THE THEORY OF EXCESS BURDEN
AND OPTIMAL TAXATION

Alan J. Auerbach

Working Paper No. 1025

The Theory of Excess Burden and Optimal Taxation

Abstract

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

The paper includes a discussion of various measures of excess burden, focusing on issues of approximation, informational requirements, aggregation over individuals, and the effects of technology. Included in the presentation of optimal tax theory is a section on tax reform, as well as an application of the theory to the case where uncertainty is present.

Alan J. Auerbach
Department of Economics
Harvard University
Cambridge, MA 02138

(617) 495-3934

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 amounts to little more than a restatement and extension, perhaps in less arcane language and terminology, of previously proven propositions. In some cases, the modern effort is less general in scope than the original.

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 Boiteaux (1956). However, this is not an isolated event. 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 Boiteaux (1951) and Debreu (1951, 1954) has until recently[2] been almost entirely ignored in the subsequent literature. Even the "Laffer curve", popular among noneconomists, might more appropriately be called the "Dupuit curve":

> 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]

One can only imagine how recent U.S. history would have been altered had Dupuit written this on a napkin.

The purpose of this chapter is to present, absent lateral or backward steps, the chronological development of the concept of excess burden and the related study of optimal tax theory. A main objective of this exercise 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 (1953-4), 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.

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. 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, as one would hope, 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 trapazoid 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.

The basic problem with consumers' surplus as a welfare measure is that 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 consumers' surpluses ΔS^i, i.e. the line integral:

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

Figure 2.1

Consumers' and Producers' Surplus
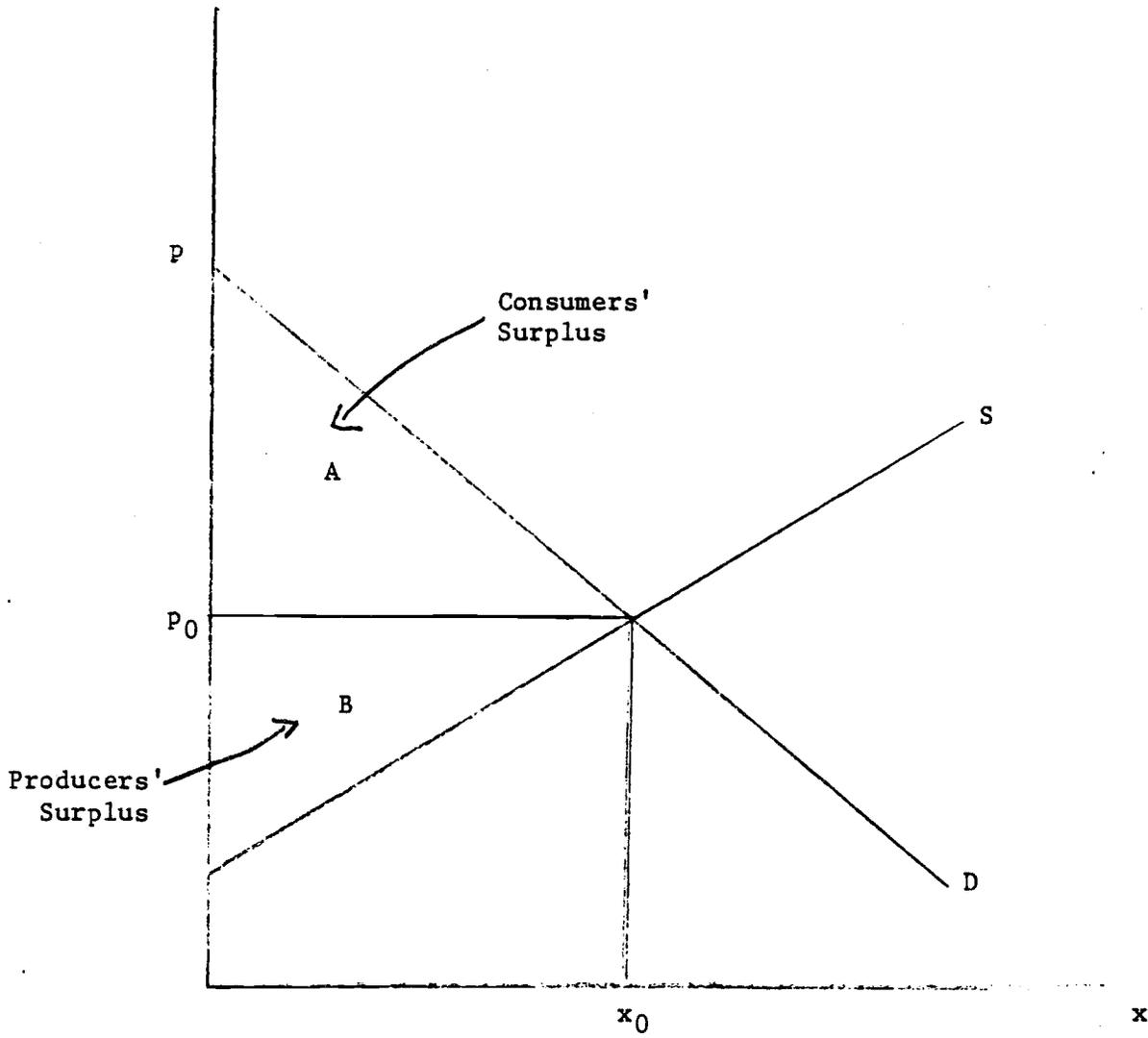
takes on different values according to the path of integration from the initial

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:

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

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

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

Subtracting ΔS_1 from ΔS_2, we obtain:

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

$$- \int_{p_0^2}^{p_1^2} [x^2(p_1^1, p^2) - x^2(p_0^1, p^2)] dp^2 \qquad (2.4)$$

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

$p^2 = p^2 + dp^2$ and $p_1^1 = p_0^1 + dp^1$, (2.4) becomes:

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

which equals zero only if the cross price derivative $\frac{\partial x^1}{\partial p^2}$ and $\frac{\partial x^2}{\partial p^1}$ are equal.[4]

Such symmetry holds for _compensated_ demands: the Slutsky matrix _is_ symmetric

(Hicks 1946). However, ordinary demand derivatives also possess income effects.

For the income effects to be equal, it is straightforward to show, budget shares

must be constant, i.e., preferences must be homothetic.

One solution to this ambiguity, then, is to define a surplus measure in terms of the 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? The problem is one of index numbers, and two natural candidates are the levels of utility prevailing before and after the price changes. Following Hicks (1942), we define the <u>compensating</u> 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 <u>equivalent</u> 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:

$$E\ (p,\ \bar{U})\ =\ \min\ (p\cdot x)\ \text{subject to}\ U(x)\ \geqslant \bar{U} \qquad (2.5)$$

we may express concisely the equivalent and compensating variations as $E\ (p,\ \bar{U})\ -\ E(p_0\cdot \bar{U})$, where $\bar{U}$ 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, 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

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

Substituting (2.6) into (2.5), we obtain for the compensating variation of a price change from p_0 to p_1:

$$CV(\underset{\sim}{p}_0, \underset{\sim}{p}_1) = E(\underset{\sim}{p}_1, V(\underset{\sim}{p}_0, y)) - E(\underset{\sim}{p}_0, V(\underset{\sim}{p}_0, y))$$

$$= E(\underset{\sim}{p}_1, V(\underset{\sim}{p}_0, y)) - y \tag{2.7a}$$

and for the corresponding equivalent variation:

$$EV(\underset{\sim}{p}_0, \underset{\sim}{p}_1) = E(\underset{\sim}{p}_1, V(\underset{\sim}{p}_1, y)) - E(\underset{\sim}{p}_0, V(\underset{\sim}{p}_1, y))$$

$$= y - E(\underset{\sim}{p}_0, V(\underset{\sim}{p}_1, y)) \tag{2.7b}$$

(where we use the identity $y = E(\underset{\sim}{p}, V(\underset{\sim}{p}, y))$)

These measures may be depicted graphically. By the envelope theorem (Shepard's Lemma), the derivative of the expenditure function with respect to an individual price p^i is simply the Hicksian or compensated demand $x_C^i(\underset{\sim}{p}, \bar{U})$. Thus, either of the Hicksian variations may be expressed (for the appropriate value of $\bar{U}$) as:

$$E(\underset{\sim}{p}_1, \bar{U}) - E(\underset{\sim}{p}_0, \bar{U}) = \int_{\underset{\sim}{p}_0}^{\underset{\sim}{p}_1} \frac{dM}{d\underset{\sim}{p}}(\underset{\sim}{p}, \bar{U}) \, d\underset{\sim}{p} = \int_{\underset{\sim}{p}_0}^{\underset{\sim}{p}_1} \underset{\sim}{x}(\underset{\sim}{p}, \bar{U}) d\underset{\sim}{p} \tag{2.8}$$

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 $D_C(\bar{U})$ is the compensated demand curve corresponding to the compensated demands $x_C(\underset{\sim}{p}, \bar{U})$, 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

Figure 2.2

Compensating and Equivalent Variations
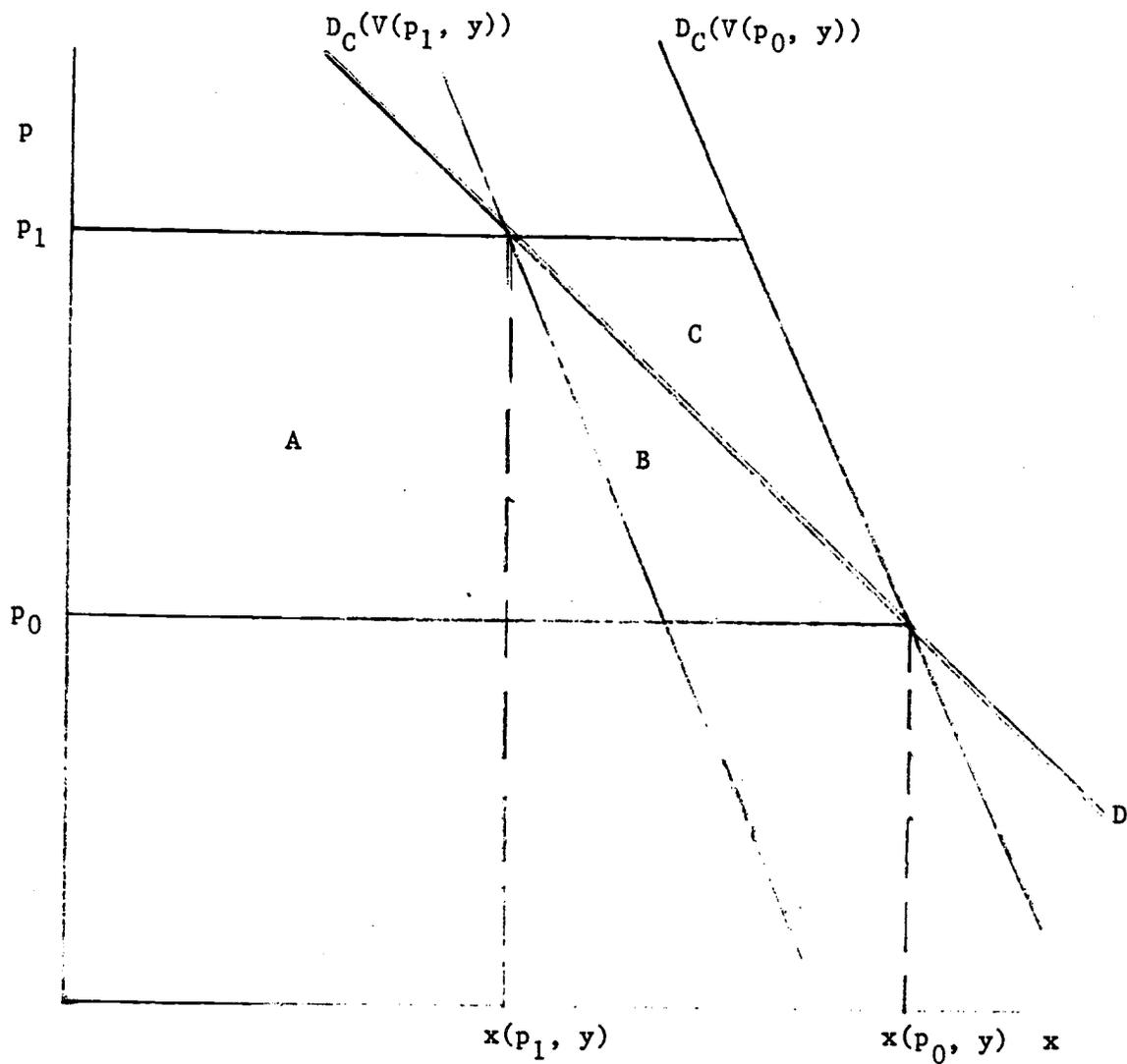
of consumers' surplus as a welfare measure. However, their argument becomes weaker when more than one price changes, for then consumers' surplus is not even single-valued. Moreover, for estimating the excess burden of a tax, it is not the entire loss to the consumer in which we are interested but rather 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.

Beginning again with the basic Marshallian approach, 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 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.

