_1, y))$$

$$+ (p_1 - p_0) \cdot (x(p_1, y) - x_C(p_2, V(p_1, y))), \qquad (2.12)$$

---

[7] This was pointed out by Hausman (1981a), among others.
[8] See Auerbach and Rosen (1980) for further discussion.

where the additional term compared to (2.10) is the revenue lost as demand declines with utility held constant at $V(p_1, y)$.

3. Evaluating the measures of excess burden

3.1. Taylor approximations and informational requirements

For purposes of exposition, it is sometimes easier to express the deadweight loss calculations above in terms of second-order Taylor approximations. For example, if we expand the exact measure EB_C around the initial price vector p_1, we obtain

$$EB_C = \left. \frac{\mathrm{d}\,EB_C}{\mathrm{d}p} \right|_{p_1} \cdot (p_2 - p_1) + \tfrac{1}{2}(p_2 - p_1)' \left. \frac{\mathrm{d}^2 EB_C}{\mathrm{d}p^2} \right|_{p_1} (p_2 - p_1) + \cdots, \quad (3.1)$$

which, ignoring all terms beyond the second order, yields

$$EB_C \approx \left[-(p_1 - p_0)' \left[\frac{\mathrm{d}x_C}{\mathrm{d}p} \right] (p_2 - p_1) \right.$$

$$\left. + \tfrac{1}{2}(p_2 - p_1)' \left[-\frac{\mathrm{d}x_C}{\mathrm{d}p} - (p_1 - p_0) \frac{\mathrm{d}^2 x_C}{\mathrm{d}p^2} \right] (p_2 - p_1) \right], \quad (3.2)$$

where x_C is evaluated at p_1 and $V(p_1, y)$. If we make a further approximation by ignoring the curvature terms of the compensated demand function $\mathrm{d}^2 x_C / \mathrm{d}p^2$, we obtain

$$EB_C \approx -\left(t'S\Delta t + \tfrac{1}{2}\Delta t'S\Delta t \right) = -\left(t'\Delta x_C + \tfrac{1}{2}\Delta t'\Delta x_C \right), \quad (3.3)$$

where $t = (p_1 - p_0)$, $\Delta t = (p_2 - p_1)$, $S = \mathrm{d}x_C / \mathrm{d}p$ is the Slutsky matrix, and $\Delta x_C = S\Delta t$. This is of a form similar to the single-market measure derived above for simple consumers' surplus, but the changes in demand are now compensated changes rather than ordinary ones. The approximation in (3.3) is that originally derived by Harberger (1964), although the procedure used to derive it here is somewhat simpler.[9]

From (3.3), we may observe a number of additional characteristics of tax-induced excess burden. First of all, when there are pre-existing taxes in other

---

[9] One can also derive higher-order approximations of EB_C. For a comparison of the errors involved in using second- and third-order approximations, see Green and Sheshinski (1979).

markets, the introduction of another tax need not worsen things. We must weigh the strictly positive term $-(\Delta t_i)^2 S_{ii}$ for the new tax in market i against the cross-effects $-t_j S_{ji} \Delta t_i$ in each other market j, which represent the loss in revenue from the tax t_j due to the drop in demand resulting from the price increase in market i. Since S_{ji} may be positive or negative, so may each of those terms. In general, if pre-existing taxes are on goods substitutable for good i ($S_{ji} > 0$), the new tax is more likely to lessen the total excess burden of the tax system.

A second observation to make from (3.3) is that excess burden is a non-linear function of tax rates. Consider, for example, a single tax t_i imposed upon a state without taxes. The excess burden is approximately $-\frac{1}{2}t_i^2 S_{ii}$, so that it increases with the *square* of the tax. This suggests that to raise a certain amount of revenue, we might reduce excess burden by using several small taxes rather than a few large ones, perhaps tilting toward those with smaller own-substitution effects for which the scale of excess burden is lower. However, once several taxes are used, the cross-effects just discussed need also be evaluated. How these aspects fit together will become clearer in Section 5 when we formally consider the optimal tax problem.

Aside from expositional purposes, the use of a Taylor approximation can only be justified on grounds of insufficient information. If we know the consumer's expenditure function, we can calculate either of the exact measures of excess burden explicitly. Even if we know only the consumer's ordinary demand function, we can solve for his indirect utility function and hence his compensated demand function (in principle) using the system of partial differential equations generated by Roy's identity,[10]

$$x(p, y) = -\frac{dU/dp}{dU/dy}. \tag{3.4}$$

Thus, we must know less than the consumer's demand function if we are to justify the use of an approximation; perhaps only its local properties. However, even in this case, it is probably preferable to construct an exact measure to the extent of one's limited knowledge of demand characteristics away from the initial equilibrium, and use confidence bounds based on the precision of our underlying parameter estimates. Alternatively, one can use revealed preference theory in conjunction with observed data to derive bounds, without ever specifying a particular demand function [Varian (1982)].

A second defense of the use of approximations or even of simple consumers' surplus measures is that the demand function as estimated is not integrable, so that we cannot use the procedure suggested above to derive the associated

---

[10] See Hausman (1981a). Vartia (1983) presents a numerical algorithm for generating utility functions from demand functions.

compensated demand function. However, lack of integrability is synonomous with the violation of the laws of demand. If such laws are violated, what interpretation can we give *any* measure we use?

3.2. Variations in producer prices

The assumption made thus far in this section that producer prices are fixed is a common one in the literature, but may do violence to our representation of the actual situation prevailing in the economy. For example, we know that a tax on a good in absolutely fixed supply is equivalent to a lump sum tax and therefore non-distortionary, regardless of how elastic the demand for the good is. Our preliminary examination of excess burden using consumers' surplus in Section 2 suggested that the excess burden of a tax is proportional to the reduction in the output of the taxed good, taking account of both demand and supply conditions. It would be useful to extend the Hicksian measures in the same direction.

The complication that arises in doing so is that it is no longer sufficient to posit a certain money value of compensation: since producer prices change, the form of compensation matters. For example, to extend the compensating variation measure of excess burden, we must specify the form in which the compensation from "outside" the system, in excess of collected revenue, will come.

To develop a compensating variation measure of the additional excess burden caused by an increase in taxes, starting at a distorted equilibrium, we let α be the compensation vector of the elements of x, and β the scalar that determines how much of the compensation bundle the consumer receives, $\beta\alpha$. If we denote producer prices by q and consumer prices by p, then the compensating variation measure of excess burden β can be defined implicitly by the equation

$$V(p_2, y_2 + R_2 - R_1 + q_2 \cdot \alpha\beta) = V(p_1, y_1), \tag{3.5}$$

where p_1 is the initial consumer price vector, p_2 the distorted price vector, q_1 and q_2 the corresponding producer price vectors, y_1 and y_2 the lump sum income in the two states, and $R_1 = (p_1 - q_1) \cdot x(p_1, y_1)$ and $R_2 = (p_2 - q_2) \cdot x_C(p_2, V(p_1, y_1))$ the revenue in the two states. The values of y are indexed by their respective states because they may vary when producer prices change. For example, if the economy's production function exhibits decreasing returns to scale in the consumer goods x, then the pure profits from competitive production are positive and change with the change in producer prices. Letting z be the vector of goods produced (negative for net factor inputs), total profits are $y = q \cdot z$. Note that production and consumption differ by the infusion of additional compensation, $\beta\alpha$.

Expression (3.5) can be transformed into another that is similar to those of the previous section. Using the fact that $U_A = U_B \rightarrow E(p, U_A) = E(p, U_B)$, and that $E(p, U(p, y)) = y$, we obtain

$$q_2 \cdot \alpha\beta = E(p_2, V(p_1, y_1)) - y_2 - (R_2 - R_1)$$

$$= [E(p_2, V(p_1, y_1)) - E(p_1, V(p_1, y_1))] + (y_1 - y_2) - (R_2 - R_1).$$

(3.6)

Compared to (2.12), there is a new term, $(y_1 - y_2)$, representing the reduction in profit between states 1 and 2. Thus, there are now three terms in the expression for excess burden, representing the changes in consumers', producers' and government surplus, as in the simple, Marshallian example depicted in Figure 2.3.

This expression for excess burden also differs in that it is not actually a *solution* for β. It will hold regardless of the choice of α, though the solution for β depends on this choice. This dependence can be demonstrated by considering the second-order approximation for β,

$$\beta \approx \frac{d\beta}{dt} \Delta t + \tfrac{1}{2} \Delta t' \frac{d^2\beta}{dt^2} \Delta t,$$

(3.7)

evaluated at the initial point 1. Total differentiation of (3.5) yields

$$\frac{dV}{dp} \cdot dp + \frac{dV}{dy} \left[\frac{dy}{dq} \cdot dq + \beta\alpha \cdot dq + d\beta\alpha \cdot q + t \cdot dx + x \cdot dt \right] = 0,$$

(3.8)

where $t = (p - q)$.

Again using the envelope theorem, one can show that $dy/dq = z$. Using this and Roy's identity [(3.4) above], we obtain from (3.8)

$$\frac{dV}{dy} [-x \cdot dp + z \cdot dq + \beta\alpha \cdot dq + d\beta\alpha \cdot q + t \cdot dx + x \cdot dt] = 0.$$

(3.9)

But since $x = z + \beta\alpha$ and $dV/dy \neq 0$, (3.9) simplifies to

$$q_2 \cdot \alpha d\beta = -t \cdot dx,$$

(3.10)

which is precisely the form of the first-order effect derived above in (3.3).

We derive the second-order term by totally differentiating (3.10). This yields

$$q_2 \cdot \alpha d^2\beta = -dt \cdot dx - d\beta\alpha \cdot dq - t \cdot d^2x,$$

(3.11)

which, even if one ignores the last curvature term, has an additional term

compared to the second-order effect in (3.3), caused by the changing value of the compensation bundle. This may be seen by substituting (3.10) and (3.11) into (3.7) to obtain

$$q_2 \cdot \alpha\beta \approx -\left(t' \Delta x + \tfrac{1}{2} \Delta t' \Delta x + \tfrac{1}{2}\beta\alpha \cdot \Delta q \right), \tag{3.12}$$

where the right-hand side of (3.12) includes the first-order approximations $(dx/dt)\Delta t$ for Δx, $(dq/dt)\Delta t$ for Δq, and $(d\beta/dt)\Delta t$ for β. Only in the case that all compensation is in the form of the numeraire commodity will (3.12) reduce to (3.3).[11]

This extra term may be represented graphically by considering the exact measure (3.6) for the case in which there are two goods, one of which is taxed. This is done in Figure 3.1. Let the untaxed good serve as numeraire, so that its price does not change. The supply curve S shows the increasing relative producer price, q, of the taxed good as its production increases. The ordinary demand curve D represents the consumer's preference, given income y_1. With an initial tax of $(p_1 - q_1)$, the initial equilibrium consumption is at x_1, where the supply curve S_1 is that facing the consumer.

As the tax is increased further, we assume the individual is maintained on the same indifference curve, so that demand for x is described by the compensated demand curve passing through the initial point. The supply curve facing the consumer now depends on the form the compensation takes. If some of the taxed good is included in α, then the supply to the consumer is described by curve S_2', rather than S_2, since total supply will exceed production. This leads to consumption at x_2, and production at z_2, rather than the single value in between that would obtain if all compensation were in the form of the numeraire commodity.

Consider now the three terms in expression (3.6). All may be represented in Figure 3.1. The first, as before, is the area to the left of the compensated demand curve between p_1 and p_2. Since $dy = z\,dq$, the second term in (3.6) equals the area to the left of the supply curve S between q_1 and q_2. Finally, R_1 and R_2 equal in area the rectangles defined by p_1, q_1 and x_1, and p_2, q_2 and x_2, respectively. The resulting area for $q_2 \cdot \alpha\beta$ is the usual trapezoid defined by the supply curve, the compensated demand curve, x_1 and x_2 (shaded in Figure 3.1), *less* that of the triangle defined by the producers' supply curve S, the social supply curve S', and prices q_1 and q_2 (cross-hatched in Figure 3.1). This new piece has an area approximately equal to $\tfrac{1}{2}(q_1 - q_2)(x_2 - z_2)$ or, since $x = z + \beta\alpha$ and only this good's price changes, $-\tfrac{1}{2}\beta\alpha \cdot \Delta q$.

Another familiar expression for the second-order effect may be derived from (3.11). Again ignoring the last curvature term, we use the fact that $x = \beta\alpha + z$ to

---

[11] In deriving a similar measure, Diamond and McFadden (1974) made this assumption.
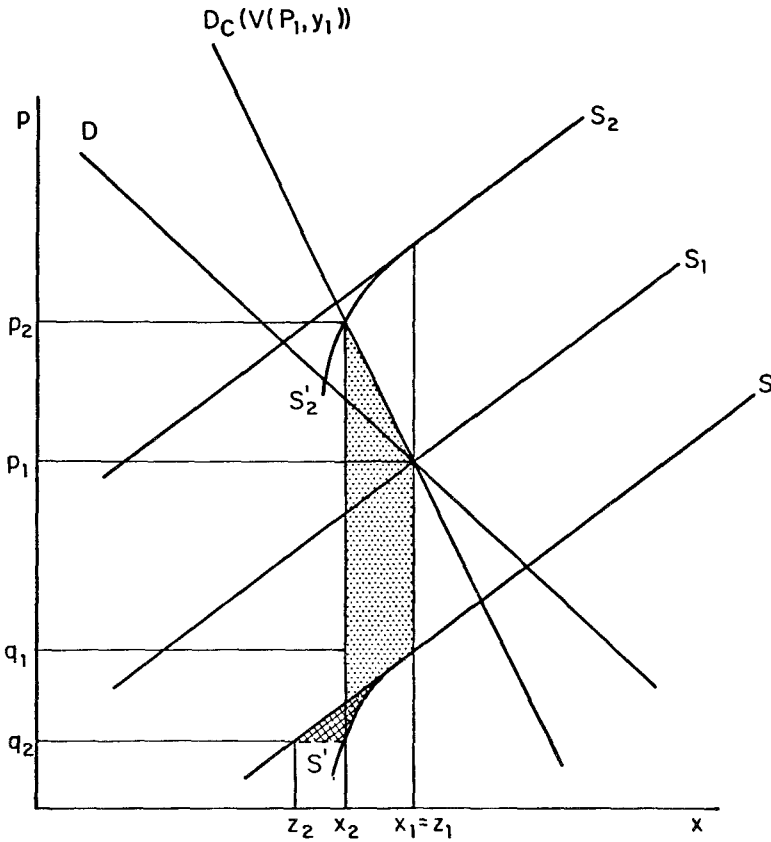
Figure 3.1. Excess burden with changing producer prices.

obtain

$$d^2\beta \approx -dp \cdot dx + dq \cdot dz = -dp'S\,dp + dq'H\,dq,\qquad(3.13)$$

where H is the Hessian of the profit function $d^2y/dq^2 = dz/dq$.

This expression for the second-order effect of a change in taxes on welfare was first developed by Boiteux (1951), although his derivation was limited to the case where the initial equilibrium is undistorted and the first-order effect $d\beta$ vanishes.

Using the notion of equivalent variation, we can construct a measure by asking what level of resources can be extracted from the consumer in excess of additional revenue to avoid an additional tax increase. This yields the following implicit

definition of β:

$$V(p_2, y_2) = V(p_1, y_1 - (R_2 - R_1) - q_1 \cdot \alpha\beta), \tag{3.14}$$

where, in this case, state 2 is the actual state with taxes at t_2, whereas state 1 is the hypothetical state in which taxes do not rise from t_1 but income is reduced to yield the same level of utility as prevails in state 2. Here, $(1 - \beta)$ is related to Debreu's (1951) *coefficient of resource utilization*, which he defines to be the proportion of society's resources that would be necessary to maintain each individual's current level of utility if all distortions were removed. Our measure differs in that we consider the marginal change, rather than removal of a distortion, and let the vector α be arbitrary. (Of course, Debreu's measure is defined relative to all kinds of distortions leading to an inefficient allocation, not just tax-induced changes in the prices of consumer goods.) As before, we cannot solve for β explicitly, but we can calculate the first-order and second-order effects $d\beta$ and $d^2\beta$ at the initial distorted point. We leave further discussion of this measure to the next subsection, which deals with aggregation over consumers.

3.3. Aggregation and welfare comparisons

Thus far, we have defined all our measures of excess burden for the case of a single individual. They are easily generalized to the case of several identical individuals. However, matters become more complicated if we wish to allow for differences in individual tastes, or even differences in income among otherwise identical individuals.

