III

DOMESTIC CONSUMPTION

AND

PRODUCTION TAXES
Theory of Optimal Commodity Taxation

Howell H. Zee

- What problems does the theory of optimal commodity taxation address?
- What are the most important optimal tax rules and their economic implications?
- What are the merits and limitations of the optimal tax rules?

The theory of optimal commodity taxation is primarily concerned with the problem of determining the structure of taxes on various commodities to raise a given amount of revenue with a minimum of the tax-induced efficiency loss as defined in Chapter II. While the problem can be analyzed in a multiperson framework to bring equity considerations to bear on the determination of such a tax structure, it is more natural and instructive to delay combining efficiency and equity concerns until the discussion of the theory of optimal income taxation (Chapter IV). Furthermore, the purely efficiency aspect of optimal commodity taxation can be of interest in and of itself, since it has important tax policy implications quite apart from equity issues. Accordingly, the entire analysis in this chapter is framed in the context of a one-person (representative-consumer) economy. 1

Defining the Problem

To focus ideas and to simplify exposition, consider a concrete example of the consumer's budget with income from two sources (earned and unearned) and expenditure on two commodities, X and Y. His earned income is the product of the wage rate \(w\) and the amount of time he devotes to work \(L\), that is, \(w \cdot L\). Let his unearned income (e.g., inherited wealth), the amount of which is assumed to be unalterable by his own behavior, be denoted by \(K\). In the absence of taxation, his budget constraint can be written as

\[ K + w \cdot (T - l) = p_x \cdot X + p_y \cdot Y \]  

(2)

Let \(\tau_w\), \(\tau_x\), and \(\tau_y\) be, respectively, the ad valorem tax rates on wages, commodity \(X\), and commodity \(Y\). With these taxes, the budget constraint becomes

\[ K + w \cdot (1 - \tau_w) \cdot (T - l) = p_x \cdot (1 + \tau_x) \cdot X + p_y \cdot (1 + \tau_y) \cdot Y \]  

(3)

or, after a slight rearrangement,

\[ K + w \cdot (1 - \tau_w) \cdot T = p_x \cdot (1 + \tau_x) \cdot X + p_y \cdot (1 + \tau_y) \cdot Y + w \cdot (1 - \tau_w) \cdot l \]  

(4)

The budget constraint as stated in equation (4) has two important implications. First, the demand for leisure is conceptually no different from that for other commodities. Second, since \(\tau_w\) enters the right-hand side of equation (4) with a negative sign, a tax on wages is seen to be partly equivalent to a subsidy on leisure (the equivalence is only partial because \(\tau_w\) also appears on the left-hand side of equation (4)). Putting it differently, if leisure is to be taxed, \(\tau_w\) must be negative (this in turn implies that labor is subsidized).

The optimal commodity tax problem can now be defined as one of determining the optimal values for the tax rates \(\tau_w\), \(\tau_x\), and \(\tau_y\) to raise a given amount of tax revenue with a minimum efficiency loss. The solution to this problem can be stated in terms of a number of tax rules, the validity of some of which requires special assumptions regarding the feasible scope of taxation and the nature of the consumer's demand curves. Four such rules are especially well-known for their important economic implications: the proportionality rule, the Ramsey rule, the inverse elasticity rule, and the Corlett-Hague rule. Each of these rules is separately discussed below. It must be emphasized that the generality of the first three rules is in no way affected by the special three-commodity \((X, Y, \text{and} \ l)\) example adopted above; they can be generalized to apply to any number of commodities in a straightforward manner.

Optimal Tax Rules with Fixed Producer Prices

This section considers optimal tax rules on the assumption that \(w\), \(p_x\), and \(p_y\) are invariant to the tax rates (i.e., the producer, or tax-exclusive, prices remain the same before and after the introduction of the tax).
taxes). Complications arising from relaxing this assumption are noted in the following discussion.

**Proportionality rule**

The proportionality rule states that if all commodities are taxable, then the optimal tax structure would be one where the tax on each commodity, expressed as a proportion of its price, is the same for all commodities. This implies that all commodities should be taxed at the same rate. The economic intuition behind this rule is easily demonstrated. As discussed in Chapter II, the efficiency loss of a tax arises from its distortion of relative prices. If all commodities are taxable and taxed at the same rate, however, then relative prices would not be affected and, as a consequence, no efficiency loss can arise.

In the context of the above example, the proportionality rule requires that all three tax rates be set equal to each other, i.e., \( \tau_x = \tau_y = -\tau_w = \tau \), where \( \tau > 0 \) is the common ad valorem tax rate (as noted earlier, \( \tau_w \) must be negative if leisure is to be taxed). It immediately follows from such a tax structure that the budget constraint in equation (4) becomes

\[
K/(1 + \tau) + w \cdot T = p_x \cdot X + p_y \cdot Y + w \cdot I,
\]

which implies that the proportionality rule in fact amounts to taxing the consumer's unearned income, \( K \), at the rate \( \tau/(1 + \tau) \). By assumption, the magnitude of \( K \) cannot be altered by the consumer's own behavior. Hence, this tax is equivalent to a lump-sum tax and entails no excess burden.

Since taxing all commodities at the same rate is the simplest tax regime implementable, at first glance, the proportionality rule seems to provide a happy coincidence between theoretical optimality and administrative simplicity. This is, unfortunately, not the case. A closer examination of equations (4) and (5) reveals that the practicality of the rule is predicated on several critical assumptions. First, the consumer must have unearned income. If \( K = 0 \), then the government receives no tax revenue. This is because the revenue raised from taxing commodities \( X \) and \( Y \) is just offset by the subsidy to labor (i.e., the tax on leisure) to preserve the proportionality of the tax structure. Second, even if \( K \) is positive, it must be sufficiently large to enable the government to raise the required amount of revenue.\(^2\)

Third, the optimality of the proportionality rule hinges on the ability of the government to tax all commodities. If, as is usually the case in realistic policy settings, not all commodities are taxable (e.g., taxing leisure, or, equivalently, subsidizing labor, is seldom a feasible policy option), then the rule is not applicable. For these reasons, the proportionality rule, though elegant, is of extremely limited practical value.

The remaining three tax rules all deal with situations in which some commodities are nontaxable. For concreteness, assume that leisure is the nontaxable commodity in the above example, so that \( \tau_w \) is identically equal to zero. The government must, therefore, meet its revenue needs solely from taxing commodities \( X \) and \( Y \) by optimally setting \( \tau_x \) and \( \tau_y \).

**Ramsey rule**

The Ramsey rule states that, for a commodity tax structure to be optimal, the proportional tax-induced reduction in the quantities demanded of a taxed commodity, as measured along its compensated demand curve, should be the same for all taxable commodities.

Two aspects of the Ramsey rule are worth emphasizing. First, it is stated in terms of tax-induced changes in quantities, not prices. This is because, ultimately, the efficiency loss associated with taxation stems from the fact that commodity taxes, unless imposed in lump sum, would induce the consumer to adopt a pattern of (compensated) demands for commodities that is different from that which he would have adopted in the absence of the taxes. It is this induced change in the quantities demanded of various taxed commodities that gives rise to the excess burden of taxation; the induced change in the relative prices of the taxed and nontaxed commodities is only the means through which the quantity changes are effected.

The second notable aspect of the Ramsey rule is that it is stated in terms of the compensated, not the ordinary, demand curves. In effect, this rule is one that minimizes the total efficiency loss from taxing the different commodities as measured by either one of the two Hickssian measures, and not by the DMH measure, of excess burden.\(^3\) If the symbol \( \Delta \) is used to denote the change in a variable, then the Ramsey rule, as applied to the above two-taxable commodities example, would imply that the optimal values for \( \tau_x \) and \( \tau_y \) are those that would produce the equality

\[
\Delta X/\Delta X_0 = \Delta Y/\Delta Y_0,
\]

where the prime "'" on a variable indicates that the change in the variable is to be measured in the compensated sense, and the subscript 0 on a variable fixes

---

\(^2\)Equation (5) indicates that the total revenue raised by the tax is \( K \cdot \tau/(1 + \tau) \).

\(^3\)See Chapter II for a discussion of the concept of compensated demand curves, as well as the difference between the Hicksian and DMH measures of excess burden.
the reference point (e.g., the pretax or posttax situation) against which the change is to be defined.

The intuition behind the Ramsey rule is straightforward. Suppose, for example, that some values for \( \tau_x \) and \( \tau_y \) are chosen so that equation (6) is violated. Then, the sum of the excess burdens on \( X \) and \( Y \) (the relevant triangular areas under the compensated demand curves of the two commodities) could be reduced simply by altering the relative values of the two tax rates, that is, the increase in the excess burden on one commodity resulting from raising its tax rate would be more than offset by the decrease in the excess burden on the other commodity resulting from lowering its tax rate. No further reduction in the sum of the excess burdens would be possible once the point is reached where, at the margin, any additional adjustment in the tax rates results in equiproportional changes in the compensated quantities demanded of the two commodities. Since the condition of equiproportional changes is stipulated in terms of quantities, the Ramsey rule would hold, in general, to nonuniform (i.e., unequal) tax rates among the taxed commodities.\(^4\)

The validity of the Ramsey rule is perfectly general; it requires no special assumptions concerning the nature of the consumer’s demand curves. Because it is stated in terms of compensated quantities, however, it shares with the Hicksian measures of excess burden the major shortcoming that it cannot be applied on the basis of readily available economic data, as compensated demand curves are not directly observable.

There is one notable circumstance under which the Ramsey rule is valid even if stated in terms of ordinary (and not compensated) quantities: the consumer’s demands for the taxed commodities are such that a change in his income produces equiproportional changes in all their quantities demanded. In this case, the rule can be equivalently stated as

\[
\frac{\Delta X}{X_0} = \frac{\Delta Y}{Y_0}
\]

(7)

where all quantities in equation (7) can now be derived from ordinary, directly observable demand curves. It will be recalled from Chapter II that this special case corresponds precisely to the condition (the so-called homothetic demands) for path independence of the DMH measure of excess burden when there are multiple price changes.

**Inverse elasticity rule**

If the ordinary demand for every taxed commodity is independent of all prices except its own, then the optimal tax rates on these taxable commodities are inversely related to the (absolute) values of their ordinary own-price elasticities of demand (\( \varepsilon \)). In the two-taxable commodities example above, this rule takes the form

\[
\varepsilon_x \frac{\tau_x}{(1 + \tau_x)} = \varepsilon_y \frac{\tau_y}{(1 + \tau_y)} = \alpha,
\]

(8)

where \( \alpha \) is some (positive) constant. It follows immediately from equation (8) that the lower the (absolute) value of a taxable commodity’s ordinary own-price elasticity, the higher should be its tax rate.

Of all the well-known tax rules, the inverse elasticity rule is perhaps the most intuitively obvious. As has been pointed out in Chapter II in the single-taxable commodity case, the excess burden of a tax varies positively with its own-price elasticity. Hence, for any given tax rate, the tax would lead to a smaller excess burden if imposed on a commodity with a low own-price elasticity, than on one with a high own-price elasticity. While this conclusion does not, in general, follow in the multiple-taxable commodity case, the assumption that the demand for every taxable commodity is independent of all prices except its own does imply that the excess burden on each taxable commodity could be analyzed as if it is the only taxable commodity.

Another appeal of the inverse elasticity rule is that it is stated in terms of ordinary, and, therefore, directly observable demand curves. This property follows from the fact that when the ordinary demand curve of a taxable commodity depends only on its own price, it will not shift when the prices of other taxable commodities are changed. Hence, minimizing the DMH measure of excess burden under an ordinary demand curve is equivalent to minimizing any one of the Hicksian measures of excess burden under the associated compensated demand curve.\(^5\) Indeed, it is equally valid to state the inverse elasticity rule in terms of the compensated demand curves.

If the consumer’s demands for all taxable commodities are independent of each other, as assumed, then the entire burden of tax-induced adjustments in his pattern of consumption would necessarily fall on the nontaxable commodities. Hence, the usefulness of the inverse elasticity rule as a guide for tax policy rests essentially on one’s evaluation of the reasonableness of

---

\(^4\)The Ramsey rule would imply the uniform taxation of taxable commodities, if the consumer’s demands for the nontaxed commodities are not affected by a change in any of the taxes.

\(^5\)Of course, the DMH and Hicksian measures would still lead to different sizes of the excess burden of a tax, for reasons explained in Chapter II.
this implication regarding the nature of the consumer's demand curves.

**Corlett-Hague rule**

Suppose that, as assumed, there are only two taxable commodities in addition to the nontaxable leisure. Then the Corlett-Hague rule states that the commodity which is a stronger complement (or weaker substitute) of leisure than the other should be taxed more heavily. 6

An intuitive explanation of the Corlett-Hague rule is simply that, even though there is, by assumption, a nontaxable commodity, this commodity can nevertheless be taxed indirectly by taxing the commodity whose consumption is complementary to it. Because its application involves the use of compensated demand curves, the Corlett-Hague rule, though highly instructive, suffers from the same limitation as the Ramsey rule.

As stated, the Corlett-Hague rule would not necessarily hold in the case of more than two taxable commodities. A slight reformulation of the rule, however, could permit it to be generalized without altering its original spirit. Suppose all taxable commodities are initially taxed at the same rate. Then, if the tax revenue is held constant, raising the tax rates on all commodities, which are stronger complements (or weaker substitutes) of leisure than those whose tax rates are being lowered, would increase welfare. The crucial difference between the earlier Corlett-Hague rule for the two-taxable commodities case and this reformulated version is that the latter need not imply that, in a pairwise comparison of commodities within each of the two groups of commodities whose tax rates are being either raised or lowered, the optimal tax rate is necessarily higher on the commodity which is a stronger complement (or weaker substitute) of leisure than that on the other commodity.

6Technically, two commodities are considered complements if a rise in the price of one commodity leads to a decrease in the compensated quantities demanded of the other commodity. They are considered substitutes if the reverse occurs.

**Optimal Tax Rules with Variable Producer Prices**

If the producer prices of the commodities adjust in response to the tax-induced changes in their demands, then it is natural to expect that the optimal tax rules would involve supply responses as well. This, in fact, turns out to be the case. Unfortunately, these supply responses do not, in general, enter into the tax rules in a neat and easily interpretable way. There are, however, two important and interesting special cases.