Figure 2.3

Excess Burden of a Tax
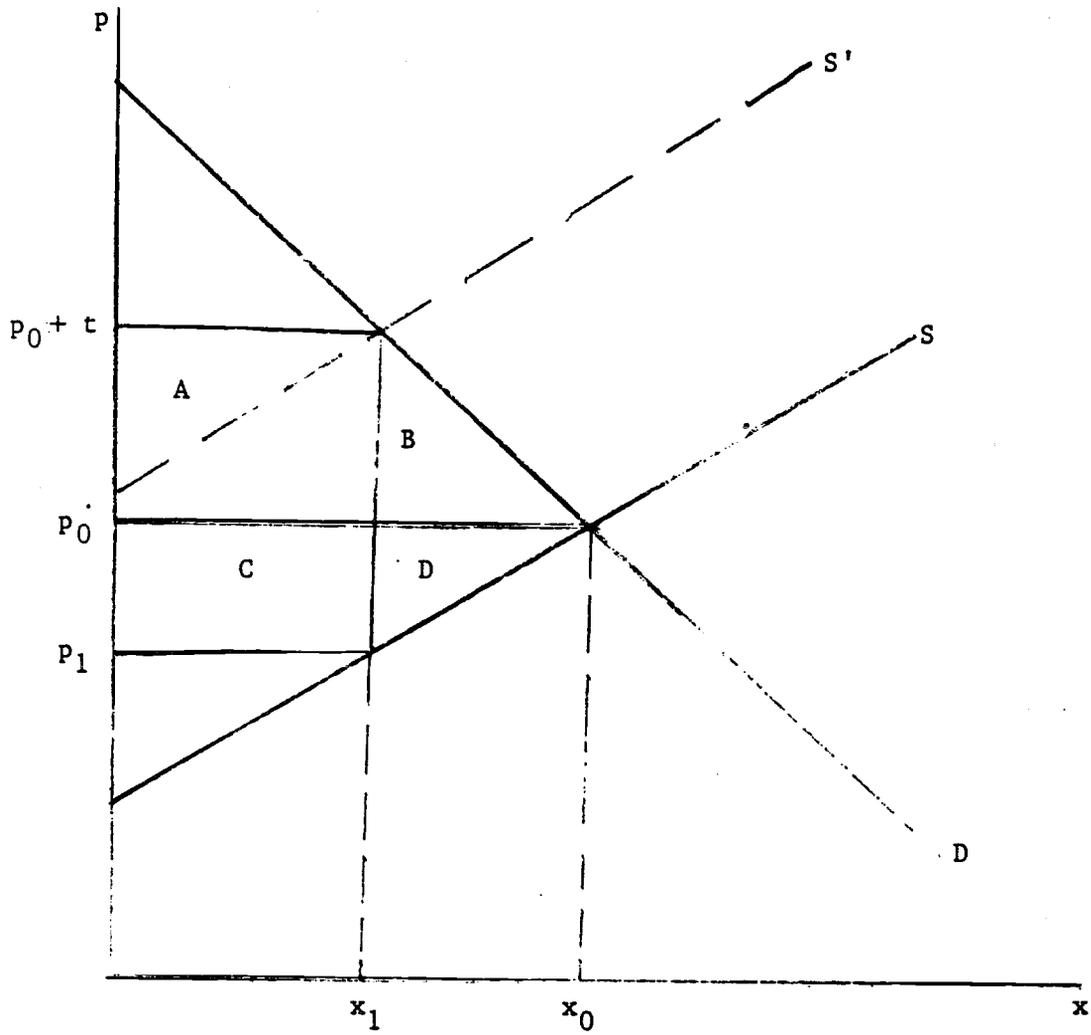
Figure 2.4

Excess Burden of a Subsidy
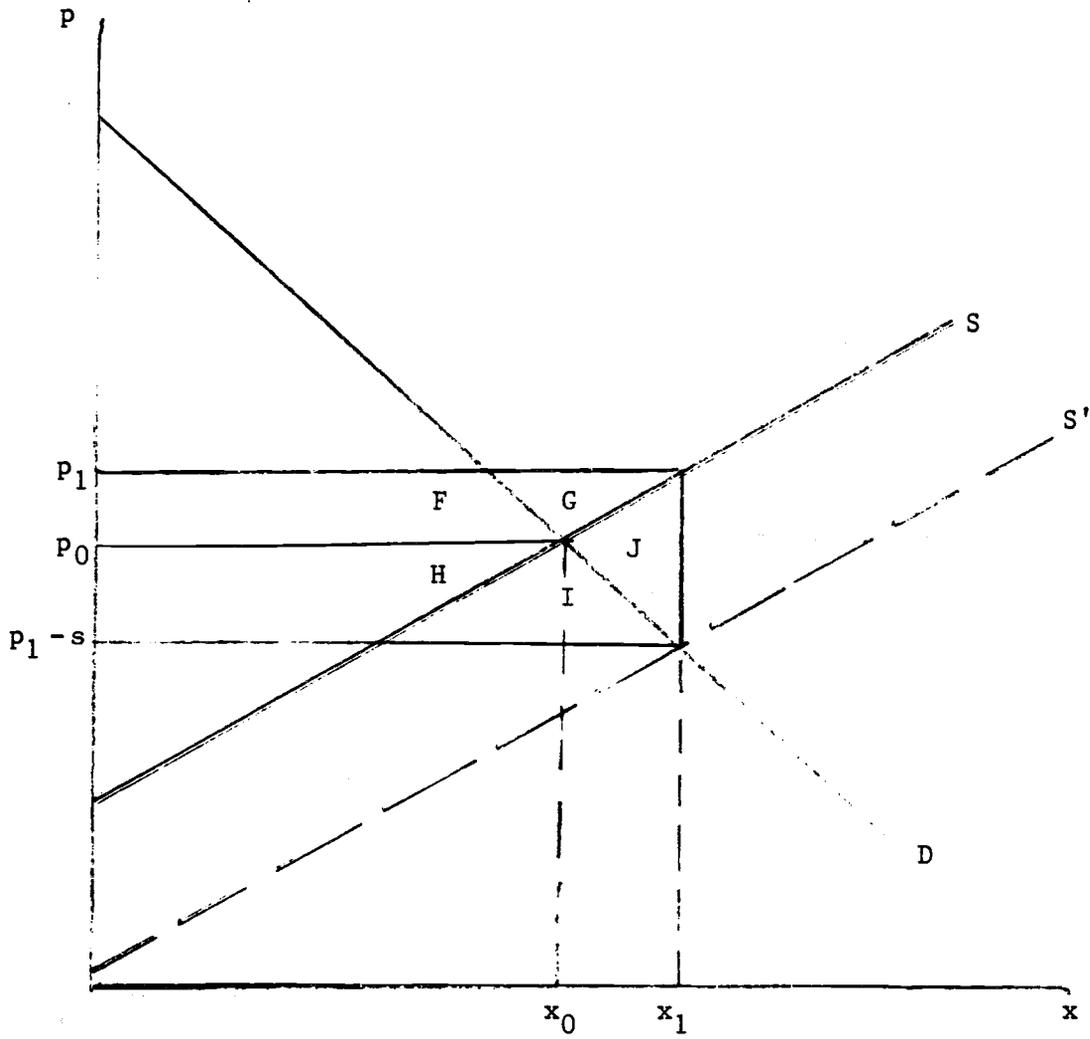
For the case where a tax already exists, we may ask what additional excess burden would be caused by a tax increase. In this case, we subtract the <u>change</u> in government revenue from the change in producers' and consumers' surplus, since revenue is positive at the initial point. The resulting measure 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 areas 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 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.

Unfortunately, we have already seen that consumers' surplus possesses problems as a welfare measure, so it is not clear how informative these results are, particularly if we wish to consider the effects of several taxes at once. It turns out that more sophisticated measures based on the Hicksian variations offer results which do not qualitatively differ from the ones we have already obtained. However, this remains to be demonstrated. 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 a tax is the amount in excess of taxes being collected that the consumer would give up to return to a no-tax state; that is, how much more could be collected from the consumer (and thrown away) than is currently being collected,

Figure 2.5

Excess Burden with a Preexisting Tax
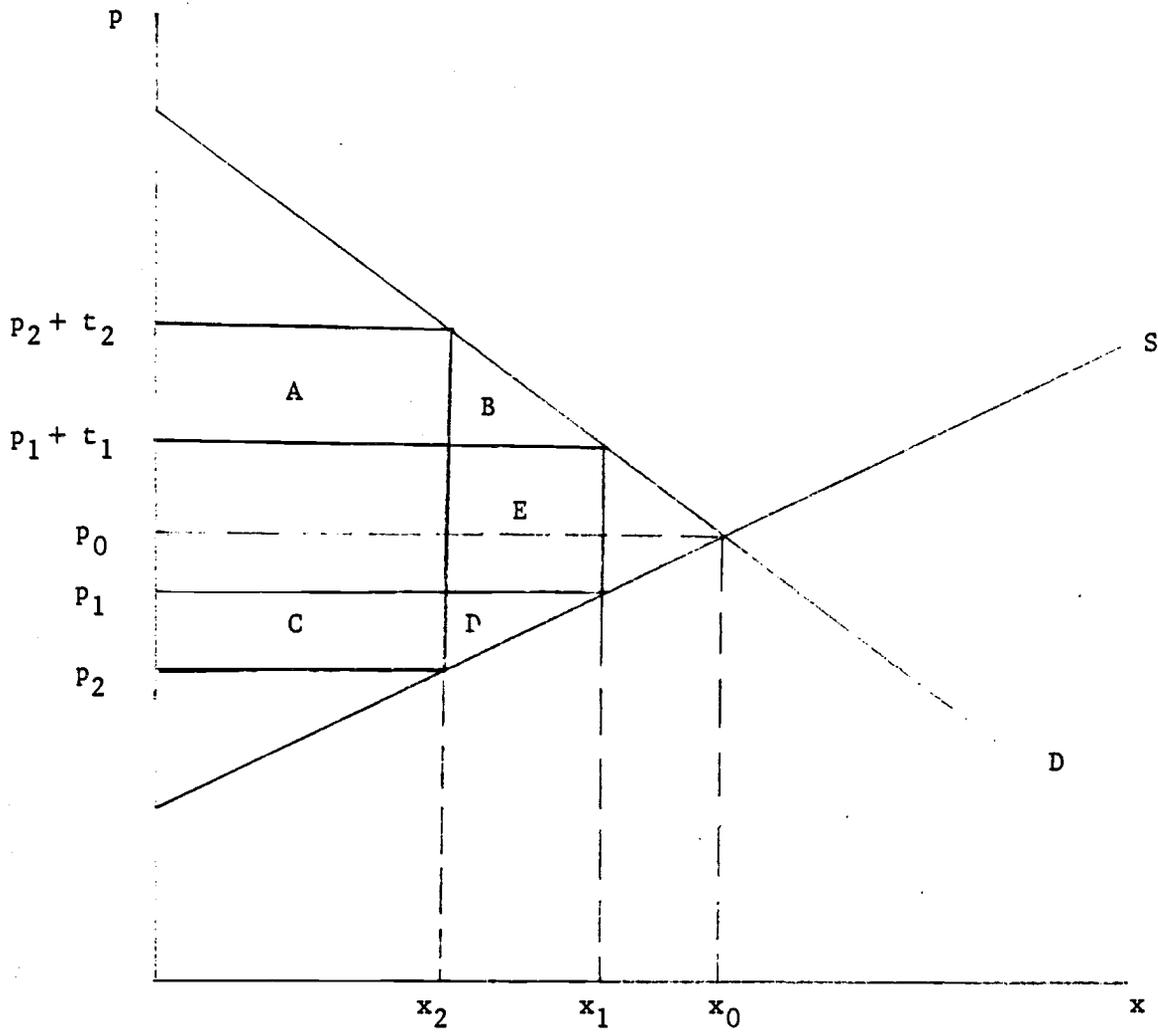
with no loss in utility, if the collection method was lump sum taxation. In the terminology used above, we may write this measure as:

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

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

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

Alternately, 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 $x(p, E(p, \bar{U})) = x_C(p, \bar{U})$).

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

Figure 2.6

A Comparison of Excess Burden Measures
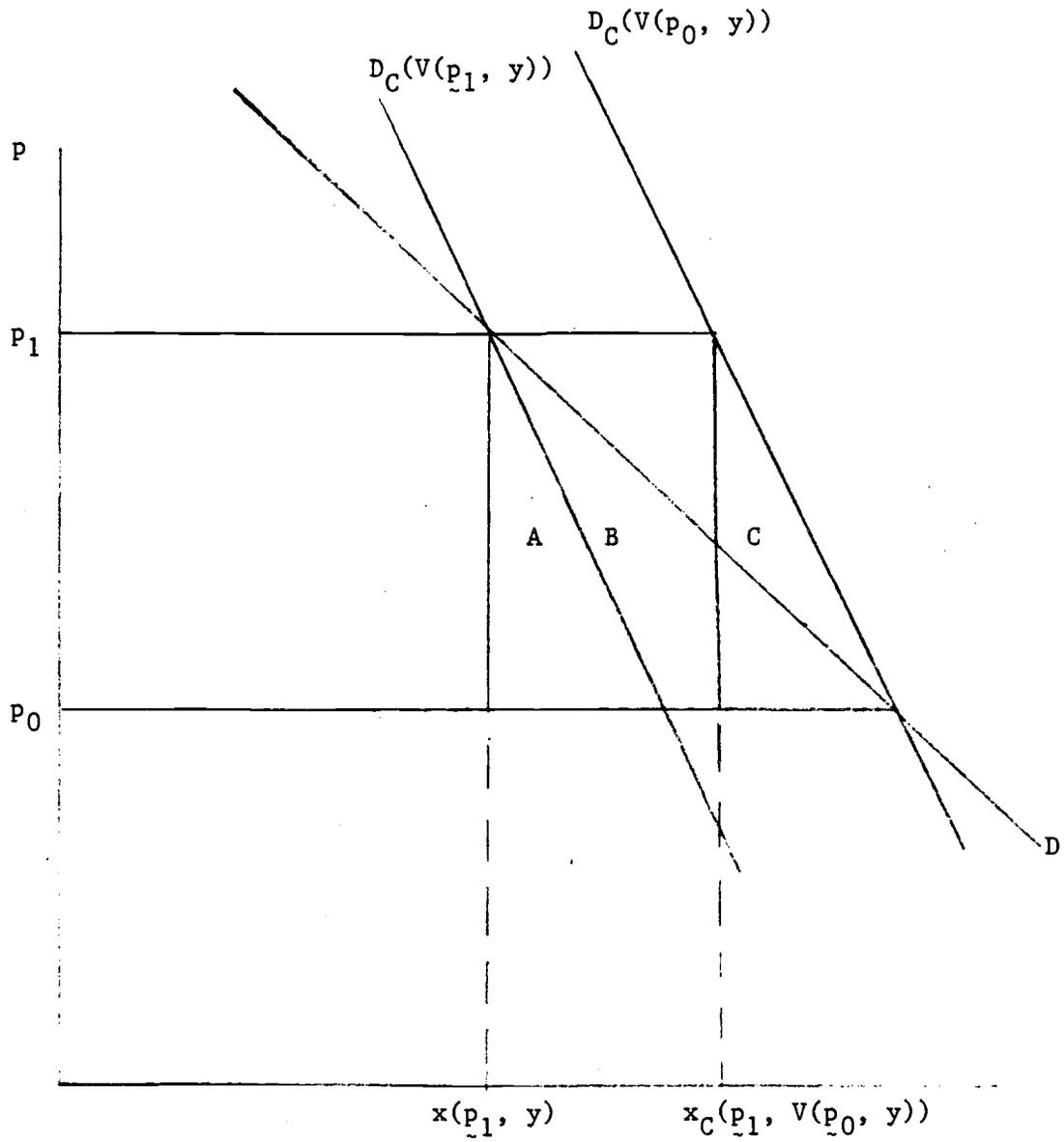
that would be collected if utility were kept at $V(p_0, y)$. This yields the areas

A, A + B, and C for the three respective measures. Note that the two Hicksian

measures no longer bracket the Marshallian one.[6] 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 com-

pensating variations may be conceived.[7] 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 addi-

tional tax revenues, that the consumer would pay to avoid the latest price

increase from p_1 to p_2:

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

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

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

$$\cdot (x_C(p_1, V(p_2, y)) - x(p_2, y)) \tag{2.11}$$

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 reve-

nues that would be required to to maintain the initial utility level, or:

$$EB_C = E\left(p_2, V(p_1, y)\right) - E\left(p_1, V(p_1, y)\right) - \left[R(p_2, E\left(p_2, V(p_1, y)\right)) - R(p_1, y)\right]$$

$$= E\left(p_2, V(p_1, y)\right) - y - (p_2 - p_0) \cdot x_C\left(p_2, V(p_1, y)\right) + (p_1 - p_0) \cdot x(p_1, y)$$

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

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

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 explosition, it is sometimes easier to express the dead-weight 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 $\underset{\sim}{p_1}$, we obtain:

$$EB_C = \frac{dEB_C}{d\underset{\sim}{p}}\Big|_{\underset{\sim}{p_1}} \cdot (\underset{\sim}{p_2} - \underset{\sim}{p_1}) + 1/2\,(\underset{\sim}{p_2} - \underset{\sim}{p_1})'\frac{d^2EB_C}{d\underset{\sim}{p}^2}\Big|_{\underset{\sim}{p_1}}(\underset{\sim}{p_2} - \underset{\sim}{p_1}) + \ldots \qquad (3.1)$$

which, ignoring all terms beyond the second order, yields:

$$EB_C \approx [-(\underset{\sim}{p_1} - \underset{\sim}{p_0})' \cdot \frac{dx_C}{d\underset{\sim}{p}}]\,(\underset{\sim}{p_2} - \underset{\sim}{p_1})$$

$$+ 1/2\,(\underset{\sim}{p_2} - \underset{\sim}{p_1})'[\frac{dx_C}{d\underset{\sim}{p}} - (\underset{\sim}{p_1} - \underset{\sim}{p_0})\frac{d^2x_C}{d\underset{\sim}{p}^2}](\underset{\sim}{p_2} - \underset{\sim}{p_1}) \qquad (3.2)$$

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

$$EB_C \approx -\,(\underset{\sim}{t}'\,S\,\Delta\,\underset{\sim}{t} + 1/2\,\Delta\,\underset{\sim}{t}'\,S\,\Delta\,\underset{\sim}{t}) = -\,(\underset{\sim}{t}'\,\Delta\,\underset{\sim}{x} + 1/2\,\Delta\,\underset{\sim}{t}'\,\Delta\,\underset{\sim}{x_C}) \qquad (3.3)$$

where $\underset{\sim}{t} = (\underset{\sim}{p_1} - \underset{\sim}{p_0})$, $\Delta\,\underset{\sim}{t} = (\underset{\sim}{p_2} - \underset{\sim}{p_1})$, $S = \frac{dx_C}{d\underset{\sim}{p}}$ is the Slutsky matrix, and and $\Delta\,\underset{\sim}{x_C} = S\,\Delta\,\underset{\sim}{t}$.