Except under very strict conditions on preferences, any measure of aggregate excess burden will depend on the initial distribution of income. Consider the case of fixed producer prices examined in Section 2, and define a measure of aggregate excess burden, using the compensating variation, as the amount that must come from outside the system to maintain each consumer at his pre-tax level of utility. For two individuals, this measure equals [compare to (2.10)]

$$L = E^1\left(p_1, V^1\left(p_0, y^1\right)\right) + E^2\left(p_1, V^2\left(p_0, y^2\right)\right) - \left(y^1 + y^2\right)$$

$$- \left(p_1 - p_0\right) \cdot \left(x_C^1\left(p_1, V^1\left(p_0, y^1\right)\right) + x_C^2\left(p_1, V^2\left(p_0, y^2\right)\right)\right), \tag{3.15}$$

where superscripts index the consumers 1 and 2.

Suppose now that the initial income distribution is changed by a small reduction in y^1 and an equal size increase in y^2. The change in L would be

$$
dL = -\frac{\partial E^1}{\partial U} \cdot \frac{\partial V^1}{\partial y} + 1 + \frac{\partial E^2}{\partial U} \cdot \frac{\partial V^2}{\partial y}
$$

$$
-1 - (p_1 - p_0) \cdot \left(-\frac{\partial x_C^1}{\partial U} \cdot \frac{\partial V^1}{\partial y} + \frac{\partial x_C^2}{\partial U} \cdot \frac{\partial V^2}{\partial y} \right), \tag{3.16}
$$

which, using the fact that $x_C(p_1, V(p_0, y)) = x(p_1, E(p_1, V(p_0, y)))$, can be rewritten as

$$
dL = -\frac{\partial E^1}{\partial U} \cdot \frac{\partial V^1}{\partial y} + \frac{\partial E^2}{\partial U} \cdot \frac{\partial V^2}{\partial y}
$$

$$
-(p_1 - p_0) \cdot \left(-\frac{\partial x^1}{\partial y} \cdot \frac{\partial E^1}{\partial U} \cdot \frac{\partial V^1}{\partial y} + \frac{\partial x^2}{\partial y} \cdot \frac{\partial E^2}{\partial U} \cdot \frac{\partial V^2}{\partial y} \right). \tag{3.17}
$$

Since $E(p_0, V(p_0, y)) = y$, we may rewrite (3.17) as

$$
dL = -\frac{\hat{\mu}^1}{\mu^1} \left(1 - (p_1 - p_0) \cdot \frac{dx^1}{dy} \right) + \frac{\hat{\mu}^2}{\mu^2} \left(1 - (p_1 - p_0) \cdot \frac{dx^2}{dy} \right), \tag{3.18}
$$

where

$$
\mu^1 = \frac{\partial E^i}{\partial U}\left(p_0, V^i(p_0, y^i) \right) \quad \text{and} \quad \hat{\mu}^i = \frac{\partial E^i}{\partial U}\left(p_1, V^i(p_0, y^i) \right)
$$

are the marginal expenditures needed per unit of increased utility at base utility level $V^i(p_0, y^i)$ and price levels p_0 and p_1, respectively. Thus, dL will equal zero, in general, only if two conditions are met:

1) $\hat{\mu}^i/\mu^i$ equals some common function of prices alone (not income) for the two individuals; and
2) the vector of income effects dx^i/dy equals some common function of prices alone.

Condition 2) implies that ordinary demand functions take the form

$$
x^i(p, y^i) = \phi^i(p) + \theta(p) y^i, \tag{3.19}
$$

for some functions $\phi^i(\cdot)$ and $\theta(\cdot)$, the latter common across individuals. [The laws

of consumer demand imply, in turn, that $\phi^i(\cdot)$ is homogeneous of degree 0 in prices and $\theta(\cdot)$ is homogeneous of degree -1 in prices, since a proportional change in p and y cannot affect $x^i(\cdot)$.] The demand function specified in (3.19) corresponds to the well-known Gorman (1953) "polar form", which plays a central role in the theory of exact aggregation.

Condition 1) implies that, for suitable transformation of the utility function, consumer i's expenditure function can be written

$$E^i(p, U^i) = \delta^i(p) + \gamma(p) \cdot U^i, \tag{3.20}$$

[with $\delta^i(\cdot)$ and $\gamma(\cdot)$ homogeneous of degree 1 in prices]. This is the expenditure function corresponding to the Gorman polar form [see Muellbauer (1976)], so that conditions 1) and 2) are each satisfied if and only if preferences satisfy this very restricted pattern that allows variations from identical homothetic preferences only through individual-specific displacements through the "basic needs" function of zero-income consumption, $\phi^i(\cdot)$.

Note that even identical preferences, unless homothetic, will not suffice. For example, suppose individuals have a price-inelastic compensated demand for a commodity at high incomes but an elastic demand at low incomes. Then the excess burden of a tax on this good will be increased if we transfer income to the poorer individual, for this will increase the overall demand elasticity for the taxed good. Thus, any measure of excess burden we envisage is not independent of the income distribution. Similarly, if we required not that each individual's utility be kept constant, but that individual 1 receive one dollar less than would be necessary, this, too, would affect the aggregate measure for the same reason.

Of course, it is still possible to *define* measures of excess burden for the multi-individual case, given the initial resource distribution. For example, we may implicitly define a compensating variation measure analogous to (3.5) by the identities

$$V^i(p_2, \omega^i(y_2 + R_2 - R_1 + q_2 \cdot \alpha\beta)) = V^i(p_1, \bar{\omega}^i y_1), \quad \forall i, \tag{3.21}$$

where i indexes the individual, $\bar{\omega}^i$ is individual i's actual profit share, and ω^i is the share needed to maintain each individual on the same indifference curve as prices rise to p_2 and the extra compensation vector $\alpha \cdot \beta$ "enters" the system. For the equivalent variation, the measure for β corresponding to (3.14) for several individuals is

$$V^i(p_2, \bar{\omega}^i y_2) = V^i(p_1, \omega^i(y_1 - R_2 + R_1 - q_2 \cdot \alpha\beta)). \tag{3.22}$$

Again, it is not generally possible to solve explicitly for β in either case, but we can derive expressions for the first-order and second-order effects $d\beta$ and $d^2\beta$ by

totally differentiating (3.21) or (3.22) for each i and then adding over i, making use of the adding-up constraint on the profit shares ω. While the resulting expressions for the compensating variation measure are essentially the same as those described in Section 3.2 (with aggregate demands replacing individual ones), an interesting result occurs in the second-order effect derived from the measure defined by (3.22). It contains an additional term reflecting the indirect impact of taxes on excess burden through the change in the income distribution in state 1 [Debreu (1954)]. Since for an equivalent variation measure state 1 is simply a hypothetical state based on the utility levels in state 2, changes in taxes, even starting at a no-tax position, influence the distribution of real income in state 1. Indeed, it should not be surprising that the condition required for this extra term to vanish is the same one required above for excess burden to be independent of the initial income distribution.

There is a temptation to respond to this dependency of excess burden on the distribution of income by conceptually separating questions of allocation and distribution, following Musgrave's (1959) framework for the different "branches" of government: let the distribution branch worry about distribution, and the allocation branch concern itself with minimizing excess burden. However, there are two problems with this approach. First, if the distribution branch is not in operation, we cannot obtain well-behaved social welfare prescriptions by comparing levels of excess burden in different allocations through the device known as the compensation principle: one state being preferred to another if winners *could* compensate losers. Unless such compensation actually occurs, the orderings coming out of such a procedure need not be well-behaved or consistent with any particular social welfare function. This is the essence of the critique of the Hicks (1940)–Kaldor (1939) approach to welfare economics [Samuelson (1947)].

A second response might be that we are only interested in efficiency, not distribution, and so will assign equal distributional weights to individuals, thereby allowing the interpretation of the aggregate measures derived above as "efficiency-only" social welfare measures. Such is the approach suggested by Harberger (1971). Unfortunately, this will not work either. We can certainly imagine a social welfare function of the form

$$w(U^1,\ldots,U^H) = \sum_{i=1}^{H} U^i, \tag{3.23}$$

and can even choose a normalization for the individual utility functions so that, in the initial state, the marginal utility of income and hence the social marginal utility of income for each individual is one ("money metric" utility). However, once prices change, as they will when taxes are introduced, the changes in real income, and hence the marginal utility of income, will generally be different.

Thus, for our measure of excess burden to correspond to a social welfare function, it would require price-dependent individual weights, even if the weights were initially equal. Only when preferences satisfy the Gorman conditions will weights initially set equal remain equal in all cases [Roberts (1980)]. Thus, it will generally not be possible to make welfare comparisons on the basis of aggregate measures of excess burden, no matter what our attitude is about the relative importance of equity and efficiency.

4. The empirical measurement of excess burden

The ultimate value of the theory developed in Sections 2 and 3 is in its application to measuring real world distortions. This section offers a brief review of some of the research that has been done in this popular area of investigation. No attempt will be made to provide an exhaustive summary of the empirical literature on the measurement of excess burden.

4.1. Measurement with Taylor approximations

The earliest empirical work on the measurement of excess burden was done by Harberger, in a series of papers. In each case, he applied a second-order Taylor approximation of the form in (3.3), expanded around the no-tax point. An example of this research may be found in Harberger (1964), which considers the welfare cost of a progressive tax on labor income by individual income classes. Treating capital as a factor supplied by households in static model, Harberger (1966) considered the deadweight loss from the production distortion caused by differential taxation of the return to capital in the corporate and non-corporate sectors. Non-tax distortions, such as those caused by monopolistic pricing, can also be analyzed using standard excess burden formulae [Harberger (1954)]. One can also analyze the intertemporal allocation distortion caused by capital income taxes by thinking of consumption in different periods as different commodities [Feldstein (1978)].

Aside from the use of the Taylor approximation, a weakness typical of most of this early work (excluding, of course, Harberger's piece on the corporate income tax) was the assumption of fixed producer prices. With a convex production frontier, changes in production prices would normally act to lessen the excess burden caused by a tax increase. An example of the sensitivity of this assumption about production parameters may be found in Chamley (1981) with respect to the excess burden of capital income taxation.

4.2. Exact measures

As stressed in Section 3, there is rarely a situation in which Taylor approxima-
tions need be used in place of exact measures based on the Hicksian variations.
This point is stressed by a number of authors [including Auerbach and Rosen
(1980) and Hausman (1981a)]. For many systems of demand functions (such as
the linear expenditure system discussed in Section 6) it is easy to recover the
parameters of the expenditure function from estimated ordinary demand func-
tions. Moreover, one can also use the standard errors of such estimates to place
confidence bounds on the excess burden measures themselves [Hausman (1981a)].
Several recent studies have used exact measures to calculate the excess burden of
taxation. For example, Rosen (1978) considered the excess burden of wage
taxation using a linear expenditure system estimated from a cross-section.

One of the additional benefits of the "exact" approach to measuring deadweight
loss is that it can readily be generalized to allow for changes in income. That is,
we can deduct from changes in the expenditure function not only changes in
revenue, but changes in income, to calculate the excess burden of a tax system
that changes individual incomes as well as the prices of some commodities. For
example, the compensating variation measure (2.10) would become

$$EB_C = E(\boldsymbol{p}_1, V(\boldsymbol{p}_0, y_0)) - y_1 - (\boldsymbol{p}_1 - \boldsymbol{p}_0) \cdot x_C(\boldsymbol{p}_1, V(\boldsymbol{p}_0, y_0)), \qquad (4.1)$$

where y_0 is income in the undistorted state and y_1 is income in the distorted
state. This tool is particularly useful for the analysis of progressive taxes, where
individuals behave *as if* they faced a proportional tax equal to the actual
marginal rate, with the inframarginal excess in collections that results being
subtracted from lump sum income. For example, consider the case of a progres-
sive labor income tax in a two-good model. The individual's before-tax and
after-tax budget lines are represented in Figure 4.1. If the individual chooses
point A, we may pretend that he did so in response to a proportional tax at rate
$(w_0 - w_A)/w_0$ and lump sum income of y_A. If he chooses point B, we could
imagine a proportional tax of $(w_0 - w_B)/w_0$ and lump sum income of y_B. This
technique has been used in labor supply estimation and excess burden calculation
by Hausman (1981b). King (1983b) has used the equivalent variation analogue of
(4.1), which he calls the "equivalent gain", to evaluate the effects of changes in
housing policy in the U.K.

An additional extension possible with exact measures is the case of discrete
choices, such as the decision to work or to purchase a durable good. Suppose
there are two regimes among which a consumer must choose. The general
methodology for calculating excess burden is, as before, to equate utility changes
from distortionary and lump sum taxation, and compare the tax revenue. How-
ever, the changes in utility take account of switches in regime that may occur in
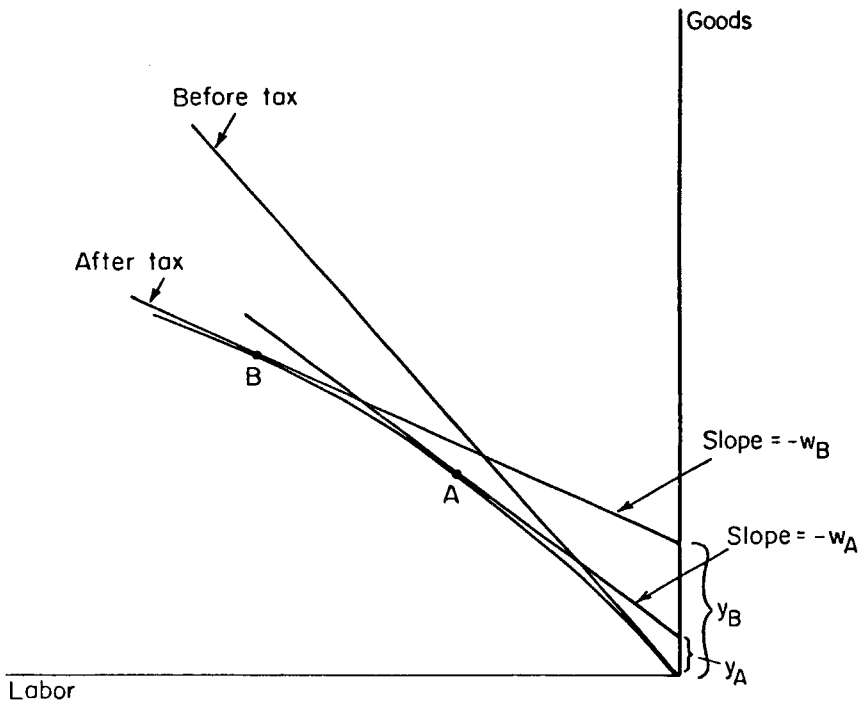
Figure 4.1. Progressive taxes and virtual income.

each case. This is a straightforward calculation when the consumer's indirect utility function is known, for it is simple to identify the regime chosen in any situation. However, if one wishes to use approximation formulae, one must take explicit account of the effect of taxes on the probability of switching regimes. [See Small and Rosen (1981).] An example of excess burden calculations with discrete decision variables is the analysis of housing subsidy programs by Venti and Wise (1984), in which individuals must decide whether to move or stay, and face different budget constraints in the two situations.

4.3. Simulation methods

Ultimately, there are limitations on the extent to which we can obtain closed form solutions for excess burden. This is particularly true of general equilibrium calculations, for we must solve explicitly for the changes in producer prices consistent with changes in consumer behavior. A solution to this problem is the

simulation model, in which explicit parameterizations of preferences and technology are made and actual equilibria calculated. It is then straightforward to estimate changes in utility caused by a change in tax regime, or the resources one could extract or must add to compensate for a given change. The latter type of calculation corresponds to the price-varying excess burden measures cited in Section 3. The use of disaggregated, static general equilibrium models to analyze the effects of taxation has now become rather common. An early example of the use of simulation technique is Shoven's (1976) reconsideration of the excess burden caused by the corporate income tax. For other applications, see the contributions in Feldstein (1983). In more recent work, Auerbach, Kotlikoff and Skinner (1983) use a perfect-foresight, overlapping-generations growth model to analyze the effects on different cohorts of individuals of various dynamic tax changes, such as an unannounced switch from income taxation to consumption taxation.

5. The theory of optimal taxation

Taxes distort behavior and cause excess burden. How can this excess burden be kept to a minimum while government simultaneously raises the revenue it requires for public expenditures? This is the optimal tax problem, solved in its basic form by Ramsey (1927).

Of course, there do exist non-distortionary taxes, at least hypothetically. Taxes on pure profits are just one form of such taxation. The optimal tax problem, in a sense, embodies the concession that such ideal taxes may be difficult to institute in practice. One might cite a number of reasons for this, including the political constraints on non-uniform taxation dependent on personal characteristics. For example, we might succeed in having a non-distortionary and progressive tax system by taxing according to genetic characteristics associated with ability, but such schemes are typically proscribed. In addition, it may be impossible to observe such characteristics.