First, if the economy's production is characterized by constant returns to scale, 7 then all of the optimal tax rules derived under the assumption of fixed producer prices would remain valid under variable producer prices. This is a particularly important result, since decentralized competitive markets are best supported by constant returns to scale technologies. Second, if both the ordinary demand for, and supply of, every taxed commodity depend only on its own price, then the optimal tax rule would only involve a simple combination of ordinary own-price elasticities of demand (\(\varepsilon\)) and supply (\(\eta\)). In the two-taxable commodities case, this rule can be stated as

\[
\frac{\tau_x \eta_x \varepsilon_x}{(1 + \tau_x) \cdot \eta_x + \varepsilon_x} = \frac{\tau_y \eta_y \varepsilon_y}{(1 + \tau_y) \cdot \eta_y + \varepsilon_y} = \bar{S}, \quad (9)
\]

where \(\bar{S}\) is again some (positive) constant. Evidently, equation (9) gives rise to a generalized version of the inverse elasticity rule: the optimal tax rate on a commodity is now seen to be inversely related to both its own-price elasticity of supply and the absolute value of its ordinary own-price elasticity of demand.

7It must be emphasized that this is not equivalent to the phenomenon of constant marginal costs.
General Sales/Turnover Tax

Tax Cascading: Concept and Measurement
HOWELL H. ZEE

• What is tax cascading, why is it undesirable, and what are its major determinants?
• How can tax cascading be illustrated through some simple analytics?
• How can the degree of tax cascading be estimated?

Whenever a commodity or service is taxed more than once under one tax as it passes through various stages of the production-distribution chain, for example, from the manufacturing to the retail stage, tax cascading results. A classic tax that gives rise to cascading is the multi-stage general turnover tax.8 Under this tax, every sales transaction is taxed, possibly at different stage-specific or transaction-specific rates, or both. Thus, a tax of \( x \) percent could be imposed on the sales of rubber, and a tax of \( y \) percent could be imposed on the sales of tires. Since the value of the rubber is incorporated into the value of tires, the former is taxed twice. It could, in fact, be taxed several more times, as would be the case if, for example, the sales of the tire wholesaler who buys the tires from the manufacturer are also subject to a tax of \( z \) percent. It is easy to surmise from the above that the effective tax burden of a cascading tax on a taxed commodity or service, by the time it reaches the final consumer, could be much higher than the tax’s nominal rate that is explicitly applied at the stage where the consumer makes his purchases. To put it differently, even if a commodity or service is exempted from tax at the retail stage, it may well be the case that its price would include tax elements stemming from taxes imposed at earlier stages on the various inputs used in its production. The true burden of a cascading tax is, therefore, frequently hidden from the consumer.

A cascading tax is universally regarded as undesirable, since by taxing transactions at stages prior to the stage of final consumption, it leads to more severe economic distortions than would a tax imposed only on final consumption, such as a retail sales tax or a full-fledged value-added tax (VAT) extended to the retail stage.9 Short of replacing a tax that cascades with one that does not, however, there are various mechanisms available to alleviate the extent of the cascading. These mechanisms are discussed in the next section entitled “Mechanisms to Alleviate Cascading.”

Determinants of the Degree of Cascading

The degree to which the true burden of a cascading tax is passed on to the final consumer depends on a number of complex but intertwined factors. Some of the more important ones are identified in this chapter.

Demand and supply elasticities

As is well known, the ability of a producer to shift the burden of the tax on his output to the purchaser depends on the relevant demand and supply elasticities. In a partial equilibrium framework of a single taxed commodity, the outcomes of four different combinations of such elasticities are illustrated in the four panels of Figure III.1. In each of the panels, \( DD' \) is the demand curve, \( SS' \) is the pretax supply curve, and \( TT' \) is the posttax supply curve. The vertical distance between the \( SS' \) and \( TT' \) curves at every price level \( P \) is simply equal to \( tP \), where \( t \) is the ad valorem tax rate. If \( P_0 \) and \( P_1 \) denote the pretax and posttax equilibrium price levels, respectively, then it is easily seen from Figure III.1 that the extent to which the price level would rise after the imposition of the tax would depend on the particular configuration of the demand and supply curves (the symbol \( \Delta \) in Figure III.1 is used to denote a change in any variable). Full forward shifting of the tax, in the sense that the purchaser bears the entire tax burden because the price level is increased by the same rate as the tax rate, would occur when either the demand curve is vertical (panel (a) of Figure III.1), that is, the elasticity of demand is zero, or the supply curve is horizontal (panel (d) of Figure III.1), that is, the elasticity of supply is infinite. No forward shifting would result if instead the demand curve is horizontal (panel (c) of Figure III.1). Under most

---

8 This tax is also known as a gross receipts tax or a transaction tax (see Due (1988) for a general discussion).

9 A general cascading tax imposed at a uniform rate would not, of course, lead to distortions in the choice of optimal factor proportions in a production process. It would, however, lead to excessive vertical integration in an economy’s industrial organization.
Figure III. 1. Demand and Supply Elasticities and the Forward Shifting of the Tax Burden

(a) Complete forward shifting

(b) Partial forward shifting

(c) No forward shifting

(d) Complete forward shifting

circumstances where neither the demand nor the supply curve displays an extreme elasticity, a partial forward shifting of the tax can be expected (panel (b) of Figure III.1).

As is evident from Figure III.1, the tax affects either the price or the quantities demanded, or both, of the taxed commodity, which in turn will have repercussions on other (even if nontaxed) commodities. Hence, the ultimate extent to which the tax is shifted forward cannot be determined until a complete general equilibrium analysis is undertaken. Voluminous literature on the (computable) general equilibrium effects of taxa-
tion has been developed (see, for example, Shoven and Whalley (1972)). Nevertheless, in most policy analyses of the economic impact of indirect taxes, it is not uncommon to adopt a short-run, partial equilibrium focus, in which case, it is appropriate to assume that the complete forward shifting of an indirect tax, corresponding to panel (d) of Figure III.1, will take place in the first instance.

**The ratio of taxed to nontaxed inputs**

The magnitude of the tax burden contained in the producer prices at each stage along the production-distribution chain depends on the ratio of taxed to nontaxed inputs at that stage, as well as such ratios at earlier stages. The lower these ratios, the smaller the tax burden. Food prices in many developing countries, for example, normally contain a relatively low tax component, as most agricultural inputs used in food production are tax-exempt.

**The degree of price pyramiding**

Price pyramiding at any given stage of production or distribution occurs when the seller raises his output price in excess of the tax burden on his inputs (see the following segment). This can occur, for example, if he has some monopolistic power and is, therefore, able to engage in the kind of pricing behavior that would allow him to derive additional profits whenever taxes are increased which are used as a convenient pretext to justify such behavior.

**The number of stages in the production-distribution chain**

The greater the number of stages of production and distribution a commodity passes through before reaching the final consumer, the greater the number of times taxed inputs would be subject to multiple taxation, and therefore, the higher the degree of the resultant cascading.

It is often difficult in practice to disentangle the complex interactions among the above factors in analyzing the extent of cascading contained in the price of any particular commodity. Nevertheless, some useful simple analytics can be developed for a systematic investigation of the problem.

**Simple Analytics of Cascading**

**Tax burden shifting in a single production-distribution stage**

To set the stage for a formal analysis, let $p$ be the producer price of a commodity, $k$ the input cost per unit of output, $v$ the commodity's per unit value added, $\tau$ the ad valorem tax rate on taxable inputs, and $\gamma$ the fraction of total inputs subject to tax. It then follows that

$$ p = v + k \cdot (1 + \gamma \cdot \tau).$$

(10)

Define the variable $\delta$ as the ratio of value added to input cost inclusive of the tax, that is

$$ \delta = v / [k \cdot (1 + \gamma \cdot \tau)].$$

(11)

Using equation (11), equation (10) can be rewritten as

$$ p = k \cdot (1 + \gamma \cdot \tau) \cdot (1 + \delta).$$

(12)

Thus, $\delta$ can be interpreted as a markup margin on the input cost inclusive of the tax. If the input cost exclusive of the tax ($k$) is held constant, then from equation (11), it is clear that any proportional change in the markup margin must stem from an excess of a proportional change in the value added over that in the tax rate, that is,

$$ \Delta \delta / \delta = \Delta v / v - \gamma \cdot \Delta \tau / (1 + \gamma \cdot \tau).$$

(13)

Equation (13) is crucial in describing the pricing behavior of the producer. To see this, first define $\phi$ as the elasticity of $p$ with respect to $\tau$:

$$ \phi = (\Delta p / p) / (\gamma \cdot \Delta \tau / (1 + \gamma \cdot \tau)).$$

(14)

$\phi$ measures, of course, the percentage change in $p$ as a result of a 1 percent change in $\tau$ (evaluated at the point $(1 + \gamma \cdot \tau)$). But directly from equation (12), the proportional change in $p$ is simply

$$ \Delta p / p = \gamma \cdot \Delta \tau / (1 + \gamma \cdot \tau) + \Delta \delta / (1 + \delta).$$

(15)

Hence, by substituting equation (15) into equation (14), $\phi$ can be restated as

$$ \phi = 1 + (\Delta \delta / \Delta \tau) \cdot [(1 + \gamma \cdot \tau) / (\gamma \cdot (1 + \delta))].$$

(16)

Equation (16) indicates that a critical factor in determining $\phi$ is the expression $(\Delta \delta / \Delta \tau)$, which measures the response of the markup margin $\delta$ to a change in the tax rate $\tau$. This response is, in turn, dependent on the extent to which the producer is willing to let the value added of his output change, as given by equation (13). Two limiting benchmark cases can be identified.

**Case A. Full price pyramiding ($\Delta \delta = 0$).**

If the producer does not allow his markup margin to change, that is, $\Delta \delta = 0$, then it follows immediately from equation (16) that $\phi = 1$, so that a 1 percent increase in the tax rate leads to a full 1 percent rise in the output price, irrespective of the proportions of taxed to nontaxed inputs and taxed inputs to value added. A producer who engages in this type of pricing behavior would clearly have his value added increased as a re-
result of the tax change. Indeed, it can be seen that, from equation (13), the proportional change in his value added would be equal to the proportional change in the tax rate, when $\Delta \delta = 0$.

**Case B. No shifting of the tax burden ($\Delta v = -\gamma k \cdot \Delta \tau$).**

If for some reason the producer is unable to shift the tax burden forward at all, then the incidence of any tax change would fall entirely on his value added, that is, $\Delta v = -\gamma k \cdot \Delta \tau$. Substituting this into equation (13) shows that his markup margin is changed according to

$$\Delta \delta = -\Delta \tau \cdot \gamma (1 + \delta)/(1 + \gamma + \tau).$$

(17)

It then follows from equation (16) that $\phi = 0$, that is, there is no change in the producer price $p$. This result is again independent of $\gamma$.

The above two cases clearly bracket all possible outcomes on the producer price in response to a change in the tax rate. An interesting intermediate case, for example, would be that the producer merely protects his value added ($\Delta v = 0$) in the face of the tax change. In this case, from equation (13), his markup margin is changed by

$$\Delta \delta = -\Delta \tau \cdot \gamma (1 + \delta)/(1 + \gamma + \tau).$$

Substituting this into equation (16) yields $\phi = 1/(1 + \delta) < 1$. Hence, the proportional change in the producer price is positive but less than the proportional change in the tax rate, and, since the magnitude of $\delta$ is inversely related to both $k$ and $\gamma$ (see equation (11)), the extent of the price change is larger the higher the ratios of taxed to nontaxed inputs and taxed inputs to value added.

**A synthetic rule**

Quite frequently, it is useful to conduct simulation exercises to analyze the price impact of replacing one tax with another under different assumed degrees of tax burden shifting. For this purpose, it would be analytically convenient if all the different possible changes in the producer price between, and inclusive of, the above two limiting cases could be captured through varying a single parameter. The simplest procedure to achieve this is to conceptualize the pricing mechanism as if it operates, not according to equation (10), but according to the synthetic rule of

$$p = A \cdot (1 + \alpha - \tau),$$

(18)

where $1 \geq \alpha \geq 0$ and $A$ is a nonzero constant. Thus, the full price pyramiding case implies $\alpha = 1$, and the case of no tax burden shifting implies $\alpha = 0$. Varying $\alpha$ between zero and unity captures all possible outcomes, including the intermediate case discussed above where

10This could be owing to, for example, the demand curve that the faces being horizontal (see panel (c) of Figure III.1).

the producer’s pricing behavior is such that his value added is not affected by the tax change.

**Cascading and multiple stages of production and distribution**

With multiple stages of production and distribution, the (possibly different) pricing behavior of producers, tax rate, ratio of taxed to nontaxed inputs, and proportion of taxed inputs to value added at each stage will all have a bearing on the ultimate impact of cascading on the price the consumer faces. The analytics developed above pertaining to a single stage cannot, therefore be generalized in a simple way to the case of multiple stages. As before, however, an interesting benchmark case can be identified under the special assumption that all producers engage in full price pyramiding behavior.

