

9

## I. INTRODUCTION AND SUMMARY

The purpose of this paper is to provide a background for those who deal with problems of economic policy, either as participants in the policy process or as its observers and interpreters. In the paper I try to elucidate some of the insights that economics (as an intellectual discipline) brings to bear, specifically on the policy problems associated with indirect taxation. For the most part, I conceive of my audience as consisting both of professional economists and of non-economists working in the policy area. For the sake of the latter I attempt to keep the level of discourse as non-technical as possible, but nonetheless some familiarity with basic economic concepts is presupposed.

For professional economists working at or near the frontiers of economics this paper may serve a different purpose--that of providing a certain perspective on the interface between the world of the policy maker and the world of economic science. This interface leaves out many matters that regularly concern each group: the practitioners and the scientists. We do not here worry about how best to bring recalcitrant taxpayers into compliance, or how to organize the timing of tax payments, or how to train a cadre of lower-level administrators. Nor do we worry, on the other side, about the niceties of optimization or suboptimization under well-defined constraints, or about implementing hypothetical systems of social values. Many of the constraints facing policy makers are political in nature, and sometimes change with amazing speed. The designers of tax systems may be aware of these constraints, but their overriding task (and ours) is to try to build tax structures that are robust, and capable of serving society well, even as the political winds rise and fall in intensity and shift in

## ISSUES OF TAX POLICY--INDIRECT TAXATION

Arnold C. Harberger

University of Chicago  
and

University of California, Los Angeles

be used to analyze indirect taxes and similar policies in a general equilibrium framework. In addition it introduces an issue that has been too little discussed in the literature: the sense in which, for the analysis of tax matters, the very definition of a commodity is a policy variable. When we determine the "coverage" of a tax, we simultaneously define the commodity subject to tax. The coverage decision is therefore an important (though neglected) part of the analysis of taxes, subsidies and similar policy instruments.

Section III deals with the efficiency costs of taxes and subsidies. It first presents the traditional analysis of the efficiency costs of a tax, and then moves on to deal with the question of coverage in terms of a specific example (a tax on automobiles and possibly other motorized vehicles). A simple (though rough) rule of thumb to guide decisions is presented: a good should generally be included in the coverage of a tax whenever at least half of the contraction in its demand that occurs as a result of its inclusion is reflected in an expansion of demand for other covered items.

Section IV treats some of the considerations governing the setting up of so-called (progressive) excise tax structures. Reasons for such a system include a) the difficulty of administering more sophisticated income or expenditure taxes, and b) the wisdom of taxing via several more or less parallel avenues when some important ones (like income taxes) are subject to widespread evasion and avoidance. The principle of including close substitutes under a single rate is explored, together with how in some societies it might conflict with a desire for a more progressive structure. An example is shown of how quickly efficiency cost can increase when the attempt is made to tax close substitutes at different rates.

direction. As to society's values, what we know best is how complex they are, how resistant to precise specification. Tax prescriptions built on a hypothetical precise specification of the social welfare function thus tend to ring hollow in the ears of those who daily struggle with policy problems.

This accounts for our effort in this paper to make only minimal implicit assumptions about society's values. We speak mainly of the tax system's effect on economic efficiency, because that is the economist's stock in trade. And contrary to what some may think, economic efficiency has never been a controversial value; controversy has instead arisen over whether economic efficiency is the primal or only value to be considered, and over how much it should be compromised to facilitate the pursuit of other goals.

With respect to non-economic values, this paper limits itself to recognizing that many societies opt for some degree of progressivity in their tax systems, and to suggesting ways and criteria for using the system of indirect taxes to help implement this desire. Beyond this we do not go, not because other values (like national defense, personal liberty, religion, ethnicity, etc.) are not very important in many societies, but rather because to recognize the specific noneconomic goals that are most pressing in one case would detract from the relevance of what we say to other audiences. Happily, most of what there is to say about indirect tax policy can be said without impinging seriously on the other values in question; and has wide applicability across countries of very different backgrounds and traditions, and at all levels of economic development.

Section II of the paper discusses some simple methodological issues. Its main purpose is to show that the familiar tools of supply and demand can

Section V is really a footnote to Section IV, but is given independent status because of the importance of its message. One of the great policy pitfalls, particularly of developing countries, has been the use of tariffs as a substitute or surrogate for luxury excise taxes. Luxury commodities are typically not, at the start, being produced at home. Thus when the idea of introducing some progression in indirect taxes comes up, it is supremely easy (and tempting) simply to impose high tariffs on such items. This all too frequently has led to the development of high-cost production at home of the very same luxury items that are protected by the tariff walls. The solution is easy and well known: make the tax an excise tax rather than a tariff from the very beginning, even when there is no existing domestic production of the good. If domestic production never would emerge under a tariff, the excise tax and the tariff amount to the same thing. If high-cost domestic production would emerge under high tariffs (as experience shows it tends to do), the use of excise taxes as an alternative policy instrument serves as a wise and prudent preventive measure, avoiding expensive policy mistakes that are very difficult later to reverse.

In Section VI the concept of the value added tax is introduced. It is there treated as a general tax, because this is what has been emphasized in most discussions of it. The great advantages of the value added tax over the turnover or "cascade" type of sales tax system (that so often preceded it) are brought out explicitly. So too is the way in which a fully general value added tax (of the consumption type) effectively exempts from tax the net of real capital accumulation.

Section VII recognizes that the value added taxes that exist in the real world are far from being fully general. The issue of its coverage thus takes on major importance and we therefore explore reasons why certain

sectors like small farmers and small retailers might be left out because of the administrative costs of including them. We also consider whether (and if so how) to include the financial sector in the light of the difficulty (for tax purposes) of defining its value added in a way comparable to that applying in other sectors.

Section VIII shows how there are at least two ways to leave out a sector--excluding it from the value added tax system, or keeping it in the system and applying a zero rate to its output. These are real alternatives in the sense that they have very different attributes and consequences. Briefly, leaving out a sector cuts the input-output chain; the value added tax on inputs into that sector remains a part of the price of its product regardless of whether the product is sold to other producers or to final consumers. Keeping a sector in, on the other hand, requires that the necessary paperwork be done, not only by the tax authorities but by all the covered enterprises (including small farmers and retailers, in case they are included). The "rewards" for this are a) that the sector's own value added is taxed, b) that the value of the final output of the sector is taxed at the desired rate, and c) that this tax on the output is automatically deducted (or offset) against other taxes (at later productive stages) to the extent that the products of the sector in question serve as inputs into other covered sectors. All of this applies, even when the final products of different sectors are subject to different rates. The value added tax system can thus be adapted to create the type of scaled (progressive) tax structure referred to in Section IV.

Finally, Section IX deals with border tax adjustments--the term used to describe the practice of a) collecting at the border on imports of a given good a levy or surcharge equivalent to the excise tax on the domestic

production of that good, and b) contemporaneously rebating at the border the indirect tax component of the cost of any export good. That section shows that while border tax adjustments might be "unnecessary" (in the sense of not affecting the real equilibrium of the economy) under a fully general value added tax, they contribute positively to the coherency of the tax structure in any plausible real-world case. Failure to implement border tax adjustments (particularly the rebating of indirect taxes embodied in exports) is most often due to the difficulty of doing so. It is shown that the value added structure can simply and effectively surmount this difficulty (for covered commodities), even if the system is not fully general and even if it incorporates multiple rates.

## II. DEMAND AND SUPPLY UNDER INDIRECT TAXES AND SUBSIDIES

It is traditional in economic writings to present the analysis of taxation using demand curves and supply curves (or, more generally, demand and supply functions) that are somehow defined independently of the taxes being analyzed. The tax or taxes to be analyzed are thrust, as it were, into a supply-demand setting that is already presumed to be there.

The traditional procedure has many advantages. Beginning economic students must be presumably taught how prices are determined in the first place, in order to set the stage for studying elements (like taxes) which enter to distort the pricing process and to modify the equilibrium that is produced. And in more advanced treatments it is natural to start with individual utility and production functions and to derive therefrom the demand and supply relationships that appear in the market. This is the way that different types of demand curves ("compensated" versus "uncompensated," those for individual commodities versus those for composite goods, etc.) and different types of supply curves (some built assuming that the quantities of certain factors of production are held fixed, others assuming that factor prices are given, some purporting to describe short-run, others middle-to-long-run behavior, etc.) are usually derived.

As one learns more about economics one comes to realize the full complexity of the concepts of demand and supply. For example, a producer will react differently to a rise in the union wage he has to pay, depending on whether the wage rise affects just his own production costs or whether it applies to his competitors as well. Both the demand and the supply of the final product will be affected differently if the price of an important raw

material rises in a fashion that is generally thought to be of short duration (say a few months) than would occur if the price rise was thought to be, for all practical purposes, permanent.

In interpreting actual data, economists have a hard time determining the signals given out by each change in prices affecting suppliers and/or demanders. One cannot tell by looking at the figures whether a particular price change is thought to be temporary or more lasting; nor can one know by looking at it what will be its likely linkage to price changes in other, related goods and services.

In theoretical analysis, however, the life of the economist is somewhat easier. Normally, he will be studying some particular "disturbance" which enters to modify a pre-existing equilibrium situation. It is thus natural for him to inquire into the normal response of the economy to the assumed disturbance. This quickly brings him to the notion of tracing the events that occur in a particular market or sector of the economy on the assumption that adjustments occur in the rest of the economy in such a way that equilibrium is maintained.

This brings us to the concept of demand and supply curves that are built on the assumption of equilibrium being maintained throughout the process. Instead of defining a demand curve assuming "other prices constant," one instead assumes that other prices move in the ways necessary to maintain equilibrium in the various markets that together constitute the economy. An arbitrary rise in the price of oil thus has the effect of causing the equilibrium prices of coal and of gas to rise. A flood of immigration of one type of unskilled labor will presumably end up by putting downward pressure on the wages of all (or nearly all) types of such labor.

Once one begins thinking along these lines one quickly sees that one might well end up with different demand curves, depending on what was the cause of the disturbance. For example, the equilibrium of the economy would likely be different if prices of housing fell because of a reduction in the costs of building materials than if housing prices were reduced by the same amount because the government adopted a policy of housing subsidies to low-income groups.