In the next subsection, we present and interpret the basic, single-individual optimal tax results, paying particular attention to the role of the "untaxed" numeraire commodity that is often a confusing part of such analysis. Section 5.2 discusses the relationship of the optimal tax solution to the measures of excess burden described above. In Sections 5.3 and 5.4, we show how the results can be extended to allow for profits and changing producer prices, and interpret the classic results of Diamond and Mirlees (1971) and Stiglitz and Dasgupta (1971) concerning the desirability of production efficiency in the presence of distortionary commodity taxes.

5.1. Basic optimal tax results

We imagine a representative consumer who has exogenous income y, and faces consumer prices $p = (p_0, p_1, \ldots, p_N)$ for the commodities $0, 1, \ldots, N$, which have fixed producer prices $q = (q_0, q_1, \ldots, q_n)$. Without any loss of generality, we may choose good zero as the numeraire and set $q_0 = 1$.

The government may use unit excise taxes $t = (t_0, t_1, \ldots, t_N)$ on the goods $0, 1, \ldots, N$, to raise a certain amount of required revenue, R. (We will relax this ignorance of the expenditure side below.) Assuming the consumer maximizes utility $U(x)$ in the goods x, subject to the prices p and income y, we may express the optimal tax problem by

$$\max_p \left[\max_x U(x) \text{ subject to } p \cdot x = y \right] \quad \text{subject to} \quad (p - q) \cdot x = R, \qquad (5.1)$$

or, using the definition of the indirect utility function $V(\cdot)$,

$$\max_p V(p, y) \quad \text{subject to} \quad (p - q) \cdot x = R. \qquad (5.2)$$

Note that we specify the price vector, p, as our control rather than t, but this is a trivial distinction when the social cost vector q is fixed since $dt/dp = I$, the identity matrix of order $N + 1$.

The first-order conditions for the Lagrangian

$$V(p, y) - \mu[R - (p - q) \cdot x] \qquad (5.3)$$

are

$$-\lambda x_i + \mu \left[\sum_j t_j \frac{\partial x_j}{\partial p_i} + x_i \right] = 0, \quad \forall i, \qquad (5.4)$$

where $\lambda = dV/dy$ is the consumer's marginal utility of income. Condition (5.4) may be rearranged in a number of ways. Perhaps the most useful involves splitting the cross-price effects $\partial x_j/\partial p_i$ using the Slutsky equation, and defining

$$\alpha = \lambda + \mu \sum_i t_j \frac{\partial x_j}{\partial y} \qquad (5.5)$$

to be the *marginal social utility* of income [Diamond (1975)], to obtain

$$-\sum S_{ij} t_j = \left(\frac{\mu - \alpha}{\mu}\right) x_i, \quad \forall i, \tag{5.6}$$

where the S_{ij}s are components of the Slutsky matrix S. The term α differs from λ because, in the presence of excise taxes, a dollar given to the individual increases his utility directly by λ *and* indirectly by the increased revenue resulting from additional expenditure. Since we can interpret the Lagrange multiplier of the revenue constraint, μ, as the shadow cost in terms of utility of raising an additional dollar of revenue, the indirect gain of revenue added by increased expenditures out of an additional dollar of income equals $\mu \sum_j t_j (\partial x_j / \partial y)$, the second term in the definition of α.

The term $(\mu - \alpha)$ represents the difference between raising a dollar of revenue at the actual margin and raising it through a direct taking of income from the consumer: the marginal excess burden of the tax. This term is always non-negative [see expression (7.8)] and hence the terms $\sum S_{ij} t_j$ are also non-negative.

There is one potential solution to (5.6) that would be particularly attractive, for it involves no distortion. If we choose equal proportional *ad valorem* taxes, or

$$t_i = \theta p_i, \quad \forall i, \tag{5.7}$$

for some constant θ, we obtain

$$-\theta \sum S_{ij} p_j = \left(\frac{\mu - \alpha}{\mu}\right) x_i, \quad \forall i. \tag{5.8}$$

But $\sum S_{ij} p_j$ equals $(1/\lambda)(dU/dp_i)|_u = 0$ for all i. (This is simply a statement of the envelope theorem.) Therefore, the system of equations in (5.8) is satisfied for $\mu = \alpha$ and hence no excess burden. Thus, proportional excise taxes would appear to be the solution.

The reason such taxes are non-distortionary, however, is the key to their limited applicability. Since $p = q + t = q + \theta p$, $p = q/(1 - \theta)$. Hence, the consumer's budget constraint becomes

$$\frac{q}{1 - \theta} \cdot x = y \quad \text{or} \quad q \cdot x = y(1 - \theta), \tag{5.9}$$

where θ is chosen to satisfy $\theta = R/y$. A system of equal excise taxes is nothing more than a tax on the consumer's exogenous income, and hence a lump sum tax.

If $y = 0$, then no finite value of θ will satisfy the revenue constraint, so we must ask when y will be non-zero.

First of all, y will be non-zero in general if there are decreasing returns to scale in production (in a more general model not assuming fixed producer prices). Even in the absence of pure profits, y will be non-zero if we interpret it as "full income" and the x vector as consumption rather than demand. For example, suppose the x vector consists of two commodities, consumption C and leisure l, and that the consumer has a labor endowment L. Without pure profits, the consumer's budget constraint in the absence of taxes may be written either as

$$qC_C + (l - L) = 0, \tag{5.10a}$$

or

$$qC_C + l = L, \tag{5.10b}$$

where labor is the numeraire and C and q_C are the amount and relative price of consumption. Interpreting the labor commodity we can tax as net *purchase* of leisure $(l - L)$, we have no income y to tax through proportional excise taxes. Interpreting the commodity as *consumption* of leisure, l, we can use the proportional tax solution on C and l to tax L indirectly. Hence, the inability to use proportional taxes to raise revenue is equivalent to restriction of taxing only explicit purchases, rather than total consumption. Under this restriction, a proportional tax raises no revenue [Baumol and Bradford (1970)]. Based on examples of this sort, various authors have equated the need to use distortionary taxes with the inability to tax leisure, but this is somewhat misleading on two counts: we *can* tax leisure *purchases* (labor supply), and this restriction applies to any commodity in which the consumer has an endowment.

Once we do restrict our taxes to net purchases, it is easiest to interpret the vector x to be such flows rather than total consumption. In exchange for the loss of non-distortionary tax scheme, we gain an additional free normalization. Since the consumer's indirect utility function is homogeneous in prices and income, and is now simply $V(p)$, it is also homogeneous of degree zero in prices. So is the revenue constraint: since $p \cdot x = 0$, it follows that for any constant ϕ,

$$(\phi p - q) \cdot x = (\phi - 1)p \cdot x + (p - q) \cdot x = (p - q) \cdot x. \tag{5.11}$$

Thus, we may choose any scale for p. It is customary to set $p_0 = 1$, thereby making the numeraire also the arbitrarily "untaxed" good. Typically, in models where there is a single factor supplied, labor, and several commodities purchased, labor is chosen as this numeraire. While such a normalization is innocuous and in no way affects the real characteristics of the outcome, it can be very confusing:

the untaxed good, labor, just happens to be the only good with an endowment, L, that we cannot tax independently of its consumption, l; hence the loss of distinction between untaxable and untaxed goods. If we chose corn as the untaxed good, labor would still have an untaxable endowment. This distinction is important when one interprets the various rules derived below.

We now have only N first-order conditions, from (5.6), having dropped that corresponding to p_0. Hence, the strategy of equal proportional taxes at rate θ (with a zero tax on good zero, of course) now gives us the terms

$$-\theta \sum_{j \neq 0} S_{ij} p_j = \theta S_{i0}, \tag{5.12}$$

on the left-hand side of (5.6). This will stand in constant proportion to x_i over i, as required for a solution, only if the compensated cross-elasticity of demand for each good i with respect to the price of good 0, $\varepsilon_{i0} = S_{i0} \cdot p_0 / x_i = S_{i0} / x_i$, is the same for all $i \neq 0$. Thus, equal proportional taxes on all *taxed* goods satisfy the first-order conditions only if all goods are equally complementary [in the sense of Hicks (1946)] to the *untaxed* good. Naturally, if these conditions are satisfied for a given choice of untaxed good, they will not generally work for another.

Our analysis of (5.6) has now generally ruled out uniform taxation. But how should the taxes diverge from uniformity? Note that the N conditions in (5.6) can be stacked to yield

$$\hat{S}\hat{t} = -\left(\frac{\mu - \alpha}{\mu}\right)x, \tag{5.13}$$

where $\hat{S}$ is the Slutsky matrix excluding good zero and $\hat{t} = (t_1, \ldots, t_N)$. Although there is no independent condition with respect to the tax on good zero (which has been normalized to zero), it is helpful to note that these N conditions imply that (5.6) also holds for good zero. This may be shown as follows. Adding a term multiplied by t_0 to each of the N first-order conditions in (5.6) has no effect, since $t_0 = 0$. Thus,[12]

$$\sum_{i=0}^{N} S_{0i} t_i = \sum_{i=0}^{N} \left(-\sum_{k=1}^{N} p_k S_{ki}\right) t_i = -\sum_{k=1}^{N} p_k \sum_{i=0}^{N} S_{ki} t_i$$

$$= -\sum_{k=1}^{N} p_k \left(-\left(\frac{\mu - \alpha}{\mu}\right) x_k\right) = -\left(\frac{\mu - \alpha}{\mu}\right) x_0. \tag{5.14}$$

[12] This uses the facts that $\sum_{k=0}^{N} p_k S_{ki} = 0$ and $p \cdot x = 0$.

Combined with (5.13), this yields

$$St = -\left(\frac{\mu - \alpha}{\mu}\right)x. \tag{5.15}$$

Suppose that the government is currently raising its revenue through lump sum taxes, and must now shift over some of the revenue collection to distortionary taxes. From above, we know that there is no first-order effect on utility of introducing distortionary taxes from a Pareto optimum, so that the effects on demand of this small change in prices will be compensated effects. Thus, to a first-order Taylor approximation, the reduction in the demand for good i will be

$$-\Delta x_i = -\sum_j S_{ij}\Delta p_j = -\sum_j S_{ij}t_j, \tag{5.16}$$

so that (5.15) calls for an equiproportional reduction in demand for each good. As suggested by Dixit (1970), this makes intuitive sense in light of the excess burden formulae calculated above. From (3.3), the introduction of small taxes t starting from a Pareto optimum induces an excess burden of approximately

$$L = \tfrac{1}{2}\sum_i \Delta t_i \Delta x_i = \tfrac{1}{2}\sum_i t_i \Delta x_i, \tag{5.17}$$

so that each small tax t_i will induce an excess burden proportional to Δx_i. On the other hand, the revenue raised by such a tax is $t_i x_i$. Thus, holding $\Delta x_i/x_i$ constant across goods results in a constant ratio of excess burden to a revenue for each tax. This is precisely the sort of marginal condition one would expect from minimizing total excess burden subject to a revenue constraint.

The actual taxes that lead to the achievement of (5.13) and (5.15) may be obtained by inverting $\hat{S}$ and multiplying both sides of (5.13) by $\hat{S}^{-1}$ to obtain

$$\hat{t} = -\left(\frac{\mu - \alpha}{\mu}\right)\hat{S}^{-1}x. \tag{5.18}$$

This yields no neat general expressions for $\hat{t}$, though for various special cases one can go a little further.

If there are only three goods, two taxed, then (5.18) yields the two equations

$$t_1 = \frac{1}{\Delta}\left(\frac{\mu - \alpha}{\mu}\right)(-S_{22}x_1 + S_{12}x_2), \tag{5.19a}$$

$$t_2 = \frac{1}{\Delta}\left(\frac{\mu - \alpha}{\mu}\right)(S_{21}x_1 - S_{11}x_2), \tag{5.19b}$$

where $\Delta = S_{11}S_{22} - S_{12}S_{21}$, which must be ≥ 0 because S is negative semi-definite. Since $S_{i0} + p_1 S_{i1} + p_2 S_{i2} = 0$ for $i = 1, 2$, we may divide (5.19a) by (5.19b) and substitute to obtain

$$\frac{t_1}{t_2} = \frac{\dfrac{1}{p_2}(S_{20} + p_1 S_{21})x_1 + S_{12}x_2}{\dfrac{1}{p_1}(S_{10} + p_2 S_{12})x_2 + S_{21}x_1},$$
(5.20)

or, defining $\theta_i = t_i/p_i$ and dividing the numerator and denominator of the right-hand side of (5.20) by $x_1 x_2$, we obtain [Corlett and Hague (1953–54) and Harberger (1964)]

$$\frac{\theta_1}{\theta_2} = \frac{\varepsilon_{20} + \varepsilon_{21} + \varepsilon_{12}}{\varepsilon_{10} + \varepsilon_{21} + \varepsilon_{12}},$$
(5.21)

where, as before, ε_{ij} is the compensated cross-elasticity $S_{ij}(p_j/x_i)$. As we discovered above, $\theta_1 = \theta_2$ is an optimal solution only if the cross-elasticities ε_{10} and ε_{20} are equal.

Because $\Delta \geq 0$, expression (5.21) calls for a higher tax on the taxed good that is the relative complement to the numeraire (ε_{i0} is smaller). This has generated the somewhat misleading explanation that we "cannot" tax good zero, so we minimize distortions by taxing more heavily its relative complement. Recall that the choice of untaxed good is arbitrary, and that (5.21) applies for any numbering of the three goods.

For a larger number of commodities, a simple result obtains if we assume that the matrix $\hat{S}$ is diagonal: all cross-effects except with respect to good zero are zero. Since $\sum_j S_{ij} p_j = 0$, this implies that, for $i = 1, \ldots, N$,

$$S_{ii}p_i + S_{i0} = 0.$$
(5.22)

Thus, this restriction *does* depend on the choice of untaxed commodity. With such a simplification, (5.18) yields the expressions

$$t_i = -\frac{1}{\Delta}\left(\frac{\mu - \alpha}{\mu}\right)\frac{x_i}{S_{ii}} \quad \text{or} \quad \theta_i \sim \frac{1}{\varepsilon_{ii}}.$$
(5.23)

This is the celebrated "inverse elasticity" rule that calls for higher proportional taxes on goods with relatively low own-price elasticities. By (5.22), this rule is

equivalent to

$$\theta_i \sim \frac{1}{\varepsilon_{i0}}, \tag{5.24}$$

as derived above for the three-good case.

Since the inverse elasticity rule results from a restriction on preferences, the choice of untaxed good becomes relevant in that it may make more sense to assume no cross-effects among taxed goods if labor is numeraire and the other goods are commodities, than to do so if one of the commodities serves as the untaxed good.

The inverse elasticity rule of (5.24) is expressed in terms of compensated elasticities. Yet in various places in the literature [Diamond and Mirrlees (1971) and Bradford and Rosen (1976)], it is expressed in terms of uncompensated elasticities. This is the result neither of a revision of demand theory nor an assumption of zero-income effects. Rather, it comes about because of a different, and equally arbitrary, restriction on preferences. We can express the optimal tax formulae in terms of ordinary uncompensated demands by rearranging (5.4),

$$-\sum_j t_j \frac{\partial x_j}{\partial p_i} = \left(\frac{\mu - \lambda}{\mu} \right) x_i, \tag{5.25}$$

which, assuming $\partial x_j / \partial p_i = 0$ unless $i = 0$ or j, yields

$$\theta_i \sim \frac{1}{\eta_{ii}}, \tag{5.26}$$

where $\eta_{ii} = -(p_i/x_i)(\partial x_i/\partial p_i)$ is the uncompensated own-elasticity of demand for good i. Expressions (5.26) and (5.24) differ because they result from different restrictions on the structure of preferences: different matrices are being assumed diagonal.

5.2. Minimizing excess burden through optimal taxation

By its definition, excess burden ought to be minimized when taxes are chosen to maximize utility. However, even for the fixed producer price case, we have at least two candidates for measuring excess burden, and they will generally take on different values. It turns out that only one of these, that based on the equivalent variation, satisfies the desirable duality property of being minimized by optimal taxes [Kay (1980)].