Using the same notations as before, and employing subscripts to denote the different stages, the pricing behavior of the producer in the first stage is given by

$$p_1 = (1 + \tau_1) \cdot k_1 \cdot (1 + \delta_1),$$

(19)

where $k_1$ represents his purchased inputs, which are taxed at the rate $\tau_1$, and $\delta_1$ is his markup margin, as defined earlier. Hence, equation (19) corresponds to the synthetic rule given by equation (18) with $\alpha = 1$ and $A = k_1 \cdot (1 + \delta_1)$. As a buyer of the output of the producer in the first stage, the producer in the second stage purchases his inputs at the producer price of $p_1$, that is, $k_2 = p_1$. If these inputs are taxed at the rate $\tau_2$, he would set his price according to

$$p_2 = (1 + \tau_2) \cdot k_2 \cdot (1 + \delta_2) = (1 + \tau_2) \cdot p_1 \cdot (1 + \delta_2).$$

(20)

Using equation (10), equation (11) can be rewritten as

$$p_2 = (1 + \tau_2) \cdot (1 + \tau_3) \cdot (1 + \delta_2) \cdot (1 + \delta_3) \cdot k_2.$$  

(21)

Equation (12) is generalizable to any number of stages in a straightforward manner. If $p_n$ is the producer price after $n$ stages, and $p'_n$ denotes the corresponding producer price in the absence of the tax, then the ratio between the two would be a measure of the degree of cascading in the producer price. From equation (21), it can be inferred that this ratio is

$$p_n/p'_n = (1 + \tau_1) \cdot (1 + \tau_2) \cdots (1 + \tau_n).$$

(22)

In the special case where all the tax rates are the same, as it would be with a flat-rate general turnover tax, i.e., $\tau_1 = \tau_2 = \cdots = \tau_n = \tau$, equation (22) takes the particularly simple form of

$$p_n/p'_n = (1 + \tau)^n.$$  

(23)

Of course, a tax is also likely to be imposed on the final consumer after the last stage in the production-
distribution chain. If $p_c$ denotes the price inclusive of the tax imposed on the consumer at the rate $\tau_c$ then after $n$ stages, the consumer price would be

$$p_c = (1 + \tau_c) \cdot p_n \tag{24}$$

By definition, the impact of $\tau_c$, being imposed on the final consumer situated at the end of the chain, has no room to cascade forward.

**A synthetic rule**

The analytics for the case of multiple stages have been developed on the special assumption of full price pyramiding behavior on the part of the producers. This assumption, however, is not as limiting as it may seem, because it sets an upper bound to the degree of cascading. If some producers engage in less-than-full price pyramiding behavior, then the degree of resultant cascading would be less than that indicated by equation (22), or that by equation (23) with a uniform tax rate. When tax rates are in fact uniform, it is again possible to capture such an outcome by a simple synthetic pricing rule, for this case of multiple stages, similar to that for the single-stage case given by equation (18).

The impact on the consumer price of a general turnover tax at the uniform rate $\tau$ can be stated, irrespective of the number of stages in the production-distribution chain, by the following synthetic rule:

$$p_c = B \cdot (1 + \delta \cdot \tau) \cdot (1 + \tau), \tag{25}$$

where $\delta \geq 0$ is a measure of the total amount of the tax burden cascaded forward based on the producers' pricing behavior through the various stages just prior to reaching the final consumer, and $B$ is a nonzero constant. If $\delta = 0$, then no producer along the production-distribution chain shifts his tax burden forward, and, therefore, there would be no cascading. The upper limit of $\delta$, denoted by $\delta^*$ and represents the outcome of full price pyramiding behavior, is dependent on the number of stages of production and distribution. From equation (23), it is easily seen that $\delta^*$ is given by

$$\delta^* = [(1 + \tau)^n - 1]/\tau. \tag{26}$$

Hence, by varying the value of $\delta$ between 0 and $\delta^*$, the impact of all possible degrees of cascading can be simulated. For example, if $n = 2$, and every producer engages in full price pyramiding, then by substituting equation (26) into equation (25), the consumer price would be

$$p_c = B \cdot (1 + \tau)^3, \tag{27}$$

with $B$ set equal to $p'_2 = (1 + \delta_2) \cdot (1 + \delta_2) \cdot k_f$.

**Estimating the Degree of Cascading**

In addition to simulating the impact of tax cascading on the consumer price, which is useful for comparing the price effects of alternative taxes, it is sometimes necessary to empirically estimate the degree of cascading associated with an existing tax. In general, this estimation cannot be carried out without detailed information on all the determinants of cascading discussed earlier in this chapter. If the existing tax has a uniform rate, however, the synthetic pricing rule, given in equation (25), can be used to derive an estimate for $\delta$ from a few, usually readily available, aggregate data on revenue and consumption.

Let $E$ and $E'$ be, respectively, total taxed and tax-exempt final consumption expenditures exclusive of the existing tax. If all taxes on capital and intermediate goods are viewed as ultimately falling on consumption through cascading at varying degrees, then it must be true that, upon utilizing the synthetic pricing rule,

$$R = [(1 + \delta \cdot \tau) \cdot (1 + \tau) - 1] \cdot E + \delta \cdot \tau \cdot E', \tag{28}$$

where $R$ is the total tax revenue. The second term on the right-hand side of equation (28) represents the amount of cascaded tax embodied in that part of consumption expenditure not subject to tax at the final consumption point. $E$ and $E'$ are, however, not directly observable, since consumption data are typically compiled inclusive of the tax. If $F$ and $F'$ denote, respectively, observable taxed and tax-exempt final consumption expenditures, then $E$ and $E'$ can be derived as follows:

$$E = F/(1 + \delta \cdot \tau), \quad \text{and}$$

$$E' = F'/(1 + \delta \cdot \tau). \tag{30}$$

Substituting equations (29) and (30) into equation (28), it is possible to solve for $\delta$ as

$$\delta = [R - F \cdot (1 + \tau)]/\tau \cdot (F + F' - R). \tag{31}$$

Equation (31) provides a simple and direct formula for estimating the degree of cascading of any tax with a uniform rate.\(^{11}\)

\(^{11}\)The proportion of tax to total consumption expenditures can usually be ascertained from an examination of the consumption basket given by either household budget surveys or the consumer price index.

\(^{12}\)A numerical example illustrating the use of the analytical framework developed here is provided in this chapter, relating to the price effects of replacing a general sales tax with a VAT.
Mechanisms to Alleviate Cascading

RONALD T. McMorran

- What mechanisms can be used to reduce cascading in sales taxes?
- What factors influence the choice of mechanism?

There are two general types of sales taxes: single-stage taxes and multi-stage taxes. A single-stage sales tax applies only at one stage of the production-distribution chain. For example, it may only apply to sales at the manufacturing, wholesale, or retail stage. In contrast, a multi-stage sales tax applies at several stages of the production-distribution chain. Examples of this type of tax include turnover taxes and the VAT. These taxes may apply to most sales of goods and services in the economy no matter at which stage of the production-distribution chain the sale is made. Alternatively, these taxes may include all sales up to a certain stage in the production-distribution chain such as the manufacturing or wholesale stages.

Table III.1. Illustration of Tax Cascading

Example 1: 10 percent multi-stage turnover tax

<table>
<thead>
<tr>
<th></th>
<th>Primary Producer</th>
<th>Manufacturer</th>
<th>Wholesaler</th>
<th>Retailer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales (excluding tax at the current stage)</td>
<td>100</td>
<td>210</td>
<td>331</td>
<td>464.1</td>
</tr>
<tr>
<td>Purchases (including tax)</td>
<td>0</td>
<td>110</td>
<td>231</td>
<td>364.1</td>
</tr>
<tr>
<td>Value added:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From current stage</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Previously taxed</td>
<td>0</td>
<td>100</td>
<td>200</td>
<td>300</td>
</tr>
<tr>
<td>Previously untaxed</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tax</td>
<td>10</td>
<td>21</td>
<td>33.1</td>
<td>46.41</td>
</tr>
<tr>
<td>Of which attributable to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current stage's value added</td>
<td>(10)</td>
<td>(10)</td>
<td>(10)</td>
<td>(10)</td>
</tr>
<tr>
<td>Previously taxed value added</td>
<td>(0)</td>
<td>(10)</td>
<td>(20)</td>
<td>(30)</td>
</tr>
<tr>
<td>Tax on tax</td>
<td>(0)</td>
<td>(1)</td>
<td>(3.1)</td>
<td>(6.41)</td>
</tr>
<tr>
<td>Previously untaxed value added</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
</tbody>
</table>

Example 2: 10 percent single-stage manufacturers' tax

<table>
<thead>
<tr>
<th></th>
<th>Primary Producer</th>
<th>Manufacturer</th>
<th>Wholesaler</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales (excluding tax at the current stage)</td>
<td>100</td>
<td>200</td>
<td>320</td>
<td>420</td>
</tr>
<tr>
<td>Purchases (including tax)</td>
<td>0</td>
<td>100</td>
<td>220</td>
<td>320</td>
</tr>
<tr>
<td>Value added:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From current stage</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Previously taxed</td>
<td>0</td>
<td>0</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Previously untaxed</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Tax</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>42</td>
</tr>
<tr>
<td>Of which attributable to:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current stage's value added</td>
<td>(0)</td>
<td>(10)</td>
<td>(0)</td>
<td>(10)</td>
</tr>
<tr>
<td>Previously taxed value added</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(20)</td>
</tr>
<tr>
<td>Tax on tax</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(2)</td>
</tr>
<tr>
<td>Previously untaxed value added</td>
<td>(0)</td>
<td>(10)</td>
<td>(0)</td>
<td>(10)</td>
</tr>
</tbody>
</table>

Source: Staff calculations.
in tax of which 10 is attributable to the current stage's value added, 10 is attributable to the previous stage's value added, and 1 is attributable to tax on tax charged on the sale at the earlier stage. In the third stage, the wholesaler makes a sale of 331 to the retailer. The wholesaler made purchases of 231 including tax and has value added of 100. Tax on the sale is 33.1 of which 10 is attributable to the current stage's value added, 20 is attributable to previous stages' value added, and 3.1 is attributable to tax on tax charged on sales at earlier stages. In the final stage, the retailer makes a sale of 464.1 to a consumer. The retailer made purchases of 364.1 including tax and has value added of 100. Tax on the sale is 46.41 of which 10 is attributable to the current stage's value added, 30 is attributable to previous stages' value added, and 6.41 is attributable to tax on tax charged on sales at earlier stages.

Generally, sales between businesses registered are not subject to tax under a single-stage sales tax. Sales from registered businesses to unregistered businesses and consumers, however, are subject to tax. Tax cascading may result when an unregistered business makes sales to a registered business and the unregistered business used taxable inputs to produce the sale.

Tax cascading under a single-stage manufacturers' tax set at a rate of 10 percent is illustrated in Example 2 in Table III.1. The example traces sales as they proceed through a production-distribution chain from a primary producer to a manufacturer to a wholesaler and then back to a manufacturer. In the first stage, the primary producer makes a sale of 100 to the manufacturer. The producer has value added of 100 and there is no tax on the sale. In the second stage, the manufacturer makes a sale of 200 to the wholesaler. The manufacturer has made purchases of 100 and has value added of 100. On the sale, the manufacturer collects 20 in tax of which 10 is attributable to the current stage's value added and 10 is attributable to the previous stage's value added, which was not previously taxed. In the third stage, the wholesaler makes a sale of 320 to a manufacturer. The wholesaler made purchases of 220 including tax and has value added of 100. There is no tax on the sale. In the final stage, the manufacturer makes a sale of 420 to an unregistered business or a consumer. The wholesaler made purchases of 320 and has value added of 100. Tax on the sale is 42 of which 10 is attributable to the current stage's value added, 20 is attributable to value added that was taxed at a previous stage, 10 is attributable to value added that was not previously taxed, and 2 is attributable to tax on tax charged on the sale at the earlier manufacturing stage.

Cascading may result in distortions in relative prices by causing the effective tax rate on final sales of various goods and services to be different from statutory rates. Cascading may also result in an increase in the cost of capital to businesses when taxes cascade on capital inputs, distorting productive efficiency. There are various mechanisms to mitigate these adverse consequences of cascading in sales taxes.

**Single-Stage Sales Taxes**

To eliminate cascading under a single-stage sales tax, sales tax generally does not apply to sales between registered traders. This is why single-stage sales taxes are sometimes referred to as suspensive taxes—tax does not apply until a sale is made by a registered vendor to an unregistered purchaser. Essentially, these taxes "ring fence" those businesses that are producing or trading taxable goods and relieve those businesses from tax on their purchases from other taxable producers or traders.

In practice, registered vendors are obliged to charge tax on all of their sales except for sales to other registered vendors. The onus is generally on the vendor to prove to the tax authorities that he was not required to charge tax on certain sales. Normally, the tax authorities will accept the purchaser's registration number as well as other information that will identify the purchaser, such as the name and location of the business, as sufficient evidence that the vendor was not required to charge tax on the sale.

On the surface, this suspensive mechanism appears simple and easy to administer for both the government and businesses. However, this is not necessarily the case. Such a system can be subject to abuse and non-compliance. Suppliers are required to evaluate evidence provided to them by purchasers. Purchasers could provide false evidence to the supplier. In such circumstances, the tax authorities may require suppliers to remit tax that was not collected even though the purchaser provided evidence that tax was not required to be paid on the sale. This potential liability to tax can introduce some uncertainty into the tax system for suppliers.

More fundamentally, however, the system is incomplete in eliminating tax cascading. Tax relief is only provided to registered producers or traders. If some part of the production or distribution chain is not permitted to register for the tax, tax cascading may result. A common example of this problem is a single-stage manufacturers' sales tax that subjects imports and manufactured goods to tax but does not subject wholesale and retail sales, or primary production to tax. This case
has been illustrated above. Under these circumstances, primary producers may produce goods using inputs which were taxable at the manufacturer or import stage. If these goods are subsequently used by manufacturers or importers to produce taxable goods, then value added that was taxed at an earlier stage in the production-distribution chain will be subject to tax again, thereby resulting in tax cascading.

**Multi-Stage Sales Taxes**

By their very nature, turnover taxes are cascade taxes because tax applies at many stages of the production-distribution chain without any mechanism to remove tax on business inputs. Some jurisdictions have attempted to alleviate cascading in turnover taxes by allowing relief for tax paid on some business inputs. This is usually accomplished by allowing businesses to earn a credit for tax paid on certain items. These items may include capital, purchases or other inputs into a production process. In some cases, relief may not be provided for inputs that are attributable to general overhead or that may not be attributable to the production of taxable goods.

Partial relief does remove some elements of cascading in a turnover tax system. Such relief, however, does not eliminate all cascading. Some jurisdictions have used partial relief as a means of experimenting with the credit mechanism to learn its operation and functioning, as a step toward the introduction of a VAT.