Taxes and subsidies (the principal policy instruments of public finance) are particularly apt devices by which one can work through the generation of supply and demand curves. In Figure 1, consider that an economy is in full equilibrium at point A, without any tax or subsidy on cement. Impose the tax  $T_b$  and the equilibrium quantity shifts to  $X_b$ , with demand and supply prices  $P_b^d$  and  $P_b^s$ . Impose the tax  $T_c$ , and the new equilibrium moves to  $X_c$ , with prices  $P_c^d$  and  $P_c^s$ . Shift to a subsidy  $Z_e$ , and one processes quantity out to  $X_e$ , supply price up to  $P_e^s$ , and demand price down to  $P_e^d$ .

In Figure 1a no demand on supply curve appears, just the equilibrium points under different tax or subsidy arrangements. When there is no tax or subsidy, A is the equilibrium position, with a tax of  $T_b$ , the equilibrium is defined by points B and B' (two points are needed because demand price differs from supply price by the amount of the tax ( $T_b$ )). Points C and C' correspond to  $T_c$ , and E and E' to a subsidy of  $Z_e$ .

In Figure 1b the natural next step is taken and the points linking quantity and demand price for different levels of tax or subsidy are joined to form the "demand curve" DD, and the points depicting quantities and supply prices are joined to form the "supply curve" SS. These can be

Fall-Adjustment Demand and Supply Curves

Price (P)

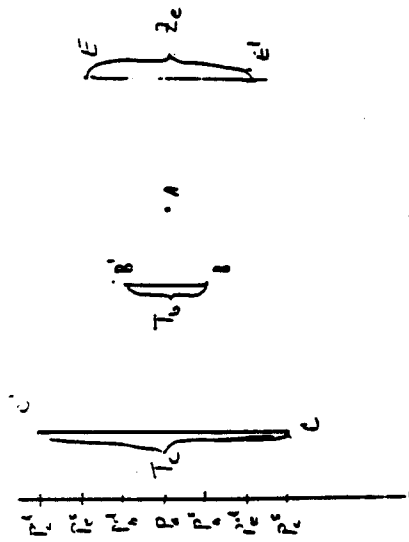


Figure 1a

Price (P)

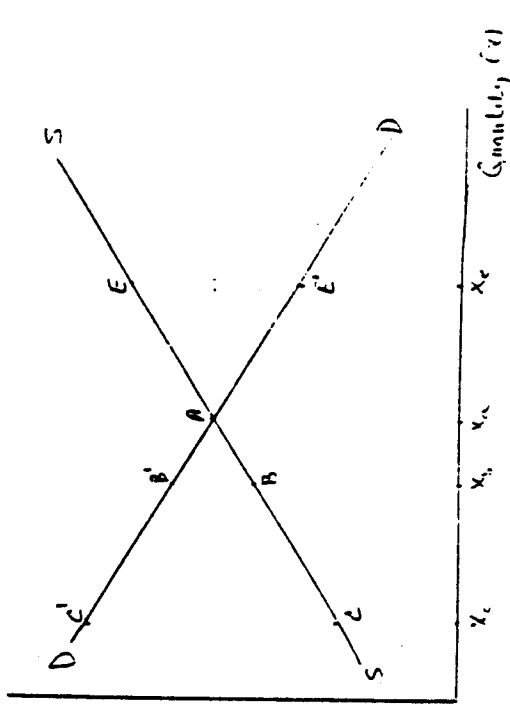


Figure 1b

<sup>1</sup> Some might prefer to call DD the locus of potential equilibrium quantity/demand-price points, and SS the locus of potential equilibrium quantity/supply-price points, but the fact remains that the two curves have as much legitimacy as demand and supply curves as do the more familiar curves of the textbooks. Moreover, DD and SS as defined here are the curves that are most immediately and directly usable in analyzing the effects of taxes and subsidies.

called "full-adjustment" or "general-equilibrium" demand and supply curves; they can be said to be defined on the policy instrument of an excise tax on product X.<sup>2</sup>

<sup>2</sup> I have not explicitly mentioned what the government does with the proceeds from the tax on X, or from where it gets the money to pay the subsidy (in cases where the equilibrium quantity is greater than  $X_b$ ). Obviously there are thousands of ways of spending public funds, and probably also thousands of ways of raising them. Public finance economists have long recognized the need for a convention to facilitate communication among them concerning the effects of taxes and subsidies. Broad consensus can be said to exist on the usefulness of the convention which postulates that tax funds are returned to the public in the form of lump-sum subsidies. Such government outlays (e.g., subsidies) as are required by the analysis are assumed to be raised from the public in the form of lump-sum taxes. The idea behind this is not at all that lump-sum taxes and subsidies are actually used in real-world situations. Rather, the lump-sum taxes and subsidies are an intellectual construct which enables us to speak separately of, say, the effects of a tax on shoes and those of a subsidy to sugar. When the tax on shoes is studied, it is considered as part of a "module," of which the other part is government spending via lump-sum subsidies. When the subsidy to sugar is analyzed, it is dealt with as part of a different module, in which the money is raised from lump-sum taxes. In this way analysts can speak of "the" effects of taxes on shoes and "the" effects of subsidies to sugar. When they want to analyze the specific case of a subsidy to sugar financed by a tax on shoes, they just merge the two modules. In the merger, the two lump-sum items cancel each other, and one is left with the combined effects of the shoe tax and the sugar subsidy.

Some writers have chosen to take some specific actual tax, like the income tax or the value added tax, and to use variations in it as the counterpoise to absorb revenue raised from the tax being analyzed (here an excise tax on shoes) or to provide money for the subsidy being studied (here a subsidy to sugar). In the view of most, however, this simply complicates the exercise. Lump-sum taxes and subsidies have the virtue of being a truly neutral alternative. With any other counterpoise (e.g., the income tax or the value added tax), one would in principle have to take account not only of the distortions involved in the tax being analyzed but also of those entailed in the counterpoise. The effect would be a significant complication of the analysis with no countervailing gain in most cases.

Defining demand and supply curves on the policy instruments of taxes and subsidies has some distinct advantages. On the whole, tax changes are typically rather long-lasting; where they are explicitly transitory, or where they are judged by the community to be likely of short duration, the analyst usually knows it. Thus, for tax changes, the nature of the price disturbance is better specified, and its duration better known, than is the case for the typical period-to-period price movements observed in the market. Moreover, the definition of the commodity is perhaps clearer in this case than in any other. The commodity is simply defined as an item that is (actually or potentially) subject to tax (or subsidy).

The simplicity of definition in this case allows one to handle easily problems that might cause considerable problems in a different framework. Consider the problem of composite commodities. Figure 2 deals with the demand for men's white cotton dress shirts. For simplicity, we assume them to be produced at constant cost, with the cost factors being unaffected by the disturbances we are analyzing. We contemplate in turn 1) a tax on men's white cotton dress shirts, 2) a tax on men's cotton dress shirts, regardless of color, 3) a tax on all men's dress shirts (regardless of the fabric), and 4) a tax on all men's shirts. There is no question but what demand for men's white dress shirts will fall most under 1) and least under 4). The principle that is involved is that underlying the theory of demand for composite commodities: in case 1) demand for men's white cotton dress shirts falls due, among other things, to people shifting to a) colored cotton dress shirts, b) to dress shirts made of fabrics other than cotton, c) to sports shirts, and d) to all sorts of other goods and services. In case 2) the substitution involved in a) is precluded; in case 3) that in b) is also curtailed, and in case 4) that in c) as well. By the time we reach

Demand for Men's White Cotton Dress Shirts

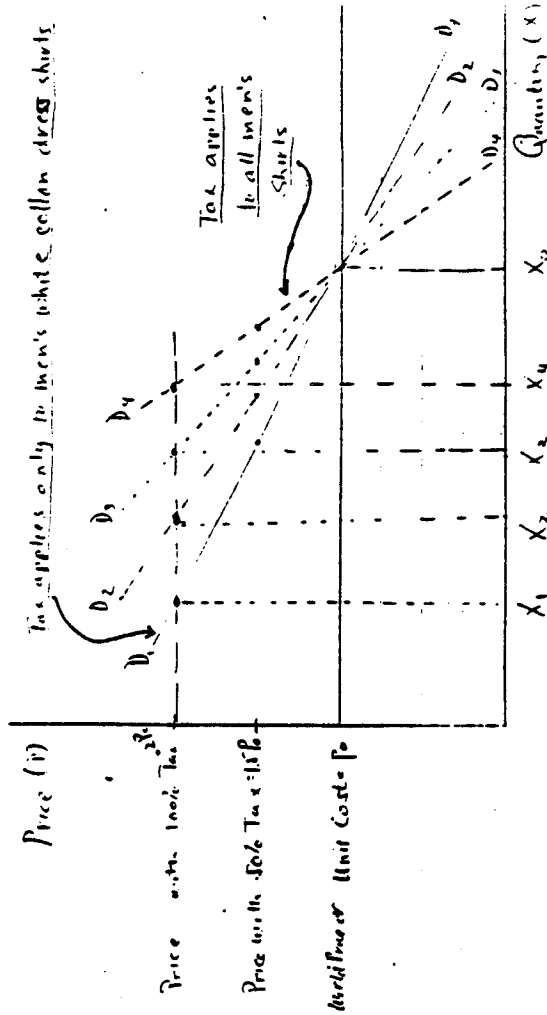


Figure 2

case a) we are dealing only with the substitution between shirts and other things; we thus see in case b) the demand for men's white dress shirts as a component of the demand for men's shirts; no substitution among shirts of different types is present.

What is illustrated in Figure 2 could have come out of a textbook chapter on the demand for composite commodities. But much is gained by looking at the problem with taxes in mind. In the first place, taxes probably provide the most plausible mechanism by which the standard assumption of composite goods demand theory--that the prices of all goods within the composite move by the same percentage, while the prices of all goods outside the composite remain constant--can be made genuinely realistic for actual cases. In an open economy the prices of tradable goods are determined by world market prices and the exchange rate; for these the assumption that the price paid by buyers rises by the amount of the tax is extremely reasonable.<sup>3</sup> Also, most manufactured goods can be said to be produced at substantially constant long-run cost, as long as productive units are of a size which exploits available economies of scale.