Recall from (2.9) that the equivalent variation measure of the excess burden of tax is

$$EB_E = E(p_1, V(p_1, y)) - E(p_0, V(p_1, y)) - R$$

$$= y - E(p_0, V(p, y)) - R. \tag{5.27}$$

Thus, minimizing this for a given value of R amounts to maximizing $E(p_0, V(p_1, y))$. But, for a given price vector, expenditure increases monotonically with the level of utility. Thus, we are maximizing $V(p_1, y)$, just as in the optimal tax problem. This is easily verified by differentiating the Lagrangian

$$E(p_0, V(p_1, y)) + \pi(R - (p_1 - p_0) \cdot x), \tag{5.28}$$

with respect to p_1.

For the compensating variation measure, which [from (2.10)] equals

$$EB_C = E(p_1, V(p_0, y)) - E(p_0, V(p_0, y)) - R$$

$$= E(p_1, V(p_0, y)) - y - R, \tag{5.29}$$

minimizing excess burden amounts to minimizing $E(p_1, V(p_0, y))$: choosing taxes to minimize the expenditure necessary to achieve the pre-tax utility level. This need not be the same price vector as the one dictated by optimal taxation. The appropriate Lagrangian here is

$$E(p_1, V(p_0, y)) - \pi(R - (p_1 - p_0) \cdot x), \tag{5.30}$$

which yields first-order conditions

$$-x_i + \pi \left[\sum_j t_j \frac{\partial x_j}{\partial p_i} + x_i \right] = 0, \tag{5.31}$$

which looks like the one derived from (5.28). However, the value of x here is at the hypothetical point at higher prices but with compensation. In the previous case, it is at the actual optimal tax point.[13]

---

[13] *A fortiori*, it can be seen that replacing p_0 with any arbitrary "reference price vector" p_0^* in the expenditure function in (5.27), to define a different concept of excess burden, i.e.,

$$EB_E^* = E(p_1, V(p_1, y)) - E(p_0^*, V(p_1, y)) - R,$$

would also yield a measure consistent with the optimal tax problem [King (1983b)].

This problem with the compensating variation also means that we cannot compare two hypothetical alternatives to a given tax situation by comparing their marginal excess burden measures. Only if preferences are homothetic [Chipman and Moore (1980)] will this problem dissappear. Of course, for pairwise comparisons, where the "initial" point is not well-defined, the equivalent variation and compensating variation are symmetrically defined, so there can be no *a priori* benefit of using one versus the other.

5.3. Changing producer prices

The simple relaxation of the fixed producer price assumption has, perhaps surprisingly, no effect at all on the optimal tax formulae in (5.18) as long as producer prices result from competitive behavior and any pure profits are taxed away by the government.

In place of the fixed producer price assumption of Section 5.1, we assume that production is governed by the production function

$$f(z) = 0, \tag{5.32}$$

where, as in Section 3, z is the production vector in the commodities $0, 1, \ldots, N$. By the assumption of competitive behavior, we know that the producer prices q are proportional to the vector of derivatives of f, $\mathrm{d}f = (f_0, f_1, \ldots, f_N)$. Without any loss of generality, we may set this proportionality constant equal to $1/f_0$ and, as before, choose good zero as numeraire, i.e., $q_0 = 1$.

The government's revenue requirement must now be specified in terms of individual commodities (as was the case of the compensation vector in Section 3), since relative producer prices can change. We refer to this as the revenue vector, R. Thus, $z = x + R$, where x is the household's vector of net purchases.

Once production has been generalized to this stage, the possibility arises of pure profits coming from decreasing returns to scale. We will consider this more general case after first solving the optimal tax problem when $f(\cdot)$ embodies constant returns to scale, i.e., is homogeneous of degree zero in all commodities. By Euler's Theorem, profits are $q \cdot z = 0$. Thus, the government's optimization problem becomes

$$\max_{p} V(p) \quad \text{subject to} \quad f(x + R) = 0, \tag{5.33}$$

where, because pure profits are zero, we can set $p_0 = q_0 = 1$ without any loss of generality, and choose only $p_1, \ldots, p_N$. To use p rather than t as the control variables, we must insure that arbitrary changes in p can be brought about by

changes in t. This is accomplished by noting that

$$d\,p = d\,t + d\,q = d\,t + d(d\,f) = d\,t + H(d\,x + d\,R),\qquad(5.34)$$

where H is the Hessian $d^2 f$ of the production function, as before. Since $d\,R = 0$ and $d\,x$ may be characterized by the Slutsky equation, we have

$$d\,p = d\,t + H\left(S - \frac{\partial x}{\partial y}x'\right)d\,p,\qquad(5.35)$$

or

$$d\,p = \left[I - H\left(S - \frac{\partial x}{\partial y}\cdot x'\right)\right]^{-1}d\,t,$$

where S is the Slutsky matrix. Moreover, since the changes in t are constrained to keep revenue constant, and hence, in the neighborhood of the optimum, utility as well, the changes in x are compensated and (5.35) simplifies to

$$d\,p = [I - HS]^{-1}d\,t = \Omega\,d\,t.\qquad(5.36)$$

As long as Ω exists (i.e., $[I - HS]$ is of full rank), we may control t indirectly through p.

The Lagrangian corresponding to (5.33) yields the first-order conditions

$$-\lambda x_i - \mu\sum_j f_j\frac{\partial x_j}{\partial p_i} = 0,\qquad i = 1,\ldots,N,\qquad(5.37)$$

where $\lambda = dV/d\,y$ and μ is the Lagrange multiplier on the production constraint. Since $p \cdot x = 0$,

$$\sum_j p_j\frac{\partial x_j}{\partial p_i} + x_i = 0.\qquad(5.38)$$

Using this and the fact that $q = d\,f$, we may express (5.37) as

$$-\lambda x_i + \mu\left[\sum_j t_j\frac{\partial x_j}{\partial p_i} + x_i\right] = 0,\qquad(5.39)$$

which is precisely condition (5.4). This result is due to Diamond and Mirrlees (1971).

In the more general case where $f(\cdot)$ is not homogeneous of degree zero, there may be pure profits, $y = q \cdot z > 0$. In this case, we know from before, equal taxes on all commodities amount to a profits tax on y, giving us $N + 1$ rather than N independent instruments. Hence, if we cannot tax one good, this represents a restriction unless we can tax profits directly. For expositional purposes, it is easiest to let the $N + 1$ instruments be the taxes on goods $1, \ldots, N$ and the profits tax, keeping $t_0 = 0$. We let τ be the rate of profits tax. The Lagrangian now is

$$V(p, (1 - \tau)y) - \mu f(x + R). \tag{5.40}$$

Using the fact that $p \cdot x = (1 - \tau)y$, we may arrange the N first-order conditions with respect to the taxes $t_1, \ldots, t_N$ to be

$$-\lambda x_i + \lambda(1 - \tau)\frac{\mathrm{d}y}{\mathrm{d}p_i} + \mu\left[\sum_j t_j \frac{\mathrm{d}x_j}{\mathrm{d}p_i} + x_i - (1 - \tau)\frac{\mathrm{d}y}{\mathrm{d}p_i}\right] = 0. \tag{5.41}$$

It is straightforward to show that if τ may be freely varied, then the $N + 1$ first-order conditions are solved for $t = 0$ and $\lambda = \mu$: no excess burden, with profits taxes being used to raise all revenue. However, if τ is constrained, we must solve the N conditions (5.41), given τ. Unless profits taxes just happen to equal $q \cdot R$, we again face an optimal tax problem.

If $\tau = 1$, so that all profits are taxed away, then (5.41) reduces to the previous optimal tax program, (5.39). Thus, pure profits do not change the picture unless they accrue at least partially to the household [Stiglitz and Dasgupta (1971)]. If τ is fixed at some value not equal to one, the formulas differ.

Since producer prices, and hence profits, change with p, the derivatives $\mathrm{d}x_j/\mathrm{d}p_i$ in (5.41) include the indirect effect of p_i on profits through changes in production,

$$\frac{\mathrm{d}x_j}{\mathrm{d}p_i} = \frac{\partial x_j}{\partial p_i} + \frac{\partial x_j}{\partial y'} \cdot (1 - \pi)\frac{\mathrm{d}y}{\mathrm{d}p_i}, \tag{5.42}$$

where $y' = (1 - \tau)y$. Using (5.42), the Slutsky equation, and the definition of α, the social marginal utility of income, from (5.5), we may rewrite (5.41) as

$$\sum_j S_{ij}t_j = -\left(\frac{\mu - \alpha}{\mu}\right)\left(x_i - (1 - \tau)\frac{\mathrm{d}y}{\mathrm{d}p_i}\right), \tag{5.43}$$

which differs from (5.6) only through the replacement of x_i with $(x_i - (1 - \tau)(\mathrm{d}y/\mathrm{d}p_i))$. One can interpret these terms as the net increase in resources needed to maintain a given level of utility with respect to an increase in p_i in the two respective cases.

If the profits tax $\tau = 0$, and if good zero is the single production factor and the sole good from which revenue is extracted, then one can show that (5.43) yields the result obtained above for fixed producer prices, that to a first-order Taylor approximation, substituting optimal taxes for lump sum taxes causes an equiproportional reduction in the output of all taxed commodities. *A fortiori*, the outcome also holds for the constant returns case just examined. This result is due to Stiglitz and Dasgupta (1971), who in turn attribute it to Ramsey (1927), though the exact equivalence is obscured by differences in methodology.

The key to the single-factor assumption is that, since the production function may be written

$$f(x) = \hat{f}(x_1, \ldots, x_N) - x_0, \tag{5.44}$$

the Hessian $H = \mathrm{d}^2 f$ is block diagonal in the untaxed good and all other goods ($H_{i0} = H_{0i} = 0$ for $i \neq 0$). Thus, the product of H and the substitution matrix S is

$$HS = \left(\begin{array}{c|c} H_{00} & 0 \\ \hline 0 & \hat{H} \end{array} \right) \left(\begin{array}{c|c} S_{00} & S_0' \\ \hline S_0 & \hat{S} \end{array} \right) = \left(\begin{array}{c|c} H_{00}S_{00} & H_{00}S_0' \\ \hline \hat{H}S_0 & \hat{H}\hat{S} \end{array} \right), \tag{5.45}$$

where $S_0' = (S_{01}, \ldots, S_{0N})$ and $\hat{H}$ and $\hat{S}$ are the blocks of H and S for goods 1 through N. This means that the changes in consumer prices of the taxed goods, $p = (p_1, \ldots, p_N)$, can be expressed [using (5.36)] in the neighborhood of the optimum as

$$\mathrm{d}\hat{p} = [I - \hat{H}\hat{S}]^{-1} \mathrm{d}\hat{t} = \hat{\Omega} \mathrm{d}\hat{t}, \tag{5.46}$$

where $\hat{t} = (t_1, \ldots, t_N)$. That is, $\mathrm{d}\hat{p}$ does not depend on the demand for x_0. From (5.46), we may express the first-order change around $\hat{t} = 0$ in $\hat{x}$, the vector of taxed goods, as

$$\Delta \hat{x} = \hat{S}\Delta\hat{p} = \hat{S}\hat{\Omega}\Delta\hat{t} = \hat{S}\hat{\Omega}\hat{t} = \hat{S}\hat{\Omega}\hat{S}^{-1}\hat{S}\hat{t}. \tag{5.47}$$

The elements of the vector $\hat{S}\hat{t}$ are described in (5.43). By the envelope theorem and the fact that $q_0 = 1$, we may solve for the term $\mathrm{d}y/\mathrm{d}p_i$,

$$\frac{\mathrm{d}y}{\mathrm{d}p_i} = \sum z_j \frac{\mathrm{d}q_j}{\mathrm{d}p_i} = \sum_{j>0} z_j \sum_k H_{jk}S_{ki} = \sum_{j>0} z_j \sum_{k>0} H_{jk}S_{ki}, \tag{5.48}$$

where the last step relies on the assumption that $H_{j0} = 0$ for $j \neq 0$. Stacking these

terms, we obtain

$$\frac{\mathrm{d}\,y}{\mathrm{d}\,\hat{p}} = SH\hat{z}, \tag{5.49}$$

where $\hat{z} = (z_1, \ldots, z_N)$. But by assumption, all revenue is spent on good zero, so $\hat{z} = \hat{x}$. Since, also by assumption, $\tau = 0$, it follows from (5.43) that

$$\hat{S}\hat{t} = \left(\frac{\mu - \alpha}{\mu}\right)(I - \hat{S}\hat{H})\hat{x}. \tag{5.50}$$

Substituting (5.50) into (5.47), we obtain

$$\Delta\hat{x} = \left(\frac{\mu - \alpha}{\mu}\right)\hat{S}\hat{\Omega}\hat{S}^{-1}(I - \hat{S}\hat{H})\hat{x}$$

$$= -\left(\frac{\mu - \alpha}{\mu}\right)\hat{S}\hat{\Omega}\hat{\Omega}^{-1}\hat{S}^{-1}\hat{x} = \left(\frac{\mu - \alpha}{\mu}\right)\hat{x}, \tag{5.51}$$

as required.

In the special case where both $\hat{H}$ and $\hat{S}$ are diagonal (i.e., there is no joint production and commodity demands are independent except with relation to the numeraire), the expression (5.49) for $\mathrm{d}\,y/\mathrm{d}\,p$ simplifies to

$$\frac{\mathrm{d}\,y}{\mathrm{d}\,p_i} = z_i H_{ii} S_{ii}, \tag{5.52}$$

which, if we again assume that all revenue raised is spent on the numeraire ($z_i = x_i$ for $i > 0$), allows us to rewrite (5.43) as

$$-S_{ii}t_i = \left(\frac{\mu - \alpha}{\mu}\right)(1 - (1 - \tau)H_{ii}S_{ii}), \tag{5.53}$$

or

$$\theta_i = -\left(\frac{\mu - \alpha}{\mu}\right) \cdot \frac{\left(\dfrac{1}{\varepsilon_{ii}} + (1 - \tau)\dfrac{1}{\sigma_{ii}}\right)}{\left(1 + (1 - \tau)\dfrac{1}{\sigma_{ii}}\right)},$$

where $\varepsilon_{ii} = -S_{ii}(p_i/x_i)$, $\sigma_{ii} = (1/H_{ii})(q_i/x_i)$ and $\theta_i = t_i/p_i$ are the demand and

supply elasticities and *ad valorem* tax for good *i*. [See Stiglitz and Dasgupta (1971) for a slightly different formulation. Also see Atkinson and Stiglitz (1980).]

5.4. Production efficiency

Thus far, we have assumed production to be efficient, with the only distortions imposed by taxes to be with respect to household decisions. However, government can induce distortions in production, either through differential taxation of factors in different uses or through the use of different shadow prices in public enterprises than those generated by coexisting competitive private markets. Should these extra policy instruments be used? Under certain well-defined conditions, they should not.

To consider the desirability of such distortions, we follow Diamond and Mirrlees (1971) and suppose there to be two production sectors, each efficient in its own production behavior. We shall refer to these as the private and public sectors, though in some cases it may be more useful to think of them both as subsectors of the private sector. The results are easily extended to several sectors.

As before, we let $f(\cdot)$ and z be the production function and output of the private sector, and introduce $g(\cdot)$ and s as the corresponding variables for the public sector. The use of distortions in the allocation of resources between the two sectors may be thought of as the direct choice of public inputs, s. Thus, the government's expanded choice problem is

$$\max_{p,s} V(p,(1-\tau)y) \quad \text{subject to} \quad f(x+R-s)=0 \quad \text{and} \quad g(s)=0,$$

$$(5.54)$$

where y is private sector profits. Attaching the Lagrange multipliers μ and ζ to the production constraints, we obtain the same first-order conditions as before with respect to p. With respect to s, we get

$$\lambda(1-\tau)\frac{\mathrm{d}y}{\mathrm{d}s_i} - \mu\left(\sum_j f_j \frac{\partial x_j}{\partial y'}(1-\tau)\frac{\mathrm{d}y}{\mathrm{d}s_i} - f_i\right) - \zeta g_i = 0. \qquad (5.55)$$

Using the normalization $q = \mathrm{d}f$ and the consumer's budget constraint, we rewrite this as

$$\lambda(1-\tau)\frac{\mathrm{d}y}{\mathrm{d}s_i} - \mu\left(-\sum_j t_j \frac{\partial x_j}{\partial y'}(1-\tau)\frac{\mathrm{d}y}{\mathrm{d}s_i} + (1-\tau)\frac{\mathrm{d}y}{\mathrm{d}s_i} - f_i\right) - \zeta g_i = 0,$$

$$(5.56)$$

or

$$g_i = \frac{\mu}{\zeta}\left(f_i - \left(\frac{\mu - \alpha}{\mu}\right)(1 - \tau)\frac{\mathrm{d}y}{\mathrm{d}s_i}\right),$$

where, as before, $\alpha = \lambda + \mu(\mathrm{d}R/\mathrm{d}y)$ is the social marginal utility of income. Thus, there are two important cases in which efficient overall production ($f_i/f_j = g_i/g_j$) will result: constant returns to scale in the private sector [Diamond and Mirrlees (1971)] and decreasing returns with 100 percent profits taxation [Stiglitz and Dasgupta (1971)]. Otherwise, inefficient production will be part of the optimal solution. The basic intuition is that as long as we can tax all but one of the commodities, we can bring about any possible configuration of relative prices consistent with a given level of revenue. When after-tax profits $(1 - \tau)y$ equal zero, these prices are the sole determinants of the consumer's decision. Thus, any attainment of a set of relative prices using a production distortion could also be obtained without one, with the simple result that the consumer could be made better off. Note that this logic only holds if all the taxes t_1 through t_N can be adjusted. With some of these held fixed, production inefficiencies may be helpful in imposing indirect taxes on the goods that cannot be freely taxed directly. We return to this point below in our discussion of tax reform.