Multi-stage noncumulative taxes or VATs are those where the trader pays net tax on value added at each stage. There are two basic methods. The first method, the credit-invoice method, allows traders to credit tax paid on their purchases against tax collected on their sales in determining the net tax payable to the treasury. The second method, the subtraction method, is used only in a few countries and requires traders to calculate value added as the difference between sales and purchased inputs and remit tax calculated on that difference. VATs are more effective in removing tax cascading on supplies between registered traders than single-stage suspension taxes because there are fewer opportunities for avoidance and noncompliance.

Unlike a suspensive system, under the credit-invoice method of VAT, all sales that are made by registered traders are subject to tax and only registered traders are permitted a credit for tax paid on inputs. Under such a system, the onus is on registered traders to prove that they are eligible to claim a credit for tax paid on inputs. Sufficient evidence of tax paid on inputs is an invoice that documents the taxable transaction for which a credit claim is being made.

The credit-invoice scheme is one of the main benefits of a VAT. It provides an audit trail that makes tax administration easier and supports voluntary compliance with the tax. Such a system is superior to the suspensive system. Under the suspensive system, the tax authorities are required to assess suppliers who did not charge tax in circumstances where tax should have been collected by a supplier on a sale. In contrast, under the credit-invoice system, the tax authorities can deny credits in circumstances where insufficient or incorrect evidence was supplied to justify a credit claim.

The credit-invoice mechanism, however, does not completely remove cascading if there are exemptions. As noted earlier, the net tax calculation of a taxable business that only purchases taxable inputs is the difference between tax on its sales and tax paid on its purchases—that is, the tax value of the business' value added. If the business, however, purchases exempt inputs, then no tax would be paid on its exempt purchases and no credit would be claimable on these purchases. Thus, the net tax calculation becomes the tax value of the business' value added plus the tax value of exempt inputs. The value of the exempt inputs will be made up of the sum of the value added from all previous stages in the production-distribution chain attributable to those exempt inputs. If some of that value added was previously subject to tax, there will be cascading. The cascading consequence of exemptions in a VAT under the credit-invoice method is numerically illustrated later in this chapter.

**A Comparison Between the Sales Tax and a VAT**

**Ronald T. McMorran**

- **What are some of the practical differences between sales taxes and VATs?**

There are many differences between a single-stage sales tax and a VAT. These differences relate to administrative and operational aspects of the two types of taxes. These aspects affect administration and compliance costs, tax evasion, and the timing of revenue flows to the treasury. Another issue is the treatment of services. In concept, single-stage sales taxes and VATs can treat services equally. In practice, this has not been...
the case. This section discusses differences between a single-stage sales tax and a VAT. 14

Location of the Tax Point in the Production-Distribution Chain

The experience of many countries provides strong evidence that any sales tax that stops short of the retail level is basically unsatisfactory and is subject to endless controversy. The problems relate to the tendency of firms to shift various activities beyond the impact of the tax, the inability to treat different distribution systems equally for tax purposes, the inability to tax imports and domestic goods equally, and the relatively high tax rate necessary for a given revenue. 15

For example, under a manufacturing level VAT, a fully integrated manufacturer/retailer would be required to calculate output tax on the value of its retail sales, whereas a manufacturer who makes sales to a middle man would only be required to calculate output tax on the value of sales to the middle man. Thus, an incentive is created to shift economic activity to the untaxed stages of distribution. At the same time, disincentives are created against the integration of the functions of manufacturing, wholesaling, and retailing. Similar problems arise with a single-stage manufacturer’s tax. Further, most manufacturer level taxes have developed complex and confusing formulas for markups from ex-factory prices or discounts from retail prices to calculate taxable wholesale prices. Over time, these rules become increasingly elaborate.

In addition, the tax base of a manufacturing stage VAT will be narrower than a retail-stage VAT. As a result, for a given amount of revenue, a retail VAT requires a lower rate than a manufacturing VAT. To the extent that the incentive for tax evasion is positively related to the tax rate, a retail VAT will thus have a lower incentive for tax evasion than a manufacturer’s VAT. Also, the share of tax paid by any one taxpayer is also lower under a retail VAT, so that such a tax system is less reliant on the compliance of a few relative to a manufacturing VAT.

In countries with a tradition of keeping business records and high tax morality, it is feasible to apply sales taxes at the retail level. For these countries, the question is whether a retail-level tax should be a single-stage retail-sales tax or a multi-stage retail VAT.

In countries without a tradition of business record keeping or with low tax morality, it is not feasible to extend sales taxes to the retail level. For these countries, there are two key tax policy questions: at what stage in the production-distribution chain should the tax be applied, and should the tax be a single-stage sales tax or a VAT? Generally, the broadest possible stage to apply a sales tax in these countries is at the manufacturing/import stage because the next stage in the production-distribution chain—the wholesale trade stage—is difficult to define. 16 Thus, the remaining policy question for these countries is to determine whether a single-stage sales tax or a VAT should be levied.

Timing of Revenue Flows

In principle, a single-stage sales tax and a VAT that have the same tax base should have the same net revenue effect and there should be little difference in the timing of their revenue flows. Tables III.2 and III.3 show the revenue flows to the government under a retail-stage VAT and a single-stage retail-sales tax, respectively. Usually, under a VAT, tax payable on sales made in a reporting period is due immediately after a reporting period ends and credit for input tax paid in a reporting period is claimable immediately after the end of the reporting period in which tax is paid. As a result, business to business sales will generally offset each other in government accounts for a reporting period and net VAT revenue in a period will only be the tax paid on final sales to consumers. This is the same as revenue flows to the government under a single-stage retail-sales tax. Differences in the timing of government revenue may arise if traders have different return periods under a VAT.

Tables III.2 and III.3 also illustrate how revenue flows differ from the attribution of tax revenue to a stage of production. Under a single-stage retail-sales tax, revenue is attributed to a single stage—the retail stage. In contrast, under a VAT, tax revenue is attributed to each stage in the production-distribution chain as tax is applied to each stage’s value added.

Administration and Compliance Costs

Administration and compliance costs under a single-stage sales tax and a VAT, which both extend to the same level in the production-distribution chain, are not likely to differ substantially for two reasons. First, tax administration has to be vigilant in its administration of either type of tax. Effective control of the tax system re-

---

15See Due (1985).
Second, traders' compliance costs may be less in aggregate under a single-stage sales tax than under a VAT because a single-stage sales tax requires less extensive records—under a single-stage sales tax, only records of sales are required to be kept, whereas under a VAT, records are required to be kept for both purchases and sales. This difference in compliance costs is trivial, however, if commercial law or other taxes require complete sets of books and records that include purchase and sale ledgers.

**Table III.3. Illustration of Government Revenue Flows Under a Retail-Sales Tax: Ten Percent Retail-Sales Tax**

<table>
<thead>
<tr>
<th>Period 1: $100 sale from primary producer to manufacturer</th>
<th>Period 2: $200 sale from manufacturer to retailer</th>
<th>Period 3: $300 sale from retailer to consumer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tax payable on sale</td>
<td>Tax payable on sale</td>
<td>Tax payable on sale</td>
</tr>
<tr>
<td>Tax credit claimed on purchase</td>
<td>Tax credit claimed on purchase</td>
<td>Tax credit claimed on purchase</td>
</tr>
<tr>
<td>Net revenue to government</td>
<td>Net revenue to government</td>
<td>Net revenue to government</td>
</tr>
<tr>
<td>Sales excluding tax</td>
<td>Purchases excluding tax</td>
<td>Tax on sales</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>300</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

**Tax Evasion**

One of the most contentious issues in this debate is the question of whether the VAT is more resistant to tax evasion. There are many factors, common to both taxes, that affect the level of tax evasion in a given country, including the level of tax morality and ability of the tax administration to enforce the tax. Neverthe-

---

17 See Bit (1986).
less, there are features of both taxes that may lead to different levels of tax evasion.

Both types of taxes are subject to evasion in the shadow economy. Under certain conditions, the two types of taxes are subject to different forms of evasion. Under a VAT, registered traders can reduce their tax liabilities by claiming, in their tax returns, credit for tax that was not paid on purchases. Under a single-stage sales tax, nonregistered purchasers may provide false information to sellers to get tax relief, thereby evading tax. The key difference between these forms of evasion is that under a VAT, the government controls the refund of input tax whereas, under a single-stage sales tax, the government never receives the tax.

Another key issue is the extent to which tax revenues are at risk. Under a single-stage sales tax, tax revenue is only collected at one stage of production or distribution in the economy. Tax evasion thus results in a loss of the entire amount of tax that would be collected at the point of taxation. For example, retail-sales taxes are collected from the weakest link in the distribution chain. Many retail sales require no invoice and the only evidence may be, for example, till rolls. Cash sales are common.

In contrast, VATs may be collected at several levels of production or distribution. Tax revenue is generally only at risk for the tax on value added at the stage in production or distribution where the tax is being evaded. And if this evasion occurs prior to the retail level under a retail VAT, then this will reduce the credit that the purchaser of the intermediate product can claim. Thus, this revenue should eventually be caught. However, "the VAT provides one opportunity for fraud not available in other forms of sales tax: fictitious claims for excessive credits and rebates, particularly through the use of counterfeit invoices." 18

A retail VAT provides a better audit trail than a single-stage retail-sales tax. The use of invoices throughout the system gives auditors an opportunity to create a better series of checks. This greatly increases the risk of detection under the VAT. 19 A large portion of tax revenue is typically collected at pre-retail stages; businesses at these stages are typically larger than retail businesses, keep better records, and are generally less prone to evasion. Evasion of a VAT at the retail level results in the loss of only the portion of the tax that is paid on the retail margin (unless the retailer manages to get credit for tax paid on purchases of goods but does not report tax on sales). 20

### The Taxation of Services

In principle, it is equally easy to tax services under a single-stage sales tax as under a VAT. In practice, however, services have been treated differently under single-stage taxes and VATs that have comparable coverage, especially at the retail level. In many cases, single-stage taxes do not apply to services. In other cases, single-stage taxes are imposed on specific services, such as hotel rooms or insurance premiums. VATs at the manufacturer-importer level often include selected services, such as telecommunications and transportation. At the retail level, VATs typically include all services except for a few exempt services, such as health care, education, social services, and financial services. These few services, however, represent a significant portion of value added attributable to services.

Traditionally, services have been difficult to tax because many services are rendered at the final stage of the production or distribution process. Usually, the labor content as a proportion of the final price is high, as is the degree of specialization. As a result, these services are often provided by many small-scale establishments, which may be difficult to reach administratively. Another complicating factor is that many services are often sold for business and personal use. At the final stage, taxable personal use is extremely difficult to separate from exempt business use, while cascading results if both uses are taxed across-the-board. Because of these reasons, services have not typically been included in pre-retail level sales taxes. 21

For retail-level taxes, there are two reasons why it is easier to accommodate services under a VAT than under a single-stage tax. First, businesses that provide services to consumers also often provide services to other businesses that would like to claim refunds of tax on their inputs so these businesses wish to be part of the VAT. Second, mixed-use services for business and private consumption pose far fewer problems under a VAT than under a single-stage tax. Under a VAT, businesses' services and private consumption are taxed in full. There is no need for a seller to determine if tax should be suspended as under a single-stage tax. Rather, the purchaser must satisfy the tax authorities that a credit can be claimed for tax paid on the purchase of the service. Relative to a single-stage tax, this removes an element of uncertainty and reduces the potential for tax evasion. 22

---

20 See Due (1985).
Value-Added Tax

HOWELL H. ZEE

Selected Issues in Designing a VAT

• What are the alternative variants of a VAT, and how do they relate to each other?
• How is the origin principle different from the destination principle, and what economic implications are entailed by each?
• What are the possible price effects of introducing a VAT, and how are they measured?

There are three possible variants of a VAT: the product-type (P-VAT), the income-type (I-VAT), and the consumption-type (C-VAT). Each variant, in turn, can be implemented under two possible principles: the origin and destination principles; and by using two main methods of computation: the credit-invoice and subtraction methods. Hence, in considering the introduction of a VAT, the choice of an appropriate VAT variant, principle, and computation method must be made. In addition, a variety of other VAT design issues must also be addressed. These include the number of tax rates, the scopes of zero-rating and exemptions, the level of exemption threshold, and possible special mechanisms to deal with businesses which are to be exempt. If the prospective VAT is to replace some existing taxes, transitional arrangements to grant relief to stocks and inventories to compensate for the existing tax elements embodied in their values might also be required. Policy decisions on the above issues will have an important bearing on both the VAT's revenue yield and its economic consequences.

This section provides a brief examination of the relative merits and shortcomings of the three VAT variants and the two tax principles. It also discusses the potential price effects of introducing a VAT. The economic implications of alternative methods of VAT computation, the number of tax rates, zero-ratings, and exemptions are considered in the next section on credit versus subtraction method. Another section, later in this chapter, illustrates how the base of a VAT can be estimated.

VAT Variants

The relationship among the three variants of VAT can be best understood by comparing the income and expenditure aggregates in the national income accounts. Expenditures on the gross domestic product (GDP) consist of final private consumption expenditure (C), gross investment expenditure (I), final government nonwage expenditure on goods and services (G_c), government expenditure on wages and salaries (G_w), and the trade balance (the values of exported (X) less imported (M) goods and nonfactor services):

$$GDP = C + I + G_c + G_w + (X - M),$$

while gross domestic income (GDI) is the sum of factor income payments (wages, interest, profits, etc.)—commonly referred to as the value added (V) of production—and depreciation (D):

$$GDI = V + D.$$  

Equality between income and expenditure gives the basic national income accounting identity:

$$GDP = GDI.$$  

Equation (34) can be stated in terms of either factor cost (i.e., exclusive of indirect taxes net of subsidies) or market prices (i.e., inclusive of indirect taxes net of subsidies).