A second advantage of the approach illustrated in Figure 2 is the way it highlights the essential arbitrariness of the definition of a commodity. If the government decides to tax white cotton dress shirts, they then and there become a commodity--at least in the sense that economists must treat them as a distinct commodity in order properly to analyze the effects of the tax. If instead of white cotton dress shirts it is all silk

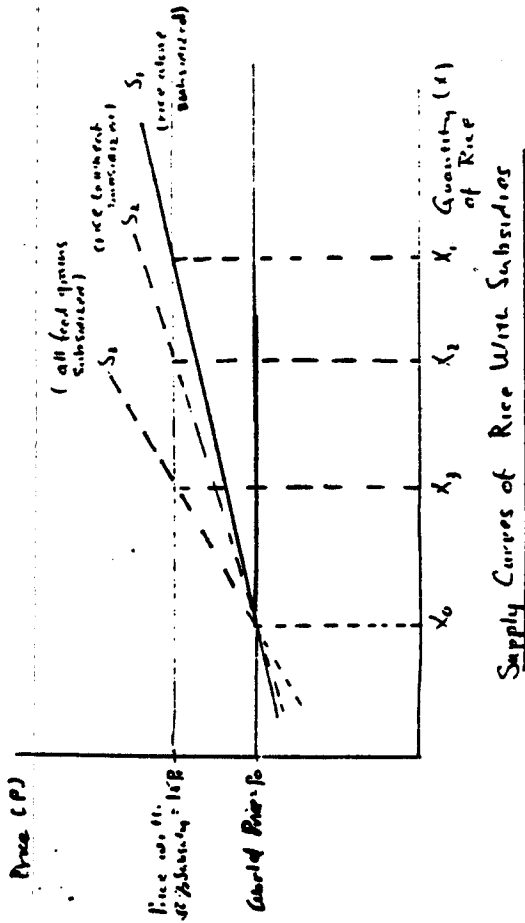
sport shirts that the government decides to tax, it is they that become a commodity in the sense used here. There accordingly are no "natural" classifications that distinguish commodities one from the other; we must constantly bear this in mind as we approach the study of economic problems.<sup>4</sup>

<sup>4</sup> Among other things, the fundamental arbitrariness (perhaps a better term is flexibility) of the definition of commodities should cause one to view with some skepticism economic propositions where validity depends on the counting of, say, commodities on the one hand and productive factors on the other, and then comparing their relative numbers. More generally, whenever particular importance is attached to some classification one should inquire as to the underlying reason why that grouping has special meaning.

A third reason why the approach of Figure 2 is useful is the ease with which it generalizes. Figure 3 illustrates how the same principles that apply on the demand side also work on the supply side. Here the case is taken of subsidizing the production of grains. The analysis is carried out in the context of an open economy, so that the effect of the subsidy is to raise the price received by domestic producers above the world price; consumers pay the world price in any case.

It is assumed that the production of one food grain competes with that of the others. Insofar as acreage is concerned. Thus, when rice alone is subsidized, its production expands (to  $X_1$ ) at the expense of wheat as well as of other food grains, in addition to whatever substitution may exist with respect to crops other than food grains and with respect to greater intensity of use of fertilizers, labor, and other variable factors in the production of rice. When rice and wheat are both subsidized, there is no longer an expansion of rice production at the expense of wheat; production therefore increases only to  $X_2$ , i.e., by less than when rice alone is subsidized (by a given amount). If all food grains are subsidized together,





Supply Curves of Rice With Subsidies

Figure 3

the expansion of rice production is still less (only to  $X_3$ ), as still further possibilities of substitution (which existed in the case where rice alone was subsidized) are now precluded.

A final example will show the general principle of defining supply and demand schedules on the instrument of a tax or subsidy in full action. Here let us assume that the economy neither imports nor exports food grains, so that they are in a technical sense nontraded goods. This yields a standard case of upward rising supply combined with downward sloping demand curves, as in Figure 1.

The added feature of Figure 1 is that both the supply and demand responses are affected when the coverage of the tax is changed. The response of the quantity of rice to the enactment of a rice subsidy alone brings substitution in demand from wheat and other food grains toward rice; likewise rice production increases because, among other reasons, farmers shift acreage from wheat and other food grains to the cultivation of rice. The consequences of the demand and supply substitutions are represented in the demand and supply curves labeled  $D_1$  and  $S_1$ .

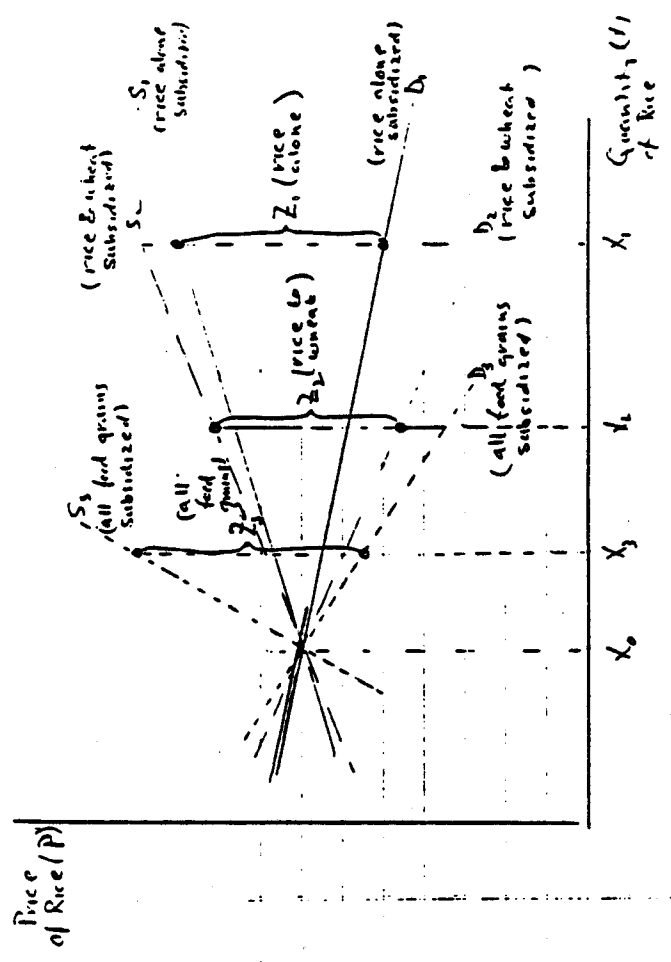
When the subsidy is extended to cover both rice and wheat we have the curves labeled  $D_2$  and  $S_2$ . And, finally, when the subsidy covers all food grains the demand and supply of rice are given by the curves  $D_3$  and  $S_3$ . The equilibrium positions under a subsidy (2) of 50 percent of cost (100 percent of demand price) are  $X_1$  when only rice is subsidized,  $X_2$  when both rice and wheat are covered, and  $X_3$  when all food grains are simultaneously subject to a 50 percent subsidy. Readers should recall that the curves in the figure show only what is happening in the market for rice--but what happens in the market for rice depends critically on how extensive is the coverage of the subsidy. As depicted, there is a vastly greater expansion

of rice production and consumption when only rice is subsidized ( $X_1$ ) than when all food grains are subsidized ( $X_3$ ).

Figure 4 contains an added subtlety. Note that as one moves from  $X_1$  to  $X_2$  the equilibrium price of rice falls, while as one moves from  $X_2$  to  $X_3$  it rises. This is due to the fact that the substitutability between rice and wheat is different in supply from what it is in demand; similarly for the substitutability between rice and other food grains. In the particular case shown, it is assumed (quite plausibly, I think) that rice and wheat are rather good substitutes for each other in demand, and less so in supply. This means that the demand curve for rice is much more stimulated by a subsidy to rice alone than to one on both rice and wheat. On the supply side it is assumed that rice and wheat are only modestly substitutable. This is reflected in the fact that the supply curve of rice ( $S_2$ ) drawn on the assumption that both rice and wheat are being subsidized is quite close to the supply curve ( $S_1$ ) generated by varying the rate of a subsidy to rice alone.

The situation is all the opposite with respect to the substitution relationship postulated between rice and other food grains. Here the distance between  $D_2$  and  $D_3$  is relatively small, indicating low substitutability in demand between rice and other food grains (not counting wheat), while the distance between  $S_2$  and  $S_3$  is substantial, reflecting quite good substitutability in production between rice and the other grains.

To summarize; we have seen in this section i) how demand and supply curves to be used in policy analysis can be defined on the instrument being studied; ii) how this greatly simplifies the analysis and at the same time keeps it compatible with a full general-equilibrium adjustment of the economy to the disturbances being studied; iii) how the definition of a



Supply and Demand Curves of Rice with Subsidies

Figure 4

commodity is fundamentally (i.e., in the nature of things) arbitrary, so that the categories that are relevant in the analysis are those that the government chooses (or might choose) to cover by tax and/or subsidy policies; 17) how there can be differential reactions of both demand and supply in a given category, depending on the coverage of the tax or subsidy being studied, and how these differential reactions are similar in nature to those dealt with in the theory of composite commodities; and v) how changing the coverage can produce very important differences in the way a given commodity (or category) reacts to a particular rate of tax or subsidy.

### III. THE EFFICIENCY COST OF TAXES AND SUBSIDIES

The efficiency cost (sometimes called welfare cost or excess burden) of taxes and subsidies stems from the fact that additional distortions are typically introduced into the economy when these policy instruments are applied. In the case of a tax demanders pay more for the product than suppliers get; in the case of a subsidy the demand price paid by demanders (and presumably representing the value that they place on the marginal units they demand) is less than the supply price received by suppliers (which in turn presumably reflects the opportunity cost of the factors of production used to produce the marginal unit).