For the case where profits are not zero, we may simplify (5.56) for the case of independent production. Considering $\mathrm{d}y/\mathrm{d}s_i$, we have (using the envelope theorem and independence assumption)

$$\frac{\mathrm{d}y}{\mathrm{d}s_i} = \sum_j z_j \frac{\mathrm{d}q_j}{\mathrm{d}z_j} \cdot \frac{\partial x_j}{\partial y'} \cdot (1 - \tau)\frac{\mathrm{d}y}{\mathrm{d}s_i} - z_i \frac{\mathrm{d}q_i}{\mathrm{d}z_i}, \tag{5.57}$$

which, using the facts that $q = \mathrm{d}f$ and $\mathrm{d}q = H$, and the assumption that all government expenditures are on the numeraire commodity ($\hat{x} = \hat{z}$), we may solve as

$$\frac{\mathrm{d}y}{\mathrm{d}s_i} = -\frac{x_i H_{ii}}{1 - (1 - \tau)\sum_j x_j H_{jj}\left(\partial x_j/\partial y'\right)} = -\frac{f_i/\sigma_{ii}}{\Gamma}, \tag{5.58}$$

where σ_{ii} is the supply elasticity for good i, and Γ must be positive for a stable solution. Thus (5.56) yields

$$\frac{g_i}{g_j} = \frac{f_i}{f_j} \cdot \left(\frac{1 + k/\sigma_{ii}}{1 + k/\sigma_{jj}}\right) \quad \text{where} \quad k > 0. \tag{5.59}$$

6. Optimal taxation and the structure for preferences

This section considers the implications of the tax formulae derived above for actual tax rates under different assumptions about the structure of preferences, and for the more general case where there are several individuals and hence distributional objectives to be satisfied. Although the results already presented expressed the optimal taxes in terms of the demands and substitution matrix of the representative consumer, these terms are not generally constant, so we have little insight into the general conditions on consumer preferences required for either uniform taxation or any other specific tax structure to be optimal. In exploring this question, we will also be able to investigate more easily the impact of distributional objectives on the optimal tax structure.

6.1. Optimal taxation from the dual perspective

To consider the role of preferences in determining optimal tax rules, it is helpful to derive such rules using the direct utility function rather than the indirect utility function. Though the derivation is less straightforward, the results are in terms of the characteristics of the utility function and, hence, preferences. This approach is taken by Atkinson and Stiglitz (1972, 1976, 1980). However, a simpler and more elegant way of arriving at their results is by transforming the optimal tax formulae themselves using duality theory. The technique described by Deaton (1979a, 1981a, 1981b) makes use of the "distance" function, sometimes referred to as the "direct" expenditure function [Cooter (1979)]. Our analysis here will generally follow that of Deaton. Because consumer preferences are defined with respect to consumption, rather than purchases, it is useful to separate these concepts by letting the vector of purchases x equal $\tilde{x} - \bar{x}$ where $\tilde{x}$ is the consumption vector and $\bar{x}$ the endowment vector. Thus, we may rewrite the indirect utility function $V(p)$, which implicitly holds $\bar{x}$ as fixed, as $V(p, p \cdot \bar{x})$, which does not. This allows us to consider the effects of changes in the consumer's lump sum income.

   In words, the distance function is the solution to the following problem: consider a consumption bundle $\tilde{x}$, and also all the combinations of price vector p and total endowment income y such that $V(p, y)$ equals (strictly speaking, at most equals) some constant utility level $\overline{U}$. Choose the vector of prices that minimizes $p^* \cdot \tilde{x}/y$, given $\tilde{x}$. The resulting value is the distance function $D(\tilde{x}, \overline{U})$. Algebraically, the problem is

$$\min_{p} (p^* \cdot \tilde{x})/y \quad \text{subject to} \quad V(p^*, y) \le \overline{U}. \tag{6.1}$$

It is explained diagrammatically in Figure 6.1, for the case of two goods. For simplicity, we assume that $\tilde{x}$ is on the indifference curve corresponding to the utility level $\overline{U}$, although only the scale of $D(\cdot)$ and not the price vector chosen would be affected by increasing or decreasing $\tilde{x}$ along the ray shown. This is easily verified from inspection of (6.1), since minimizing $(p^*, \tilde{x})/y$ is equivalent to minimizing $(p^* \cdot \lambda \tilde{x})/y$ for any $\lambda > 0$. By choosing $\tilde{x}$ to be just feasible, given $\overline{U}$, we will obtain a value $D(\tilde{x}, \overline{U}) = 1$.

The figure depicts two different combinations of p^* and y, indexed 1 and 2, that satisfy $V(p^*, y) = \overline{U}$. Since the price vector p_2^* results in a tangency away from $\tilde{x}$, purchase of $\tilde{x}$ would require a greater expenditure than y_2. This is not the case with p_1^*, since it is tangent to the indifference curve at $\tilde{x}$. (A flatter budget line would again necessitate an increase in expenditure to purchase $\tilde{x}$.) Thus, the price vector chosen, given $\tilde{x}$ and $\overline{U}$, is tangent to the indifference curve corresponding to $\overline{U}$ at point $\tilde{x}$ (or, more generally, if $\tilde{x}$ is not on the indifference curve, at the point on the indifference curve on the ray through $\tilde{x}$ from the
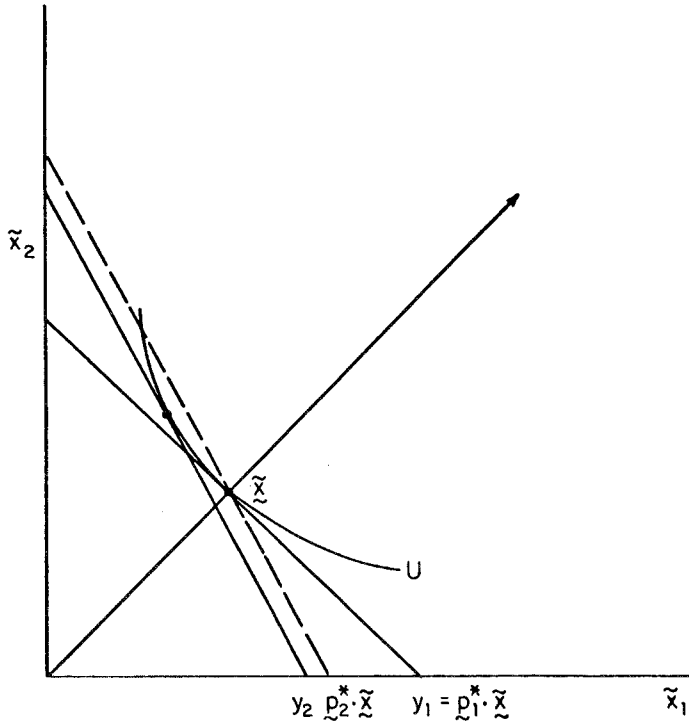

Figure 6.1. The distance function.

origin). Just as the indirect expenditure function chooses consumption, given prices and utility, the distance function chooses normalized prices, given consumption and utility. Since these prices are based on the consumer's indirect utility function, we may interpret them as points on the consumer's inverse compensated demand curve, expressing willingness to pay. By the envelope theorem, the partial derivatives of the distance function with respect to the elements of $\tilde{x}$ are those normalized inverse demands:

$$\frac{\partial D}{\partial \tilde{x}_i} = a_i(\tilde{x}, \overline{U}) = \frac{p_i^*}{y}. \tag{6.2}$$

The Hessian of the distance function is referred to as the Antonelli matrix $A = (a_{ij})$.[14]

Now, consider the actual price vector that prevails, p, and choose $\tilde{x}$ such that $\tilde{x} = \tilde{x}_C(p, U)$. Then by construction, $p^* = p$ and $y = E(p, \overline{U})$ solve (6.1), and we have the identity [from (6.2)]

$$a_i(\tilde{x}(p, \overline{U}), \overline{U}) = \frac{p_i}{E(p, \overline{U})}. \tag{6.3}$$

Multiplying (6.3) through by $E(p, \overline{U})$, and differentiating with respect to each price, we obtain conditions which can be stacked to yield

$$E(p, \overline{U}) AS = I - ax_C(p, \overline{U}), \tag{6.4}$$

where $a = (a_0, \ldots, a_N)$. Evaluating at $\overline{U} = V(p, p \cdot \tilde{x})$, this yields

$$(p \cdot \tilde{x}) AS = I - ax(p, p \cdot \tilde{x}). \tag{6.5}$$

Now, let us return to the optimal tax results described in Section 5. Multiplying both sides of (5.15) by $(p \cdot \tilde{x})A$, and using the fact that since a is homogeneous of degree zero with respect to $\tilde{x}$, $A\tilde{x} = 0$, we obtain

$$t = a(x + \overline{x})'t - \left(\frac{\mu - \alpha}{\mu}\right)(p \cdot \tilde{x}) A(x - \overline{x})$$

$$= a(R + t \cdot \overline{x}) + \left(\frac{\mu - \alpha}{\mu}\right)(p \cdot \tilde{x}) A\overline{x}, \tag{6.6}$$

---

[14]See Deaton (1979a) for further discussion of the properties of the function $D(\cdot)$ and the matrix A.

where $R = t \cdot x$ is tax revenue. Using the fact that $t_0 = 0$ to eliminate $(\mu - \alpha)/\mu$, we obtain [Deaton (1981b)]

$$\theta_i = \frac{t_i}{p_i} = \left(\frac{R + t \cdot \bar{x}}{p \cdot \bar{x}} \right) \left(1 - \frac{(A\bar{x})_i}{(A\bar{x})_0} \cdot \frac{a_0}{a_i} \right) = \nu \left(1 - \frac{\sum_j \bar{x}_j \partial \ln a_i / \partial x_j}{\sum_j \bar{x}_j \partial \ln a_0 / \partial \tilde{x}_j} \right), \qquad (6.7)$$

which, in turn, implies that, for any i and j,

$$\theta_i - \theta_j = \nu' \sum_k \bar{x}_k \frac{\partial \ln(a_j / a_i)}{\partial \tilde{x}_k}, \qquad (6.8)$$

where $\nu' = \nu / \left(\sum \bar{x}_j \mathrm{d} \ln a_0 / \mathrm{d} \tilde{x}_j \right)$. From (6.8), we see that a sufficient condition for the taxes to be the same is that the ratio of marginal valuations (a_j / a_i) be independent of the consumption of commodities in which the consumer has an endowment. This is equivalent to the distance function being separable, or capable of being expressed as

$$D(\tilde{x}, \bar{U}) = f(\tilde{x}_1, \tilde{x}_2, \bar{U}, \phi(\tilde{x}_3, \bar{U})), \qquad (6.9)$$

where $\tilde{x}_1$ are the commodities in which there is an endowment and $\tilde{x}_3$ are the goods on which taxes are uniform.[15] It also follows that the normal or indirect expenditure function is separable in the corresponding prices [Gorman (1976)]. This separability of the expenditure function is referred to as implicit separability and differs from the separability of the direct and indirect utility functions.[16] Indeed, they are the same only if the utility function is homogeneous in $\tilde{x}_3$ as well [Deaton (1981a)], and it is easy to construct counter-examples for the case where preferences are just weakly separable [Auerbach (1979a)].

In the special case where the consumer's only endowment is in the numeraire commodity, good zero (presumably leisure), the sufficient (and now necessary, as well) condition for uniform taxation of commodities is implicit separability from leisure. It is also possible in this case to say more about which goods will be taxed more heavily if weak separability but not homogenity is satisfied. We begin by

---

[15] Because $D(\cdot)$ is homogeneous of degree 1 in $\tilde{x}$, f must be homogeneous of degree 1 in $\tilde{x}_1$, $\tilde{x}_2$ and ϕ, and ϕ homogeneous of degree 1 in $\tilde{x}_3$.

[16] (Weak) separability of the direct utility function, for example, would allow the utility function $U(x)$ to be written $f(x_1, x_2, \phi(x_3))$, i.e., the marginal rate of substitution between elements of x_3 is independent of the levels of x_1 and x_2.

rewriting (6.8) as

$$\theta_i - \theta_j = \nu' \bar{x}_0 \frac{\partial \ln(a_j/a_i)}{\partial \tilde{x}_0}, \tag{6.10}$$

where $\nu' = \nu/(x_0 a_{00}/a_0)$.

By the convexity of $D(\cdot)$, ν' has the opposite sign of ν and hence is negative. Since $a_j/a_i = p_j/p_i = U_j/U_i$,

$$\frac{d \ln(U_j/U_i)}{d \tilde{x}_0} = \frac{d \ln(a_j/a_i)}{\tilde{d} \tilde{x}_0} = \frac{\partial \ln(a_j/a_i)}{\partial \tilde{x}_0} + \frac{\partial \ln(a_j/a_i)}{\partial \bar{U}} \cdot \frac{dU}{d \tilde{x}_0} \tag{6.11}$$

[Deaton (1981a)]. Thus, when utility is separable into goods and leisure, (6.10) becomes

$$\theta_i - \theta_j = -\nu' \bar{x}_0 \frac{\partial \ln(a_j/a_i)}{\partial \bar{U}} \cdot \frac{dU}{d \tilde{x}_0}, \tag{6.12}$$

so that taxes will be higher on those goods that are necessities, if these are defined by those whose valuation by the consumer declines relatively with an increase in real income. This is particularly important if we use empirical demand estimates based on restricted functional forms to estimate optimal taxes. For example, the linear expenditure system

$$\tilde{x}_i(p, p_0 \tilde{x}_0) = c_i + \left(\frac{b_i}{\sum_j b_j} \right) \cdot \frac{\left(p_0 x_0 - \sum_j p_j c_j \right)}{p_i}, \tag{6.13}$$

often used in empirical work, comes from the Stone–Geary utility function

$$U(\tilde{x}) = \prod_i (\tilde{x}_i - a_i)^{b_i}, \tag{6.14}$$

which is strongly separable, but not homogeneous unless the terms a_i equal zero (in which case it is simply Cobb–Douglas).

6.2. *Distributional objectives*

Once we allow for the presence of several individuals with different tastes or income, distributional considerations become an issue.[17] As stressed in Section 3, these considerations must be represented by the specification of an explicit social welfare function based on individual utilities. This cannot normally be achieved by the direct choice of distributional weights on individual income unless the weights are allowed to change with prices in a complicated fashion. There are two problems we consider in this subsection. First, when and how are the previously derived optimal tax rules influenced by equity considerations? Second, if we choose leisure as numeraire and admit lump sum taxes that *cannot* vary across individuals, when will uniform commodity taxes be optimal?

We begin by specifying a social welfare function of the form

$$W = W(U^1, \ldots, U^H), \tag{6.15}$$

which, maximized subject to the usual revenue constraint under the assumption of zero profits in the private sector, yields the following N first-order conditions for optimal commodity taxes $\hat{t} = (t_1, \ldots, t_N)$:

$$-\sum_h W_h \lambda^h x_i^h + \mu \left[\sum_j t_j \sum_h \frac{\partial x_j^h}{\partial p_i} + x_i \right] = 0, \qquad i = 1, \ldots, N, \tag{6.16}$$

where $W_h = \partial W / \partial U^h$, $\lambda^h = dU^h / dy^h$ and $x_i = \sum_h x_i^h$. Defining α^h, as before, to be the social marginal utility of individual h's income,

$$\alpha^h = W_h \lambda^h + \mu \frac{dR}{dy^h}, \tag{6.17}$$

we may express the conditions (6.16) as

$$\sum_j t_j S_{ij} = -\left(\frac{\mu - \tilde{\alpha}_i}{\mu} \right) x_i, \qquad i = 1, \ldots, N, \tag{6.18}$$

where $S_{ij} = \sum_h S_{ij}^h$ and

$$\tilde{\alpha}_i = \sum_h \left(\frac{x_i^h}{x_i} \right) \alpha^h \tag{6.19}$$

is the average value of α, weighted by individual consumption shares of good i.