P-VAT

Broadly speaking, a P-VAT taxes all expenditures (except government wage expenditure, which is infeasible to tax under any variant of VAT) on GDP if implemented on the origin principle, and on GDP adjusted for the trade balance if implemented on the destination principle. Assuming for the time being that the origin principle is employed, so that exports, being of domestic origin, are taxed but imports, whose value originated from abroad, are not (see below for a detailed discussion). Then the base of a P-VAT is simply the sum of all expenditures on GDP (hence its name as a product-type VAT) net of government wage expenditure. From equation (32), this base can be expressed as

$$\text{base of P-VAT} = GDP - G_w = C + I + G_c + (X - M).$$  

While the C-VAT is the most prevalent variant adopted in international practice, a discussion of all three variants is important in understanding their relative merits and shortcomings.
Gross investment expenditure, which is part of the base of a P-VAT, reflects an economy's actual aggregate expenditure on capital goods (i.e., gross capital formation) in a given period. Part of this expenditure, however, is used to compensate for capital goods that have been consumed or depreciated. While depreciation as such is merely a bookkeeping entry and does not represent an actual economic transaction, it does affect the computation of profitability, and thus the value added, of businesses. An I-VAT excludes depreciation from its base; it therefore taxes the net, rather than the gross, investment expenditure:

$$\text{base of I-VAT} = \text{GDP} - G_w - D = C + (I - D) + G_c + (X - M).$$

(36)

From equations (33) and (34), however, it is clear that the base of an I-VAT can be alternatively expressed as

$$\text{base of I-VAT} = \text{GDI} - G_w - D = V - G_w. \quad (36')$$

The second equality in equation (36') indicates that the base of an I-VAT is simply the sum of factor income payments (net of government wage expenditure), hence its name as an income-type VAT.24

**C-VAT**

If, in addition to depreciation, expenditures on capital goods which contribute to a net augmentation of the capital stock are also not taxed, then the entire gross investment expenditure would in effect be excluded from the tax base. The resultant base would be the base of a C-VAT:

$$\text{base of C-VAT} = \text{GDP} - G_w - I = C + G_c + (X - M). \quad (37)$$

**A Comparison of the Three Variants**

It is clear from the above that the P-VAT has the broadest base among the three variants; the C-VAT has the narrowest. The broadness of the P-VAT base is bought, however, at a potentially high economic price. By imposing a tax burden on gross purchases of capital goods without giving any relief even on depreciation, the P-VAT exerts a strong discouragement to investment. Furthermore, to the extent that businesses succeed in shifting at least part of their capital costs forward, taxing capital goods would result in cascading if the P-VAT uses the credit-invoice method.25 By similar reasoning, these shortcomings, albeit to a lesser extent, are also present with the I-VAT.

While providing the narrowest of the three tax bases, the C-VAT, being a general tax on consumption (inclusive of net exports if implemented on the origin principle, as shown above), is economically the most neutral (as it generates no distortion in the production process between capital and other inputs) and is, therefore, generally regarded as the superior variant among the three. It is also the most widely adopted variant in countries that have a VAT. It does, however, require the highest tax rate to achieve a given revenue yield. For this reason, some countries have attempted to expand the base of the C-VAT by taxing capital goods partially.26 It is easy to see that, depending on the fraction of capital goods that is deemed taxable, such an expansion of the C-VAT base amounts to a modified version of I-VAT.

If the trade balance, that is, $X - M$, is removed from each of the three tax bases expressed above, then all three VAT variants would be converted to the destination principle. In this case, it follows immediately from equation (28) that the base of the C-VAT becomes $C + G_c$ which corresponds to the total sales at the retail level. Hence, in terms of tax policy, a C-VAT implemented under the destination principle is equivalent to a retail-sales tax. The administrations and collection mechanisms of the two taxes are, of course, quite different.27

**Origin Versus Destination Principle**

**Basic concepts**

A VAT can be implemented under either the origin or the destination principle.28 Under the former, the VAT is imposed on the value added of all taxable products (henceforth taken to encompass both goods and services) that are produced domestically; under the latter, the VAT is imposed on the value added of all taxable products that are consumed domestically. Obviously, the two principles are identical in a closed

---

24 In national income accounting terminology, subtracting depreciation from gross domestic product (income) results in net domestic product (income). When net factor income from abroad (abstracted from the discussion here for simplicity) is added to net domestic income, the resultant sum is commonly referred to as simply national income, which in turn can be stated on the basis of either factor cost or market prices.

25 See earlier section on tax cascading: concept and measurement for a discussion of forward shifting and cascading, and the following section for a discussion of the credit-invoice method.

26 In practice, this is usually achieved, in a VAT that uses the credit-invoice method, by phasing the provision of tax credits on capital goods over a number of years (as opposed to providing the credits immediately).

27 See an earlier section of this chapter for a more detailed comparison between a retail-sales tax and a VAT.

28 A third principle, known as the restricted origin principle, is discussed below.
In an open economy, the difference between them lies solely in their treatments of imports and exports: exports are taxed but imports are not under the origin principle, while just the converse holds under the destination principle.

It is important to note that the distinction between the two principles is based on the location of production and consumption, and not on the type of products being produced or consumed—the latter being the basis for distinguishing between the P-VAT and the C-VAT, as discussed above. Hence, contrary to popular belief, there is no conceptual incompatibility in implementing a C-VAT under the origin principle or a P-VAT under the destination principle. With a C-VAT, for example, capital goods, whether imported or locally produced, are not taxed under either principle as long as they are purchased for domestic use, but will be taxed if exported under the origin principle. Exported capital goods are, of course, not taxed with a C-VAT under the destination principle. A matrix of the tax treatments of capital goods under the four combinations of the two VAT variants and two tax principles is given below.

<table>
<thead>
<tr>
<th></th>
<th>C-VAT</th>
<th>P-VAT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Origin</strong></td>
<td>Taxed if exported; not taxed otherwise.</td>
<td>Not taxed if imported; taxed otherwise.</td>
</tr>
<tr>
<td><strong>Destination</strong></td>
<td>Not taxed.</td>
<td>Not taxed if exported; taxed otherwise.</td>
</tr>
</tbody>
</table>

Hence, capital goods completely escape taxation only under the combination of C-VAT and destination principle.

**Border tax adjustments and valuation of traded products**

Since a destination-based VAT taxes imports but not exports, it requires border tax adjustments, that is, the VAT must be applied on (removed from) products as they enter (leave) the country. Surprisingly, these border tax adjustments require no special attention be paid to the valuation of traded products for VAT purposes. Irrespective of the method of computation (credit-invoice or subtraction) used, exporters have no incentive to under-declare export values, as exports are not taxed. On the imports side, any revenue lost owing to under-declaration of import values is fully recovered at the later stages of the production-distribution chain, unless the importing activities are directly undertaken by consumers (e.g., cross-border shopping). The situation is exactly reversed, however, with an origin-based VAT.

As imports are not taxed but exports are taxed just like domestic sales, no border tax adjustments are necessary under the origin principle. But since an origin-based VAT taxes the domestic value added of taxable products, and the values of imported products are embedded in the sales of domestic producers, import values must be properly ascertained so that they can be removed from the tax base. The proper valuation of exports is equally important, as any revenue lost resulting from incorrect export declarations is never recovered. For tax purposes, then, the valuation problem under the origin principle is primarily one of safeguarding against over-declarations by importers and under-declaration by exporters.

**A comparison of the two principles**

- **Distribution of the global tax base.** There is clearly an important implication regarding the distribution of the global tax base among trading countries in choosing between the origin and destination principles, as the value added of exports is part of the tax bases of the exporting countries under the former, but is part of the tax bases of the importing countries under the latter. Hence, under the origin principle, a trade surplus (deficit) would expand (reduce) a country's tax base; exactly the reverse would hold under the destination principle. Only if every country's trade account is balanced (though not necessarily with each other) would such distributive consequences be absent.

- **Production and consumption efficiencies.** Apart from the distribution of the global tax base, would the two principles entail different efficiency implications for global production and consumption? Obviously, if all countries have the same VAT rate, the two principles would yield identical outcomes, since the tax rate on any taxable product would be the same regardless of where it is produced or consumed. If countries have

---

29In this sense, the distinction is fully analogous to that between the residence and source principles in income taxation.

30For a demonstration of how under-declared values of products in one stage would be caught in subsequent stages in the production-distribution chain, see the section on credit vs. subtraction method. That consumers could benefit from under-declaring the values of products they directly imported is due entirely to their status as nonregistered VAT payers. Clearly, such benefits could also accrue to others who are specifically exempt by the VAT system.

31An origin-based VAT using the credit-invoice method would require an imputation of tax credits applicable to imports.

32See Berglas (1981); Coomans (1987); Frenkel, Razin, and Sadka (1991); Shibata (1967); and Sinn (1990) for various discussions.

33For simplicity, it is assumed here that the VAT systems in different countries can differ, if at all, only in terms of their VAT rates, with each country having a single rate internally. Complications arising from violating this assumption are noted below.
different VAT rates, however, the two principles would, in general, have different efficiency implications.

Under the destination principle, producer prices of every traded product are equalized across countries through free trade (since a product entering a country is free from the VAT of the exporting country), but with different VAT rates, its consumer prices, and therefore the relative consumer prices between it and nontraded products, would be different in each country, thus resulting in inefficient global consumption. \(^3^4\) Under the origin principle, consumer prices of every traded product are equalized across countries through free trade (since a product entering a country bears the VAT of the exporting country), but with different VAT rates, its underlying producer prices, and therefore the relative producer prices between it and nontraded products, would be different in each country, thus resulting in inefficient global production. Hence, a movement from one tax principle to the other essentially entails a trade-off between global consumption and production inefficiencies, when VAT rates are not uniform across countries.

Note that the relative producer and consumer prices between any pair of traded products within a country is always the same with a single internal VAT rate. This implies that the equalization of relative producer prices among tradable products across countries under the destination principle is the same as equalizing their relative consumer prices. Similarly, under the origin principle, the equalization of relative consumer prices among tradable products across countries is the same as equalizing their relative producer prices. Therefore, if all products are tradable, neither tax principle would produce any global consumption or production inefficiency.

- **Equivalence and nonequivalence under flexible prices.** A presumption in the above efficiency analyses is, of course, that all prices are flexible so that the relevant producer and consumer prices can be equalized through trade. Hence, in the absence of nontraded products, both the origin and destination principles would entail no inefficiency even when VAT rates differ across countries. A related (but different) proposition is that, as long as all prices (and exchange rates) are flexible, a switch from one tax principle to the other would entail no change in the cross-country production patterns. This well-known equivalence result is, however, predicated on two crucial assumptions.

First, it assumes that all tradable products are taxed at the same VAT rate within a country. If this does not hold, a switch in tax principles would produce complex responses in production and consumption patterns. In reality, of course, VAT systems among countries can differ not only in terms of rates, but in other structural aspects as well, thus rendering the outcomes even murkier.

Second, different VAT rates across countries would impose different burdens on factor incomes, through their differential impacts on producer prices. For example, by equalizing the consumer prices of a tradable product across countries, the origin principle, as noted earlier, would give rise to different underlying producer prices. Under the destination principle, differences in consumer prices of a tradable product across countries, generated by differences in their VAT rates, could also be translated into differences in its producer prices through adjustments in exchange rates. In either case, factor incomes would not be the same across countries, which in turn would induce international factor movements and consequently alter countries' production patterns. Hence, unless factors are immobile, the equivalence result would not hold in general.\(^3^5\)

- **International practice.** Almost all countries in the world that presently have a VAT have implemented it on the destination principle.\(^3^6\) The reason is probably twofold. First, the origin principle, which allows imports entering a country to bear the VAT burdens of exporting countries, is likely to lead to undesirable tax competition, with the clear implication that flexible prices and exchange rates are not being viewed as adequate mechanisms for alleviating such behavior among trading countries. The destination principle, in contrast, is regarded as effective in ensuring that traded products contain no VAT elements of the exporting countries. Second, while the destination principle requires border tax adjustments, such adjustments can be carried out with relative ease, and in any case, they seem to be a small price for circumventing the potential valuation problems associated with the origin principle.\(^3^7\)

- **Restricted origin principle.** The restricted origin principle is primarily of interest only to countries that are members of a customs or economic union, within which border controls on trades among union members...
are either absent or ineffective by default, or their removal is a stated objective of the union. Because of the required border tax adjustments, the implementation of the destination principle is generally regarded as infeasible without border controls. If union members implement their VAT systems under the origin principle regarding intra-union trades, but under the destination principle regarding trades with nonunion members, the restricted origin principle is said to apply. This is the principle on which the VAT systems in the republics of the former Soviet Union are presently based, and toward which the VAT systems in the European Community (EC) countries are aiming.\(^{38}\)

Under the restricted origin principle, VAT rate divergence among union members is, of course, very much an issue (just as it is under the general origin principle). This problem can be apparently overcome, however, by a clearinghouse mechanism whereby the importer in a member country is given credit, under the regular credit-invoice method, for the VAT paid on his imports in the exporting member countries. Such foreign credits given would be tallied by every union member against every other union member, and the net claims on each member would be settled by a clearinghouse. In effect, such a mechanism would restore the substantive content of a destination-based VAT, with border tax adjustments being shifted from border controls to the clearinghouse. A proposal along these lines has been outlined by the EC for use with the definitive VAT system after the transitional period expires.

In addition to the need for setting up a new bureaucratic apparatus to administer the clearinghouse, however, it has been noted that this approach has a number of other shortcomings, including the existence of incentives for member countries to under-declare their exports and over-declare their imports, and the necessity for some rate conformity among members is not entirely obviated. Since cross-border purchases by nonregistered VAT payers, that is, consumers, are effectively taxed at the rates prevailing in the origin countries, their volume could be substantial if rate differentials across member countries are significant.\(^{39}\)

In terms of economic efficiency, it can be shown that a VAT implemented under a genuine restricted origin principle (i.e., one without the clearinghouse mechanism) would be identical to that implemented under the general origin principle only if all member countries adopt the same rate, even if all prices and exchange rates are flexible and all trade accounts are bilaterally balanced among members.

**Price Effects of a VAT**\(^{40}\)

**Impact on the rate of inflation versus the price level**

A common concern in many countries contemplating the introduction of a VAT is that it may produce an inflationary impact, even if it is designed to replace one or more existing taxes in a revenue-neutral manner. This concern usually stems from the prospective VAT typically having a much broader base than the taxes it replaces, so that many untaxed items under the existing system will be taxed under the VAT. The validity of this concern can be examined under two alternative interpretations of the term "inflationary impact."