Figure 5 illustrates the measurement of efficiency cost using a traditional textbook example. When a tax of  $T$  per unit is introduced on the product  $X$ , the equilibrium quantity falls from  $X_0$  to  $X_1$ . One question to be asked is, what was this "lost" amount worth to demanders? Another is, what saving in costs can be attributed to the reduction in the quantity of  $X$ ? In the absence of complicating factors (such as pre-existing distortions in the markets for related goods), the answers are: a) the value to demanders of the forgone output ( $X_0$  minus  $X_1$ ) is equal to the area  $(ABX_1X_0)$  under the demand curve between these two quantities, and b) the reduction in output of  $X$  makes possible a saving of economic costs equal to the area  $(ACK_1K_0)$  under the supply curve between the same two quantities. In reaching these measures of forgone benefits and of costs saved, the demand price (i.e., the height of the demand curve) is taken to reflect the economic value placed by demanders on each successive unit, and the supply price (the height of the supply curve) is taken to reflect the effective economic (opportunity) cost

The Efficiency Costs of a Tax

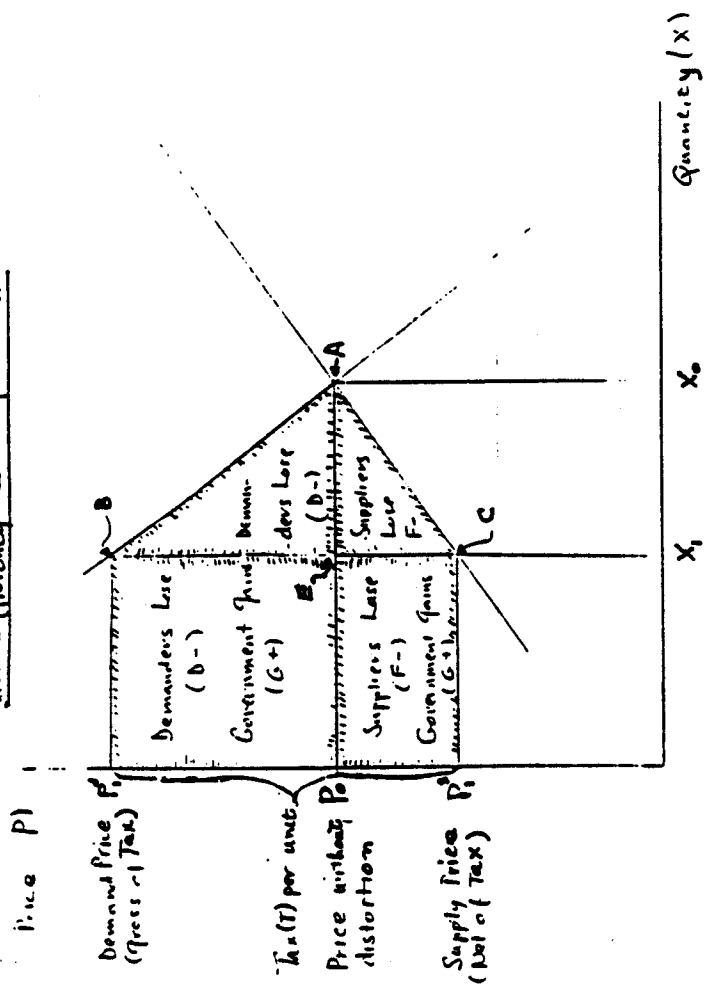


Figure 5

of each successive unit. The difference between the amount of forgone benefit ( $ABX_1X_0$ ) and the amount of costs saved ( $ACX_1X_0$ ) is the triangle ABC. This triangle measures the efficiency cost (or welfare cost or deadweight loss) associated with the tax T.

There is a second method of evaluating the efficiency cost of the tax. This method does its measuring taking as a base the price (rather than the quantity) axis. It looks more directly at the gains and losses perceived by different groups, rather than concentrating on the economic valuation of quantities demanded and supplied. When, as a consequence of the tax, the price paid by demanders goes up from  $P_0$  to  $P_1^d$ , those demanders who in the presence of the tax continue to buy the product (in a quantity equal to  $X_1$ ) have to pay for these units a total extra amount equal to  $BEF_0P_1^d$ . In addition they (as demanders) give up a benefit equal to  $ABX_1X_0$  on the units ( $X_0$  minus  $X_1$ ) that they no longer demand. As partial compensation for this benefit (as it were) they now no longer have to lay out the amount  $ACX_1X_0$  that they would have spent to buy  $X_0$  minus  $X_1$  in the absence of the tax. Their net loss with respect to the quantity  $X_0$  minus  $X_1$  is thus the triangle ABE.

An exactly corresponding analysis applies to the factors of production engaged in producing X. The price they receive goes down from  $P_0$  to  $P_1^s$ . They end up receiving this reduced price on the quantity  $X_1$ . Their loss with respect just to this quantity (and owing to the reduction of the price they receive) is equal to  $F_0P_1^sCE$ . In addition they cease to produce the quantity  $X_0$  minus  $X_1$ . On this quantity they were initially (i.e., in the absence of the tax) receiving a total amount equal to  $ACX_1X_0$ . This amount is now no longer received, but the resources that were initially used to produce  $X_1$  minus  $X_0$  have value in alternative uses which is (again in the

Figure 6 shows how the concept of efficiency cost can be put to use in the design of a system of indirect taxes. The diagram refers to automobiles, but the taxes contemplated are  $T_1$ , a tax striking automobiles alone;  $T_2$ , a tax striking automobiles and trucks; and  $T_3$ , a tax striking motorcycles and scooters as well as automobiles and trucks.

If automobiles alone are subject to a tax (in this case of 60 per cent), there is a drastic cutback in their quantity. This occurs in part because of what we can call generalized substitution--against other goods and services quite generally--but in considerable measure it is due to substitution vis-a-vis quite specific alternative products. In many countries (LDCs especially) high taxes on cars induce the widespread use of trucks as family passenger vehicles. (In many cases whole industries have come into existence to adapt the trucks to make them more closely equivalent to cars). The same is true with respect to motorcycles and scooters. Not only are they direct substitutes for cars, in the simple sense that more motorcycles and scooters will be bought in the presence of a high (versus a low) tax on cars; but also, just as with trucks, the tax on automobiles can cause new industries to arise which then convert motorcycles and scooters into vehicles that are more closely equivalent to cars, adding seats and canopies, making enclosures to surround the passengers and even the driver, etc. This type of response enhances the substitutability of other motor vehicles for cars. It helps to create an important drop in the demand for cars when the tax falls only on them.

In Figure 6, the tax on autos alone causes their quantity to fall from  $X_0$  to  $X_1$ . When the tax falls on trucks as well as cars, the incentive to use trucks in place of cars disappears. Now the tax causes the demand for cars to fall only to  $X_2$ . Finally, when the tax strikes all motor

absence of complicating factors such as distortions in those other markets) equal to the area under the supply curve  $AC_1X_0$ . The net loss to suppliers (factors of production) on the quantity  $X_0$  minus  $X_1$  is thus the triangle ABC.

We have yet to count the government's gains. Revenues from the tax are equal to the tax rate (T) times the quantity ( $X_1$ ) that is produced (and demanded) in the presence of the tax. The government's gain therefore just cancels the losses  $BDP_0^d$  plus  $P_0^dPCG$  sustained by demanders and suppliers (taken together) on the quantity  $X_1$ .<sup>5</sup> When the gains and losses of all par-

<sup>5</sup>It is simplest just to treat the government as an additional actor on the economic scene, experiencing gains in some circumstances and losses in others. We are thinking in this way when we say that the government's gain just cancels certain losses sustained by demanders and/or suppliers. It would be more precise, instead of assigning the increased tax revenue to the government as such, to consider it as an agent of the people as a whole. Thus the revenue from the tax under study could be used to reduce other tax rates (in which case the beneficiaries would be the taxpayers who perceive those reductions) or alternatively to increase government payments or other spending (in which case the beneficiaries would be those who receive the payments or who otherwise gain from the uses to which the government puts the tax revenue). This might indeed be done in specific actual cases for which the government's likely disposition of the proceeds can be known (or guessed at). At the level of an exposition like the present one, however, there is no point in complicating the analysis in this way. We here stick with the convention that the added tax revenues are returned to the people via lump-sum subsidies; they are labeled as "government gains" to reflect the government's role in the process as well as to distinguish them from the gains of demanders and the losses of suppliers.

ties are offset one against the other, the net loss is (just as in the first method) the triangle ABC. We now see that this is composed of two smaller triangles--the first (ABC) representing the loss to demanders (D-), uncompensated by any corresponding gain to the government, and the second (BDC) representing a similar loss (F-) sustained by suppliers (factors of production).

vehicles it precludes all substitution among them; there remains only the generalized substitution between motor vehicles on the one hand and the rest of goods and services on the other. It is this type of substitution that is responsible for the reduction in the demand for cars from  $X_0$  to  $X_3$  in the case where all motor vehicles are subject to the tax.

Figure 6 demonstrates the wisdom of making an indirect tax more rather than less inclusive. At the very least, efforts should be made to bring under the umbrella of a single tax rate packages of goods and/or services that are good substitutes for each other. The point is stronger than it may at first appear. Not only do the triangles of efficiency cost ( $R_0Q_1R_1$ ,  $R_0Q_2R_2$ ,  $R_0Q_3R_3$ ) get progressively smaller as we move from  $T_1$  to  $T_2$  to  $T_3$ ; simultaneously there are increased receipts from the tax that falls on automobiles. In the case depicted, these receipts rise from  $P_0P_1Q_1R_1$  to  $P_0P_2Q_2R_2$  to  $P_0P_3Q_3R_3$ .

In addition to the revenue derived from taxing automobiles increasing, and its efficiency cost declining as close substitutes for cars are brought within the tax net, there is also of course the fact that still further tax revenue is generated from the markets for any such substitutes (trucks, scooters and motorcycles, in this case). Thus, if the purpose were to raise only a given amount of revenue the rate of tax could be cut as the coverage was extended, still further reducing the efficiency cost as measured in a market like that of automobiles in the present example.