---

[17] Indeed, even if all individuals are identical, the optimal tax system need not dictate identical treatment. This is discussed in Section 7.

This neat formulation [due to Diamond (1975)] shows that the "equal propor-
tional reduction" rule is amended to call for a greater proportional reduction in
the purchase of commodities for which $\tilde{\alpha}_i$ is small. The implication of this result is
more clearly seen if we note [following Feldstein (1972)] that

$$\tilde{\alpha}_i = \text{cov}\left(\left(\frac{x_i^h}{x_i}\right), \alpha^h\right) + \frac{1}{H}\sum_h \alpha^h, \tag{6.20}$$

so that $\tilde{\alpha}_i$ exceeds the unweighted mean of α^h if and only if purchases of
commodity i are positively correlated with α over individuals. Normally, this
would define a necessary good, whose budget shares fall with income and hence
rise with α. Note, however, that (6.18) applies to proportional reductions in
purchases of different commodities, and does not offer an explicit solution for
individual tax rates, unless we assume aggregate commodity demands to be
independent ($S_{ij} = 0$ for $i \neq j$). This yields

$$\frac{\theta_i}{\theta_j} = \frac{\varepsilon_{jj}}{\varepsilon_{ii}} \cdot \left(\frac{\mu - \tilde{\alpha}_i}{\mu - \tilde{\alpha}_j}\right), \tag{6.21}$$

which says that the normal inverse elasticity rule is changed by the addition of a
second term expressing distributional concerns. Note that as marginal excess
burden, and hence the size of μ relative to $\tilde{\alpha}$, increases, efficiency considerations
come to dominate these optimal tax rules [Feldstein (1972)].

The addition of the possibility of lump sum taxation increases the generality of
the problem without much additional complexity. If individuals have one source
of income, then the combination of N commodity taxes and a lump sum tax may
be thought of as a linear income tax plus $N - 1$ additional commodity taxes. The
ability to use lump sum taxation simply adds a constant tax term T to each
consumer's indirect utility function and a term HT to the revenue constraint.
Differentiating the expanded Lagrangian with respect to T, we obtain the ad-
ditional first-order condition

$$-\sum_h W_h \lambda^h + \mu\left[\sum_j t_j \sum_h \frac{\partial x_j^h}{\partial y^h} - H\right] = 0, \tag{6.22}$$

to be added to the N conditions in (6.16). This new condition simplifies to

$$\mu = \frac{1}{H}\sum_h \alpha^h = \bar{\alpha}. \tag{6.23}$$

thus, (6.18) becomes

$$\sum_j t_j S_{ij} = \frac{\text{cov}\left(\left(\frac{x_i^h}{x_i}\right), \alpha^h\right)}{\bar{\alpha}} x_i, \qquad i = 1, \ldots, N. \tag{6.24}$$

Now, there should be reductions in commodity purchases only to the extent that the good in question is consumed relatively more by people with low values of α; purchases of some goods will increase. With equal distributional weights, α^h, each of these reductions would be zero, and hence pure lump sum taxation would be optimal.

An interesting question to ask here is under what conditions proportional taxes $\theta = (t_1/p_1, \ldots, t_N/p_N)$ will be equal? In other words, since such uniform taxes are equivalent to a single, proportional tax on the numeraire, labor, when is a linear income tax optimal? A sufficient condition [Deaton (1979b)] is that each individual h have a utility function weakly separable into goods and leisure, with the subfunction in goods possessing linear Engel curves with common slopes across individuals. The intuition behind this result is that the restriction on goods is that preferences obey the Gorman polar form required for exact aggregation of commodity demands. If we can perform such aggregation, then we cannot use differential taxation to distinguish among individuals for purposes of redistribution: a linear income tax exhausts our capacity in this regard.

Note the similarity of this result to that of the case of *non*-linear income taxation [Atkinson and Stiglitz (1976)], where weak separability alone is sufficient for the optimality of income taxation. There is a clear relationship here between the relaxation of the restriction on the linearity of taxes, on the one hand, and that of the linearity of preferences, on the other.

Empirical studies of optimal taxation are not very common, perhaps because the information needed concerning various cross-substitution terms is difficult to obtain without a restriction on preferences that prejudges the result. Two studies, by Atkinson and Stiglitz (1972) and Deaton (1977), utilize the linear expenditure system, which calls for higher taxes on necessities in the single-consumer case (as discussed above) and, in the multi-consumer case with lump sum taxes available, calls for no differential commodity taxes at all, since the Gorman conditions are satisfied. Nevertheless, these calculations are still instructive. Deaton, for example, calculates the optimal taxes on commodities under the assumption that labor is fixed and there are no lump sum taxes. Obviously, with fixed labor supply, uniform taxes on commodities are non-distortionary, but may have undesirable distributional effects. For a demand system estimated for the U.K., he calculated optimal tax rates for eight groups of commodities under various assumptions about the degree of inequality in the social welfare function. Perhaps the most

interesting result obtained was that optimal tax rates do not behave monotonically with respect to the degree of inequality aversion implicit in the social welfare function.

A recent application of the optimal tax results in the context of developing countries (India) may be found in Heady and Mitra (1982). Still another approach has been to infer from an existing indirect tax structure what the government's preferences would have to be for the structure to be optimal [Christiansen and Jansen (1978) for Norway, Ahmad and Stern (1981) for India].

7. Further topics in optimal taxation

There are a number of particular problems involving taxation generally to which optimal tax theory has been applied. This section presents some of these.

7.1. Public goods provision

The classic conditions for efficiency in the provision of public goods were derived by Samuelson (1954). Aside from the standard requirement that, for private (rival) goods, each consumer's marginal rate of substitution between two goods should equal the social marginal rate of transformation, there was the new condition that, between a private and a public good, the marginal rate of transformation should equal the *sum* of individual marginal rates of substitution. This is because every consumer partakes of each additional unit of the public good.

Pigou (1947) argued that in considering the benefits of a new public project, the government should recognize that its undertaking may require the introduction of additional deadweight loss through the tax system. The implication that this increases the social cost of public goods has been addressed by a number of authors, including Diamond and Mirrlees (1971), Stiglitz and Dasgupta (1971) and Atkinson and Stern (1974).

Even to examine the question of public goods, we must allow for the presence of several individuals. Since we are not directly interested in distributional issues here, we assume all H individuals to be identical in all respects. If we let G be a public good on which all government revenue is spent and which all consume, then each individual's indirect utility function becomes

$$V(p; G) = \max_{x} U(x; G) \quad \text{subject to} \quad p \cdot x = 0, \tag{7.1}$$

with $\partial V/\partial G = \partial U/\partial G|_{x=x_C(p; G)}$. The production function is $f(x; G) = 0$. The

government maximizes the welfare of the representative individual by maximizing the sum of individual utilities, since all individuals are the same. This gives rise to the Lagrangian

$$L = HV(p; G) - \mu f(x; G),$$ (7.2)

with first-order conditions with respect to each price (except that of the untaxed numeraire)

$$-H\lambda x_i^h - \mu \sum_j f_j \frac{\partial x_j}{\partial p_i} = 0, \qquad i = 1, \dots, N,$$ (7.3)

where λ and μ are defined in the usual way. As in Section 5.3, we use the fact that $p \cdot x^h = 0$ for each individual h to obtain

$$\lambda x_i + \mu \left[\sum t_j \frac{\partial x_j}{\partial p_i} + x_i \right] = 0, \qquad i = 1, \dots, N,$$ (7.4)

where

$$x_i = \sum_h x_i^h = H x_i^h.$$

As before, this may be rewritten as

$$St = -\left(\frac{\mu - \alpha}{\mu} \right) x,$$ (7.5)

where S is the aggregate Slutsky matrix and α is the social marginal utility of each individual's income.

The first-order condition with respect to the choice of public good G is

$$H \frac{\partial U}{\partial G} - \mu \left[\sum_i f_i \frac{\partial x_i}{\partial G} + f_G \right] = 0,$$ (7.6)

which yields (since $\lambda \equiv dU/dx_0^h$, $q_i \propto f_i$, $q_0 = 1$ and $p \cdot x^h = 0$)

$$\sum_h \frac{dU^h/dG}{dU^h/dx_0} = \left(\frac{\mu}{\lambda} \right) \left(\frac{f_G}{f_0} - \frac{dR}{dG} \right),$$ (7.7)

where R is the revenue collected (equal to the public goods purchased, in

equilibrium). This result says that the appropriate social cost of the public good G in terms of the numeraire, x_0, to which the sum of marginal rates of substitution should be set equal, differs from the marginal rate of transformation f_G/f_0 for two reasons. First, if public goods are complementary to taxed goods, increasing G may reduce excess burden by increasing consumption of taxed goods, making $dR/dG > 0$ [Diamond and Mirrlees (1971)]. The other term μ/λ, equals the ratio of the marginal disutility of raising a dollar of revenue divided by the marginal utility of income, and exceeds one to the extent that an increase in revenue increases excess burden. This corresponds to the point raised by Pigou. However, it need not be the case that μ/λ exceeds one. Again, there is an income effect at work.

This possibility is demonstrated (following Atkinson and Stern) by multiplying both sides of (7.5) by the vector t to obtain

$$ t'St = -\left(\frac{\mu - \alpha}{\mu} \right) R, \tag{7.8} $$

which, by the negative semi-definiteness of S, implies that $\mu > \alpha$ for positive revenue. But $\alpha \geq \lambda$ [see equation (5.5)] only if dR/dy is positive. If taxed goods are, on average (weighted by tax rates) inferior, $dR/dy < 0$ and $\lambda > \alpha$. Hence, λ may actually exceed μ, meaning that raising an additional dollar to pay for public goods may actually lessen excess burden by causing a shift toward the consumption of taxed goods.

7.2. Externalities

Referring again to Pigou, we know that the appropriate response by the government (under conditions of perfect information) to an externality is the imposition of a tax that causes producers of the externality to internalize the additional social cost (or benefit) of their action. Suppose, however, that all commodities, including the one possessing the externality, are subject to distortionary taxation. How is the Pigouvian prescription affected? Following Sandmo (1975), we assume identical individuals, fixed producer prices and let the externality be a symmetric consumption externality related to total consumption of good N. Thus, individual utility for the representative individual h is $U(x^h; x_N)$, where $x_N = Hx_N^h$. The partial derivative of U with respect to x_N may be positive or negative. Assuming for convenience that each individual takes x_N as given (as will be approximately true for H large), we may express the corresponding indirect utility function as $V(p; x_N)$, parallel to the public good example, with $\partial V/\partial x_N = \partial U/\partial x_N|_{x_C(p; x_N)}$.

Maximizing the sum of utilities with respect to p subject to the need to raise revenue R through distortionary taxes yields the N first-order conditions

$$-\lambda x_i + H \frac{\partial U}{\partial x_N} \cdot \frac{\partial x_N}{\partial p_i} + \mu \left[x_i + \sum_j t_j \frac{\partial x_j}{\partial p_j} \right] = 0, \quad i = 1, \dots, N, \tag{7.9}$$

or

$$-\lambda x_i + \mu \left[x_i + \sum_j t_j^* \frac{\partial x_j}{\partial p_i} \right] = 0, \quad i = 1, \dots, N, \tag{7.10}$$

where

$$t_i = t_i^*, \qquad\qquad i = 1, \dots, N-1,$$
$$= t_i^* - H \frac{\partial U}{\partial x_N} / \mu, \quad i = N.$$

Equation (7.10) is the standard optimal tax result, but it applies to the vector t^* rather than t. The difference between them implies that the optimal tax on good N equals that dictated by the standard formula plus the externality imposed by additional consumption of the good: the Pigouvian tax. Thus, the optimal tax and Pigouvian taxes are separable, in a sense; we may imagine choosing the two independently. However, this independence is only present analytically, since the actual level of the externality, and hence the Pigouvian tax, depends on the actual equilibrium and hence the optimal tax rates; the same is true in the other direction.

7.3. Pre-existing distortions

If the government faces pre-existing distortions (of which the preceding example of externalities is a specific kind), it may wish to alter its choice of optimal taxes. Following Green (1961), let us assume that lump sum taxes are available, but certain prices are distorted and cannot be influenced directly. This could be the result of non-competitive behavior, but we shall assume it to be due to some tax that must be maintained, perhaps for political purposes. Assuming that the representative individual's only lump sum income is from the government, we have the problem

$$\max_{p^*, T} V(p, -T) \quad \text{subject to} \quad (p - q) \cdot x + T = R, \tag{7.11}$$

where p^* is the subset of p that may be adjusted. Note that unless at least two prices are fixed, equiproportional, non-distortionary taxation is possible.

Differentiating the Lagrangian corresponding to (7.11) with respect to p_i and T yields

$$-\lambda x_i + \mu \left[\sum_j t \frac{\partial x_j}{\partial p_i} + x_i \right] = 0, \qquad \forall p_i \in p^*, \tag{7.12a}$$

$$-\lambda + \mu \left[-\sum_j t_j \frac{\partial x_j}{\partial y} + 1 \right] = 0, \tag{7.12b}$$

which may be written as

$$\sum_j S_{ij} t_j = -\left(\frac{\mu - \alpha}{\mu} \right) x_i, \qquad \forall p_i \in p^*, \tag{7.13a}$$

$$\mu = \alpha, \tag{7.13b}$$

for α defined as above. These conditions are quite familiar, and yield the requirement that

$$\sum_j S_{ij} t_j = 0, \qquad \forall p_i \in p^*. \tag{7.14}$$

This does not result in uniform taxes unless at most one tax is fixed (in which case the zero degree homogeneity of S allows us to choose any level of proportional taxes). In particular, suppose all taxes but t_1 are fixed, and $t_0 = t_3 = \cdots = t_N = 0$. Then there is one condition, corresponding to the choice of t_1. Using compensated elasticities $\varepsilon_{ij} = S_{ij}(p_j/x_i)$, we may express this as

$$\theta_1 = -\theta_2 \varepsilon_{12}/\varepsilon_{11}, \tag{7.15}$$

where $\theta_i = t_i/p_i$ is the proportional tax on good i. Since $\varepsilon_{11} < 0$, this calls for a tax on good 1 (assuming $\theta_2 > 0$) if $\varepsilon_{12} > 0$, and a subsidy if $\varepsilon_{12} < 0$. If the distorted good is a substitute to good 1, a tax on good 1 will shift consumption into good 2, lessening the original distortion. Taxing a complement, however, would worsen the distortion. (Compare butter and margarine vs. left shoes and right shoes.)

In the wider case in which there are several pre-existing distortions and a single free instrument, t_1, the condition is

$$\theta_1 = - \sum_{j \neq 1} \theta_j \varepsilon_{1j} / \varepsilon_{11}, \tag{7.16}$$

so that the complement–substitute rule now applies to the tax-weighted commodity average. More generally, when several instruments can be set, the results are more complicated.

Several other authors have considered particular restrictions on commodity taxation and profits taxation [for example, Dasgupta and Stiglitz (1972) and Mirrlees (1972)] and the effect of such restrictions on the desirability of production efficiency. Auerbach (1979b) considers the particular production distortion of differential capital income taxation, obtaining a uniform taxation result about separability of factors in production that closely parallels those on the consumption side already discussed in Section 6.

7.4. Taxation and risk

There are many interesting questions that concern the interaction between taxes and risk-bearing. A particular one that fits into the current discussion is the optimal taxation of risky assets. This problem was first examined by Stiglitz (1972) and extended by Auerbach (1981). The basic insight is that the optimal tax results already derived can be applied directly to the case of risky assets by imagining the commodities being taxed to be Arrow–Debreu state-contingent ones. The differences that arise come from the fact that we normally make different assumptions about the structure of utility functions and the completeness of markets when we deal with risk.

The basic model we consider, following Stiglitz (1972), is a two-period model in which the representative individual may consume a certain good (leisure) out of some endowment, and may purchase one of two linearly independent assets yielding returns in two states at date 1. Because the two assets span the states of nature, the consumer may purchase any combination of state-contingent commodities at date 1, and there is a well-defined implicit price for each. A corollary of this is that there is a unique pair of tax rates on commodities in the two states corresponding to each tax regime that applies to the assets themselves. This is helpful, because though our optimal tax results apply to the former, actual tax rules normally apply to the latter. In the more general case without asset spanning, the optimal tax problem becomes more complicated, just as it would if individual commodities in a riskless world could not be purchased independently. Stiglitz (1972) obtained his main result concerning the relative taxation of a risky

and a riskless asset from a direct consideration of the effects of taxation on asset demands. It is, perhaps, easier to see the connection with previous results, and the effects of particular assumptions, if we begin with the state-contingent commodities themselves [following Auerbach (1981)].