If inflationary impact is taken to mean a sustained increase in the rate of inflation, then the concern would be conceptually misguided. The introduction of a VAT (or any tax, for that matter) can never, by itself, lead to a sustained increase in the rate of change in the price level; such a change in the inflation rate can only be produced by an expansionary monetary policy under all circumstances. If, however, the term is interpreted as a once and for all increase in the price level (or a one-period increase in the inflation rate), then

---

\(^{38}\)The directive on the removal of fiscal frontiers in the EC (Directive 91/660/EEC), adopted by the Council of the European Communities in December 1991 and went into force on January 1, 1993, provides for a period of transitional arrangements whereby the VAT on intra-EC trades would still be based on the destination principle, but their monitoring would no longer be based on physical controls at internal borders. Instead, the monitoring would be carried out on the basis of a new reporting system of statistics on intra-EC trades (Regulation (EEC) No. 3350/91) and new requirements on administrative cooperation among EC countries (Regulation (EEC) No. 218/92). In effect, a version of the so-called deferred payment scheme (see below) would be adopted.

The transition period is envisioned to last until December 31, 1996 (with possible extensions) after which a definitive system of origin-based VAT would be implemented for intra-EC trades (i.e., a VAT based on the restricted origin principle), together with a clearinghouse mechanism. As explained below, however, the adoption of such a mechanism would result in a VAT that resembles a destination-based system in its economic impact, except for those intra-EC trades, e.g., sales of cars and distance (mail-order) sales, explicitly excluded from the mechanism.

\(^{39}\)It should perhaps be noted that there is an alternative method, known as the deferred payment scheme and has been used in the Benelux countries, for implementing a destination-based VAT without border controls. It essentially involves taxing imports, not at the point when they physically enter the country, but at the next stage of their production-distribution chain, which is automatically accomplished through the catching-up property of the credit-invoice method (see section on credit vs. subtraction method for a discussion of this property). Exporters are given tax refunds on the basis not of documents from customs authorities, but of other proper documentary proofs that products have in fact been exported.

As noted earlier, a version of the above deferred payment scheme has now, in effect, been adopted by the EC during the transitional period until a definitive VAT system is implemented.

\(^{40}\)Tait (1988) contains a good discussion on this topic.
whether the VAT is inflationary in this sense would depend on a number of factors.

The conventional conclusion is that, ceteris paribus, a VAT that is introduced in a revenue-neutral manner would have no impact on the aggregate price level, since the aggregate demand in this case is unchanged. Similar reasoning would then suggest that a revenue-enhancing (revenue-losing) VAT would lead to a decline (rise) in the aggregate price level, since it corresponds to a contractionary (expansionary) fiscal policy stance. There are, however, two problematic aspects to these conclusions: conceptually, they neglect a possible supply response to the VAT's introduction; and operationally, they overlook that measuring any change in the price level is often based on a price index, such as the consumer price index (CPI), with historically fixed weights.

Supply response

The conventional conclusion of no aggregate price change following the introduction of a revenue-neutral VAT is predicated on the assumption that the VAT induces no aggregate supply response and, therefore, the budget constraint for the private sector remains unchanged. If, however, the VAT entails a lower degree of distortion in production efficiency than the taxes it replaces—which is most likely to be the case as this is commonly the primary reason for introducing the VAT in the first place—then there is a strong presumption that the tax switch would elicit a supply response that would exert a downward pressure on the price level even under the condition of initial revenue neutrality.

Historically fixed weights in a price index

The replacement of an existing tax with a VAT will almost always change the structure of relative prices in the economy, even if the revenue yield remains the same after the tax switch. For the no price effect to hold, the expenditure patterns of the private sector must shift immediately and in such a way within the same overall budget constraint so as to produce an unchanged aggregate price level. Operationally, however, the interest of the policymaker is typically focused on measuring the price effect based on a widely understood price index, such as the CPI. Since an index of this type invariably has historically fixed weights, price calculations based on it cannot capture the equilibrium adjustments in expenditure patterns necessary to produce the constant price outcome. Hence, as long as a fixed-weight price index has to be used to calculate aggregate price effects, the direction in which the index would change after the tax switch is a priori uncertain, depending as it must be on the pattern of the change in relative prices.

Simulating the change in the CPI: a numerical example

To illustrate the nature of calculating the price effect of a tax switch based on the CPI by means of a concrete numerical example, consider the case of replacing an existing general sales tax (GST) of 11 percent with a VAT of 13 percent. Suppose that 52 percent of the consumption basket in the CPI is exempt from the GST, but the scope of exemptions would go down to 31 percent of the basket under the prospective VAT. Under either the GST or the VAT, the consumer prices of tax-exempt items would contain some tax elements due to cascading. For the taxed items, however, cascading would be present under the GST but largely eliminated under the VAT. Hence, the outcome of the price calculations would critically depend on the degree of cascading present in the economy.

It will be recalled from the section on tax cascading: concept and measurement that, irrespective of the number of stages in the production-distribution chain and the producer's pricing behavior, the degree of cascading, 8, can be embodied in the following synthetic rule for simulation purposes:

\[ p_c = B \cdot (1 + 8 \cdot \tau) \cdot (1 + \tau), \]

where \( \tau \) is the relevant tax rate, \( p_c \) denotes the consumer price inclusive of the tax, and \( B \) is a positive constant. The impact of a tax at the rate \( \tau \) on the consumer price can then be simulated with alternative values of \( 8 \). Table III.4 provides illustrative calculations, based on such a synthetic rule with \( B \) normalized to unity, of the impacts on the CPI of the GST and the VAT in the above example.

Column 8 of Table III.4 shows that the GST element in the CPI ranges from 8.2 percent with \( 8 = 0.25 \) to 16.9 percent with \( 8 = 1 \), while column 15 of Table III.4 shows that the VAT will increase the CPI from 10 percent to 13 percent over the same range of \( 8 \) values. An essentially neutral price impact would be achieved if \( 8 = 0.5 \), where the removal of the GST is shown to reduce the CPI by 11.1 percent and the VAT's introduction to increase the CPI by 11 percent. Of course, the precise price impact of replacing the GST with the VAT can be determined only after the true value for \( 8 \) is ascertained. A method for estimating \( 8 \) based on available revenue data is provided in the section on tax cascading entitled "concept and measurement."

---

*In reality, some cascading would be present if the VAT credit chain associated with any taxed item is broken at any stage along its production-distribution chain. See section on credit vs. subtraction method for a discussion.
Table III.4. Simulated Impact of Replacing a GST with a VAT on the CPI
(In percent)

<table>
<thead>
<tr>
<th>Simulated CPI Effect of GST at 11 Percent</th>
<th>Tax-exempt Items (52% of CPI)</th>
<th>Weighted average of tax effects on overall CPI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Taxed items (48% of CPI)</td>
<td>(8)=(5)·(7)x69%+(6)x31%</td>
</tr>
<tr>
<td>Degree of cascading (1)</td>
<td>Cascaded tax rate</td>
<td>Compounded effect on CPI from tax (4)=(1)x11%</td>
</tr>
<tr>
<td>100</td>
<td>11.0</td>
<td>11.0</td>
</tr>
<tr>
<td>75</td>
<td>8.3</td>
<td>23.2</td>
</tr>
<tr>
<td>50</td>
<td>5.5</td>
<td>20.2</td>
</tr>
<tr>
<td>25</td>
<td>2.8</td>
<td>14.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Simulated CPI Effect of VAT at 13 Percent</th>
<th>Tax-exempt Items (31% of CPI)</th>
<th>Overall impact on CPI of replacing GST by VAT (16)=(15)-(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Taxed items (69% of CPI)</td>
<td></td>
</tr>
<tr>
<td>Cascaded tax rate (9)</td>
<td>Final-stage tax rate (10)</td>
<td>Compounded effect on CPI from tax (14)=(12)</td>
</tr>
<tr>
<td>0.0</td>
<td>13.0</td>
<td>13.0</td>
</tr>
<tr>
<td>0.0</td>
<td>13.0</td>
<td>9.8</td>
</tr>
<tr>
<td>0.0</td>
<td>13.0</td>
<td>6.5</td>
</tr>
<tr>
<td>0.0</td>
<td>13.0</td>
<td>3.3</td>
</tr>
</tbody>
</table>

Source: Staff calculations.

The Credit Versus Subtraction Method

- What are the mechanics of computation under the credit method and what economic implications does it entail?
- What are the mechanics of computation under the subtraction method and what economic implications does it entail?
- What are the comparative merits and limitations of the two methods?

This section compares, by means of a simple numerical example, the merits and shortcomings of two alternative methods of computing the value added of a taxable transaction under a VAT: the credit (also known as the invoice) method and the subtraction method. The two methods give rise to different economic consequences whenever a VAT contains multiple rates, zero-ratings, and exemptions. The numerical example assumes a three-stage production-distribution process in which a manufacturer sells to a wholesaler who in turn sells to a retailer. The value added at each stage is assumed to be 100 and the VAT rate is assumed to be 10 percent on a tax-exclusive basis (or 9.1 percent on a tax-inclusive basis) at all stages. Implications from using the two methods when the VAT has multiple rates are also noted.

The Credit Method

The various aspects of the credit method are illustrated in the four panels of Table III.5.

The mechanics of computation

The credit method requires that the amount of VAT charged be explicitly stated on the invoice associated with any taxable transaction. The amount of tax a merchant submits to tax authorities is simply the difference between the tax he collected on his sales and the tax he paid on his purchases. Panel A of Table III.5 shows how a VAT of 10 is collected at each of the three stages using this method. The total VAT collected is thus 30—precisely 10 percent of the total value added of 300. Note that it is the consumer who bears the ultimate burden of the VAT, since the retailer’s sales cum VAT is 330, the merchant at each stage only acts as a collection agent.
### Table III.5. The Credit Method

<table>
<thead>
<tr>
<th></th>
<th>Manufacturer</th>
<th>Wholesaler</th>
<th>Retailer</th>
<th>Total VAT to the Final Consumer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. All businesses subject to tax</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Sales (excluding VAT)</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>(2) Purchases (excluding VAT)</td>
<td>0</td>
<td>100</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>(3) VAT on sales [10 percent of (1)]</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>(4) Credit on purchases [10 percent of (2)]</td>
<td>0</td>
<td>10</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>(5) Net VAT payments [(3)-(4)]</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td><strong>B. Retailer is zero-rated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Sales (excluding VAT)</td>
<td>100</td>
<td>200</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>(2) Purchases (excluding VAT)</td>
<td>0</td>
<td>100</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>(3) VAT on sales [10 percent of (1)]</td>
<td>10</td>
<td>20</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(4) Credit on purchases [10 percent of (2)]</td>
<td>0</td>
<td>10</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>(5) Net VAT payments [(3)-(4)]</td>
<td>10</td>
<td>10</td>
<td>-20</td>
<td>0</td>
</tr>
<tr>
<td><strong>C. Retailer is exempt</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Sales (excluding VAT)</td>
<td>100</td>
<td>200</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>(2) Purchases (excluding VAT)</td>
<td>0</td>
<td>100</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>(3) VAT on sales [10 percent of (1)]</td>
<td>10</td>
<td>20</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>(4) Credit on purchases [10 percent of (2)]</td>
<td>0</td>
<td>10</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>(5) Net VAT payments [(3)-(4)]</td>
<td>10</td>
<td>10</td>
<td>na</td>
<td>20</td>
</tr>
<tr>
<td><strong>D. Wholesaler is exempt</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Sales (excluding VAT)</td>
<td>100</td>
<td>210</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>(2) Purchases (excluding VAT)</td>
<td>0</td>
<td>100</td>
<td>210</td>
<td></td>
</tr>
<tr>
<td>(3) VAT on sales [10 percent of (1)]</td>
<td>10</td>
<td>na</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>(4) Credit on purchases [10 percent of (2)]</td>
<td>0</td>
<td>na</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>(5) Net VAT payments [(3)-(4)]</td>
<td>10</td>
<td>na</td>
<td>31</td>
<td>41</td>
</tr>
</tbody>
</table>

Source: Staff calculations.

Note: na means not applicable.

1 Sales cum VAT to the final consumer is the sum of (1) and (3) in each panel under the retailer's column.

2 For the retailer, 0 percent of (1).

---

### Zero-rating the retailer

Under a VAT system, a zero-rated merchant charges no VAT on his sales but can claim a refund on the VAT he paid on his purchases. Panel B of Table III.5 illustrates the case of a zero-rated retailer, who is able to claim a refund of 20. This refund turns out to be exactly equal to the sum of the VAT collected from the manufacturer and the wholesaler; the total VAT collected is thus zero. As no VAT is charged on the retailer's sales of 300, there is no tax burden borne by the consumer.

There are two important implications from the above result. First, zero-rating a merchant effectively removes the tax elements in all the sales of taxed merchants situated before the zero-rated merchant along the production-distribution chain. In the above example, the tax elements in the sales of the manufacturer and wholesaler—merchants situated before the retailer—are removed when the retailer is zero-rated. Second, a VAT implemented with the credit method effectively taxes a product at the rate that is applied at the final stage of the product’s sale, even if different rates were applied at earlier stages. Again referring to the numerical example above, the illustrated sales bear no tax because the retailer—the seller at the final stage—happens to apply (by assumption) a tax rate of zero. Alternatively, had the retailer been subject to a VAT rate of, say, 5 percent (while the rate of 10 percent continues to apply to the manufacturer and wholesaler), his refund would have been 15, and the total VAT collected would have been 5; the retailer's sales cum VAT in this case would have been 305.

### Exempting the retailer

Like a zero-rated merchant, an exempted merchant does not charge a VAT on his sales; unlike a zero-rated merchant, however, the exempted merchant is out of the VAT net and therefore cannot claim a refund on the VAT he paid on his purchases. Panel C of Table III.5 illustrates the case of an exempted retailer, who pays a VAT of 20 on his purchases but obtains no refund. The total VAT collected in this case is 20—10 from the manufacturer and 10 from the wholesaler. Compared with the case where all businesses are taxed (Panel A of Table III.5), exempting the retailer lowers the VAT collection—and therefore the tax element in the sales to the consumer—by 10, which is precisely equal to 10 percent of the retailer’s value added.
Breaking the credit chain

In contrast to exempting a merchant situated at the final stage of the production-distribution chain, which, as illustrated in Panel C of Table III.5, has the effect of removing his value added from the VAT base, exempting a merchant at any stage prior to the final stage has a drastically different effect under a VAT using the credit method. Panel D of Table III.5 illustrates the case where the wholesaler—instead of the retailer—is exempt from tax.