There is, however, a sense in which the efficiency cost of a tax might increase when its coverage is extended. This stems from the substitutability between the newly-taxed item(s) and the rest of goods and services. Assume that bicycles are a relatively poor substitute for automobiles (and other motorized vehicles). In that case, when the coverage of

Efficiency Costs in the Market for Autos

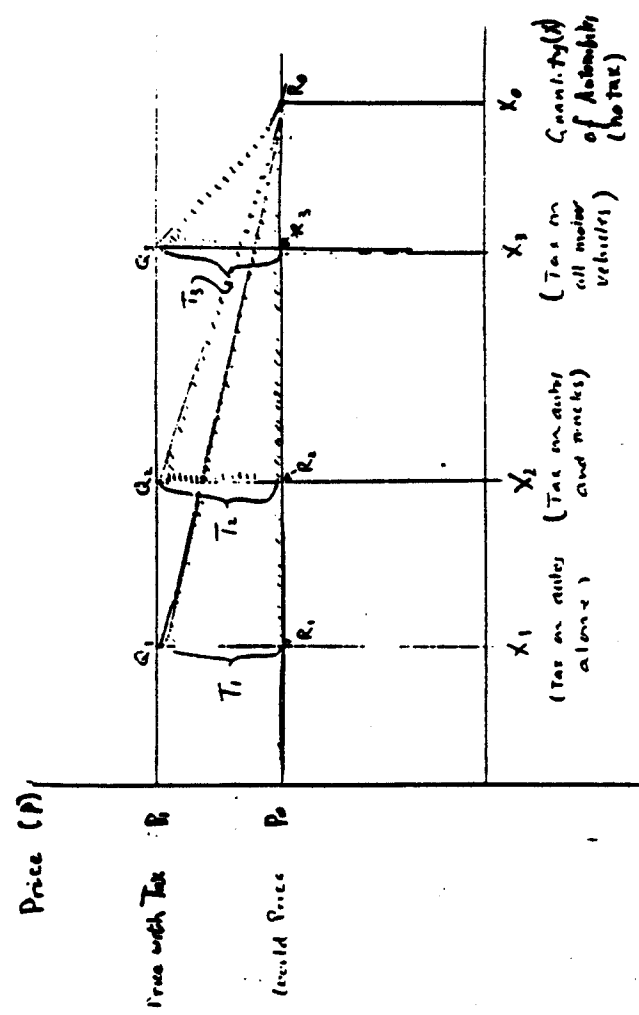


Figure 6

the tax on autos is extended to bicycles, there is little or no increase in the quantity of cars (i.e., the equilibrium volume does not shift sharply to the right as it did from  $X_1$  to  $X_2$  and from  $X_2$  to  $X_3$  in Figure 6). But at the same time a new triangle of efficiency cost is generated due to the introduction of the tax in the market for bicycles. This can under the postulated circumstances easily outweigh the small (if any) gains in efficiency due to the small (or nonexistent) rightward shift of the equilibrium level of supply and demand in the automobile market.

The general principle is that one should (on efficiency grounds at least) try to broaden the coverage of an indirect tax when the newly added items are good substitutes for those already in the covered groups. A rough guide is that it is good to include an added item if at least half of the expected contraction in its quantity (as a consequence of its being taxed) is reflected in increased spending on other covered items.<sup>6</sup> This is indeed

<sup>6</sup>Let  $C_i$  represent the marginal resource cost of the  $i$ th commodity and  $T_i$  the tax on it. A new tax  $T_j$  on any good or service  $X_j$  will change the quantities of each good  $i$  by the amount  $S_{ij}T_j$ , where  $S_{ij} = \partial X_i / \partial T_j$ . With a given amount of total resources in the economy, we have that  $\sum_i C_i S_{ij} = 0$ ; that is to say, the marginal resource costs saved in the activities that contract will be matched by the marginal resource costs newly incurred in the activities that expand. Units may be chosen so that all  $C_i = 1$ ; if this is done we may speak of  $\sum_i S_{ij} = 0$ . The own-price term,  $S_{jj}$  is thus equal to  $-\sum_{i \neq j} S_{ij}$ . For the case at hand we can represent by  $X_1$  the items already covered by the tax, by  $X_2$  an item that is a candidate to be added for coverage, and by  $X_3$ , "all other goods and services." Then we have  $S_{12} + S_{32} = -S_{22}$ . The requirement that "at least half of the contraction in quantity is reflected in increased spending on other covered items" is thus equivalent to  $S_{12} > S_{32}$ . The efficiency cost engendered in the market for  $X_2$  is a triangle whose base (of negative sign) is  $S_{22}T_2$ , and

whose height is  $T_2$ . Its area (negative, reflecting a cost) is  $\frac{1}{2} S_{22}T_2$ ; the gain in efficiency in the market for  $X_1$  is equal to  $T_1$  (the tax already existing on the product  $X_1$ ), times  $S_{12}T_2$ , the change in quantity of  $X_1$  brought about by the addition of  $X_2$  to the coverage. We must compare  $\frac{1}{2} S_{22}T_2$  (a loss, of negative sign) with  $T_1 S_{12}T_2$  (a gain, of positive sign). When we speak of extending the coverage of the same tax, this means that  $T_1$  and  $T_2$  will be the same. The fact that  $S_{12} + S_{32} = -S_{22}$  signifies that our assumption of  $S_{12} > S_{32}$  is equivalent to  $S_{12} > \frac{1}{2} S_{22}$ . That is to say, the area of the rectangle of gain  $T_1 S_{12}T_2$  exceeds that ( $\frac{1}{2} S_{22}T_2$ ) of the triangle of loss. Thus, so long as  $S_{12} > S_{32}$ , the efficiency gain from extending coverage to  $X_2$  will exceed the cost.

no more than a guide; what it does is to guarantee that in addition to raising more tax revenue from the items previously covered, and to raising additional revenue from the items newly covered, there will be an absolute reduction in the efficiency cost associated with the whole package of covered items (old and new together). Any time this condition is in fact met, there is a very strong argument for extending the coverage. What we cannot do is reverse the rule; when the condition is not met, we cannot say there is a strong argument against extending the coverage (indeed, to do so would in effect argue against broad, relatively general taxes like those on sales, value added, manufacturing activity, and the like).

The message up to this point is, then, that when designing a system of indirect taxes, we should make an effort to include in a single package any group of goods and services that are close substitutes for each other. We have seen that failing to include a close substitute as part of the "package" to be taxed at a given rate has the dual effect of increasing the efficiency cost of the tax and of reducing the amount of revenue collected. If we then end up going against this message, it should only be after long

and careful thought. It should not be as a consequence of oversight, or caprice, or simple negligence at the time a tax law is drafted.<sup>7</sup>

<sup>7</sup> I have neglected discussing complementarity between and among commodities because it is not a phenomenon of great moment in the design of tax systems. There are many instances of complementarity in the world--left and right shoes, chassis and engines, motors and carburetors, houses and windows --but there is unlikely ever to be a sensible reason for opting to put separate taxes on the different members of such pairs. Without attempting to delve into the analytical nuances that are involved in the differential taxation of the different members of a pair or set of close complements, let me simply state that the course of good judgment is to treat the members of such a pair or set as a unit, and to consider that unit (e.g., the finished car, the house with windows, the pair of shoes) as the object of study. Once this is done there will be relatively little (if any) complementarity remaining in the system, as between plausible tax packages as defined above.

IV. ON SCALED (PROGRESSIVE) EXCISE TAX STRUCTURES

There are many reasons why a country might choose a scaled or progressive excise tax structure as one component of its overall tax strategy. Most of these reasons have to do with the difficulty or impossibility of implementing other taxes, targeted more directly at the desired progressivity. Thus, when one contemplates the idea of a progressive tax system, one typically thinks of total income, total consumption and total wealth (net worth) as plausible measures, each of which could serve as a useful "base" for a progressive tax. That is to say, as one or another of these measures grew in size, the taxpayer would become liable for tax at a higher marginal rate.

The trouble is that each of these likely bases for progressive taxation is difficult to administer, more so as one moves from the richer and more industrialized to the poorer and less developed countries of the world. The difficulties are of at least three types: administration, compliance, and evasion. On administration one needs for any of the three types of tax--on income, consumption, or net wealth--a fairly sophisticated cadre of professionals. This can be difficult to assemble even in the most advanced countries, where typically a highly skilled and experienced tax accountant can earn in the private sector three or four times the maximum he could obtain as a tax administrator. The disparity is even greater in less developed countries, which moreover are often hampered by a short supply (nationwide) of technically trained personnel.

On compliance one must worry about the sheer capacity of the population to do what is required by a specific type of tax. It is folly to ask illiterate peasant farmers or street vendors to keep accurate accounts for



tax purposes when they cannot even keep them for themselves. Both within and between countries, the problem of compliance, as I view it, is concentrated at the poorer levels, with masses of people who are for practical purposes incapable of doing the type of financial reporting that a modern system of income, consumption or wealth taxation would require. As economic development occurs, and education and literacy increase, the population outgrows this problem, but that of evasion and avoidance (in which human capacities are fully put to work to frustrate the intent of the tax law) rises in importance.

It is an important fact that access to most types of evasion and avoidance is available to some groups and not to others, and when access is available groups also differ in the ease with which and the degree to which they can take advantage of it. Much has been written about how farmers and independent professionals almost routinely understate their actual incomes for tax purposes. Especially in recent years, there has also been much discussion of the so-called "underground economy," which covers a wide range of services in addition to the activities already mentioned. Yet side by side with those who readily avoid and evade income taxes, there are the earners of wages and salaries, which are not only fully reported and taxed, but on which the taxation is made virtually instantaneous through withholding at the source.

One of the arguments for using the excise tax structure as a supplement to the income tax is that excise taxes are typically much less subject to evasion than income taxes, and that to the degree that some evasion or avoidance in any event occurs, it is unlikely that the same groups will be involved as those who succeed in frustrating the income tax collector. Individual taxpayers and groups will therefore come closer to paying their

"fair share" in a mixed income-cum-excise tax system than in an income tax system alone.

Thus, for the reasons indicated, one may want to reduce reliance on other forms of progressive taxation because of limitations of administrative personnel, or because of problems of compliance by large groups of potential taxpayers, or because avoidance and evasion create problems of equity as well as of efficiency. Regardless of which of these reasons is dominant, it is natural in such circumstances to think of a scaled structure of excise taxes as doing a part of the job (of introducing some elements of progression into the tax system). We now address some of the problems involved in creating such a structure.

For a scaled structure to accomplish the purpose at hand, one would want to select for taxation at the highest rates a set of commodities purchased predominantly by the very rich. For taxation at more moderate rates one would have a subset of commodities purchased rather widely among well-to-do people, but not much by those less well off. Then perhaps there would be a third level, at which the goods and services subject to tax would be rather widely consumed by the so-called middle classes.

The technical problem posed by the idea of a scaled structure of excise taxes is that it may be hard to distinguish distinct sets of commodities consumed principally by the different socioeconomic strata. To the extent that the commodities chosen are rather widely consumed in the different strata a general tax on a set of them will not contribute a progressive component to the tax structure. Usually, some rough sort of differentiation can be made with the goal of progressivity in mind, but it may run the risk of inflicting high efficiency costs upon the economy. In such

cases it may be prudent to limit the degree to which progressivity is sought via the structure of excise taxes.

Examples will follow which illustrate the nature of the problem.

Fortunately, fate seems to have been kind in the sense that the problems appear to be least for the poorest countries, which on the whole are those that face the greatest difficulties of assembling a solid, relatively progressive direct tax structure. The issue is quite straightforward. Where the automobile, the refrigerator, the air conditioner, etc. already, because of the low level of income of the country, identify their owners or consumers as being in the top socioeconomic bracket, it is fairly easy to make the indirect tax structure progressive by imposing a special tax on those items (together, perhaps, with other modern household appliances).