Letting the good consumed in period 0 be good zero, and the other two commodities be labelled 1 and 2, and taking good 0 to be numeraire, we have the basic optimal tax rule (5.21), which we write here for convenience

$$\frac{\theta_1}{\theta_2} = \frac{\varepsilon_{12} + \varepsilon_{21} + \varepsilon_{20}}{\varepsilon_{12} + \varepsilon_{21} + \varepsilon_{10}}. \tag{7.17}$$

This result can be simplified if we adopt the axioms necessary for the consumer to engage in expected utility maximization. In this case, the consumer's objective function becomes

$$U(x_0, x_1, x_2) = \pi_1 U^1(x_0, x_1) + \pi_2 U^2(x_0, x_2), \tag{7.18}$$

where $U^1(\cdot) = U^2(\cdot)$, π_i is the possibility of state i occurring, and ε_{10} and ε_{20} may be expressed as

$$\varepsilon_{i0} = M\left[-\left(\frac{U_{22}^j x_j}{U_2^j}\right) + p_j x_j \frac{\mathrm{d}\ln(U_2^j/U_2^i)}{\mathrm{d}x_0} \right], \qquad i = 1, 2, \quad j = 2, 1, \tag{7.19}$$

where M is a positive constant and U_i and U_{ij} are first and second derivatives of utility. The second term in brackets in (7.19) is familiar from Section 6, and equals zero if preferences are weakly separable between periods. If this is so (in which case, utility is also strongly separable, since it is already assumed separable between states), then the tax on good 1 should be higher than that on good 2 if and only if $-(U_{22}^1 x_1/U_2^1) > -(U_{22}^2 x_2/U_2^2)$, but these are just the Arrow (1965)–Pratt (1964) measures of relative risk-aversion in the two states. Intuitively, as an individual becomes more risk-averse, his behavior becomes less responsive to differences in rates of return. Thus, a tax is less distortionary.

That taxes should be equal when relative risk aversion is constant is not suprising, even without knowledge of the basic optimal tax results. It is for this class of preferences that the basic results of Samuelson (1969) and Merton (1969) concerning the separation of portfolio and savings decision apply. If we cannot influence the amount of savings, and hence leisure consumed, by inducing portfolio shifts, then such a relative distortion has no benefit.

To convert these results to the taxes on the two assets themselves, which we label A and B, we use the fact [see Auerbach (1981)] that

$$\mathrm{sgn}(\theta_A - \theta_B) = \mathrm{sgn}(r_A^1 r_B^2 - r_A^2 r_B^1)\mathrm{sgn}(\theta_1 - \theta_2), \tag{7.20}$$

where r_j^i is the return in state i of asset j. Assuming one asset, which we take to be asset A without loss of generality, is risk-free, then the tax should be greater (smaller) on the risky asset B if relative risk-aversion is higher (lower) in the state with the higher (lower) return. In other words, the risky asset should face a higher or lower tax than the safe asset according to whether relative risk-aversion is increasing or decreasing [Stiglitz (1972)]. More generally, if both assets are risky, then one can apply any standard notion of increasing risk [Rothschild and Stiglitz (1970)] to argue that if asset B is riskier than asset A, its return will be more dispersed and hence $(r_A^1 r_B^2 - r_A^2 r_B^1)$ will be positive. This will yield a similar result for taxation of the *riskier* asset.

It is important to recognize that these results assume complete, competitive markets. While a common assumption without risk, it is less acceptable when the commodities concerned are state-contingent. (The same critique also applies to intertemporal problems with date-indexed goods.) In particular, we are implicitly assuming that the government cannot increase the diversification of risk by collecting risky taxes and pooling them. In a real world context where many assets are not traded, this may be a highly questionable restriction to impose.

A second issue of taxation and risk concerns the question of whether the government can increase the welfare of the representative individual by *inducing* risk through the tax system. Normally, risk averse individuals are made worse off by being forced to bear risk. However, the optimal taxation equilibrium is a distorted one, and the famous dictum of Lipsey and Lancaster (1956–57) applies here: once one condition for a Pareto optimum is violated, there is no reason to expect that the violation of others will necessarily worsen matters.

There are two general strands in the literature that deal with the use of induced risk as a policy tool. Weiss (1976) and Stiglitz (1982) show that a random tax system, or one in which there is tax evasion with a probability of detection, may be superior to a certain tax system because, under specified conditions with respect to individual preferences, such risk may lessen the labor supply distortion of the income tax. [Also see Sandmo (1981) on the subject of tax evasion.]

A second issue relates to the case of several individuals, and arises from the possibility that in the presence of indirect taxation, the utility possibility frontier may be non-convex. Even with identical individuals, then, we might wish to tax the consumption of the same good by different individuals at different rates [Atkinson and Stiglitz (1976, 1980), Stiglitz (1982), Balcer and Sadka (1982)]. This is depicted in Figure 7.1. Suppose two individuals, 1 and 2, have identical preferences and consume goods and leisure. If we seek to maximize $(U_1 + U_2)$ by choosing individual-specific excise taxes on consumption, the first-order condition will be zero with equal taxes at $U_1 = U_2 = U^E$, by the symmetry of the problem. But this may represent a local minimum, as shown. Social welfare may be improved by choosing either point A or point B. This represents an unequal treatment of equal individuals and may violate proscriptions of such horizontal inequity. However, suppose the tax system were randomized so that point A were
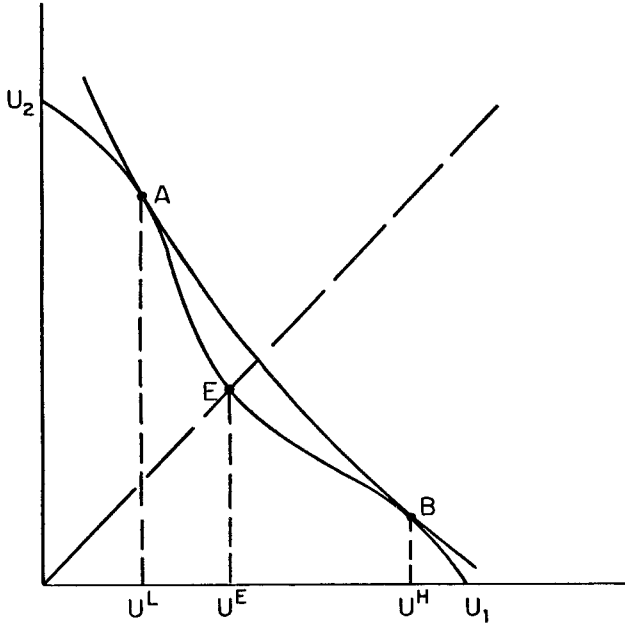
Figure 7.1. Optimal taxation and non-convexities.

chosen half the time, and point, B the other half. This would give the same expected utility to each individual. Moreover, it would yield the same value of the social welfare function, defined on individual expected utilities, as before at either A or B,

$$EU_1 + EU_2 = \tfrac{1}{2}[U^L + U^H] + \tfrac{1}{2}[U^H + U^L] = U^H + U^L. \tag{7.21}$$

Thus, randomization may be desirable.

8. Tax reform

All of the optimal tax problems analyzed thus far share in common the fact that global optima are sought. There are a number of new issues arising from a consideration of tax *reform*, rather than tax design.

One problem of tax reform derives from the existence of an initial allocation. Though a new tax system may be more efficient and more equitable than the existing one, the transition from old to new may cause a redistribution of

resources to occur than in itself is undesirable. For example, it has often been suggested in the U.S. that the tax subsidy for state and municipal bonds be removed. If this were done unexpectedly, it would cause a capital loss for the holders of such bonds, but not for other, otherwise identical individuals. Such treatment may be thought of as a violation of horizontal equity [Feldstein (1976)] which may be explicitly accounted for in an expanded social welfare function [King (1983a)]. This problem undoubtedly is one of the reasons why tax reform is so difficult to achieve.

A second general problem of tax reform, which shall be the main focus of this section, is that the *direction* in which to move from the current system is not always evident. Even if all distortions can be reduced somewhat, this may not increase economic efficiency. The basic difficulty is that we can only be sure that movement in the direction of a global optimum will improve matters if we are sufficiently close to that optimum initially. A related problem is whether one can increase economic efficiency in a piecemeal fashion, by removing distortions one at a time. In general, such a scheme for tax reform may decrease welfare along the transition path to a global optimum. Restrictions on preferences and production sufficient to prevent this are extremely restrictive [Boadway and Harris (1977)].

8.1. Moving to lump sum taxation

Lump sum taxes are non-distortionary, but it need not follow that partially reducing distortions and replacing them with lump sum taxes will improve efficiency. One case in which it *will* is when the distortionary tax rates are set at each point of the transition at the optimal tax rates for the revenue being collected by non-lump sum taxes. That is, if a certain amount of revenue, R, is collected initially by the distortionary taxes, and a lump sum tax T is introduced, the new taxes should be those optimal for collecting $R - T$. As T increases, this sequence of optimal tax rates insures a monotonic increase in utility. This result is due to Atkinson and Stern (1974), and demonstrated as follows. Consider the optimal tax problem

$$\max_{T,p} V(p, -T) \quad \text{subject to} \quad (p-q)\cdot x + T \geq R, \tag{8.1}$$

where T is the lump sum tax faced by the individual. Differentiating the corresponding Lagrangian with respect to T yields the effect of an increase in T on utility, given that p is chosen optimally,

$$\frac{d\mathscr{L}}{dT} = -\lambda + \mu\left[-\sum t_j \frac{\partial x_j}{\partial y} + 1\right] = (\mu - \alpha), \tag{8.2}$$

where λ, α and μ are defined in the usual way to be the marginal utility of income, the social marginal utility of income and the Lagrange multiplier on the revenue constraint. However, we know from expression (7.8) that $\mu > \alpha$, so utility must increase as T does: when the tax vector t is chosen optimally, there is always a positive marginal excess burden to revenue collection.

Unfortunately, this is not a very realistic assumption to make in the current context. The taxes we may wish to reform may cause unnecessarily large distortions, and we may be restricted to a proportional reduction formula, or some other constraint on how they are to be lowered.

Consider the case of an arbitrary change in the levels of excise taxes t and lump sum taxes T for the case of a single individual and fixed producer prices. [This latter assumption can be relaxed; see Dixit (1975).] We have [following Atkinson and Stiglitz (1980)]

$$dU = \sum_i \frac{\partial V}{\partial p_i} dt_i - \frac{\partial V}{\partial y} dT = -\lambda(x \cdot dt + dT),$$

(8.3a)

and

$$dR = d(t \cdot x + T) = x \cdot dt + t \cdot dx + dT = 0,$$

(8.3b)

which yields

$$dU = \lambda t \cdot dx.$$

(8.4)

Utility is increased by the tax change if consumption changes to increase revenue from the existing taxes, thereby reducing the associated excess burden.

From the Slutsky equation, we have

$$dx = \frac{\partial x}{\partial p} dt - \frac{\partial x}{\partial y} dT = S dt - \frac{\partial x}{\partial y}(x' dt + dT),$$

(8.5)

which, combined with (8.3b) and (8.4), yields

$$dU = \frac{\lambda}{1 - t \cdot \partial x / \partial y} \cdot t' S dt.$$

(8.6)

This holds for any change in t and T, and can be useful in analyzing particular kinds of tax reforms. For example, suppose all distortions are reduced proportionally, i.e., $dt = -bt$. Then because S is negative semi-definite, $dU \geq 0$ if and only if $(1 - t \cdot \partial x / \partial y) \geq 0$ [Dixit (1975)]. This condition says that a dollar increase in income causes the consumer to pay less than a dollar in additional excise taxes.

Since $p = q + t$, it is equivalent to the requirement that $q \cdot x$ increase with y: as the consumer spends more, the *social* cost of the goods purchased also increases. If this condition is violated, then it is possible that multiple equilibria exist, and the tax reduction may move the economy away from the undistorted optimum [Foster and Sonnenschein (1970)].

This may be demonstrated graphically [following Hatta (1977)] for the simple case in which there are only two goods. Suppose that a certain revenue R (measured in units of commodity 1) must be raised, and that the consumer has an endowment $\bar{x}_1$. The possible equilibria lie along the social production constraint M in Figure 8.1. Superimposed on this constraint are a series of indifference curves, the highest feasible one passing through point A, the undistorted optimum. Normally, we would expect that as we travel along M from point A toward either axis, decreasing the feasible utility level, the marginal rate of substitution between x_1 and x_2 changes monotonically. (This is true, of course, for movements along an indifference curve and, hence, for *local* movements away from A along M, where there is no first-order income effect.) If this is the case, then a
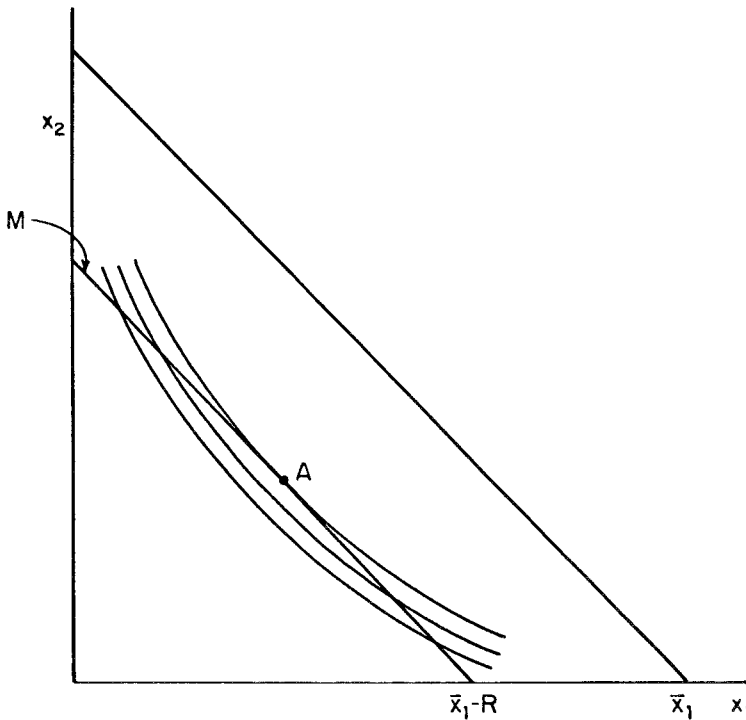

Figure 8.1. Prices and utility.

revenue-preserving reduction in the divergence between the relative price of x_2 and its social cost, in terms of x_1, must increase utility, for it will induce a movement along M toward point A. However, there may be cases in which there is no such monotonicity, and a given slope may occur at an odd number of different points on M, not just one. In this case, reductions in the price distortion may actually move the consumer *away* from point A.

That this possibility is equivalent to the condition derived from (8.6) is demonstrated graphically in Figure 8.2, where an increase in lump sum income above $\bar{x}_1$ causes the consumer to shift from point B to point C, inside the production constraint M. Since the indifference curve slopes at B and C are the same, the slope at D must be flatter than at B. Thus, a steepening of the consumer's budget line resulting from a reduction in the price distortion will cause a movement away from B, along M, toward the x_1 axis rather than toward D and A, thereby lowering the consumer's utility.

A particular application of this result is that when equilibrium is unique, a consumption tax is superior to a wage tax in the presence of pure rents, since the
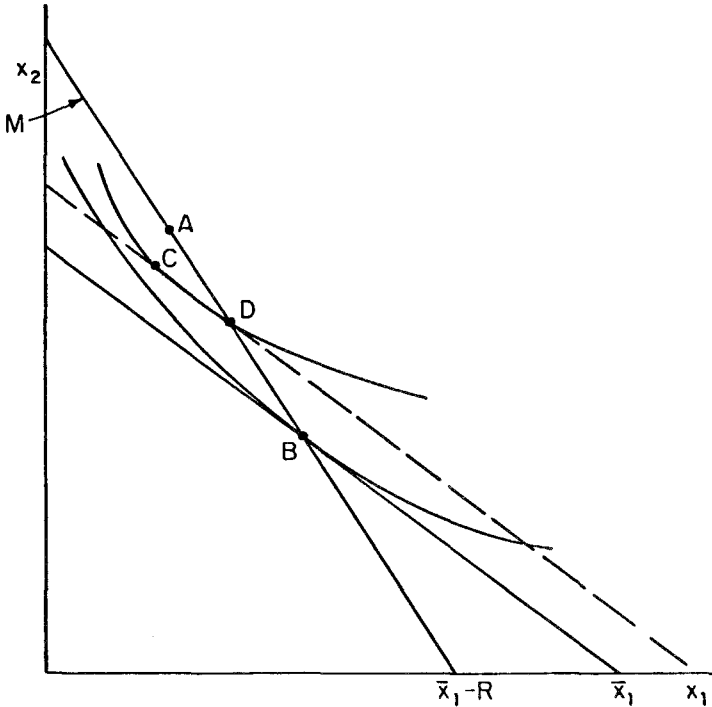

Figure 8.2. Multiple equilibria with taxes.

former tax is equivalent to the latter in conjunction with a lump sum rent tax [Helpman and Sadka (1982)].