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.[8]

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 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 should use several small taxes rather than 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:[9]

$$\underset{\sim}{x}(\underset{\sim}{p}, y) = - \frac{dU/d\underset{\sim}{p}}{dU/dy} \qquad\qquad (3.4)$$

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.

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 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 nondistortionary, 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 $\underset{\sim}{\alpha}$ be the compensation vector of the elements of $\underset{\sim}{x}$, normalized so that at initial prices $\underset{\sim}{\alpha}$ has a cost of unity, and β the scalar that determines how much of the compensation bundle the consumer receives, $\beta\underset{\sim}{\alpha}$. If we denote producer prices by $\underset{\sim}{q}$ and consumer prices by $\underset{\sim}{p}$, then the compensating variation measure of excess burden β can be defined implicitly by the equation

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

where $\underset{\sim}{p_1}$ is the initial consumer price vector, $\underset{\sim}{p_2}$ the distorted price vector, $\underset{\sim}{q_1}$ and $\underset{\sim}{q_2}$ the corresponding producer price vectors, y_1 and y_2 the lump sum income in the two states, and $R_1 = (\underset{\sim}{p_1} - \underset{\sim}{q_1}) \cdot \underset{\sim}{x}(\underset{\sim}{p_1}, y_1)$ and $R_2 = (\underset{\sim}{p_2} - \underset{\sim}{q_2}) \cdot x_C(\underset{\sim}{p_2}, V(\underset{\sim}{p_1}, y_1))$ 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 $\underset{\sim}{x}$, then the pure profits from competitive production are positive and change with the change in producer prices. Letting $\underset{\sim}{z}$ be the vector of goods produced, total profits are $y = \underset{\sim}{q} \cdot \underset{\sim}{z}$. Note that production and consumption differ by the infusion of additional compensation, $\beta \cdot \underset{\sim}{\alpha}$.

It is not generally possible to find a closed form solution for β from its implicit definition in (3.5). However, we can consider the change in β resulting from an increase in taxes starting at the initial equilibrium by calculating

the total differential of (3.5), evaluated at that point and then solving for

$d\beta$. Doing so, we obtain:

$$\frac{dV}{dp} \cdot dp + \frac{dV}{dy} [\frac{dy}{dq} \cdot dq + \beta\alpha \cdot dq + d\beta\alpha \cdot q + t \cdot dx + x \cdot dt] = 0 \quad (3.6)$$

where $t = (p - q)$.

Again using the evelope theorem, one can show that $\frac{dy}{dq} = z$. Using this and

Roy's identity ((3.4) above), we obtain from (3.6):

$$\frac{dV}{dy} [-x \cdot dp + z \cdot dq + \beta\alpha \cdot dq + d\beta\alpha \cdot q + t \cdot dx + x \cdot dt] = 0 \quad (3.7)$$

But since $x = y + \beta\alpha$ and $\frac{dV}{dy} \neq 0$, and by the normalization $q \cdot \alpha = 1$, (3.7)

simplifies to

$$d\beta = -t \cdot dx \quad (3.8)$$

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

However, here dx is the change in x as taxes rise, holding utility constant and

compensating with the bundle α. Because the choice of α affects the equili-

brium, it affects dx and hence $d\beta$. For example, consider the case in which

there are two goods, one of which is taxed. Let the untaxed good serve as nu-

meraire, so that its price does not change. Then we can represent the effect of

the compensation bundle in a graph showing the demand for the taxed good. This

is done in Figure 3.1. The supply curve S represents the relative producer

price of x, and depends only on the amount of x produced. The demand curve tra-

ces out the compensated demand for the taxed good, which is a function of the

good's own price alone in this two-good case. With the introduction of a tax t

to a previously undistorted equilibrium, the change in x depends on whether the

compensation is in the numeraire good or the taxed good. In the former case, x

declines to x_2. In the latter case, less of the taxed good is produced <u>because</u>

Figure 3.1

Excess Burden with Changing Producer Prices
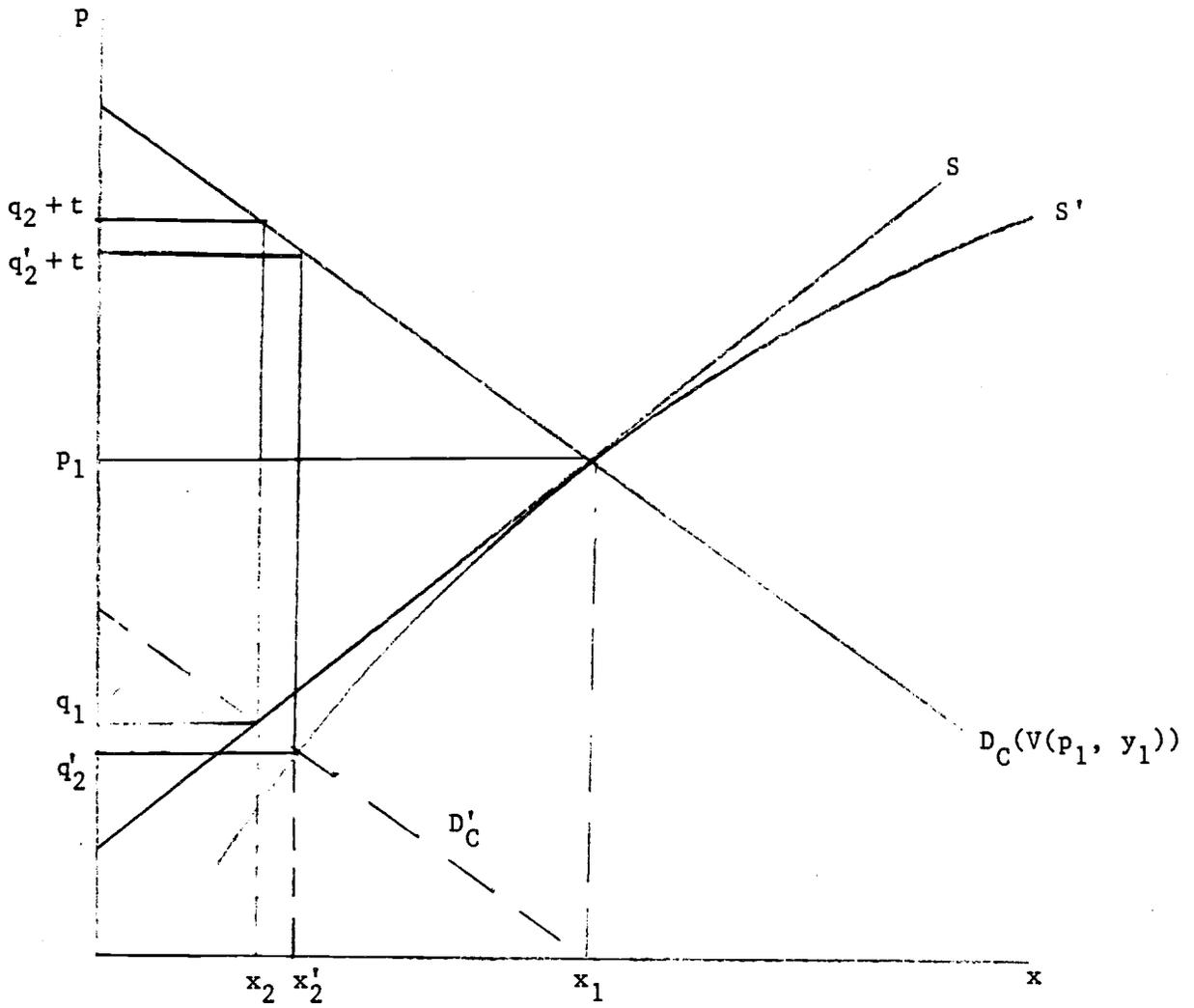
of the excess burden. This can be represented by a second supply curve S',

which accounts for the fact that total supply exceeds production. (Note that S'

will lie to the right of S except at the undistorted price.) The result is a

smaller decline in x, to x_2'. Thus, with producer prices changing, an additional

ambiguity is added to the measurement of excess burden.

We can also derive the second-order effect on β of a change in the tax vec-

tor $\underset{\sim}{t}$ by taking the total differential of (3.7). This yields

$$d^2\beta = -d\underset{\sim}{t} \cdot d\underset{\sim}{x} - d\underset{\sim}{\beta\alpha} \cdot d\underset{\sim}{q} - \underset{\sim}{t} \cdot d^2\underset{\sim}{x} \qquad (3.9)$$

which, even if we ignore the last curvature term, has an additional term, com-

pared to the second-order effect in (3.3), caused by the changing value of the

normalized compensation bundle. If all compensation is in the form of the

numeraire good, or if there are no taxes initially, this term also vanishes and

we are left with the more familiar second order effect.[10]

Another familiar expression for the second-order effect may be derived from

(3.9). Again ignoring the last curvature term, we use the fact that $\underset{\sim}{x} = \underset{\sim}{\beta\alpha} + \underset{\sim}{z}$

to obtain:

$$\partial^2\beta \approx - d\underset{\sim}{p} \cdot d\underset{\sim}{x} + d\underset{\sim}{q} \cdot d\underset{\sim}{z} = -d\underset{\sim}{p} S d\underset{\sim}{p} + d\underset{\sim}{q} H d\underset{\sim}{q} \qquad (3.10)$$

where H is the Hessian of the profit function $\dfrac{d^2y}{d\underset{\sim}{q}^2} = \dfrac{d\underset{\sim}{z}}{d\underset{\sim}{q}}$.

This expression for the second-order effect of a change in taxes on welfare was

first developed by Boiteaux (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:

$$V(\underset{\sim}{p_2}, y_2) = V(\underset{\sim}{p_1}, y_1 - (R_2 - R_1) - \underset{\sim}{q_1} \cdot \underset{\sim}{\alpha}\beta) \qquad (3.11)$$

where, in this case, state 2 is the actual state with taxes at $\underset{\sim}{t_2}$, whereas state 1 is the hypothetical state in which taxes do not rise from $\underset{\sim}{t_1}$ 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) <u>coefficient of resource utilization</u>, 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 $\underset{\sim}{\alpha}$ 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):

$$L = E^1(\underset{\sim}{p}_1, V^1(\underset{\sim}{p}_0, y^1)) - y^1 + E^2((\underset{\sim}{p}_1, V^2(\underset{\sim}{p}_0, y^2)) - (y^1 + y^2)$$

$$- (\underset{\sim}{p}_1 - \underset{\sim}{p}_0) \cdot (x_c^1(\underset{\sim}{p}_1, V^1(\underset{\sim}{p}_0, y^1)) + x_c^2(\underset{\sim}{p}_1, V^2(\underset{\sim}{p}_0, y^2)) \qquad (3.12)$$

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:

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

$$- 1 - (\underset{\sim}{p}_1 - \underset{\sim}{p}_0) \cdot (- \frac{\partial x_c^1}{\partial U} \cdot \frac{\partial V^1}{\partial y} + \frac{\partial x_c^2}{\partial U} \cdot \frac{\partial V^2}{\partial y}) \qquad (3.13)$$

which, using the fact that $x_c(\underset{\sim}{p}_1, V(\underset{\sim}{p}_0, y)) = \underset{\sim}{x}(\underset{\sim}{p}_1, E(\underset{\sim}{p}_1, V(\underset{\sim}{p}_0, y)))$, can be rewritten:

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

$$- (\underset{\sim}{p}_1 - \underset{\sim}{p}_0) \cdot (- \frac{\partial x^1}{\partial y} \cdot \frac{\partial E^1}{\partial U} \cdot \frac{\partial V^1}{\partial y} + \frac{\partial x^2}{\partial y} \cdot \frac{\partial E^2}{\partial U} \cdot \frac{\partial V^2}{\partial y}) \qquad (3.14)$$

Since $E(\underset{\sim}{p}_0, V(\underset{\sim}{p}_0, y)) = y$, we may rewrite (3.14) as:

$$dL = -\frac{\hat{\mu}^1}{\mu^1}(1 - (\underset{\sim}{p}_1 - \underset{\sim}{p}_0) \cdot \frac{dx^1}{dy}) + \frac{\hat{\mu}^2}{\mu^2}(1 - (\underset{\sim}{p}_1 - \underset{\sim}{p}_0) \cdot \frac{dx^2}{dy})$$

where $\mu^i = \frac{\partial E^i}{\partial U}(\underset{\sim}{p}_0, V^i(\underset{\sim}{p}_0, y^i))$ and $\hat{\mu}^i = \frac{dE^i}{dU}(\underset{\sim}{p}_1, V^i(\underset{\sim}{p}_0, y^i))$ are the marginal

expenditures needed per unit of increased utility at base utility level

$V^i(\underset{\sim}{p}_0, y^i)$ and price levels $\underset{\sim}{p}_0$ and $\underset{\sim}{p}_1$, respectively. Thus, dL will equal zero,

in general, only if two conditions are met:

(1) $\dfrac{\hat{\mu}^i}{\mu^i}$ equals some common function of prices alone (not income) for

the two individuals; and

(2) the vector of income effects $\dfrac{dx^i}{dy}$ equals some common function of

prices alone.

Condition (2) implies that ordinary demand functions take the form:

$$\underset{\sim}{x}^i(\underset{\sim}{p}, y^i) = \underset{\sim}{\phi}^i(\underset{\sim}{p}) + \underset{\sim}{\theta}(\underset{\sim}{p})y^i \tag{3.16}$$

for some functions $\underset{\sim}{\phi}^i(\cdot)$ and $\underset{\sim}{\theta}(\cdot)$, the latter common across individuals. (The

laws of consumer demand imply, in turn, that $\underset{\sim}{\phi}^i(\cdot)$ is homogeneous of degree 0 in

prices and $\underset{\sim}{\theta}(\cdot)$ is homogeneous of degree -1 in prices, since a proportional

change in $\underset{\sim}{p}$ and y can't affect $\underset{\sim}{x}^i(\cdot)$.) The demand function specified in (3.16)

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 a suitable transformation of the utility

function, consumer i's expenditure function can be written:

$$E^i(\underset{\sim}{p}, U^i) = \delta^i(\underset{\sim}{p}) + \gamma(\underset{\sim}{p}) \cdot U^i \tag{3.17}$$

(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, $\underset{\sim}{\phi}^i(\cdot)$.