The problem is, most severe in, say, the OECD countries, in many of which the ownership of durable goods of the types indicated extends well below the median of the distribution of income and wealth. Here it may even be true that groups at or around the median spend a higher fraction of their income on items like those listed than do the significantly richer strata of society. In order to obtain any significant degree of progression in the excise tax structure in such countries, one would have to impose a substantial amount of tax-rate differentiation, within each "product." One could tax the Mercedes Bens and the Cadillacs at a higher rate than the Mitsubis and the Oldsmobiles, and these in turn at a higher rate than the Hondas Civicas and the Fiats. One could tax large and possibly more "sophisticated" refrigerators and air conditioners at higher rates than smaller and simpler ones, etc. No doubt by such a strategy one could indeed achieve some degree of progression in excise taxation, even in the richer industrial countries.

But it would come at the expense of very substantial efficiency costs, since taxes would strike close substitutes at significantly different rates.<sup>8</sup>

<sup>8</sup> Consider a case in which, in response to a 20 percent tax on automobiles, the demand for each of three "classes" of cars would fall from 100 to 80. The triangle of efficiency cost in each class would be  $1/2 \times (.2) \times (20) = 2$ . Efficiency cost would thus be 2 percent of the amount spent on cars, and would be 12 1/2 percent of the tax proceeds (tax receipts are  $80 \times .20 = 16$  in each class). Now let there be a 30 percent tax on the class I cars, a 20 percent tax on class II, and a 10 percent tax on class III. Let the quantities demanded now be 60 for class I, 80 for class II and 100 for class III. The total demand is the same; it has just been reshuffled due to intraclass substitution. Efficiency cost is now  $1/2 \times (.3) \times (40) = 6$  for class I,  $1/2 \times (.2) \times (20) = 2$  for class II, and  $1/2 \times (.1) \times (0) = 0$  for class III. The total efficiency cost is therefore 8, or over 16 percent of tax revenue, which in this case is equal to  $44 = .30 \times 60 + .20 \times 80 + .10 \times 100$ .

To get approximately the same revenue as would be produced by a uniform 20 percent tax, the average tax rate would have to be raised to about 24 percent. With a similar pattern of tax differentiation (36 percent for class I, 24 percent for class II, and 12 percent for class III) and a similar pattern of reaction of quantities demanded (52 for class I, 76 for class II, 100 for class III) the welfare cost would be  $1/2 \times (.36) \times (48) = 9.6$  for class I,  $1/2 \times (.24) \times (24) = 2.88$  for class II, and  $1/2 \times (.12) \times (0) = 0$  for class III. Here the efficiency cost would be 11.52, amounting to about 24 percent of the tax revenue of 48.96.

Readers should appreciate how modest are the tax rates and how reasonable are the demand reactions assumed here. Yet in order to obtain approximately the same tax revenue, the efficiency cost nearly doubles (from about 12 percent to about 24 percent of tax revenue) simply by moving from a uniform rate of tax to a moderately progressive structure on a set of commodities that are quite good substitutes. The message is clear: differential rates of tax applied to close substitutes can carry quite high efficiency costs and should be avoided where possible.

## V. TARIFFS AS POSSIBLE COMPONENTS OF A SCALED

### EXCISE TAX STRUCTURE

One of the natural temptations in the design of a progressive excise tax structure, especially in countries at a relatively early stage of development, is to tax imported consumer goods at significantly higher rates than their domestic counterparts. Lying behind this temptation is the fact that for a substantial range of commodities imported consumer goods tend to be more "luxurious," and to be more widely consumed by the upper income groups, than similar domestically-produced goods. In many cases there may be no home production at all of consumer goods of a particular type.

In the judgment of most informed professional observers, behind this quite natural temptation there very often lurks a trap. The trap in question has already been analyzed in the preceding section--that of taxing at separate rates close substitutes which might better be "packaged" and taxed at a single rate.

The old-fashioned story--dating from the wave of import substitution in many less developed countries during the 1950s and 1960s--is as follows. Motivated by a desire to tax luxury goods at high rates, and reinforced by a determination not to use "scarce" foreign exchange for luxury purposes, countries placed extremely heavy tariffs on all kinds of durable goods--automobiles, refrigerators, air conditioners, television and hi-fi sets, etc.

In most of the cases in question these high tariffs (which sometimes reached 200, 300, and even 400 percent) appeared to serve their purpose for a time. The goods subject to duty were indeed luxury goods in the countries concerned (in the sense that they were predominantly bought by upper and

middle income families), and the taxes on them raised significant revenues from the consuming classes.

The problem manifested itself only gradually over time--it consisted of the growth of local production of the tariffed luxury items, behind the protection of huge tariff walls. This production was typically not in the comparative advantage of the countries concerned. It took place at a real cost much higher than the world market price of the good in question. But as the production of domestic substitutes grew, demand was shifted toward them from the highly taxed imported items. The local production of substitute goods in these cases not only entailed inefficiency in itself, so to speak, but it also involved a correspondingly large loss of tariff revenue as demand was shifted away from the tariffed items.

The standard recommendation in cases like the one described above is that instead of a tariff, an excise tax ("luxury tax") should be put on in the first place. In that way no artificial incentive is introduced to stimulate the inefficient production of domestic substitutes, but by the same token, no barrier is placed in the way of the efficient domestic production of such substitutes. If a capability for domestic production arises in the future, at less than world prices, such an opportunity can be exploited.

The simplest rule to follow is that the desire to tax luxuries should never be the motive for a tariff. If there is already domestic production of the commodity in question, then a tariff would violate the principle of placing close substitutes in the same category for tax purposes. If there is no domestic production of the good initially, the placing of a tariff on it will give new and added incentives to the business of making local substitutes. So what began as a luxury tax on all refriger-

erators (say) will end up as a tariff protecting the domestic production of those luxury goods. The easiest way to avoid the problem is to have the law itself institute a tax (not a tariff) on refrigerators. This tax can be collected at the border by the customs authorities in the same way a tariff would be collected, so far as imported items are concerned. If there is no production of domestic refrigerators initially, no collection problem will exist. But when such production comes into being, tax would be collected on it, presumably at the level of the factory.

#### VI. THE VALUE ADDED TAX AS A GENERAL TAX

Few events in the history of taxation have been so dramatic as the rise of the value added tax to its present level of importance as a source of revenue in many countries. From a position of essentially zero significance in 1951 or 1952, it had by the late 1970s, become one of the most important generators of tax revenue in the world. France was the first country to implement a value added tax, but it was soon followed by other European countries and before long by a number of others in the Western Hemisphere, in Asia, and elsewhere. By now countries with probably half the total gross domestic product of the world outside the Soviet sphere of influence have adopted the value added tax as a major revenue source.

To understand how a "conquest" of this magnitude could occur in so many places and in so short a time, one must appreciate both a) the conceptual simplicity and consequent intellectual appeal of the value added tax, and b) the pre-existing situations that were common in the countries that adopted it.

In concept, the value added tax is what its name implies--a tax on the value that is added to a good or service as it passes through a particular stage of production. The tax is usually levied by taxing the entire value of the output of a productive stage (usually defined as the output of a firm or enterprise), and then deducting (i.e., taking as a tax credit) the tax that has been already paid on the goods or services that entered as inputs into that productive stage.<sup>9</sup> Enforcement is relatively easy because no

<sup>9</sup> This method is known as the "abstraction" method, so-called because it taxes the whole and then subtracts the tax already paid on inputs. The so-called "addition method" defines the tax base by adding up the separate components of factor payments at the stage which is being subject to tax. Under this method, wages and salaries, interest and possibly rent payments,

and profits would be added together; their sum for a particular firm would define its value added. In practice the subtraction method can be said to completely dominate the field.

tax credit is granted without proof that tax has in fact been paid at the prior stage; failure to collect tax at one stage thus does not imply that the tax has been "lost" as far as the authorities are concerned. Far from it; if tax is not paid at stage A, the producer at stage B who uses inputs from stage A will presumably have no receipts to prove that the tax has been paid earlier. Without these receipts he has nothing to deduct, so he (the producer at stage B) in effect picks up the tax that producers at stage A failed to pay.<sup>10</sup>

<sup>10</sup>This property of picking up at later stages the tax that earlier stages fail to pay has sometimes been "institutionalized" by the tax law. Thus, small farmers may be left out of the system, with the conscious idea that their output will in any event be taxed so long as it passes through a subsequent taxable stage (food processing, spinning and weaving, etc.).

In practice, the value added tax won its quickest support in countries that previously used sales taxes of the "cascade" or "turnover" type. These taxes would be levied each time a good or service changed hands. The undesired consequence of cascade-type taxes was that the value added at the early stages of a productive process was often taxed, in effect, several times. This in turn created an undesired incentive for the vertical integration of the entire productive chain.

Table 1a illustrates how a cascade or turnover tax works. We trace here the evolution of a product like bread through four successive production stages. In Table 1a a turnover tax of 10 percent of the value of

TABLE 1a

## ILLUSTRATION OF TURNOVER TAX

	Value Added	Turnover Tax 10 Percent	Price Paid by Buyer
Farmer	20.00	2.00	22.00
Miller	20.00	4.20	46.20
Baker	20.00	6.62	72.82
Retailer	20.00	9.28	102.10
Total	80.00	22.10	
With Vertical Integration	80.00	8.00	88.00

TABLE 1b  
ILLUSTRATION OF VALUE ADDED TAX

	Value Added	Sale Value of Product Less Tax Already Paid	Tax Paid at this Stage (\$ 25%)	Tax Paid at Previous Stages	Net Tax Paid at this Stage
Farmer	20	20	5	0	5
Miller	20	40	10	5	5
Baker	20	60	15	10	5
Retailer	20	80	20	15	5
Total	80				20
With Vertical Integration	80	80	20	0	20

product applies at each stage. The base of the tax is equal to the price paid by the buyer of the product of the previous stage, plus the value added of the present stage. Thus, at the miller stage, the tax of 4.20 is equal to 10 percent of (22 plus 20); at the baker stage the tax of 6.62 is 10 percent of (46.20 + 20.00), and so on. By the time the final, retailer stage has been passed, the total tax paid is 22.10,<sup>11</sup> on a cumulated value added

<sup>11</sup>The sum of 22.10 can be broken down in the following way. On the farmer's value added of 20, tax was paid four times. First 2 of tax was paid on the farmer's 20, then, as a part of the miller's 42, tax of 2.20 was paid on the basis of the farmer's 22 (the initial 20 plus the initial 2 of tax). Then at the baker's stage, as part of the baker's 66.20, there was 24.20 stemming from the farmer, on which 2.42 of tax was paid. Finally, at the retailer stage the farmer's value added plus the cumulated taxes on it amounted to 26.62 (24.20 plus 2.42), on which 2.66 of tax was due. The total tax paid on the farmer's value added of 20 was thus 2.00 plus 2.20 plus 2.42 plus 2.66, for a total of 9.28. On the miller's value added, tax was paid only three times for a total of 2.00 + 2.20 + 2.42 = 6.62. The baker's stage paid 2.00 + 2.20 + 4.20, and finally the retailer's stage paid 2.00. When one breaks down the tax paid in this way, one obtains the same total as before—i.e., 9.28 + 6.62 + 4.20 + 2.00 = 22.10.

of only 80. One of the great defects of a turnover-type system is the incentive to vertical integration. This is revealed in the final row of Table 1a, where turnover tax is calculated on the assumption that all four processes are integrated in a single enterprise. As can be seen, the amount of tax is reduced dramatically, from 22.10 to 8.00 by the simple "trick" of vertical integration.