The wholesaler, being exempted from the VAT, cannot claim a tax credit of 10 for the VAT he paid on his purchases from the manufacturer. In response, he compensates by charging a correspondingly higher price on his sales—210 now as opposed to 200 in the case where he is taxed (Panel A of Table III.5). Moreover, since the wholesaler does not charge a VAT on his sales to the retailer, the latter also has no tax credit to claim. Hence, the retailer simply collects a VAT of 31 from the consumer, which is 10 percent of his sales of 310 (210 plus his value added of 100). The total VAT collected is now 41—31 from the retailer and 10 from the manufacturer. Compared with Panel A of Table III.5, exempting the wholesaler thus causes the VAT revenue to increase by 11, or the VAT base to increase by 110. This happens because, by breaking the credit chain, the exemption not only allows the value added of the exempted merchant to be recovered at a later stage (the retail stage in the example) where it is taxed, but it also causes the value added of the merchant at the earlier stage (the manufacturing stage in the example) to be taxed again at a later stage. The resulting double taxation of the value added of all merchants situated before the exempted merchant along the production-distribution chain is exactly the same as a cascading turnover tax would have caused. In the example, the value added of the manufacturer is taxed twice, thus accounting for 10 of the additional VAT collected. The remaining 1 is simply the result of the tax on tax.

The Subtraction Method

The various aspects of the subtraction method are illustrated in the four panels of Table III.6, which uses the same numerical example as in Table III.5.

The mechanics of computation

Under the subtraction method, each merchant's tax liability is computed by applying the applicable VAT rate to the difference between his total sales (inclusive of the VAT element in his sales price) and his total purchases (inclusive of the VAT element in his purchase price). Hence, unlike the credit method, the amount of VAT connected with a taxable transaction is not required to be explicitly stated on the associated invoice. Panel A of Table III.6 shows how a VAT with the same rate structure as the one considered earlier under the credit method is collected under the subtraction method.

The first row in each panel of Table III.6, labeled "sales (excluding VAT)," refers to the amount of sales at each stage along the production-distribution chain excluding the VAT charged at that stage, but it includes any VAT charged at earlier stages. For example, the wholesaler in Panel A of Table III.6, his sales excluding VAT (210) can be determined by adding his value added (100) to his purchases from the manufacturer (110), the latter being an amount inclusive of the VAT charged by the manufacturer. The wholesaler's sales including the VAT (220) are obtained simply by adding the VAT on his value added (10 percent of 100, or 10) to his sales excluding the VAT (210).

The above steps serve only the purpose of illustrating the mechanics of how a merchant would arrive at the proper pricing decision inclusive of the VAT. For purposes of computing the merchant's VAT liability, only the difference between his sales and his purchases (both on a gross-of-tax basis) matters. For the wholesaler in the example, his VAT base (110) is calculated by subtracting his purchases (110) from his sales (220), both inclusive of all VAT elements. His VAT liability is then determined by applying the VAT rate of 9.1 percent (which is equivalent to 10 percent on a tax-exclusive basis) to his base of 110, resulting in a VAT payment of 10.

As illustrated in Panel A of Table III.6, where all businesses are subject to tax, the subtraction method yields the same outcome as the credit method: a VAT of 10 is collected at each stage along the production-distribution chain, but it is the consumer who bears the entire tax burden.

Zero-rating the retailer

Just like the credit method, all tax elements are removed under the subtraction method if the merchant at the end of the production-distribution chain is zero-rated. As illustrated in Panel B of Table III.6, zero-rating the retailer effectively sets his sales to zero for purposes of calculating his tax base, thus resulting in a negative base of 220 and a tax refund of 20, which is exactly equal to the total VAT element on his purchases. Since this refund is anticipated, the retailer's sales to the final consumer will (or should) not contain any VAT element at all.
### Table III.6. The Subtraction Method

<table>
<thead>
<tr>
<th>Sales cum VAT to the Final Consumer</th>
<th>Manufacturer</th>
<th>Wholesaler</th>
<th>Retailer</th>
<th>Total VAT</th>
</tr>
</thead>
</table>

#### A. All businesses subject to tax

1. Sales (excluding VAT) 100 210 320
2. Purchases (including VAT) 0 110 220
3. Sales (including VAT) [110 percent of ((1)-(2))+(2)] 110 220 330
4. VAT base [(3)-(2)] 110 110 110
5. Net VAT payments [9.1 percent of (4)] 10 10 10 30 330

#### B. Retailer is zero-rated

1. Sales (excluding VAT) 100 210 300
2. Purchases (including VAT) 0 110 220
3. Sales (including VAT) [110 percent of ((1)-(2))+(2)] 110 220 300
4. VAT base [(3)-(2)] 110 110 110 30
5. Net VAT payments [9.1 percent of (4)] 10 10 10 30 300

#### C. Retailer is exempt

1. Sales (excluding VAT) 100 210 320
2. Purchases (including VAT) 0 110 220
3. Sales (including VAT) [110 percent of ((1)-(2))+(2)] 110 220 320
4. VAT base [(3)-(2)] 110 110 na
5. Net VAT payments [9.1 percent of (4)] 10 10 0 20 320

#### D. Wholesaler is exempt

1. Sales (excluding VAT) 100 210 310
2. Purchases (including VAT) 0 110 210
3. Sales (including VAT) [110 percent of ((1)-(2))+(2)] 110 210 310
4. VAT base [(3)-(2)] 110 na 110
5. Net VAT payments [9.1 percent of (4)] 10 10 0 20 320

#### E. Both wholesaler and retailer zero-rated

1. Sales (excluding VAT) 100 200 282
2. Purchases (including VAT) 0 110 200
3. Sales (including VAT) [110 percent of ((1)-(2))+(2)] 110 200 282
4. VAT base [(3)-(2)] 110 na 110 20
5. Net VAT payments [9.1 percent of (4)] 10 10 0 20 282

Source: Staff calculations.

Note: na means not applicable.

*Sales cum VAT to the final consumer is (3) in each panel under the retailer's column.
*For the retailer (1).
1 For the wholesaler (1).
2 For both the wholesaler and retailer (1).
3 For both the wholesaler and retailer, their sales are set to zero in their VAT base calculations.

---

**Exempting the retailer**

When the retailer is out of the VAT system, he does not have to compute his VAT base (since he cannot claim any refund) and does not charge any VAT on his sales to the consumer (Panel C, Table III.6). His sales (320) are simply the sum of his purchases inclusive of the VAT (220) and his value added (100). The total tax burden on the consumer is now 20—10 less than the case when all businesses are taxed (Panel A, Table III.6). This reduction in VAT collection corresponds exactly to the shrinkage of the VAT base as a result of the removal of the retailer's value added. As can be seen from Panel C, Table III.6, the subtraction method also yields the same result as the credit method in this case.

**Exempting the wholesaler**

In stark contrast to the credit method, under which breaking the credit chain by exempting the wholesaler gives rise to cascading, no cascading results from exempting the wholesaler under the subtraction method (Panel D, Table III.6). This is so because the subtraction method computes the VAT base at each stage as the difference between the sales and purchases at that stage.
stage, and therefore, no value added of earlier stages can ever cascade forward and be taxed again. In Panel D, Table III.6, the wholesaler, being tax exempt, charges no VAT on its sales (210)—the sum of its purchases (110) and its value added (100)—to the retailer. Since the entire amount of the retailer's purchases is deducted from the retailer's own sales in determining his VAT base, that base cannot include the value added of the manufacturer and the wholesaler. Hence, exempting a merchant from tax under the subtraction method always removes the exempted merchant's value added from the tax base, regardless of where he is situated along the production-distribution chain (compare Panels B and C, Table III.6). Unlike the credit method, this loss of the tax base is never recovered down the chain, nor will any value added up the chain be taxed again down the chain.

Zero-rating both the wholesaler and the retailer

Under the credit method, the zero-rating of a merchant does not constitute a break in the credit chain, and therefore cannot cause any unintended effects, regardless of where and how often it occurs along the production-distribution chain. In contrast, under the subtraction method, if the zero-rating occurs at two or more stages, the result is that the VAT base will be reduced by more than what the zero-rating is intended to achieve. Panel E of Table III.6 illustrates the case where both the wholesaler and the retailer are zero-rated. The zero-rating of the wholesaler produces the desired effect: he gets a VAT refund of 10 to compensate for the VAT he paid on his purchases from the manufacturer. Moreover, he does not charge any VAT on his sales to the retailer. The retailer, however, being zero-rated also, likewise obtains a refund (18, or 9.1 percent on his purchases of 200 from the wholesaler). In the example, the benefit of the retailer's refund is assumed to be passed on to the consumer. Thus, instead of being charged for a total of 300 on his purchases (the intended amount of the zero-rating), the consumer ends up paying only 282. The VAT has become a value-added subsidy.

This peculiar outcome underscores a shortcoming of the subtraction method, which contains a built-in assumption that a VAT has been paid on all purchases made by a merchant as long as he is in the VAT system (as opposed to one who is tax-exempt), even if the VAT element had already been removed because of the zero-rating(s) somewhere earlier along the production-distribution chain. By similar reasoning, the subtraction method can be seen to be unsuitable for any VAT system that has multiple rates, for it will assume that the purchases of any merchant had been all taxed at the same rate that applies to the merchant in question, even if such purchases had been taxed at different rates earlier.43

Summary of Comparison

If a VAT has a single rate and has no exemptions and zero-ratings, the credit and subtraction methods will produce identical outcomes, and the choice between the two must be made on grounds other than tax policy.44 When the VAT contains multiple rates and extensive exemptions, however, neither method is capable of completely overcoming the undesirable effects of these imperfections in the VAT system itself, and the choice between the two comes down to a comparative trade-off of costs and benefits in each.

With the credit method, the important implication is that exempting any merchant situated anywhere before the final stage along the production-distribution chain breaks the credit chain and results in cascading, which reduces one of the fundamental benefits of having a VAT in the first place. With the subtraction method, no such cascading can result. It is, however, incapable of handling multiple rates. Since a perfect VAT system does (and will) not exist in practice, the preferred method clearly depends on the nature and severity of the imperfections. For example, on the one hand, the use of the subtraction method is ruled out in a VAT with multiple rates.45 On the other hand, for a single-rate VAT with extensive exemptions, the subtraction method should clearly be preferred.

Estimating the VAT Base

- **What is the base of a VAT, and how can it be estimated, using a step-by-step illustration, from various data sources?**

The applicable base of a VAT depends on a number of aspects related to its design, for example, whether it is origin- or destination-based, of the income- or cons-

---

43For a VAT system that has only a zero rate in addition to one positive rate, the subtraction method can still be used, but it requires that all merchants must separate their purchases which are not taxed from those which are taxed. This requirement in effect turns the subtraction method into the credit method in substance.
44 For example, the credit method leaves behind an "audit trail," in the sense that the invoice of any taxable transaction can easily be used to check the tax paid by the buyer against the tax collected by the seller. Hence, the credit method could be preferred on this administrative ground.
45This is probably the most important reason for the prevalence of the credit method in the VAT systems around the world. This limitation of the subtraction method, however, could be exploited to the advantage of a country's tax design, such as that in Japan, by ensuring that its adoption would necessarily entail a single-rate VAT.
Table III.7. A Framework for Estimating the Base of a VAT

<table>
<thead>
<tr>
<th>Starting point of estimation:</th>
<th>Remarks</th>
<th>Likely Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP (market prices)</td>
<td>Sum of value added of domestic production</td>
<td>NIA</td>
</tr>
<tr>
<td>Adjustment A: trade balance</td>
<td>For destination-based VAT</td>
<td>NIA</td>
</tr>
<tr>
<td>1. Minus exports</td>
<td></td>
<td>NIA</td>
</tr>
<tr>
<td>2. Plus imports</td>
<td></td>
<td>NIA</td>
</tr>
<tr>
<td>Adjustment B: capital formation</td>
<td>For consumption-type VAT</td>
<td>NIA</td>
</tr>
<tr>
<td>1. Minus gross domestic capital formation</td>
<td></td>
<td>NIA</td>
</tr>
<tr>
<td>2. Add residential buildings</td>
<td></td>
<td>NIA</td>
</tr>
<tr>
<td>3. Add capital formation in exempt sectors</td>
<td></td>
<td>NIA</td>
</tr>
<tr>
<td>Adjustment C: exempt sectors</td>
<td>E.g., nonexportable agriculture, financial services, owner-occupied dwellings, wholesale and retail</td>
<td>NIA</td>
</tr>
<tr>
<td>1. Minus value added of exempt sectors (factor costs)</td>
<td></td>
<td>NIA, GOT</td>
</tr>
<tr>
<td>2. Minus indirect taxes in exempt sectors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjustment D: cascading</td>
<td>Caused by credit-invoice method</td>
<td>I/O, GOT</td>
</tr>
<tr>
<td>1. Add purchases of output from exempt sectors by taxed sectors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Add taxed inputs in exportable agriculture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjustment E: government expenditure</td>
<td>Nontaxable expenditure component in NIA</td>
<td>NIA, GOT</td>
</tr>
<tr>
<td>Minus expenditure on wages and salaries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjustment F: final private expenditure</td>
<td>E.g., rents, education, and health services</td>
<td>NIA, CPI</td>
</tr>
<tr>
<td>1. Minus exempt expenditures</td>
<td>Due to cascading</td>
<td>I/O</td>
</tr>
<tr>
<td>2. Add taxed inputs in exempt expenditures</td>
<td></td>
<td>NIA</td>
</tr>
<tr>
<td>3. Add foreign expenditures in local markets</td>
<td></td>
<td>NIA</td>
</tr>
<tr>
<td>4. Minus expenditures abroad by residents</td>
<td></td>
<td>NIA</td>
</tr>
<tr>
<td>Adjustment G: exemption threshold</td>
<td>For reducing administrative costs</td>
<td></td>
</tr>
<tr>
<td>1. Minus sales of firms below threshold</td>
<td></td>
<td>GOT</td>
</tr>
<tr>
<td>2. Add taxed inputs in above sales</td>
<td></td>
<td>GOT</td>
</tr>
<tr>
<td>Adjustment H: tax replacement</td>
<td>Minus sales taxes to be replaced by the VAT</td>
<td>GOT</td>
</tr>
<tr>
<td>Minus estimated extent of leakage</td>
<td></td>
<td>GOT</td>
</tr>
<tr>
<td>Remarks</td>
<td>Sum of value added of domestic production</td>
<td>NIA</td>
</tr>
</tbody>
</table>

1The estimation framework assumes a destination-based, consumption-type VAT implemented with a credit-invoice method.
2NIA = national income accounts; I/O = input/output tables; CPI = consumer price index basket; and GOT = various government and tax departments.