Note that even identical preferences, if not homothetic, will suffice. For example, suppose individuals have an inelastic 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 <u>define</u> measures of excess burden for the multi-individual case. For example, we may implicitly define a compensating variation measure analogous to (3.5) by the identities:

$$v^i(\underset{\sim}{p}_2, \omega^i(y_2 + R_2 - R_1 + \underset{\sim}{q}_2 \cdot \underset{\sim}{\alpha\beta}) = v^i(\underset{\sim}{p}_1, \bar{\omega}^i y_1) \qquad \forall i \qquad (3.18)$$

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 $\underset{\sim}{p}_2$ and the extra compensation vector $\underset{\sim}{\alpha} \cdot \beta$ "enters" the system. For the equivalent variation, the measure for β corresponding to (3.11) for several individuals is

$$v^i(\underset{\sim}{p}_2, \bar{\omega}^i y_2) = v^i(\underset{\sim}{p}_1, \omega^i(y_1 - R_2 + R_1 - \underset{\sim}{q}_2 \cdot \underset{\sim}{\alpha\beta})) \qquad (3.19)$$

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 (3.18) or (3.19) 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 Subsection 3.2 (with aggregate demands replacing individual ones), an interesting result occurs in the second-order effect derived from the measure defined by (3.19). 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 won't work either. We can certainly imagine a social welfare function of the form:

$$w(U^1, \ldots, U^H) = \sum_{i=1}^{H} U^i \tag{3.20}$$

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. 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 reqmain equal in all cases (Roberts 1980). Thus, we are free to consider our aggregate measures of excess burden to be measures of efficiency. However, it will generally not be possible to make welfare comparisons on the basis of such measures, 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 the second order Taylor approximation (3.3), implicitly derived from the compensating variation measure of excess burden, for the case in which there are no pre-existing taxes. 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 noncorporate sectors. Nontax 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 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 approximations 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 functions. 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:

$$EB_C = E\big(\underset{\sim}{p_1}, V(\underset{\sim}{p_0}, y_0)\big) - y_1 - (\underset{\sim}{p_1} - \underset{\sim}{p_0}) \cdot \underset{\sim}{x_C}\big(\underset{\sim}{p_1}, V(\underset{\sim}{p_0}, y_0)\big) \qquad (4.1)$$

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 <u>as</u> <u>if</u> they faced a proportional tax equal to the actual marginal rate, with the inframarginal excess in collections that results being assigned to lump sum income. For example, consider the case of a progressive 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 (1982b) 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. However, the changes in utility take account of switches in regime that may occur in 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 expclicit 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 (1982), in which indivudals must decide whether to move or stay, and face different budget constraints in the two situations.

Figure 4.1

Progressive Taxes and Virtual Income
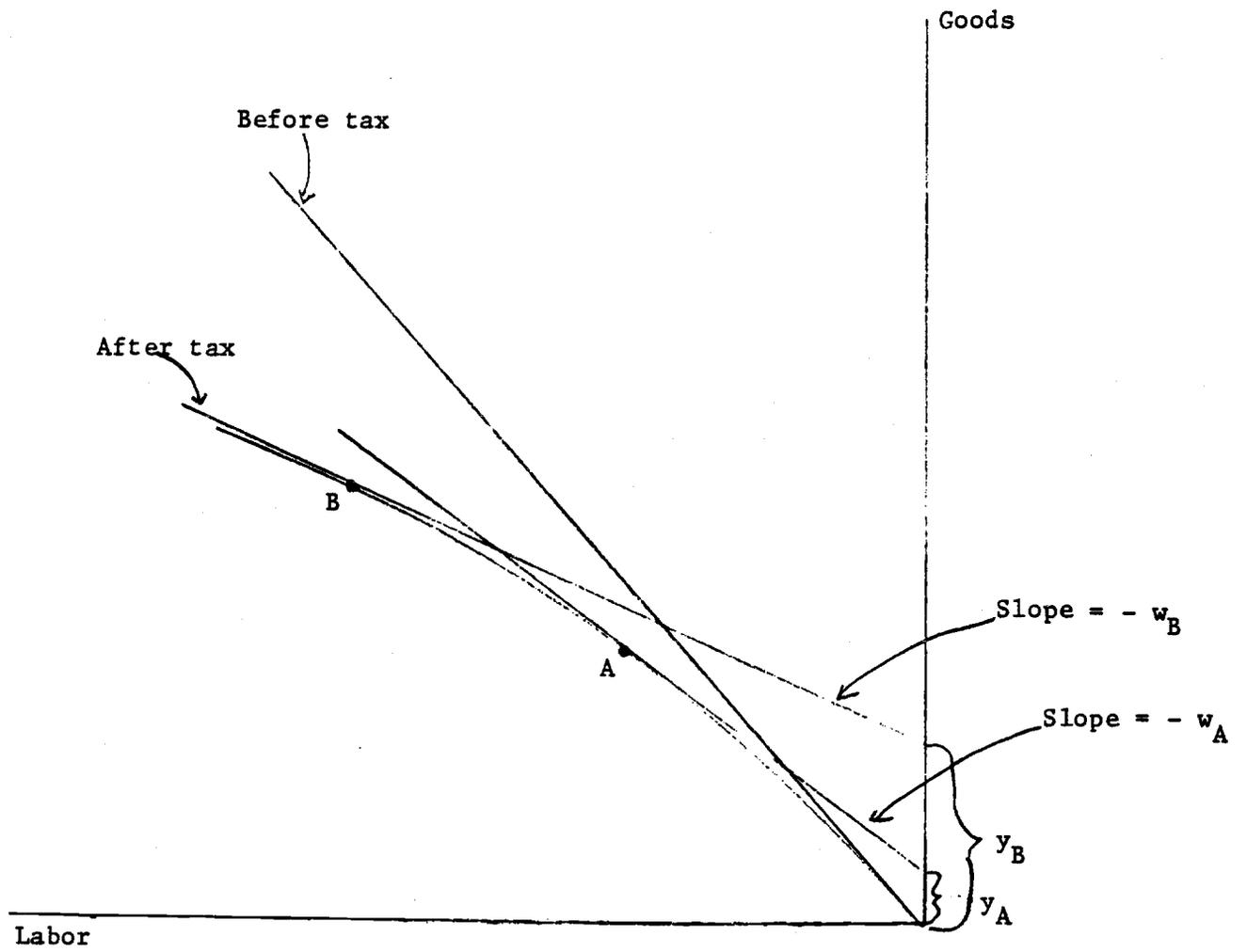
4.3 Simulation Methods

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. An early example of the use of simulation technique is Shoven's (1976) reconsideration of the excess burden caused by the corporate income tax. In more recent work, Auerbach, Kotlikoff and Skinner (1981) use a perfect foresight, overlapping generations growth model to analyze the effect 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 nondistortionary 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 nonuniform taxation dependent on personal characteristics. For example, we might succeed in having a nondistortionary and progressive tax system by taxing according to genetic characteristics associated with ability, but such schemes are typically proscribed.

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. In Subsections 5.2 and 5.3, 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. Subsection 5.4 discusses the relationship of the optimal tax solution to the measures of excess burden described above.

5.1 Basic Optimal Tax Results

We imagine a representative consumer who has exogenous income y, and faces consumer prices $\underset{\sim}{p} = (p_0, p_1, \ldots, p_N)$ for the commodities $0, 1, \ldots, N$, which have fixed producer prices $\underset{\sim}{q} = (q_0, q_1, \ldots, q_N)$. 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, ..., 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 ex-
press the optimal tax problem by:

$$\max_{p} [\max_{x} U(x) \text{ subject to } p \cdot x = y] \text{ subject to } (p - q) \cdot x = R \qquad (5.1)$$

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

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

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 $\frac{dt}{dp} = I$
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

$$- \lambda x_i + \mu[\sum_j t_j \frac{dx_i}{dp_i} + x_i] = 0 \qquad \nabla_i \qquad (5.4)$$

where $\lambda = \frac{dV}{dy}$ 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 $\frac{dx_i}{dp_i}$ using the Slutsky equation, and defining

$$\alpha = \lambda + \mu \sum_j t_j \frac{dx_i}{dy} \qquad (5.5)$$

to be the <u>marginal social utility</u> of income (Diamond 1975), to obtain:

$$- \sum_j S_{ij} t_j = \left(\frac{\mu - \alpha}{\mu}\right) x_i \qquad \forall_i \qquad (5.6)$$

where the S_{ij}s are components of the Slutsky matrix S. The term α differs from λ because, in the presence of taxes, a dollar given to the individual increases his utility directly by λ <u>and</u> indirectly by the increased revenue. 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 $\mu \sum_j t_j \frac{dx_j}{dy}$, 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 concern: the marginal excess burden of the tax. This term is always non-negative (see Section 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

$$t_i = \theta p_i \qquad \forall_i \qquad (5.7)$$

for some constant θ, we obtain

$$-\theta \sum S_{ij} p_j = \left(\frac{\mu - \alpha}{\mu}\right)_i \qquad \forall_i \qquad (5.8)$$

But $\sum S_{ij} p_j$ equals $\frac{1}{\lambda} \frac{dU}{p_i} \Big|_u = 0$ 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 nondistortionary, however, is the key to their limited applicability. Since $p = q + t = q + \theta p$, $p = \frac{q}{1 - \theta}$. Hence, the consumer's budget contraint becomes:

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

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 exogeneous 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 nonzero.

First of all, y will be nonzero 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 nonzero 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

$$q_C C + (\ell - L) = 0 \tag{5.10a}$$

or

$$q_C C + \ell = L \tag{5.10b}$$

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, $(\ell - L)$, we have no income y to tax through proportional excise taxes. Interpreting the commodity as _consumption_ of leisure, ℓ, we can use the proprotional tax solution on C and ℓ to tax L indirectly. Hence, the inability to use

proportional taxes to raise revenue is equivalent to the restriction to taxing flows, 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 <u>can</u> 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 $\underset{\sim}{x}$ to be such flows rather than total consumption. In exchange for the loss of a nondistortionary 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 $V(\underset{\sim}{p})$, it is also homogeneous of degree zero in prices. So is the revenue constraint, since, for any constant ϕ,

$$(\phi\underset{\sim}{p} - \underset{\sim}{q}) \cdot \underset{\sim}{x} = (\phi - 1)\underset{\sim}{p} \cdot \underset{\sim}{x} + (\underset{\sim}{p} - \underset{\sim}{q}) \cdot \underset{\sim}{x} = (\underset{\sim}{p} - \underset{\sim}{q}) \cdot \underset{\sim}{x} \qquad (5.11)$$

Thus, we may choose any scale for $\underset{\sim}{p}$. 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 can't 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 the untaxable endowment. This distinction is important when one interprets the various rules now derived.

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

$$- \theta \sum_{j \neq 0} S_{ij} p_j = \theta S_{i0} \qquad (5.12)$$

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? A clue comes from considering the prescription given by (5.6) for the case of small taxes. 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

$$- \Delta x_i = - \sum_i S_{ij} \Delta p_j = - \sum_i S_{ij} t_j \qquad (5.13)$$

so that (5.6) calls for an equiproportional reduction in demand for each taxed 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 $\underset{\sim}{t}$ starting from a Pareto optimum induces an excess burden of
approximately:

$$L = - \tfrac{1}{2} \sum_i \Delta t_i \; \Delta x_i = - \tfrac{1}{2} \sum_i t_i \; \Delta x_i \tag{5.14}$$

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
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 this result may be obtained by stacking con-
ditions (5.6) to obtain:

$$\hat{S}\hat{\underset{\sim}{t}} = (\frac{\mu - \alpha}{\mu})\underset{\sim}{x} \tag{5.15}$$

where $\hat{S}$ in the Slutsky matrix excluding good zero, and $\hat{\underset{\sim}{t}} = (t_1, \ldots, t_N)$, and
multiplying through by $\hat{S}^{-1}$:

$$\hat{\underset{\sim}{t}} = (\frac{\mu - \alpha}{\mu})\underset{\sim}{x} \; \hat{S}^{-1}\underset{\sim}{x} \tag{5.16}$$

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

If there are only three goods, two taxed, then (5.16) yields the two equations:

$$t_1 = \tfrac{1}{\Delta}(\frac{\mu - \alpha}{\mu})(S_{22}x_1 - S_{12}x_2) \tag{5.17a}$$

$$t_1 = \tfrac{1}{\Delta}(\frac{\mu - \alpha}{\mu})(-S_{21}x_1 + S_{11}x_2) \tag{5.17b}$$

where $\Delta = S_{11}S_{22} - S_{12}S_{21}$. Since $S_{i0} + p_1 S_{i1} + p_2 S_{i2} = 0$ for $i = 1,2$, we may divide (5.17a) by (5.17b) and substitute to obtain:

$$\frac{t_1}{t_2} = \frac{-\frac{1}{p_2}(S_{20} + p_1 S_{21})x_1 - S_{12}x_2}{-\frac{1}{p_1}(S_{10} + p_2 S_{12})x_2 - S_{21}x_1} \tag{5.18}$$

or, defining $\theta_i = t_i/p_i$ and dividing the numerator and denominator of the right-hand side of (5.18) by $x_1 x_2$, we obtain (Harberger 1964)

$$\frac{\theta_1}{\theta_2} = \frac{\varepsilon_{20} + \varepsilon_{21} + \varepsilon_{12}}{\varepsilon_{10} + \varepsilon_{21} + \varepsilon_{12}} \tag{5.19}$$

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

If the cross effects ε_{12} and ε_{21} are zero, then (5.19) calls for a higher tax on the taxed good that is the relative complement to the numeraire (ε_{10} is smaller). This has generated the somewhat misleading explanation that we "can't" 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 (5.19) applies for any numbering of the three goods. If the cross-elasticities between goods 1 and 2 are nonzero, we may even observe the anomalous result that the relative complement to good zero should receive a lighter tax.

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 $i = 1, \ldots, N$,

$$S_{ii} p_i + S_{i0} = 0 \tag{5.20}$$

Thus, this restriction <u>does</u> depend on the choice of untaxed commodity. With such a simplification, (5.16) yields the expressions:

$$t_i = -\frac{1}{\Delta}\left(\frac{\mu - \alpha}{\mu}\right)\frac{x_i}{S_{ii}} \tag{5.21}$$

or $\quad \theta_i \sim 1/\varepsilon_{ii}$

where $\Delta = S_{11}S_{22} - S_{21}S_{12} > 0$.

This is the celebrated "inverse elasticity" rule that calls for higher proportional taxes on goods with relatively new own price elasticities. By (5.20), this rule is equivalent to

$$\theta_i \sim 1/\varepsilon_{i0} \tag{5.22}$$

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 the restrictions 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 (5.21) is expressed in terms of compensated elasticities. Yet in various places in the literature (Diamond and Mirrlees 1971; 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 (5.3):

$$-\sum_j t_j \frac{dx_j}{dp_i} = \left(\frac{\mu - \lambda}{\mu}\right)x_i \tag{5.23}$$

which, assuming $\dfrac{dx_j}{dp_i} = 0$ unless $i = 0$ or j, yields:

$$\theta_i \sim 1/\eta_{ii} \qquad (5.24)$$

where $\eta_{ii} = -\dfrac{p_i}{x_i}\dfrac{dx_i}{dp_i}$ is the uncompensated own elasticity of demand for good i. Expressions (5.24) and (5.21) differ because they result from different restrictions on the structure of preferences: different matrices are being assumed diagonal.

5.2 Changing Producer Prices

The simple relaxation of the fixed producer price assumption has, perhaps suprisingly, no effect at all on the optimal tax formulae in (5.16) as long as producer prices result from competitive behavior and there are no pure profits that are not taxed away by the government.

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

$$f(\underset{\sim}{z}) = 0 \qquad (5.25)$$

where, as in Section 3, $\underset{\sim}{z}$ is the production vector in the commodities $0, 1, \ldots, N$. By the assumption of competitive behavior, we know that the producer prices $\underset{\sim}{q}$ are proportional to the vector of derivatives of f, $df = (f_0, f_1, \ldots, f_N)$. Without any loss of generality, we may set this proportionality constant equal to one 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 homogenous of degree zero in all commodities. By Euler's Theorem, profits are $q \cdot z = 0$. Thus, the government's optimization problem becomes:

$$\max_{p} V(p) \text{ subject to } f(x + R) = 0 \qquad (5.26)$$

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 t can be brought about by changes in p. This is accomplished by noting that

$$dp = dt + dq = dt + d(df) = dt + H(dx + dR) \qquad (5.27)$$

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

$$dp = dt + H(S - \frac{dx}{dy} x')dp \qquad (5.28)$$

or

$$dp = [I - H(S - \frac{dx}{dy} \cdot x')]^{-1}dt$$

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 (5.28) simplifies to:

$$dp = [I - HS]^{-1} dt = \Omega \, dt \qquad\qquad (5.29)$$

As long as Ω is of full rank, we may control t indirectly through p.