Contrast the above case with that illustrated in Table 1b. Here a value added tax of 25 percent takes a net amount equal to exactly 5 at each stage. No stage's value added is taxed more than once, and when vertical integration is assumed, the tax payment is exactly equal to the sum of the net taxes paid by the different stages in the absence of integration. Thus under a value added tax there is neither a differential treatment of the

various stages, nor is there an artificial incentive favoring vertical integration.

These examples show why it was so easy to effectuate the shift from turnover or cascade-type taxation to the value added tax. The value added tax has had much less momentum in countries which did not have a turnover tax to begin with. In such countries, a value added tax is sometimes justified as being more rational and equitable than a hodgepodge of separate excise taxes, imposed on different products at different rates in response to different situations at different points in the country's history. The neat, newly woven blanket (the value added tax) is taken as a substitute for an old and ragged patchwork quilt (of separate excises on many items) inherited from the past. Indeed, this is a good reason to adopt a value added tax, but the precondition is the existence of a widespread and differentiated pattern of separate excises with little or no coherency or rationale of its own.

Where the value added tax has made least headway is in countries that place substantial reliance on a single-stage sales tax. Most often the single-stage tax is at the retail level, but it might also be at the wholesale or manufacturer's level. In such cases the existing single stage tax is a moderate-to-good substitute for the value added tax, and the pressure to adopt a new system is correspondingly less.

Much of the enthusiasm for the value added tax comes from its being a general and neutral tax. We shall see that it is in practice neither of the above. But on the other hand it certainly has the potential to be, and undoubtedly is in many actual cases, more nearly general and closer to truly neutral than any of its plausible competitors in the indirect tax field.

The usual form of actual value added taxes has been the so-called "consumption type" of tax. Under the variant, an entity's sales are subject to tax; from the tax thus computed there is then a deduction for the taxes already paid on inputs purchased by the firm. Under the "consumption type" of tax no distinction is made between current inputs and capital inputs; the firm deducts in identical fashion the tax already paid on either type of input.

Where the value added tax is applied everywhere at an equal rate, this treatment of capital inputs is equivalent to the immediate expensing (or writeoff) of capital goods purchases. The value added tax strikes the gross income from capital when it is earned, but it permits the deduction of capital acquisitions themselves. This in effect leaves the income from capital untaxed, and turns a value-added tax of the consumption type into something quite close to a tax on labor income.<sup>12</sup>

<sup>12</sup>The gross income from capital is equal to its net income plus depreciation. If an investment just barely yields the opportunity cost of capital, the investment will have a net present value of zero when all flows of benefits and costs are discounted at the opportunity cost rate. The present value of capital outlays (PVKO) will thus equal the present value of income from capital (PVIK) plus the present value of depreciation allowances (PVDA). Providing for a tax offset of the tax rate  $\tau$  times PVKO, and then subsequently taxing YK and DA, so that the present value of the tax is  $\tau(PVYK + PVDA)$  will in the normal case where  $(PVKO - PVYK + PVDA)$  not entail any net tax on capital at all. When a value added tax is first instituted, there is typically a net tax on the income from capital, because YK plus DA typically exceeds capital outlays for the economy as a whole. Put another way, with a newly instituted value added tax, tax is paid on the income from new and old capital, while only the new capital acquisitions are subject to deduction. As time goes on, a larger and larger fraction of the reproducible capital stock of the system will have been subject to deduction (tax offset) at the time of its purchase. When the point is reached where the entire reproducible capital stock has been so treated, we can then say that a value added tax of the consumption type, in the net, taxes the present value of extraordinary income from reproducible capital (PVYK + PVDA - PVKO) where (PVYK + PVDA - PVKO) is negative. In addition, of course, the value added tax will naturally tax the income from nonreproducible assets (land, etc.) to the extent that their costs are not counted as costs of capital outlays in the first place.

### VII. COVERAGE OF A VALUE-ADDED TAX

Textbook treatments of the value-added tax are most often built on the assumption that the tax is fully general (or nearly so). Such treatment<sup>13</sup> often focus on the margins of non-generality that may nonetheless exist. Thus, it is really not possible for a value-added tax to strike the leisure time enjoyed by members of the labor force. Hence, to the degree that the labor supply elasticity is non-zero, some economic inefficiency will result even from a value added tax covering all marketed goods and services. Textbook treatments may go on to mention that a fully general value-added tax may be designed so as not to interfere with decisions between saving and consumption. This is accomplished when the tax adopted is of the "consumption type."

<sup>13</sup>As mentioned previously, a "consumption type" value added tax permits enterprises to deduct not only the tax already paid on raw materials and intermediate inputs but also the tax embodied in capital goods purchased by the firm. Thus in any period of time value added tax on capital goods will be paid by those who make them, but this tax will be offset (in the context of a general value-added tax of the consumption type) by the tax credits claimed by the purchasers of these capital goods. To the extent that all investment activity done in the economy can claim credits for the value-added tax embodied in the capital goods it acquires, no value-added tax ends up being paid on such goods, and capital formation in that economy is left undistorted by the tax.

The textbook treatments just mentioned are very important in helping us to understand the attributes of general taxes, as well as the idea of "neutrality" in taxation. But it must be emphasized that the role of such discussions is to instruct, to impart understanding, to elucidate principles. In actual fact, no value added tax in the world comes anywhere near to full generality. Rarely, in fact, is much more than half of the total value added in the economy captured by the value-added tax system.

Accordingly, we do not further elaborate on the properties that a value-added tax might have if it were truly general. We instead concentrate on the issues involved in dealing with value added taxes whose coverage falls within the ranges we encounter in actual practice. At the one extreme we might consider a system in which the activities comprising agriculture, manufacturing, mining, public utilities and transport, and wholesale and retail trade are quite successfully covered by the value-added tax network. This would be as close to complete coverage as one could expect to get, but even such a broad system would almost certainly omit the great bulk of the nation's educational and medical establishments, plus many types of services that are individually provided (household servants, repairsmen, etc.).

Working down from this degree of generality, one would find systems that consciously left small farmers and small retailers out of the tax net but included larger farmers and larger retailers who could be expected to keep orderly records, perhaps for business purposes (as, for example, corporations) or to comply with another tax (e.g., a personal or business income tax). Still less general would be a value added tax that specifically excluded farmers and retailers of all types--a coverage that is sometimes said to be at the "manufacturer's level," this level being itself defined to include mining and perhaps electricity, transport and other public utilities.

Regardless of the choice of alternative general coverage from somewhere along the spectrum just described, policymakers would probably have to deal separately with the financial sector (banks, savings and loan companies, insurance companies, etc.). The reason lies in the difficulty of arriving, for this sector, at a precise definition (for tax purposes) of the



items that will be called sales (on which tax would presumably be levied) and those that will be called purchases (which would potentially give rise to tax offsets). Is the interest received by a bank on its loans to be called a "sale" in the same sense as the sale of cloth by a textile maker or of beef by a meat packer? Is the interest paid by a bank to its depositors to be treated analogously to the purchase of component parts by an auto maker? In a similar vein, should insurance premiums (most of which go into a contingency reserve to pay claims) be treated as sales in the same fashion as the output of a food processing firm? These questions are simply raised here to reveal to readers how perplexing and knotty can be the problems of applying a value added tax to the financial sector. To my knowledge, the problem has no simple answer, which undoubtedly accounts for the frequency with which the financial sector is left out of the value added tax system.<sup>14</sup>

<sup>14</sup> If the choice is made to include the financial sector within the coverage of the value added tax, the simplest way to do so is to tax it by the "addition method," even though for the rest of the system tax is calculated by the "subtraction method." This would entail each firm's paying the statutory rate of its tax on the value added accruing to labor and to capital. A pragmatic definition here is simply the entire wage and salary bill of the enterprise, plus the profits accruing to its owners.

The preceding discussion should make it clear that one cannot draw up a blueprint which would tell a country how broad a coverage it should seek to establish for its value added tax. Insofar as the economist can put forward a rule on his own criteria, it would be to make the coverage as broad as possible. Another, perhaps better way to state the same thing is to say that each curtailment of the coverage of the tax involves an additional cost in terms of economic efficiency looking solely at the value added tax itself. This cost should be offset by countervailing benefits in order to justify any given curtailment from generality. The offsetting

benefits may take the form of a saving of administrative costs (such as might be involved in extending the tax to small farmers and/or small retailers), or because of the "merit good" nature of the product produced (which is perhaps the case for hospitals and educational establishments), or because of some additional policy objective (e.g., the stimulation of production in a backward region).

Strictly economic arguments for curtailing the coverage of a value added tax are rare, except when the tax has already lost any semblance of generality. At such a point, one has to treat the value added tax like any other tax on a subset of goods, and one can revert to the rule of thumb presented earlier, that an item should be included "so long as at least half of the decrement in the item's own consumption is reflected in increased spending on other covered items."

## VIII. MULTIPLE RATES IN A VALUE ADDED TAX

The first item to consider under the heading of multiple rates in a value added tax is the distinction between "excluded" and "zero rated" products. This is an issue that comes up even when all the tax that is in fact collected is levied at a common rate (say 20 percent). Consider the cases of farmers and retailers, mentioned earlier. Assume that it has been decided not to collect the 20 percent tax from either group. This decision does not end the story, for the question of exclusion versus zero rating must be dealt with.