Another result that follows from (8.6) is for the case where the tax distortion is zero for one good (arbitrarily, good zero) and equiproportional for other goods. That is, in our previous notation, $\hat{t} = \theta \hat{p}$. Since $p'S = 0$, (8.6) may be rewritten as

$$\mathrm{d}U = -\frac{\lambda}{1 - t \cdot \mathrm{d}x/\mathrm{d}y} S_0 \cdot \mathrm{d}\hat{t}, \tag{8.7}$$

where $S_0 = (S_{01}, S_{02}, \ldots, S_{0N})$. A sufficient condition for this to be positive (assuming $\mathrm{d}R/\mathrm{d}y < 1$) is that taxes be decreased on substitutes for good zero ($S_{0i} > 0$) and increased on complements [Dixit (1975)].

8.2. Reform without lump sum taxation

This problem is harder, because there is no obvious "first-best" looming in the distance to guide our movement. General characterization of the direction in which taxes should be changed is a difficult problem, and while progress has been made [Guesnerie (1977), Diewert (1978)], there is little we can say of a concrete nature without further assumptions.

One approach that sidesteps this problem is to characterize observable changes in equilibrium that will result *if* welfare is improved. Following Pazner and Sadka (1981), we can use revealed preference theory to evaluate a balanced budget change in distortionary taxes. Let $t_0 = p_0 - q$ be the initial set of taxes (with producer prices fixed) and $t_1 = p_1 - q$ be the prospective change. If $p_1 \cdot x_1 > p_1 \cdot x_0$ (where x_0 and x_1 are the purchases in the two situations), then x_1 is preferred by the consumer. Hence, utility has increased. However, since $\mathrm{d}(t \cdot x) = 0$, $q \cdot x_1 = q \cdot x_0$, so that $t_1 \cdot x_1 > t_1 \cdot x_0$, or $t_1 \cdot \Delta x > 0$. [Note the similarity of this discrete condition to (8.4).] Likewise, if $t_0 \cdot \Delta x < 0$, the original situation is preferred. Unfortunately, there is an indeterminate range in which neither of these conditions is satisfied.

If we assume producer prices to be fixed (here this restriction is necessary) and that all goods but the numeraire are taxed uniformly, then we can characterize a utility increasing tax change. The three-good case was analyzed by Corlett and Hague (1953–54), with a generalization provided by Dixit (1975), whose analysis we follow. Note that (8.4) still is valid in determining whether a tax change increases utility. However, since lump sum taxes are unavailable, $t \cdot x = 0$. Using (8.5), for $\mathrm{d}T = 0$, we have

$$0 = \mathrm{d}(t \cdot x) = x \cdot \mathrm{d}t + t \cdot \mathrm{d}x = x \cdot \mathrm{d}t + t'\frac{S}{\Delta}\mathrm{d}t = \left(x' + t\frac{S}{\Delta}\right)\mathrm{d}t, \tag{8.8}$$

where $\Delta = 1 - t \cdot \partial x / \partial y$. For the case where $t_0 = 0$ and $\hat{t} = \theta \hat{p}$, we use the homogeneity of S to rewrite this as

$$\left(x' - \frac{\theta}{\Delta} S'_0 \right) \mathrm{d}t = 0, \tag{8.9}$$

which, using the definition of compensated elasticities $\varepsilon_{ij} = S_{ij}(p_j / x_i)$, may be written

$$\sum_i x_i \left(1 - \frac{\theta}{\Delta} \varepsilon_{i0} \right) \mathrm{d}t_i = 0.$$

From (8.4), we have (for $\mathrm{d}T = 0$)

$$\mathrm{d}U = -\lambda x \cdot \mathrm{d}t. \tag{8.10}$$

If we assume that

$$x_i \left(1 - \frac{\theta}{\Delta} \varepsilon_{i0} \right) = \frac{\mathrm{d}R}{\mathrm{d}t_i}$$

is positive, and make the related assumption that Δ is positive, then [comparing (8.9) and (8.10)], in changing two taxes, we should decrease the one for which

$$\frac{x_i \left(1 - \dfrac{\theta}{\Delta} \varepsilon_{i0} \right)}{x_i} \tag{8.11}$$

is smaller, or ε_{i0} is larger – increase the tax on the relative complement. This extends in an obvious way if we choose pairs of taxes successively.

References

Ahmad, E. and N. Stern, 1981, On the evaluation of indirect tax systems: An application to India, Feb. (University of Warwick, Coventry).

Arrow, K., 1965, Essays in the theory of risk-bearing (North-Holland, Amsterdam).

Atkinson, A.B. and N.H. Stern, 1974, Pigou, taxation and public goods, Review of Economic Studies 41, 119–128.

Atkinson, A.B. and N.H. Stern, 1980, Taxation and incentives in the UK, Lloyds Bank Review, 43–46.

Atkinson, A.B. and J.E. Stiglitz, 1972, The structure of indirect taxation and economic efficiency, Journal of Public Economics 1, 97–119.

Atkinson, A.B. and J.E. Stiglitz, 1976, The design of tax structure: Direct versus indirect taxation, Journal of Public Economics 6, 55–75.

Atkinson, A.B. and J.E. Stiglitz, 1980, Lectures on public economics (McGraw-Hill, New York).

Auerbach, A.J., 1979a, A brief note on a non-existent theorem about the optimality of uniform taxation, Economics Letters 3, 49–52.

Auerbach, A.J., 1979b, The optimal taxation of heterogeneous capital, Quarterly Journal of Economics 93, 589–612.

Auerbach, A.J., 1981, Evaluating the taxation of risky assets, HIER working paper 857.

Auerbach, A.J. and H. Rosen, 1980, Will the real excess burden please stand up?, HIER working paper 767.

Auerbach, A.J., L.J. Kotlikoff and J. Skinner, 1983, The efficiency gains from dynamic tax reform, International Economic Review 24, 81–100.

Balcer, Y. and E. Sadka, 1982, Horizontal equity, income taxation and self-selection with an application to income tax credits, Journal of Public Economics 19, 291–309.

Baumol, B. and D. Bradford, 1970, Optimal departures from marginal cost pricing, American Economic Review 60, 265–283.

Boadway, R. and R. Harris, 1977, A characterization of piecemeal second best policy, Journal of Public Economics 8, 169–190.

Boiteux, M., 1951, Le "revenue distribuable" et les pertes économiques, Econometrica 19, 112–133.

Boiteux, M., 1956, Sur le gestion des monopoles publics astreints à l'equilibre budgetaire, Econometrica 24, 22–40. Translated and reprinted in 1971, Journal of Economic Theory 3, 219–440.

Bradford, D. and H. Rosen, 1976, The optimal taxation of commodities and income, AEA Papers and Proceedings 66, 94–101.

Chamley, C., 1981, The welfare cost of capital income taxation in a growing economy, Journal of Political Economy 89, 468–496.

Chipman, J.S. and J.C. Moore, 1980, Compensating variation, consumer's surplus, and welfare, American Economic Review 70, 933–949.

Christiansen, V. and E.S. Jansen, 1978, Implicit social preferences in the Norwegian system of indirect taxation, Journal of Public Economics 10, 217–245.

Cooter, 1979, A new expenditure function, Economics Letters 2, 103–110.

Corlett, W.J. and D.C. Hague, 1953–54, Complementarity and the excess burden of taxation, Review of Economic Studies 21, 21–30.

Dasgupta, P.S. and J.E. Stiglitz, 1972, On optimal taxation and public production, Review of Economic Studies 39, 87–103.

Deaton, A.S., 1977, Equity, efficiency and the structure of indirect taxation, Journal of Public Economics 8, 299–312.

Deaton, A.S., 1979a, The distance function in consumer behaviour with applications of index numbers and optimal taxation, Review of Economic Studies 46, 391–406.

Deaton, A.S., 1979b, Optimally uniform commodity taxes, Economics Letters 2, 357–361.

Deaton, A.S., 1981a, Optimal taxes and the structure of preferences, Econometrica 49, 1245–1260.

Deaton, A.S., 1981b, Numeraires and endowments in optimal tax formulae: A note, Discussion paper 81/113, Sept. (University of Bristol).

Debreu, G., 1951, A classical tax-subsidy problem, Econometrica 22, 14–22.

Debreu, G., 1954, The coefficient of resource utilization, Econometrica 19, 273–292.

Diamond, P.A., 1975, A many-person Ramsey rule, Journal of Public Economics 4, 335–342.

Diamond, P.A. and D.L. McFadden, 1974, Some uses of the expenditure function in public finance, Journal of Public Economics 3, 3–21.

Diamond, P.A. and J. Mirrlees, 1971, Optimal taxation and public production I: Production efficiency, and II: Tax rules, American Economic Review 61, 8–27 and 261–278.

Diewert, W.E., 1978, Optimal tax perturbations, Journal of Public Economics 10, 139–177.

Diewert, W.E., 1981, The measurement of deadweight loss revisited, Econometrica 49, 1225–1244.

Dixit, A.K., 1970, On the optimum structure of commodity taxes, American Economic Review 60, 295–301.

Dixit, A.K., 1975, Welfare effects of tax and price changes, Journal of Public Economics 4, 103–123.

Dupuit, J., 1844, De la mesure de l'utilité des travaux publics, Annales des Ponts et Chaussées 8. Translated and reprinted in: K. Arrow and T. Scitovsky, eds., 1969, AEA readings in welfare economics, 255–283.

Feldstein, M.S., 1972, Distributional equity and the optimal structure of public prices, American Economic Review 62, 32–36.

Feldstein, M.S., 1976, On the theory of tax reform, Journal of Public Economics 6, 77–104.

Feldstein, M.S., 1978, The welfare cost of capital income taxation, Journal of Public Economics 86, 29–51.

Feldstein, M.S., ed., 1983, Behavioral simulation methods in tax policy analysis (University of Chicago Press, Chicago, IL).

Foster, E. and H. Sonnenschein, 1970, Price distortion and economic welfare, Econometrica 38, 281–297.

Gorman, W.M., 1953, Community preference fields, Econometrica 21, 63–80.

Gorman, W.M., 1976, Tricks with utility functions, in: M.J. Artis and A.R. Nobay, eds., Essays in economic analysis (Cambridge University Press, London).

Green, H.A.J., 1961, The social optimum in the presence of monopoly and taxation, Review of Economic Studies 29, 66–78.

Green, J. and E. Sheshinski, 1979, Approximating the gains from welfare reforms, Journal of Public Economics 11, 179–195.

Guesnerie, R., 1977, On the direction of tax reform, Journal of Public Economics 7, 179–202.

Harberger, A., 1954, Monopoly and resource allocation, American Economic Review 42, 77–87.

Harberger, A., 1964, Taxation, resource allocation and welfare, in: The role of direct and indirect taxes in the Federal Reserve system (Princeton University Press, Princeton, NJ, for the NBER and Brookings Institution).

Harberger, A., 1966, Efficiency effects of taxes on income from capital, in: M. Krzyzaniak, ed., Effects of corporation income tax (Wayne State University Press, Detroit, MI) 107–117.

Harberger, A., 1971, Three postulates for applied welfare analysis, Journal of Economic Literature 9, 785–797.

Hatta, T., 1977, A theory of piecemeal policy recommendations, Review of Economic Studies 44, 1–21.

Hausman, J.A., 1981a, Exact consumer's surplus and deadweight loss, American Economic Review 71, 662–676.

Hausman, J.A., 1981b, Labor supply, in: H. Aaron and J. Pechman, eds., How taxes affect economic behavior (Brookings Institution, Washington, DC) 27–84.

Heady, C.J. and P.K. Mitra, 1982, Restricted redistributive taxation, shadow prices and trade policy, Journal of Public Economics 17, 1–22.

Helpman, E. and E. Sadka, 1982, Consumption versus wage taxation, Quarterly Journal of Economics 97, 363–372.

Hicks, J.R., 1940, The valuation of social income, Economica N.S. 7, 105–124.

Hicks, J.R., 1942, Consumer's surplus and index numbers, Review of Economic Studies 9, 126–137.

Hicks, J.R., 1946, Value and capital, 2nd ed. (Oxford University Press, Oxford).

Hotelling, H., 1938, The general welfare in relation to problems of taxation and of railway utility rates, Econometrica 6, 242–269.

Kaldor, N., 1939, Welfare propositions in economics, Economic Journal 49, 549–552.

Kay, J.A., 1980, The deadweight loss from a tax system, Journal of Public Economics 13, 111–120.

King, M.A., 1983a, An index of inequality: With applications to horizontal equity and social mobility, Econometrica 51, 99–115.

King, M.A., 1983b, Welfare analysis of tax reforms using household data, Journal of Public Economics, forthcoming.

Lipsey, R.G. and R.K. Lancaster, 1956–57, The general theory of second best, Review of Economic Studies 24, 11–32.

Marshall, A., 1920, Principles of economics (Macmillan, London).

Merton, R.C., 1969, Lifetime portfolio selection under uncertainty: The continuous-time case, Review of Economic Statistics 51, 247–257.

Mirrlees, J.A., 1972, On producer taxation, Review of Economic Studies 39, 105–111.

Mohring, H., 1971, Alternative welfare gain and loss measures, Western Economic Journal 9, 349–368.

Muellbauer, J.N.J., 1976, Community preferences and the representative consumer, Econometrica 44, 979–999.

Musgrave, R.A., 1959, The theory of public finance (McGraw-Hill, New York).

Pazner, E.A. and E. Sadka, 1981, Welfare criteria for tax reforms: Efficiency aspects, Journal of Public Economics 16, 113–122.

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

Pratt, J.W., 1964, Risk aversion in the small and in the large, Econometrica 32, 122–136.

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

Roberts, K.W.S., 1980, Price-independent welfare prescriptions, Journal of Public Economics 13, 277–297.

Rosen, H., 1978, The measurement of excess burden with explicit utility functions, Journal of Political Economy 86, S121–136.

Rothschild, M. and J.E. Stiglitz, 1970, Increasing risk I: A definition, Journal of Economic Theory 2, 225–243.

Samuelson, P.A., 1947, Foundations of economic analysis (Harvard University Press, Cambridge, MA).

Samuelson, P.A., 1954, The pure theory of public expenditure, Review of Economics and Statistics 36, 387–389.

Samuelson, P.A., 1969, Lifetime portfolio selection by dynamic stochastic programming, Review of Economic Statistics 51, 239–246.

Sandmo, A., 1975, Optimal taxation in the presence of externalities, Swedish Journal of Economics 77, 86–98.

Sandmo, A., 1981, Income tax evasion, labour supply, and the equity–efficiency tradeoff, Journal of Public Economics 16, 265–288.

Shoven, J.B., 1976, The incidence and efficiency effects of taxes on income from capital, Journal of Political Economy 84, 1261–1284.

Small, K.A. and H.S. Rosen, 1981, Applied welfare economics with discrete choice models, Econometrica, 105–130.

Stiglitz, J., 1972, Taxation, risk taking and the allocation of investment in a competitive economy, in: M. Jensen, ed., Studies in the theory of capital markets (Praeger, New York) 294–361.

Stiglitz, J., 1982, Utilitarianism and horizontal equity: The case for random taxation, Journal of Public Economics 18, 1–33.

Stiglitz, J. and P.S. Dasgupta, 1971, Differential taxation, public goods and economic efficiency, Review of Economic Studies 38, 151–174.

Varian, H.R., 1982, The nonparametric approach to demand analysis, Econometrica 50, 945–973.

Vartia, Y.O., 1983, Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions, Econometrica 51, 79–98.

Venti, S.F. and D.A. Wise, 1984, Moving and housing expenditure: Transactions costs and equilibrium, Journal of Public Economics 23, 207–243.

Weiss, L., 1976, The desirability of cheating incentives and randomness in the optimal income tax, Journal of Political Economy 84, 1343–1352.

Willig, R.E., 1976, Consumer's surplus without apology, American Economic Review 66, 589–597.