The Lagrangian corresponding to (5.26) yields the first order conditions:

$$- \lambda \, x_i - \mu \sum_j f_j \frac{dx_j}{dp_i} = 0 \qquad\qquad i = 1, \ldots, N \qquad (5.29)$$

where $\lambda = \frac{dV}{dy}$ and μ is the Lagrange multiplier on the production constraint. Since $p \cdot x = 0$,

$$\sum_j p_j \frac{dx_j}{dp_i} + x_i = 0 \qquad\qquad (5.30)$$

Using this and the fact that $q = df$, we may express (5.29) as:

$$\lambda \, x_i + \mu \left[\sum_j t_j \frac{dx_j}{dp_i} + x_i \right] = 0 \qquad\qquad (5.31)$$

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:

$$V(p, (1-\tau)y) - \mu \, f(x + R) \qquad\qquad (5.32)$$

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

$$- \lambda x_i + \lambda(1-\tau) \frac{dy}{dp_i} + \mu[\sum_j t_j \frac{dx_j}{dp_i} + x_i - (1-\tau) \frac{dy}{dp_i}] = 0 \qquad (5.33)$$

It is straightforward to show that if τ may be freely varied, then the N + 1 first-order conditions are solved for $\underset{\sim}{t} = 0$ 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 in (5.33), given τ. Unless profits taxes just happen to equal $\underset{\sim}{q} \cdot \underset{\sim}{R}$, we again face an optimal tax problem.

If $\tau = 1$, so that all profits are taxed away, then (5.33) reduces to the previous optimal tax program, (5.31). 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 $\underset{\sim}{p}$, the derivatives $\frac{dx_j}{dp_i}$ in (5.33) include the indirect effect of p_i on profits through changes in production:

$$\frac{dx_j}{dp_i} = \frac{dx_j}{dp_i}\Big|_{\underset{\sim}{q}} + \frac{dx_i}{dy'} \cdot (1-\tau) \frac{dy}{dp_i} \qquad (5.34)$$

where $y' = (1-\tau)y$.

Using (5.34), the Slutsky equation, and the definition of α, the social marginal utility of income, from (5.5), we may rewrite (5.33) as:

$$\sum_j S_{ij} t_j = (\frac{\mu - \alpha}{\mu})(x_i - (1-\tau) \frac{dy}{dp_i}) \qquad (5.35)$$

which differs from (5.6) only through the replacement of x_i with $(x_i - (1-\tau)\frac{dy}{dp_i})$.

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 in which revenue is extracted, then one can show that (5.35) 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:

$$f(\underset{\sim}{x}) = \hat{f}(x_1, \ldots, x_N) - x_0 \qquad (5.36)$$

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

$$HS = \left(\begin{array}{c|c} H_{00} & 0 \\ \hline 0 & \hat{H} \end{array}\right)\left(\begin{array}{c|c} S_{00} & S_0' \\ \hline S_0 & \hat{S} \end{array}\right) = \left(\begin{array}{c|c} H_{00}S_{00} & 0 \\ \hline 0 & \hat{H}\hat{S} \end{array}\right) \qquad (5.37)$$

where $\underset{\sim}{S_0'} = (S_{01}, \ldots, S_{0N})$ 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, $\underset{\sim}{p} = (p_1, \ldots, p_N)$, can be expressed (using 5.29) in the neighborhood of the optimum as:

$$d\hat{\underset{\sim}{p}} = [I - \hat{H}\hat{S}]^{-1} d\hat{\underset{\sim}{t}} = \hat{\Omega} d\hat{\underset{\sim}{t}} \quad . \tag{5.38}$$

where $\hat{\underset{\sim}{t}} = (t_1, \dots, t_N)$. That is, $d\hat{\underset{\sim}{p}}$ doesn't depend on the demand for x_0. From (5.38), we may express the first-order change around $\hat{\underset{\sim}{t}} = 0$ in $\hat{\underset{\sim}{x}}$, the vector of taxed goods, as:

$$\Delta \hat{\underset{\sim}{x}} = \hat{S} \Delta \hat{\underset{\sim}{p}} = \hat{S} \hat{\Omega} \Delta \hat{\underset{\sim}{t}} = \hat{S} \hat{\Omega} \hat{\underset{\sim}{t}} = \hat{S} \hat{\Omega} \hat{S}^{-1} \hat{S} \hat{\underset{\sim}{t}} \tag{5.39}$$

The elements of the vector $\hat{S} \hat{\underset{\sim}{t}}$ are described in (5.35). By the envelope theorem and the fact that $q_0 = 1$, we may solve for the term $\dfrac{dy}{dp_i}$:

$$\frac{dy}{dp_i} = \sum_j z_j \frac{dq_j}{dp_i} = \sum_{j>0} z_j \sum_k H_{jk} S_{ki} = \sum_{j>0} z_j \sum_{k>0} H_{jk} S_{ki} \tag{5.40}$$

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

$$\frac{dy}{d\hat{\underset{\sim}{p}}} = S H \underset{\sim}{z} \tag{5.41}$$

where $\hat{\underset{\sim}{z}} = (z_1, \dots, z_N)$. But by assumption, $\underset{\sim}{z}$ is zero, except in good zero, so $\hat{\underset{\sim}{z}} = \hat{\underset{\sim}{x}}$. Since, also by assumption, $\tau = 0$, it follows from (5.41) that

$$\hat{S}\hat{\underset{\sim}{t}} = - \left(\frac{\mu - \alpha}{\mu}\right)(I - \hat{S}\hat{H})\hat{\underset{\sim}{x}} \tag{5.42}$$

Substituting (5.42) into (5.39), we obtain:

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

$$\tag{5.43}$$

$$= - \left(\frac{\mu - \alpha}{\mu}\right) \hat{S} \hat{\Omega} \hat{\Omega}^{-1}\hat{S}^{-1}\hat{\underset{\sim}{x}} = \left(\frac{\mu - \alpha}{\mu}\right)\underset{\sim}{x}$$

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 (5.41) for $\frac{dy}{d\underset{\sim}{p}}$ simplifies to

$$\frac{dy}{dp_i} = z_i H_{ii} S_{ii} \qquad\qquad (5.44)$$

which, if we again asume that all revenue raised is spent on the numeraire ($z_i = x_i$ for $i > 0$), allows us to rewrite (5.35) as

$$- S_{ii} t_i = \left(\frac{\mu - \alpha}{\mu}\right)\left(1 - (1-\tau)H_{ii}S_{ii}\right) \qquad\qquad (5.45)$$

or

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

where $\varepsilon_{ii} = - S_{ii}\frac{p_i}{x_i}$, $\sigma_{ii} = \frac{1}{H_{ii}}\frac{q_i}{x_i}$ 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.)

5.3 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 $\underset{\sim}{z}$ be the production function and output of the private sector, and introduce $g(\cdot)$ and $\underset{\sim}{s}$ 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, $\underset{\sim}{s}$. Thus, the government's expanded choice problem is:

$$\underset{\underset{\sim}{p},\underset{\sim}{s}}{\max} \; V(\underset{\sim}{p}, (1-\tau)y) \text{ subject to } f(\underset{\sim}{x} + \underset{\sim}{R} - \underset{\sim}{s}) = 0 \text{ and } g(\underset{\sim}{s}) = 0 \qquad (5.46)$$

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 $\underset{\sim}{p}$. With respect to $\underset{\sim}{s}$ we get:

$$\lambda \, (1-\tau) \, \frac{dy}{ds_i} - \mu \, (\underset{j}{\Sigma} \, f_j \, \frac{dx_j}{dy'} \, (1-\tau) \, \frac{dy}{ds_i} - f_i) - \zeta \, g_i = 0 \qquad (5.47)$$

Using the normalization $\underset{\sim}{q} = df$ and the consumer's budget constraint, we rewrite this as:

$$\lambda \, (1-\tau) \, \frac{dy}{ds_i} - \mu \, (- \underset{j}{\Sigma} \, t_j \, \frac{dx_j}{dy'} \, (1-\tau) \, \frac{dy}{ds_i} + (1-\tau) \, \frac{dy}{ds_i} - f_i) - \zeta \, g_i = 0$$

or $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.48)$

$$g_i = \frac{\mu}{\zeta} f_i - (\frac{\mu - \alpha}{\zeta})(1-\tau)\frac{dy}{ds_i}$$

where, as before, $\alpha = \lambda + \mu \frac{dR}{dy}$ is the social marginal utility of income. Thus, there are two important cases in which efficient overall production $(\frac{f_i}{f_j} = \frac{g_i}{g_j})$ 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 rela-

tive 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 de-

cision. 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 can-

not 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 (5.48) for the

case of independent production. Considering $\frac{dy}{ds_i}$, we have (using the envelope

theorem and independence assumption):

$$\frac{dy}{ds_i} = \sum_j z_j \frac{dq_j}{dz_j} \cdot \frac{dx_j}{dy'} \cdot (1-\tau)\frac{dy}{ds_i} - z_i \frac{dq_i}{dz_i} \qquad (5.49)$$

which, using the facts that $\underset{\sim}{q} = df$ and $d\underset{\sim}{q} = H$, and the assumption that all

government expenditures are on the numeraire commodity, $(\hat{x} = \hat{z})$ we may solve as:

$$\frac{dy}{ds_i} = - \frac{x_i H_{ii}}{1 - (1-\tau) \sum_j x_j H_{jj}\frac{dx_j}{dy'}} = - \frac{f_i/\sigma_{ii}}{\Gamma} \qquad (5.50)$$

where σ_{ii} is the supply elasticity for good i, and Γ must be positive for a

stable solution. Thus (5.48) yields:

$$\frac{g_i}{g_j} = \frac{f_i}{f_j} \cdot \left(\frac{1 + k/\eta_{ii}}{1 + k/\eta_{jj}}\right) \qquad (5.51)$$

where $k = (1-\tau)\left(\frac{\mu - \alpha}{\mu}\right)/\Gamma > 0$. This result says that, relative to production efficiency, private sector production should be subsidized in that commodity in relatively inelastic supply.

5.4 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, 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 a tax is:

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

$$= y - E(p_0, V(p_1, y)) - R \tag{5.52}$$

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 monitonically 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

$$E(p_0, V(p_1, y)) + \pi\left(R - (p_1 - p_0) \cdot x\right) \tag{5.53}$$

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

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

$$= E(p_1, V(p_0, y)) - y - R \tag{5.54}$$

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:

$$E(p_1, V(p_0, y)) - \pi (R - (p_1 - p_0) \cdot x) \qquad (5.55)$$

which yields first-order conditions:

$$- x_i + \pi [\sum_j t_j \frac{dx_j}{dp_i} + x_i] = 0 \qquad (5.56)$$

which looks like the one derived from (5.53). 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.

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 disappear. 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.

6. Optimal Taxation and the Structure of 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 express 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.

Before we continue our analysis, we note that equation (5.6), which expresses the optimal taxes $\hat{t} = (t_1, \ldots, t_N)$ in terms of the substitution matrix $\hat{S}$ and the purchases $\hat{x}$ in the commodities $1, \ldots, N$, may be extended to include the untaxed numeraire good, zero. Obviously, adding a term multiplied by t_0 to each of the N first-order conditions has no effect, since $t_0 = 0$. Moreover, since[11]

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

(6.1)

$$= -\sum_{k=0}^{N} p_k \left(-\left(\frac{\mu - \alpha}{\mu} \right) x_k \right) = -\left(\frac{\mu - \alpha}{\mu} \right) x_0$$

we may rewrite (5.6) as

$$S\underset{\sim}{t} = -\left(\frac{\mu - \alpha}{\mu} \right) \underset{\sim}{x}$$

(6.2)

This means that the optimal tax formula may be interpreted as calling for the proportional reduction in the purchases of _all_ goods, not just taxed goods.

This symmetry is required by the fact that the numeraire is arbitrarily chosen.

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 $\underset{\sim}{x}$ equal $\underset{\sim}{\tilde{x}} - \underset{\sim}{\bar{x}}$ where $\underset{\sim}{\tilde{x}}$ is the consumption vector and $\underset{\sim}{\bar{x}}$ the endowment vector. Thus, we may rewrite the indirect utility function $V(\underset{\sim}{p})$, which implicitly holds $\underset{\sim}{\bar{x}}$ as fixed, as $V(\underset{\sim}{p}, \underset{\sim}{p} \cdot \underset{\sim}{\bar{x}})$, 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 $\underset{\sim}{\tilde{x}}$, and also all the combinations of price vector $\underset{\sim}{p}$ and total endowment income y such that $V(\underset{\sim}{p}, y)$ equals (strictly speaking, at most equals) some constant utility level $\bar{U}$. Choose the vector of prices that minimizes $\underset{\sim}{p}*/y$, given $\underset{\sim}{\tilde{x}}$. The resulting value is the distance function $D(\underset{\sim}{\tilde{x}}, \bar{U})$. Algebraically, the problem is:

$$\min_{\underset{\sim}{p}} \; (p^* \cdot \underset{\sim}{\tilde{x}})/y \quad \text{subject to} \quad V(\underset{\sim}{p^*}, y) \leqslant \bar{U} \qquad (6.3)$$

It is explained diagrammatically in Figure 6.1, for the case of two goods. For simplicity, we assume that $\underset{\sim}{\tilde{x}}$ is on the indifference curve corresponding to the utility level $\bar{U}$, although only the scale of $D(\cdot)$ and not the price vector chosen would be affected by increasing or decreasing $\underset{\sim}{\tilde{x}}$ along the ray shown. This is easily verified from inspection of (6.3), since minimizing $(\underset{\sim}{p^*} \cdot \underset{\sim}{\tilde{x}})/y$ is equivalent to maximizing $(\underset{\sim}{p^*} \cdot \underset{\sim}{\tilde{x}})/y$ for any $\lambda > 0$. By choosing $\underset{\sim}{\tilde{x}}$ to be just feasible, given $\bar{U}$, we will obtain a value $D(\underset{\sim}{\tilde{x}}, \bar{U}) = 1$.