That question is a real one as far as farmers are concerned. They are likely to use a number of inputs (seeds, fertilizer, fencing materials, perhaps gasoline and farm machinery) whose prices embody a 20 percent value added tax. If the farmers are left out of the value added system, this tax will end up as a component of the prices of the final products they produce. Their own value added will be exempt from tax, but the value added by the suppliers of their inputs will not. If, on the other hand, they are included within the value added system, but at a zero rate, they will pay no tax on the sales of their products, but will be eligible to receive (either in cash or in a credit against other taxes they may owe) a rebate for the value added tax on the inputs embodied in those products. In this case, assuming the production of farm products is competitive, the buyers of these products will pay correspondingly lower prices for them. Not only will the farmers' value added be exempt from tax, but also all the value added entering into the inputs that they buy.

In the case cited, the choice of treatment to apply is not self-evident; it will depend on the circumstances of the case at hand and on the

objectives of the policy makers. If the purpose of policy is to avoid the administrative costs of dealing with lots of farmers, then the best choice is to leave them out of the system. In that case their own value added is not taxed at this stage, but will be taxed at a later stage as other members of the system (say food processors) use the farmers' output in their own productive processes (e.g., canning or freezing). As far as product sold directly by farmers to final consumers (or to other customers outside the value added system) is concerned, the value added by the farmers themselves will go untaxed, but the tax paid at all prior stages of production will still be present as a part of the price borne by consumers.

Alternatively, suppose that the object is not to reduce the administrative burdens of the tax authorities but to exempt from tax some or all of the food products that consumers buy. In this case one might contemplate in the first instance applying a zero rate to food processors, leaving farmers entirely within the system, paying the full tax (which then would be later rebated to the food processors). The only reason to zero-rate the farmers would be for the purpose of seeing to it that no tax would be incorporated in the prices of the products that they sold directly to final consumers (at open-air markets, etc.). But this objective could be better accomplished by leaving the farmers within the system generally, and simply zero rating that fraction of their output estimated to be sold directly to final consumers. This treatment would ensure that all taxes paid by the seed and fertilizer producers and other suppliers of inputs into agriculture would be rebated, and hence would not remain as part of the price paid by consumers.<sup>15</sup>

<sup>15</sup> Leaving the farmers out of the system would result in the tax embodied in their inputs being carried forward to the food processors. They would have no receipt for taxes paid, hence could claim no tax credit. Consumers would bear this cost. Where the objective in excluding the

farmers is to streamline tax administration, this "flaw" is less objectionable than it is in the case where the whole idea is to keep the food expenditures of consumers out of the tax net.

It should be obvious from the above discussion that zero rating would make no sense for small retailers or for retailers taken as a whole. Zero rating for all retailers would end up annulling the entire value added tax (if it is of a consumption type and assuming that all taxed items pass through the retail stage before reaching the final consumer). Zero rating for small retailers would annul all the tax paid previously on whatever items passed through their hands. Hence if administrative or other considerations dictating leaving some substantial group of retailers (or all of them) out of the tax net, the only plausible way to do it is to leave them out of the value added tax system altogether.

It should be now also be clear that the possibility of zero rating adds considerable flexibility to a value added tax system. As mentioned, one need not think of it only in terms of zero rating an entire group (like farmers or food processors). One might zero-rate certain products (e.g., food grains and simple breads made therefrom), or certain channels of sale (direct sales by farmers to consumers), without in the process distorting the internal coherency of the entire value added system. (If a zero-rated product is bought by a producer within the system, he pays full tax on his output, but gets zero tax credit on the input in question, hence the tax that was taken away by zero-rating is automatically put back once the product is used as an input.)

The same degree of simplicity applies to multiple rates of value added tax. The use of multiple rates makes sense only at the level of the final consumer, so I will confine my discussion to this case. It is clear that if a particular product is taxed on sale at 40 percent, with all taxes

previously paid on its inputs being rebated to the producer at the final stage, then the ultimate consumer ends up paying a tax of 40 percent on the full value of that product. But if some of the output of the product in question is in turn used as an input by other producers, they in turn will pay tax at their own regular rate of, say, 20 percent, while receiving a tax credit for the 40 percent rate paid on their highly-taxed input. Thus they will end up being essentially unaffected by the fact that in an earlier productive stage, tax was levied at 40 rather than 20 percent.

In this way it is possible to use the value-added tax as a vehicle for constituting the type of scaled (presumably progressive) excise tax structure that was dealt with earlier in this paper. Indeed, considering that the intelligent implementation of such a scheme typically calls for rather broad categories of commodities being taxed at similar rates to final consumers, but not when they are used as inputs into subsequent production, the value added system seems to be an eminently appropriate way to do so.

o.i.f. that country) by exactly the amount of the tax in question; this difference will be specifically represented by the border tax adjustments. To see how the country's economy can attain an equivalent equilibrium in the absence of border tax adjustments, assume that the fixed exchange rate is suddenly devalued, and the border tax adjustments eliminated. The high internal price of exportable goods, which was previously offset for foreigners via the rebating of taxes at the border, is now offset by a higher exchange rate. Similarly, the "unfair" competition of foreign sources, which previously was counterbalanced by surcharges levied on imports, is now counterbalanced by the increased cost of imports brought about by the country's currency devaluation. Nothing in the postulated situation will cause any domestic prices, wages, or costs to be any different from what they were previously. The devaluation, in this example, simply takes the place of the pre-existing border tax adjustments. This example makes it easy to see also how the first and second situations would differ if the value added tax did not fully cover all tradable goods and services. If, say, wheat were left outside the tax net its internal price would in the first case be below that of the tradables covered by the tax, while in the second case it would be the same (per dollar's worth of tradables). In such cases of discrepancy the border tax adjustments give the more appropriate way of fitting a given set of indirect taxes into the world economic setting.

The best way to look at border tax adjustments in a real-world setting is to consider any excise tax levied on a particular good. If that tax is levied at the consumer level (say, when a suit is sold in an apparel store) there is no need for border tax adjustments. Imported as well as domestically-produced suits will automatically fall within the tax et when they are sold, and suits that are exported will not be struck by a consumer-level tax. Suppose, however, that the government decides to levy the tax at the producer level. Now a price advantage will be accorded to imported suits that would sell at the same price as a domestic suit without the tax. This advantage can, however, be easily neutralized by applying the appropriate border tax adjustment (in this case, imposing the tax in question on imported suits as they enter the country). Similarly, if domestically produced suits were exported, the disadvantage represented by the tax would be neutralized by rebating the tax at the border.

#### IX. BORDER TAX ADJUSTMENTS: THE GATT RULES

Whenever indirect taxes deal with tradable commodities, the question arises of how to deal with exports of goods produced at home under conditions where their domestic selling price includes the tax, and how to deal with imports that compete with domestic products subject to the tax. There exists a standard set of rules and procedures, approved and monitored under the General Agreement on Tariffs and Trade (GATT), by which countries are authorized to rebate to the exporter the amount of any locally-imposed indirect taxes (including import duties) that are included as components of the cost of the product being exported. By the same token, when goods are imported whose domestically-produced counterparts are subject to an indirect tax, the imported goods can be required to pay this same tax, in addition to any tariff to which they may be subject. "Border tax adjustment" is the term used to describe the practice of levying countervailing taxes on imported goods (to match indirect taxes paid by their locally-produced counterparts) and of rebating on exports the indirect taxes embodied in their costs.

In theoretical analyses of value added taxes, it is possible to reach the conclusion that it does not matter whether border taxes adjustments are made or not. But this conclusion holds only in the case where the tax is truly general<sup>16</sup>; hence it does not apply to actually prevailing systems of value added taxation.

<sup>16</sup> A fully general value added tax has the effect of introducing a wedge between domestic resource costs on the one hand and the sale value of domestically-produced goods and services on the other. Border tax adjustments impose this same wedge on imported goods, and take the wedge away when commodities are exported. This is easy to visualize by thinking of a country with a fixed exchange rate. All tradable goods will have prices within the country which are higher than the international prices (f.o.b. or

Firm will have value added tax receipts for the tax they has up to his own productive stage been embodied in the inputs. He will receive a credit for these taxes regardless of whether he sells the product at home or abroad. No other credit or any other treatment is needed. Items that are exported can simply be directly "zero rated" under the value added tax. No further action is necessary. <sup>17</sup>

<sup>17</sup> If some administrative consideration should require that the seller initially pay tax on the exported items (omitting them with the rest of his production), a rebate or tax credit would then be in order. But the implementation would be scarcely harder than in the case treated in the text. It would not entail any special investigation into the value added taxes paid on the specific components, nor into the weight of these components in the total cost of the exported items. All these costs would be offset through the natural operation of the value added system. The only calculation involved would be (in the case of a multiple rate system) to make sure that the rebate was at the appropriate rate for the items being exported.

A great deal of ignorance as well as misunderstanding prevails in relation to border tax adjustments. Many countries fail to apply these adjustments; others view them as being interferences with free or fair trade. Economic analysis would suggest that neither attitude is appropriate. The application of border tax adjustments is and should remain a reasonable policy action, fully within a country's sovereign rights. What it does, in effect, is to convert into a consumer tax a levy that applies formally to the producer level. Since any country clearly has the right to levy taxes on its consumers, it seems fully legitimate for it to do so in an indirect way (taxes at the producer level combined with border tax adjustments) as against a more direct way (taxes at the consumer level without such adjustments) if this is more convenient in the light of administrative or other considerations.

Just as it is wrong to condemn border tax adjustments as trade interferences, so it is generally unwise for a country to fail to apply them. Given that all real-world systems of indirect taxes are far from general, and that few indirect taxes are levied with the objective of discouraging domestic production of the goods in question, it seems reasonable that the tax authorities would indeed want to neutralize any artificial disadvantage that its indirect tax system placed on domestic producers of tradable goods vis-a-vis their foreign counterparts.

Most of the time when border tax adjustments are not implemented it is because the tax component of the cost of an export good is difficult to estimate. Where the estimation is easy, it is usually done. One way for a country to facilitate the calculation and implementation of border tax adjustments is for it to adopt a value added tax system. Whether such a system has a unified rate or multiple rates does not matter; the exporting