The figure depicts two different combinations of p^* and y, indexed 1 and 2, that satisfy $V(\underset{\sim}{p^*}, y) = \bar{U}$. Since the price vector $\underset{\sim 2}{p^*}$ results in a tangency away from $\underset{\sim}{\tilde{x}}$, purchase of $\underset{\sim}{\tilde{x}}$ 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 $\underset{\sim}{\tilde{x}}$. (A flatter budget line would again necessitate an increase in expenditure to purchase $\underset{\sim}{\tilde{x}}$.) Thus, the price vector chosen, given $\underset{\sim}{\tilde{x}}$ and $\bar{U}$, is tangent to the indifference curve corresponding to $\bar{U}$ at point $\underset{\sim}{\tilde{x}}$ (or, more generally, if $\underset{\sim}{\tilde{x}}$ is not on the indifference curve, at the point on the indifference curve on the ray through $\underset{\sim}{\tilde{x}}$ from the origin). Just as the indirect expenditure function chooses consumption, given prices and utility, the distance function chooses 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 $\underset{\sim}{\tilde{x}}$ are those inverse demands:

$$\frac{\partial D}{\partial \underset{\sim}{\tilde{x}}_i} = a_i(\underset{\sim}{\tilde{x}}, \bar{U}) = \frac{p_i^*}{y} \qquad (6.4)$$

Figure 6.1

The Distance Function
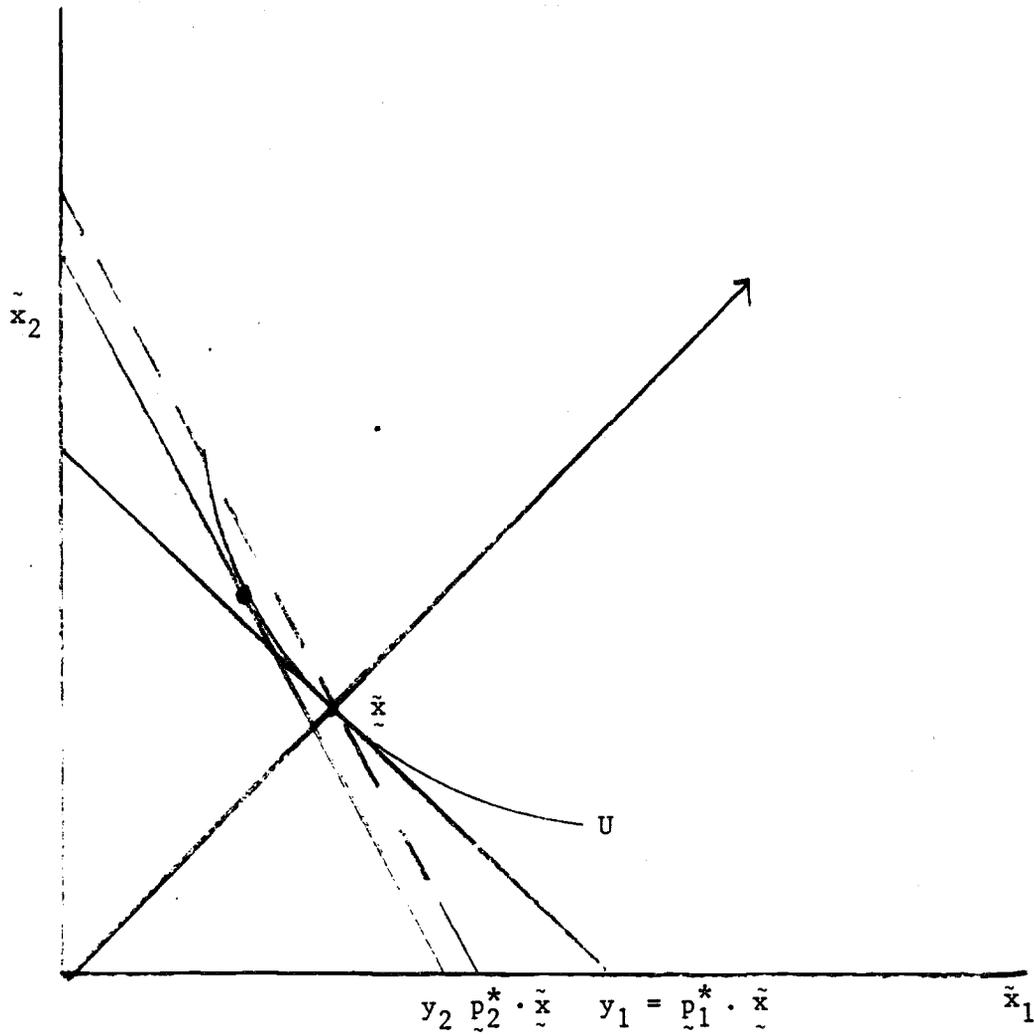
The Hessian of the distance function is referred to as the Antonelli matrix $A = (a_{ij})$.[12]

Now, consider the actual price vector that prevails, $\underset{\sim}{p}$, and choose $\tilde{\underset{\sim}{x}}$ such that $\tilde{\underset{\sim}{x}} = \tilde{\underset{\sim}{x}}_C(\underset{\sim}{p}, \bar{U})$. Then, by construction, $\underset{\sim}{p}^* = \underset{\sim}{p}$ and $y = E(\underset{\sim}{p}, \bar{U})$ solve (6.3), and we have the identity (from (6.4))

$$a_i\left(\tilde{\underset{\sim}{x}}(\underset{\sim}{p}, \bar{U}), \bar{U}\right) = \frac{p_i}{E(\underset{\sim}{p}, \bar{U})} \tag{6.5}$$

Multiplying (6.5) through by $E(\underset{\sim}{p}, \bar{U})$, and differentiating with respect to each price, we obtain conditions which can be stacked to yield:

$$E(\underset{\sim}{p}, \bar{U})\, AS = I - a\hat{x}_C'(\underset{\sim}{p}, \bar{U}) \tag{6.6}$$

where $\underset{\sim}{a} = (a_0, \ldots, a_N)$. Evaluated at $\bar{U} = V(\underset{\sim}{p}, \underset{\sim}{p} \cdot \bar{x})$, this yields:

$$(\underset{\sim}{p} \cdot \bar{x})AS = I - a\tilde{x}(\underset{\sim}{p}, \underset{\sim}{p} \cdot \bar{x}) \tag{6.7}$$

Multiplying both sides of (6.2) by $(\underset{\sim}{p} \cdot \bar{x})A$, and using the fact that since $\underset{\sim}{a}$ is homogeneous of degree zero with respect to $\tilde{\underset{\sim}{x}}$, $A\tilde{\underset{\sim}{x}} = 0$, we obtain:

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

$$= \underset{\sim}{a}(R + \underset{\sim}{t} \cdot \bar{x}) + \left(\frac{\mu - \alpha}{\mu}\right)(\underset{\sim}{p} \cdot \bar{x})A\bar{x} \tag{6.8}$$

where $R = \underset{\sim}{t} \cdot \underset{\sim}{x}$ is tax revenue. Using the fact that $t_0 = 0$ to eliminate $\left(\frac{\mu - \alpha}{\mu}\right)$, we obtain (Deaton 1981b):

$$\theta_i = \frac{t_i}{p_i} = \left(\frac{R + \underset{\sim}{t} \cdot \underset{\sim}{\bar{x}}}{\underset{\sim}{p} \cdot \underset{\sim}{\bar{x}}} \right) \left(1 - \frac{(A\bar{x})_i}{(A\bar{x})_0} \cdot \frac{a_0}{a_i} \right)$$

(6.9)

$$= \nu \left(1 - \frac{\sum_j \bar{x}_j \, \partial \ln a_i / \partial \tilde{x}_j}{\sum_j \bar{x}_j \, \partial \ln a_0 / \partial \tilde{x}_j} \right)$$

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

$$\theta_i - \theta_j = \nu' \sum_j \bar{x}_j \frac{\partial \ln (a_j/a_i)}{\partial \tilde{x}_j}$$

(6.10)

where $\nu' = \nu / (\sum_j \bar{x}_j \, d \ln a_0 / d \tilde{x}_j)$. From (6.10), we see that a sufficient condi-
tion 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:

$$D(\tilde{x}, \bar{U}) = f\left(\tilde{x}_1, \tilde{x}_2, \bar{U}, \Phi(\tilde{x}_3, \bar{U}) \right)$$

(6.11)

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.[13] 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 separ-
ability and differs from the separability of the direct and indirect utility
functions. Indeed, they are the same only if the utility function is homogen-
eous 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 (presumably leisure), the sufficient (and now necessary, as well) condition for uniform taxation of commodities is implicit separability from leisure (Sandmo 1974). It is also possible in this case to say more about which goods will be taxed more heavily if weak separability but not homogeneity is satisfied. We begin by rewriting (6.10) as:

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

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 $\left(\frac{\mu - \alpha}{\mu}\right) > 0$: marginal excess burden is positive). Since $a_j/a_i = p_j/p_i = U_j/U_i$,

$$\frac{d \, \ln (U_j/U_i)}{d\tilde{x}_0} = \frac{d \, \ln (a_j/a_i)}{d\tilde{x}_0} = \frac{\partial \, \ln (a_j/a_i)}{\partial \tilde{x}_0} + \frac{\partial \, \ln (a_j/a_i)}{\partial \bar{U}} \cdot \frac{dU}{\partial \tilde{x}_0} \tag{6.13}$$

(Deaton 1981a). Thus, when utility is separable into goods and leisure, (6.12) becomes:

$$\theta_i - \theta_j = -\nu' \, \bar{x}_0 \, \frac{\partial \, \ln (a_j/a_i)}{d\bar{U}} \cdot \frac{dU}{\partial \tilde{x}_0}. \tag{6.14}$$

so that taxes will be higher on those goods that are necessities, as 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:

$$\tilde{x}_i(\underset{\sim}{p},\ p_0\bar{x}_0) = c_i + \left(\frac{b_i}{\underset{j}{\Sigma}\ b_j}\right) \cdot \frac{(p_0 x_0 - \underset{j}{\Sigma}\ p_j c_j)}{p_i} \tag{6.15}$$

often used in empirical work comes from the Stone-Geary utility function:

$$U(\underset{\sim}{\tilde{x}}) = \underset{i}{\pi}\ (\tilde{x}_i - a_i)^{b_i} \tag{6.16}$$

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.[14] As stressed in Section 3, these consideratons 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 optional tax rules influenced by equity considerations? Second, if we choose leisure as numeraire and admit lump sum taxes that <u>cannot</u> vary across individuals, when will uniform commodity taxes be optimal? That is, when will linear income taxation be optimal?

We begin by specifying a social welfare function of the form

$$W = W(U^1,\ \ldots,\ U^H) \tag{6.17}$$

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 $\underset{\sim}{\hat{t}} = (t_1,\ \ldots,\ t_N)$:

$$- \sum_h W_h \; \lambda^h x_i^h + \mu \; [\sum_j t_j \; \sum_h \frac{x_j^h}{dp_i} + x_i] = 0 \qquad i=1,\ldots,N \qquad (6.18)$$

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

$$\alpha^h = W_h \; \lambda^h + \mu \frac{dR}{dy^h} \qquad (6.19)$$

we may express the conditions (6.18) as:

$$\sum_j t_j \; S_{ij} = -(\frac{\mu - \tilde{\alpha}_i}{\mu})x_i \qquad i=1,\ldots,N \qquad (6.20)$$

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

$$\tilde{\alpha}_i = \sum_h (\frac{x_i^h}{x_i}) \; \alpha^h \qquad (6.21)$$

is the average value of α, weighted by individual consumption shares of good i. This neat formulation (due to Diamond 1975) shows that the "equal proportional 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

$$\tilde{\alpha}_i = \text{Cov} \; ((\frac{x_i^h}{x_i}), \; \alpha^h) + \frac{1}{H} \sum_h \alpha^h \qquad (6.22)$$

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 (6.20) 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

$$\frac{\theta_i}{\theta_j} = \frac{\varepsilon_{jj}}{\varepsilon_{ii}} \cdot \left(\frac{\mu - \tilde{\alpha}_i}{\mu - \tilde{\alpha}_j}\right) \tag{6.23}$$

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 additional first order condition

$$- \sum_h W_h \lambda^h + \mu \left[\sum_j t_j \sum_h \frac{dx_j^h}{dy^h} - H\right] = 0 \tag{6.24}$$

to be added to the N condition in (6.18). This new condition simplifies to

$$\mu = \frac{1}{H} \sum_h \alpha^h = \bar{\alpha} \tag{6.25}$$

Thus, (6.20) becomes

$$\sum_j t_j S_{ij} = \frac{\text{cov}\left(\frac{x_i}{\bar{x}_i}\right), \, \alpha^h)}{\bar{\alpha}} \, x_i \qquad i=1,\ldots,N \qquad (6.26)$$

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. 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 $\underset{\sim}{\theta} = (\frac{t_1}{p_1}, \ldots, \frac{t_N}{p_N})$ 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 on 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 causes 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 nondistortionary, 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.

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 <u>sum</u> 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:

$$V(\underset{\sim}{p}; G) = \max_{\underset{\sim}{x}} U(\underset{\sim}{x}; G) \text{ subject to } \underset{\sim}{p} \cdot \underset{\sim}{x} = 0 \qquad (7.1)$$

with $\frac{\partial V}{\partial G} = \frac{\partial U}{\partial G}\Big|_{\underset{\sim}{x} = \underset{\sim}{x}_C(\underset{\sim}{p}; G)}$. The production function is $f(\underset{\sim}{x}; G) = 0$. 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:

$$L = HV(\underset{\sim}{p}; G) - \mu f(\underset{\sim}{x}; G) \tag{7.2}$$

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

$$- H \lambda x_i^h - \mu \sum_j f_j \frac{dx_j}{dp_i} = 0 \qquad\qquad i=1,\ldots,N \tag{7.3}$$

where λ and μ are defined in the usual way. As in Subsection 5.2, we use the fact that $\underset{\sim}{p} \cdot \underset{\sim}{x}^h = 0$ for each individual h to obtain:

$$\lambda x_i + \mu \left[\sum_j t_j \frac{dx_j}{dp_i} + x_i \right] = 0 \qquad\qquad i=1,\ldots,N \tag{7.4}$$

where $x_i = \sum_h x_i^h = H x_i^h$. As before, this may be rewritten:

$$S\underset{\sim}{t} = - \left(\frac{\mu - \alpha}{\mu} \right) \underset{\sim}{x} \tag{7.5}$$

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:

$$H \frac{\partial U}{\partial G} - \mu \left[\sum_i f_i \frac{dx_i}{dG} + f_G \right] = 0 \tag{7.6}$$

which yields (since $\lambda \equiv dU/dx_0^h$, $q_0 = f_0 = 1$ and $\underset{\sim}{p} \cdot \underset{\sim}{x}^h = 0$):

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

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

$\dfrac{f_G}{f_0}$ 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 $\dfrac{dR}{dG} > 0$ (Diamond and Mirrlees 1971). The other term, $\dfrac{\mu}{\lambda}$, 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 $\dfrac{\mu}{\lambda}$ 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 $\underset{\sim}{t}$ to obtain

$$\underset{\sim}{t}'S\underset{\sim}{t} = - \left(\frac{\mu - \alpha}{\mu}\right)R \qquad (7.8)$$

which, by the negative semi-definiteness of S, implies that $\mu > \alpha$ for positive revenue. But $\alpha > \lambda$ (see equation 5.5) only if $\dfrac{dR}{dy}$ is positive. If taxed goods are, on average (weighted by tax rates) inferior, $\dfrac{dR}{dy} < 0$ 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 taxes xN 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 $\frac{\partial V}{\partial x_N} = \frac{\partial U}{\partial x_N}\Big| x_C(p; x_N)$.

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:

$$- \lambda x_i + H\frac{\partial U}{\partial x_N} \cdot \frac{dxN}{dp_i} + \mu[x_i + \sum_j t_j\frac{dxj}{dp_i}] = 0 \qquad (i=1,\dots,N) \qquad (7.9)$$

or

$$- \lambda x_i + \mu[x_i + \sum_j t_j^* \frac{dx_j}{dp_i}] = 0 \qquad (i=1,\dots,N) \qquad (7.10)$$

where $t_i = t_i^*$ $\qquad$ i = 1, ..., N-1

$$t_i^* - H\frac{\partial U}{\partial x_N} \qquad i = N$$

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 Pigouian 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 noncompetitive 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

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

where $\underset{\sim}{p^*}$ is the subset of $\underset{\sim}{p}$ that may be adjusted. Note that unless at least two prices are fixed, equiproportional, nondistortionary taxation is possible.

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

$$- \lambda x_i + \mu [\sum_j t_j \frac{dx_j}{dp_i} + x_i] = 0 \qquad\qquad p_i \in \underset{\sim}{p^*} \qquad (7.12a)$$

$$- \lambda + \mu [\sum_j t_j \frac{dx_j}{dy} + 1] = 0 \qquad\qquad (7.12b)$$

which may be written as:

$$\sum_j S_{ij} t_j = - (\frac{\mu - \alpha}{\mu}) \qquad\qquad p_i \in \underset{\sim}{p^*} \qquad (7.13a)$$

$$\mu = \alpha \qquad\qquad (7.13b)$$

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

$$\sum_j S_{ij} t_j = 0 \qquad\qquad p_i \ \epsilon \ \underset{\sim}{p}^* \qquad (7.14)$$

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 $t_0 = t_3 = \ldots, t_N = 0$. Then there is one condition, corresponding to the choice of t_1. Using compensated elasticities $\epsilon_{ij} = - S_{ij} \dfrac{x_j}{p_i}$, we may express this as:

$$\theta = -\theta_2 \ \epsilon_{12}/\epsilon_{11} \qquad (7.15)$$

where $\theta_i = t_i/p_i$ is the proportional tax on good i. Since $\epsilon_{11} < 0$, this calls for a tax on good 1 (assuming $\theta_2 > 0$), if $\epsilon_{12} > 0$, and a subsidy if $\epsilon_{12} < 0$. 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:

$$\theta_1 = - \sum_{j \neq 1} \theta_j \ \epsilon_{1j}/\epsilon_{11} \qquad (7.16)$$

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 result about
separability of factors in production that closely parallels those on the con-
sumption 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 com-
pleteness 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 indepen-
dent 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 com-
modities 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 (5.19), which we write here for convenience

$$\frac{\theta_1}{\theta_2} = \frac{\epsilon_{12} + \epsilon_{21} + \epsilon_{20}}{\epsilon_{12} + \epsilon_{21} + \epsilon_{10}} \qquad (7.17)$$

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

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

where $U^1(\cdot) = U^2(\cdot)$ and ϵ_{10} and ϵ_{20} may be expressed as

$$\epsilon_{i0} = M[- (\frac{U_{22}^j x_j}{U_2^j}) + p_j x_j \frac{d \ln(U_2^j/U_2^i)}{dx_0}] \qquad i=1,2, \ j=2,1 \qquad (7.19)$$

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 $- (\frac{U_{22}^1 x_1}{U_2^1}) > - (\frac{U_{22}^2 x_2}{U_2^2})$, 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 surprising, 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 decisions apply. If we can't 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

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

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 risker 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 <u>inducing</u> 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 (1956-7) 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) shows 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 nonconvex. 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). This is depicted in Figure 7.1. Suppose two individuals, 1 and 2, have identical preferences and consume

Figure 7.1

Optimal Taxation with Nonconvexities
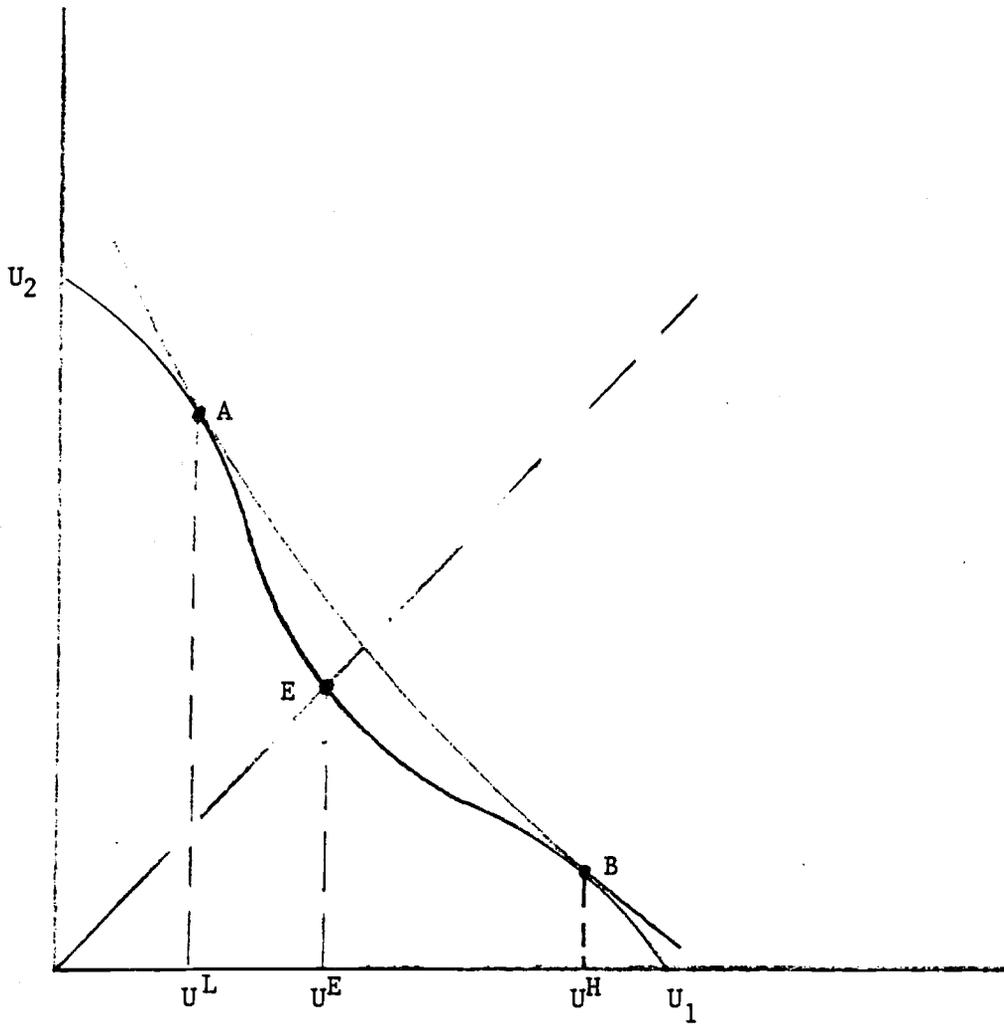
goods and leisure. If we seek to maximize $(U_1 + U_2)$ by choosing individual-specific excise taxes on consumption, the first-order condition will be zero with equal taxes at $U_1 = U_2 = U^E$, by the symmetry of the problem. But this may represent a local minimum, as shown. Social welfare may be improved by choosing either point A or point B. This represents an unequal treatment of equal individuals and may violate proscriptions against such horizontal equity. However, suppose the tax system were randomized so that point A were 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 alloca-
tion.   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
1982a).   This problem undoubtedly is one of the reasons why tax reform is so dif-
ficult 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 remain-
ing distortion, 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 nondistortionary, 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:

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

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 $\underset{\sim}{p}$ is chosen optimally:

$$\frac{\partial \mathcal{L}}{dT} = - \lambda + \mu \left[-\sum_{j} t_{j} \frac{dx_{j}}{dy} + 1\right] = (\mu - \alpha) \tag{8.2}$$

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 $\underset{\sim}{t}$ 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 $\underset{\sim}{t}$ 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):

$$dU = \sum_i \frac{\partial V}{\partial p_i} dt_i - \frac{\partial V}{\partial y} dT = - \lambda \, (\underset{\sim}{x} \cdot d\underset{\sim}{t} + dT) \tag{8.3a}$$

and $\quad dR = d(\underset{\sim}{t} \cdot \underset{\sim}{x} + T) = \underset{\sim}{x} \cdot d\underset{\sim}{t} + \underset{\sim}{t} \cdot d\underset{\sim}{x} + dT = 0 \tag{8.3b}$

which yields

$$dU = \lambda \, \underset{\sim}{t} \cdot d\underset{\sim}{x} \tag{8.4}$$

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:

$$d\underset{\sim}{x} = \frac{d\underset{\sim}{x}}{d\underset{\sim}{p}} d\underset{\sim}{t} - \frac{d\underset{\sim}{x}}{dy} dT = S d\underset{\sim}{t} - \frac{d\underset{\sim}{x}}{dy} (x'd\underset{\sim}{t} + dT) \tag{8.5}$$

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

$$dU = \frac{\lambda}{1 - \underset{\sim}{t} \cdot \dfrac{d\underset{\sim}{x}}{dy}} \cdot \underset{\sim}{t}'S \, d\underset{\sim}{t} \tag{8.6}$$

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, $dU > 0$ if and only if $(1 - t \cdot \frac{dx}{dy}) > 0$ (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 <u>social</u> 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 <u>local</u> movements away from A along M, where there is no first order income effect.) If this is the case, then a revenue-preserving reduction in the divergence between the relative price of x_2 and its social cast, 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 be had by an odd number of different points on M, not just one. In this case, distortions in the price reduction may actually move the consumer <u>away</u> from point A.

Figure 8.1

Prices and Utility
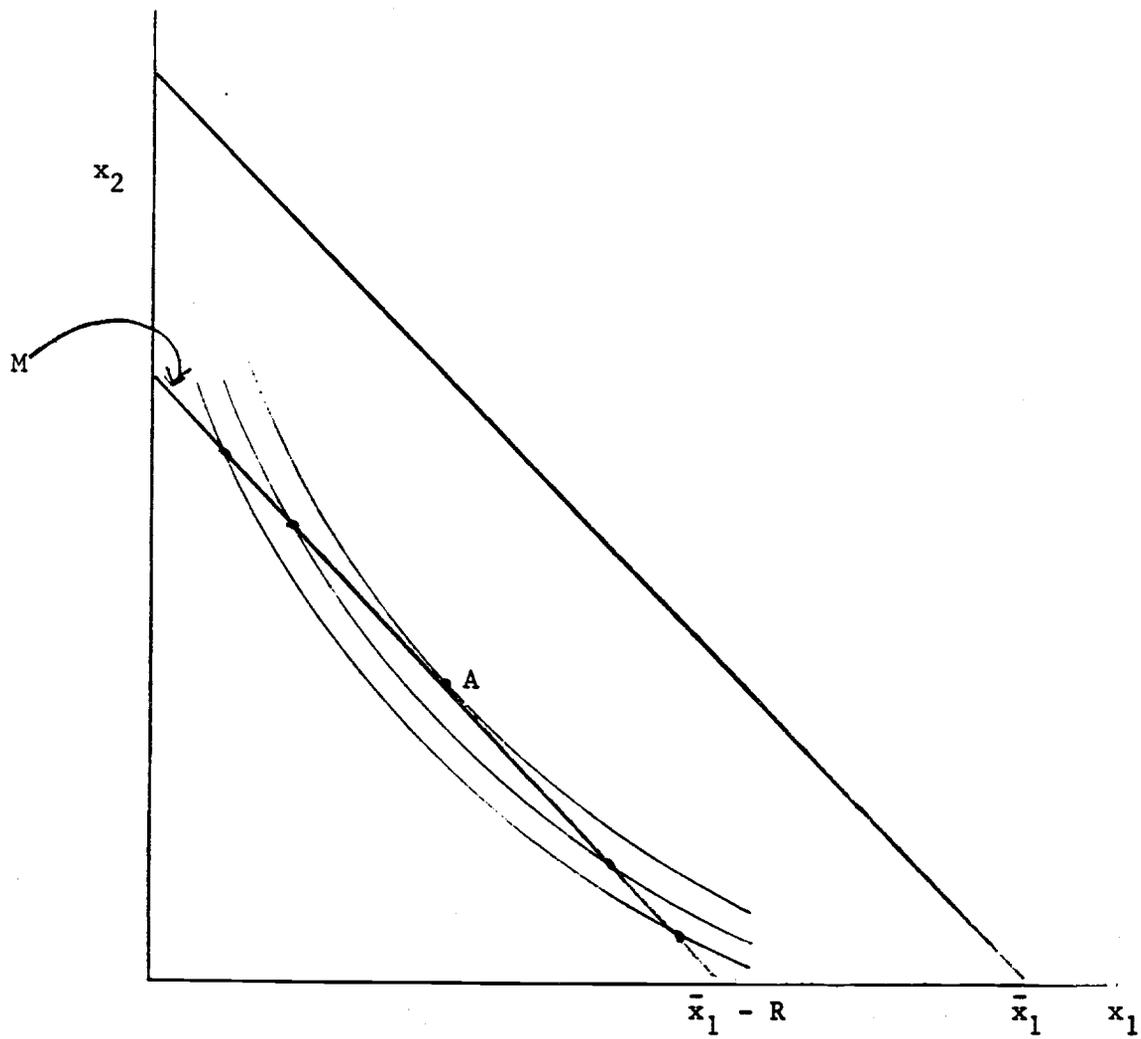
That this possibility is equivalent to the condition derived from (8.6) is demonstrated graphically in Figure 8.2, where an increase in lump sum income above $\bar{x}_1$ causes the consumer to shift from point B to point C, inside the production constraint M. Since the indifference curve slopes at B 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 aplplication 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 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 $\underset{\sim}{p}'S = 0$, (8.6) may be rewritten

$$dU = - \frac{\lambda}{1 - \underset{\sim}{t} \cdot \dfrac{d\underset{\sim}{x}}{dy}} \cdot \underset{\sim}{S_0} \cdot d\underset{\sim}{\hat{t}} \qquad (8.7)$$

where $\underset{\sim}{S_0} = (S_{01}, S_{02}, \ldots, S_{0N})$. A sufficient condition for this to be positive (assuming $\frac{dR}{dy} < 1$) 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

Figure 8.2

Multiple Equilibria with Taxes
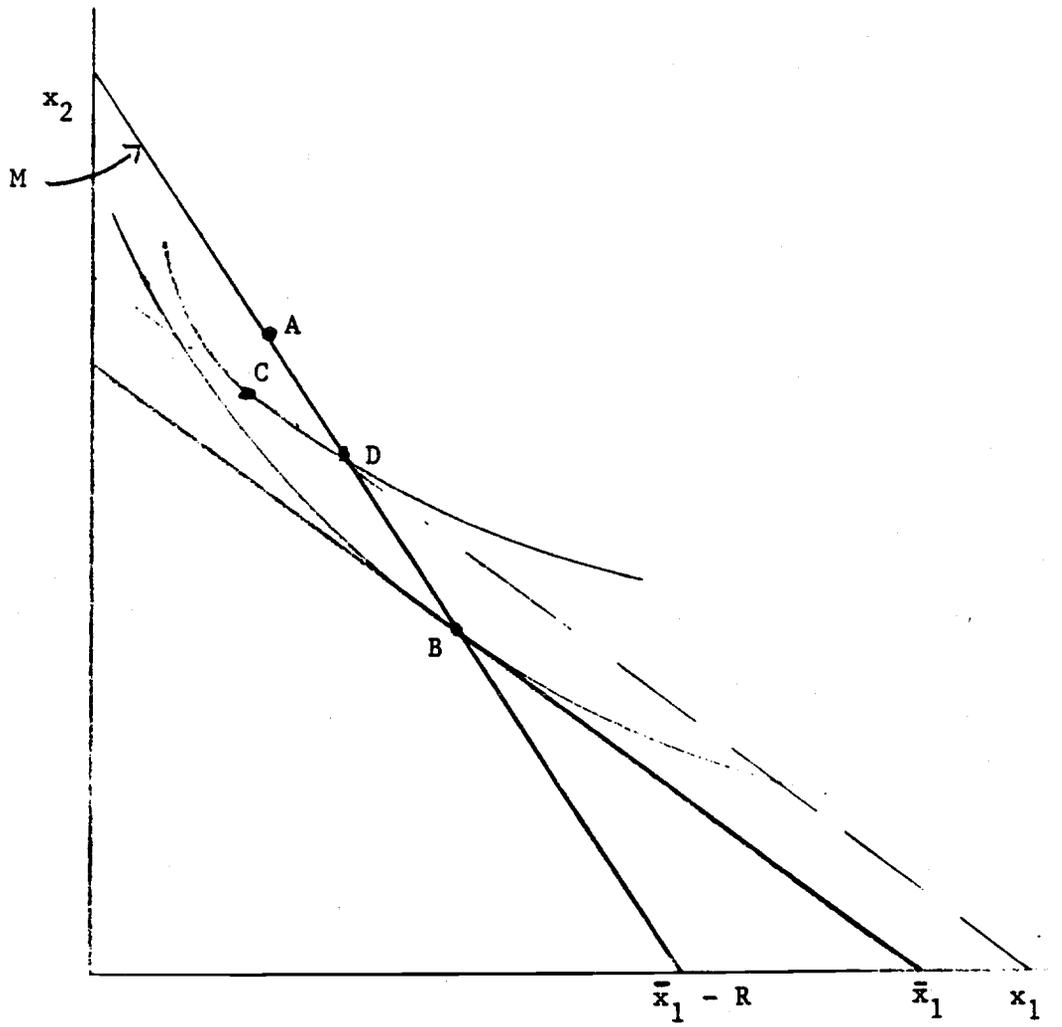
been made (Guesnerie 1977, Diewert 1978), there is little we can say of a concrete nature without further assumptions.

One approach that sidesteps this problem is to characterize observable changes in equilibrium that will result $\underline{if}$ 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 $d(t \cdot x) = 0$, $q \cdot x_1 = q \cdot x_0$, 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 (1953-4), with a generalization provided by Dixit (1975), whose anlaysis we follow. Note that (8.4) still is valid in determining whether a tax change icnreases utility. However, since lump-sum taxes are unavailable, $t \cdot x = 0$. Using (8.5), for $dT = 0$, we have:

$$0 = d(t \cdot x) = x \cdot dt + t \cdot dx = x \cdot dt + t' \frac{S}{\Delta} dt$$

$$= (x' + t' \frac{S}{\Delta}) \, dt \tag{8.8}$$

where $\Delta = 1 - \underset{\sim}{t} \cdot \dfrac{d\underset{\sim}{x}}{dy}$. For the case where $t_0 = 0$ and $\hat{t} = \theta \hat{p}$, we use the homogeneity of S to rewrite this as:

$$(x' - \frac{\theta}{\Delta} \underset{\sim 0}{S'}) \, d\underset{\sim}{t} = 0 \qquad (8.9)$$

which, using the definition of compensated elasticities $t_{ij} = S_{ij} \dfrac{R_j}{x_i}$

$$\underset{i}{\Sigma} \, x_i \, (1 - \frac{\theta}{\Delta} \epsilon_{i0}) \, dt_i = 0$$

From (8.4), we have (for $dT = 0$):

$$dU = -\lambda \, \underset{\sim}{x} \cdot d\underset{\sim}{t} \qquad (8.10)$$

If we assume that $x_i \, (1 - \frac{\theta}{\Delta} \epsilon_{i0}) = \dfrac{dR}{dt_i}$ 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

$$\frac{x_i(1 - \frac{\theta}{\Delta} \epsilon_{i0})}{x_i} \qquad (8.11)$$

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.

Footnotes

1. Dupuit (1844).

2. See, for example, Diewert (1981).

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

4. See Hotelling (1938) for the original statement of this result.

5. We use here the identity $x(\underset{\sim}{p}, E(\underset{\sim}{p},u)) = x_c(\underset{\sim}{p}, u)$.

6. This was pointed out by Hausman (1981a), among others.

7. See Auerbach and Rosen (1980) for further discussion.

8. One can also derive higher order approximations of EB_c. For a comparison of second and third order approximations, see Green and Sheshinski (1979).

9. See Hausman (1981a).

10. This may be seen by substituting (3.8) into (3.9), and considering the Taylor approximation of the change in β:

$$\Delta \beta \approx \frac{d\beta}{d\underset{\sim}{t}} \cdot \Delta \underset{\sim}{t} + 1/2 \, \Delta \underset{\sim}{t}' \left(\frac{d^2\beta}{d\underset{\sim}{t}^2}\right) \Delta \underset{\sim}{t}$$

which yields

$$\Delta \beta \approx - \left(\underset{\sim}{t}' \, \Delta \underset{\sim}{x}(1 + 1/2 \, \underset{\sim}{\alpha} \cdot \Delta \underset{\sim}{q}) + 1/2 \, \Delta \underset{\sim}{t}' \, \Delta \underset{\sim}{x}\right)$$

In deriving a similar measure, Diamond and McFadden (1976) assumed all compensation to be in the form of the numeraire good, in which case $\underset{\sim}{\alpha} \cdot \Delta \underset{\sim}{q} = 0$.

11. This uses the facts that $\sum_{k=0} p_k S_{ki} = 0$ and $\underset{\sim}{p} \cdot \underset{\sim}{x} = 0$, and equation (5.15).

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

13. Because $D(\cdot)$ is homogeneous of degree 1 in $\underset{\sim}{x}$, f must be homogeneous of degree 1 in $\underset{\sim}{x}_1, \underset{\sim}{x}_2$ and Φ and Φ homogeneous of degree 1 in $\underset{\sim}{x}_3$.

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

References

K. Arrow. 1965. Essays in the Theory of Risk-Bearing. Amsterdam: North Holland.

A.B. Atkinson and N.H. Stern. 1974. "Pigou, Taxation and Public Goods." Review of Economic Studies 41, 119-128.

_____ and _____. 1980. "Taxation and Incentives in the UK." Lloyds Bank Review, 43-46.

A.B. Atkinson and J.E. Stiglitz. 1972. "The Structure of Indirect Taxation and Economic Efficiency." Journal of Public Economics 1, 97-119.

_____ and _____. 1976. "The Design of Tax Structure: Direct Versus Indirect Taxation." Journal of Public Economics 6, 55-75.

_____ and _____. 1980. Lectures on Public Economics. New York, N.Y. and Maidenhead, England: McGraw-Hill.

A.J. Auerbach. 1979a. "A Brief Note on a Non-Existent Theorem About the Optimality of Uniform Taxation." Economics Letters 3, 49-52.

_____. 1979b. "The Optimal Taxation of Heterogeneous Capital." Quarterly Journal of Economics, 93, 589-612.

_____. 1981. "Taxes, Firm Financial Policy and the Cost of Capital: An Empirical Analysis." HIER Working Paper No. 915, August.

_____ and H. Rosen. 1980. "Will The Real Excess Burden Please Stand Up?" HIER Working Paper No. 767.

_____, L.J. Kotlikoff and J. Skinner. 1981. "The Efficiency Gains from Dynamic Tax Reform." HIER Working Paper No. 870, December.

B. Baumol and D. Bradford. 1970. "Optimal Departures from Marginal Cost Pricing." American Economic Review 60, 265-283.

R. Boadway and R. Harris. 1977. "A Characterization of Piecemeal Second Best Policy." Journal of Public Economics 8, 169-190.

M. Boiteaux. 1951. "Le 'Revenue Distribuable' et les Pertes Economiques." Econometrica 19, 112-133.

_____. 1956. "Sur le gestion des monopoles publics astreints a l'equilibre budgetaire." Econometrica 24, 22-40. Translated and reprinted in 1971, Journal of Economic Theory 3, 219-40.

D. Bradford and H. Rosen. 1976. "The Optimal Taxation of Commodities and Income." AER Papers and Proceedings 66, 94-101.

J.S. Chipman and J.C. Moore. 1980. "Compensating Variation, Consumer's Surplus, and Welfare." AER 70, 933-940.

References (Continued)

R.Cooter. 1979. "A New Expenditure Function." Economics Letters 2, 103-110.

W.J. Corlett and D.C. Hague. 1953-4. "Complementarity and the Excess Burden of Taxation." Review of Economic Studies 21, 21-30.

P.S. Dasgupta and J.E. Stiglitz. 1972. "On Optimal Taxation and Public Production." Review of Economic Studies 39, 87-103.

A.S. Deaton. 1977. "Equity, Efficiency and the Structure of Indirect Taxation." Journal of Public Economics 8, 299-312.

_____. 1979a. "The Distance Function in Consumer Behaviour with Applications of Index Numbers and Optimal Taxation." Review of Economic Studies 46, 391-406.

_____. 1979b. "Optimally Uniform Commodity Taxes." Economics Letters 2, 357-361.

_____. 1981a. "Optimal Taxes and the Structure of Preferences." Econometrica 49, 1245-1260.

_____. 1981b. "Numeraires and Endowments in Optimal Tax Formulae: A Note," University of Bristol Discussion Paper No. 81/113, September.

G.Debreu. 1954. "A Classical Tax-Subsidy Problem." Econometrica 22, 14-22.

_____. 1951. "The Coefficient of Resource Utilization." Econometrica 19, 273-92.

P.A. Diamond. 1975. "A Many-Person Ramsey Rule." Journal of Public Economics 4, 335-42.

_____ and D.L. McFadden. 1974. "Some Uses of the Expenditure Function in Public Finance." Journal of Public Economics 3, 3-21.

_____ and J. Mirrlees. 1971. "Optimal Taxation and Public Production I: Production Efficient and II: Tax Rules." American Economic Review 61, 8-27 and 261-278.

W.E. Diewert. 1978. "Optimal Tax Perturbations." Journal of Public Economics 10, 139-177.

_____. 1981. "The Measurement of Deadweight Loss Revisited." Econometrica 49, 1225-1244.

A.K. Dixit. 1970. "On the Optimum Structure of Commodity Taxes." American Economic Review 60, 295-301.

_____. 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 Ponte et Chaussées 8. Translated and reprinted in K. Arrow and T. Scitovsky (eds.), 1969, AEA Readings in Welfare Economics, 255-283.

M.S. Feldstein. 1972. "Distributional Equity and the Optimal Structure of Public Prices. American Economic Review 62, 32-36.

_____. 1976. "On the Theory of Tax Reform." Journal of Public Economics 6, 77-104.

_____. 1978. "The Welfare Cost of Capital Income Taxation." Journal of Public Economics 86, S29-51.

E. Foster and H. Sonnenschein. 1970. "Price Distortion and Economic Welfare." Econometrica 38, 281-297.

W.M. Gorman. 1953. "Community Preference Fields." Econometrica 21, 63-80.

_____. 1976. "Tricks with Utility Functions." In Essays in Economic Analysis, M.J. Artis and A.R. Nobay (eds.), London: Cambridge University Press.

H.A.J. Green. 1961. "The Social Optimum in the Presence of Monopoly and Taxation." Review of Economic Studies 29, 66-78.

J. Green and E. Shoshinsni. 1979. "Approximating the Gains from Welfare Reforms." Journal of Public Economics 11, 179-195.

R. Guesnerie. 1977. "On the Direction of Tax Reform." Journal of Public Economics 7, 179-202.

A. Harberger. 1954. "Monopoly and Resource Allocation." American Economic Review 42, 77-87.

_____. 1964. "Taxation, Resource Allocation and Welfare." In The Role of Direct and Indirect Taxes in the Federal Reserve System, Princeton: Princeton University Press for the NBER and Brookings Institution.

_____. 1966. "Efficiency Effects of Taxes on Income From Capital." In Effects of Corporation Income Tax, J. Krzyaniak (ed.).

_____. 1971, "Three Postulates for Applied Welfare Analysis." Journal of Economic Literature 9, 785-97.

T. Hatta. 1977. "A Theory of Piecemeal Policy Recommendations." Review of Economic Studies 44, 1-21.

J.A. Hausman. 1981a. "Exact Consumer's Surplus and Deadweight Loss." American Economic Review 71, 662-676.

_____. 1981b. "Labor Supply." In How Taxes Effect Economic Behavior, H. Aaron and J. Pechman (eds.). Washington, D.C.: Brookings Institution, pp. 27-84.

E. Helpman and E. Sadka. 1982. "Consumption Versus Wage Taxation." Quarterly Journal of Economics 97, 363-372.

J.R. Hicks. 1940. "The Valuation of Social Income." Economica 26

_____. 1942. "Consumers' Surplus and Index Numbers." Review of Economic Studies 9, 126-37.

_____. 1946. Value and Capital. 2nd edition. Oxford University Press.

N. Kaldor. 1939. "Welfare Propositions in Economics." Economic Journal 49, 549-552.

J.A. Kay. 1980. "The Deadweight Loss From a Tax System." Journal of Public Economics 13, 111-120.

M.A. King. 1982a. "An Index of Inequality: With Applications to Horizontal Equity and Social Mobility." Econometrica 50, forthcoming.

_____. 1982b. "Welfare Analysis of Tax Reforms Using Household Data." Journal of Public Economics, forthcoming.

R.G. Lipsey and R.K. Lancaster. 1956-7. "The General Theory of Second Best." Review of Economic Studies 24, 11-32.

A. Marshall. 1920. Principles of Economics. London: Macmillan.

R.C. Merton. 1969. "Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case." Review of Economic Statistics 51, 247-57.

J.A. Mirrlees. 1972. "On Producer Taxation." Review of Economic Studies 39, 105-111.

H. Mohring. 1971. "Alternative Welfare Gain & Loss Measures." Western Economic Journal 9, 349-368.

J.N.J. Muellbauer. 1976. "Community Preferences and the Representative Consumer." Econometrica 44, 979-999.

R.A. Musgrave. 1959. The Theory of Public Finance. New York: McGraw-Hill.

E.A. Pazner and E. Sadka. 1981. "Welfare Criteria for Tax Reforms: Efficiency Aspects." Journal of Public Economics 16, 113-122.

A.C. Pigou. 1947. A Study in Public Finance. 3rd edition. London: Macmillan.

J.W. Pratt. 1964. "Risk Aversion in the Small and in the Large." Econometrica 32, 122-36.

F.P. Ramsey. 1927. "A Contribution to the Theory of Taxation." Economic Journal 37, 47-61.

K.W.S. Roberts. 1980. "Price-Independent Welfare Prescriptions." Journal of Public Economics 13, 277-297.

H. Rosen. 1978. "The Measurement of Excess Burden with Explicit Utility Functions." Journal of Political Economy 86, No. 2, Part 2, S121-136.

M. Rothschild and J.E. Stiglitz. 1970. "Increasing Risk I: A Definition." Journal of Economic Theory 2, 225-243.

P.A. Samuelson. 1947. Foundations of Economic Analysis. Cambridge: Harvard University Press.

_____. 1969. "Lifetime Portfolio Selection by Dynamic Stochastic Programming." Review of Economic Statistics 51, 239-46.

A. Sandmo. 1974. "A Note on the Structure of Optimal Taxation." American Economic Review 64, 701-706.

_____. 1975. "Optimal Taxation in the Presence of Externalities." Swedish Journal of Economics 77, 86-98.

_____. 1981. "Income Tax Evasion, Labour Supply, and the Equity-Efficiency Tradeoff." Journal of Public Economics 16, 265-288.

J.B. Shoven. 1976. "The Incidence and Efficiency Effects of Taxes on Income from Capital." Journal of Political Economy 84, 1261-1284.

K.A. Small and H.S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." Econometrica 105-130.

J. Stiglitz. 1972. "Taxation, Risk Taking and the Allocation of Investment in a Competitive Economy." In Studies in the Theory of Capital Markets, M. Jensen (ed.).

_____ and P.S. Dasgupta. 1971. "Differential Taxation, Public Goods and Economic Efficiency." Review of Economic Studies 38, 151-1974.

S.F. Venti and D.A. Wise. 1982. "Moving and Housing Expenditure: Transactions Costs and Equilibrium." Paper presented at SSRC/NBER Conference on Microdata and Public Economics, Oxford, June.

L. Weiss. 1976. "The Desirability of Cheating Incentives and Randomness in the Optimal Income Tax." Journal of Political Economy 84, 1343-1352.

R.E. Willig. 1976. "Consumer's Surplus Without Apology." American Economic Review 66, 589-597.