Consumption-type, implemented with a credit-invoice or subtraction method, and contains many or few exemptions. This section develops a general framework for estimating the base of the most widely adopted form of the VAT—a destination-based, consumption-type system implemented with a credit-invoice method. Table III.7 describes the framework with a step-by-step illustration of the various computations involved in the estimation process. Remarks about the computations and the likely sources where required data can be found at each step are also provided.

GDP as Starting Point of Estimation

A logical starting point for estimating the VAT base is the gross domestic product (GDP) of an economy, since it represents the sum total of the value added in the domestic production of goods and services. From this, various adjustments can then be made to account for the different features that have been incorporated into the VAT’s design. For a destination-based, consumption-type VAT, however, a legitimate question can be raised as to whether final consumption expenditure, which represents the sum total of value added of domestic consumption, is not a more direct starting point in estimating the VAT base. The answer, as it turns out, depends to a large extent on the scope and nature of exemptions of the VAT under consideration.

To see the issue related to the choice of the starting point of estimation more clearly, consider the following familiar national income accounting identity at the most aggregative level:
\[ \text{GDP} = C + I + G_c + G_w + (X - M), \] (38)

where \( C \) is final private consumption expenditure, \( I \) is investment expenditure, \( G_c \) is final government expenditure on goods and services (net of wages and salaries), \( G_w \) is government expenditure on wages and salaries, and \( X \) and \( M \) are, respectively, the values of exported and imported goods and nonfactor services (the term \( (X - M) \) represents, therefore, the trade balance). Since a destination-based, consumption-type VAT is basically a tax on final consumption expenditure on goods and services within the domestic economy, its base must correspond to the sum of final private consumption expenditure \( C \) and government expenditure on goods and services (net of wages and salaries) \( G_c \). But a rearrangement of equation (38) yields

\[ C + G_c = \text{GDP} - (X - M) - I - G_w, \] (39)

which indicates that to use GDP as the starting point in estimating the VAT base indeed entails a roundabout way of arriving at the sum of \( C \) and \( G_c \) as data on the latter consumption categories are usually directly available in national income accounts.

There is, however, a technical difficulty in bypassing data on the production side of the economy: it is seldom feasible to implement a VAT without exemptions, and if such exemptions do not have a one-to-one correspondence with final consumption items (as in sector-specific exemptions, such as those granted to the wholesale and retail sectors), adjustments to the base estimation would then have to be made using sectoral value-added data, that is, data from which GDP from the production side is computed in the first place. Furthermore, in many developing countries, consumption data are often less reliably compiled than data on production. Indeed, it is not uncommon to find that the former are simply derived on the basis of the accounting relationship depicted in equation (39). Hence, under most circumstances, there is no avoiding the use of GDP data on the production side in estimating the VAT base.

**Adjustments to GDP**

This section describes the necessary adjustments to the GDP figure—the starting point of estimation—to arrive at the base of a destination-based, consumption-type VAT implemented with a credit-invoice method. The order of the adjustments follows that given on the right-hand side of equation (39).

**Adjustment A: trade balance**

Since a destination-based VAT taxes consumption at the point where consumption occurs, it covers imports but excludes exports. Hence, the trade balance, that is, exports less imports \( (X - M) \), is a reduction of the VAT base and must, therefore, be subtracted from GDP.

**Adjustment B: capital formation**

For the consumption-type VAT, gross capital formation, representing the sum total of purchases of capital goods (i.e., investment expenditure \( I \)) in the economy, is not taxed and should, therefore, be removed from the VAT base. That part of capital formation which consists of new residential buildings, however, will be in the base (since consumers cannot claim tax credits); it should, therefore, be added back. Capital goods purchased by exempt sectors (see below) are also an addition to the base, as tax credits would not be available to these sectors.

**Adjustment C: exempt sectors**

The VAT base is reduced by the sum of the value added of exempt sectors. Certain sectors of an economy, such as the financial and insurance sector and the sector representing the imputed value of owner-occupied dwellings, are either difficult or infeasible to tax and are almost always exempted in a VAT system. In many developing countries, it is often considered administratively infeasible to subject the wholesale and retail sectors to the VAT, in which case the value added of these sectors must also be removed from the VAT base.

The treatment of the agricultural sector under a VAT is also problematic. In most developing countries, for equity as well as administrative reasons, farmers are generally exempted from indirect taxation. It is, therefore, realistic to assume that this sector would be exempt also under the VAT. In deducting the value added of the agricultural sector, however, a distinction must be made between the value added of exportable and that of nonexportable agricultural output. The loss related to the former has already been accounted for in the treatment of exports. Thus, only the loss of the latter needs to be subtracted in arriving at the VAT base.

Since sectoral value-added data in national income accounts are usually given on the basis of factor costs,

---

48Formally, it zero-rates exports. See the section on credit vs. subtraction method for a discussion of the zero-rating concept in a VAT.

49More formally, if the VAT is first taxed but receives a full tax refund afterwards in the form of tax credits under the credit-invoice method.

50In many EC countries, for example, farmers are exempted from the VAT, but are provided various forms of compensation to alleviate their burden arising from their taxed inputs. Some countries, however, such as New Zealand, Sweden, and the United Kingdom, treat farmers just like other taxed producers. It must be noted that, even if the agricultural sector is taxed, small farmers could be exempted by the exemption threshold (see below).

51If a portion of the exportable agricultural output is consumed domestically in an unprocessed form, it should also be subtracted.
indirect taxes in the exempt sectors must also be deducted to convert the adjustment to a market-price basis.

**Adjustment D: cascading**

Whenever an exempted business sells its products to any taxed business under a VAT using the credit-invoice method, cascading would result in much the same way a turnover tax does, and consequently would cause the VAT base to be higher than that had the exemption not been granted. Hence, the sum of the total value (not simply the value added) of that part of the output of the exempt sectors that is sold to taxed sectors as inputs must be added to the VAT base. An interesting implication of this adjustment is that it is theoretically possible (although unlikely in practice) to end up with a larger VAT base when exempt sectors are present than that when they are not. This result, however, cannot be interpreted as an indication that exemptions are therefore desirable, since the increase in the tax base is entirely due to cascading—the elimination of which is usually one of the primary reasons for considering the introduction of a VAT.

Another necessary adjustment related to cascading involves taxed agricultural inputs. If farmers are exempt, they cannot claim credits for the VAT they paid on their purchases. This tax element cannot be removed even with the zero-rating of exports (except for farmers who are direct exporters). Hence, the value of the taxed inputs used in the production of exportable agricultural output is an addition to the VAT base. This adjustment does not apply, however, to those inputs used in the production of nonexportable agricultural output, since the preceding adjustment—the adding back of the output of exempt sectors sold to taxed sectors—has already accounted for them.

**Adjustment E: government expenditure**

While all government purchases of goods and services should be taxed, that portion of its expenditure which represents wages and salaries \((G_w)\) is not taxable, and hence must be deducted from the VAT base.

**Adjustment F: final private expenditure**

The completion of Adjustments A through E above yields the sum of final private and (nonwage) government consumption expenditures, that is, \(C + G_c\). From this, exempted final private consumption expenditures must be subtracted from the base, such as those on rents, and education and health services. Since, for reasons identical to those discussed in Adjustment D above relating to cascading, these expenditures would contain tax elements as long as inputs used in their production are taxed at earlier stages anywhere along the production-distribution chain, the value of such inputs should, therefore, be added to the VAT base. Finally, an adjustment must be made for consumption by foreigners in local markets (an increase to the base) and purchases by domestic residents abroad (a decrease to the base).

**Adjustment G: exemption threshold**

Small businesses are usually large in number but collectively yield little revenue for the tax system. A VAT that includes all such businesses would, therefore, unduly impair its administrative efficiency. Methods that have been used by countries around the world to deal with the small business problem are varied. By far the simplest and most frequent one involves setting a turnover threshold below which businesses are exempt. The total sales of the businesses below the threshold are, therefore, not part of the VAT base and should be deducted from it. Taxed inputs used in these sales, however, must be added back owing to the cascading discussed above.

**Adjustment H: tax replacement**

The estimate of the VAT base arrived at after completion of Adjustment G would be on the basis of market prices, that is, inclusive of all existing indirect taxes in the base. If, as is frequently the case, the VAT under consideration is intended to replace one or more such taxes, then the total revenue from the taxes to be replaced must be subtracted from the above base estimate.

**Adjustment I: collection leakage**

This adjustment is necessitated by the recognition that the VAT base, computed according to the above procedure, is a potential figure, which may not be the same as the recoverable VAT base. The gap between the two could be due to a number of possible collection leakages, most notably that of taxpayer noncompliance. The extent of these leakages must, however, be judged on a country-by-country basis.

---

52 This consequence would not arise if the VAT is implemented under the subtraction method. See the section on credit vs. subtraction method for a numerical demonstration.

53 This is demonstrated in the section on credit vs. subtraction method.

54 The proper choice of the threshold, which involves the optimal balancing of increased administrative costs against additional revenue losses, requires information concerning the frequency distribution of businesses by sales brackets.
Excise Taxes

WILLIAM J. McCARTEN AND JANET STOTSKY

- What is the role of excise taxes in a tax system?
- What goods do excise taxes typically apply to and why?
- What are the major design issues in an excise tax?
- How do excises correct for negative externalities?

Excise taxes are taxes imposed on the consumption of selected goods, such as alcoholic beverages, tobacco products, and petroleum products. Goods selected for excise coverage typically exhibit one or more of the following characteristics: first, their production and sales are closely supervised by the government, that is, they are sumptuary goods or services; second, they are characterized by price-inelastic demand schedules; third, they have an income elasticity of demand greater than unity, that is, they are luxury goods or services; or fourth, their consumption is regarded by the government as lacking merit or as likely to cause negative externalities. Excise tax rates may be defined in either specific or ad valorem terms and they are usually much higher than the rates applied under the general sales tax or VAT. Because of their relatively narrow focus, excise taxes can have a powerful impact on consumer decision making and resource allocation and are therefore potentially very effective instruments for the attainment of policy goals beyond revenue generation.

The Rationale for Excise Taxation

Revenue generation

Why should a selected set of commodities and services be subject to taxation in excess of a general sales tax or VAT? The most compelling reason for the use of excise taxes is that they can potentially raise a great deal of revenue with little distorting effect, generating little excess burden. They also require relatively little administrative effort and provide limited opportunities for tax evasion. Large sales volumes, few producers, inelastic demand, easy definability, and a lack of close substitutes favor the use of excises where revenue generation is the goal. If excises are applied only to goods and services that exhibit inelastic own-price demand within the observed range of price changes, the percentage change in consumption will be smaller than the percentage change in price. If the compensated own-price elasticities for excisable commodities are also low, the excess burdens arising from the imposition of an excise will be less than for equal revenue-yielding taxes imposed on commodities or factors with more elastic demands. The theory of optimal taxation postulates that when supply is infinitely elastic, the indirect tax structure which generates the smallest loss in efficiency is one where the tax rates for different commodities are set at differential rates so that the tax rate expressed as a proportion of the after-tax price is inversely proportional to the compensated demand elasticity.

The case for relying on excises for a significant proportion of consumption tax revenues is bolstered in many developing countries by the administrative weaknesses and opportunities for evasion which often characterize other consumption taxes, such as the general sales tax, the VAT, and even customs duties.

Correcting for negative externalities

Two commodity groups are perceived as appropriate targets for excise taxation on the basis of the negative externalities associated with their use. These groups are sumptuary or regulated goods, such as tobacco and alcohol products, and inputs into motor vehicle use, including both fuel and the vehicles themselves. Excises on such commodities are intended to internalize negative externalities generated by the consumer. High sumptuary taxes are often justified in terms of the social and medical costs of alcoholism and the medical hazards of tobacco use. Even if medical treatment for those with lung cancer is nonexistent or borne entirely by patients, cancer caused by secondary smoke is clearly a negative externality. High rates of

---

55 In developing countries, levies on tobacco products, alcoholic beverages, and petroleum production may typically account for over two thirds of excise tax receipts. When fuel taxes, motor vehicle taxes, tire taxes, and other road use charges are considered in total, motor vehicle use emerges as the most important source of revenue, while tobacco products and alcoholic beverages vie for second place. Other commonly found excises are soft drinks, textiles, cement, sugar, and gambling.

56 The conclusion follows from the well-known result that the excess burden of a tax is a positive function of the compensated elasticity of demand.

57 See the section on the theory of optimal commodity taxation.
taxation on tobacco and alcohol can also be justified on the grounds that consumers, particularly the young, may not be fully aware of the long-run health consequences of using these products.

High excises may not be effective deterrents to consumption for addictive substances. For instance, studies on alcohol and tobacco use have found price inelastic demand for these products. There are, however, two important caveats to these results. First, most of the evidence has been assembled in low-tax jurisdictions and most of the studies have adopted empirical specifications which constrain elasticities to be constant over the entire range of demand. Second, recent studies on cigarette demand by teenagers in industrial countries have found price elasticities in excess of one. These findings suggest that high excises may be effective in deterring teenagers from smoking, in contrast to their impact on adults, who may have already acquired an addiction to tobacco. If high tax rates on tobacco are sustained over a long period, the higher prices may depress the demand of successive cohorts of teenagers and have a lasting impact on aggregate consumption. These findings are also likely to be valid in developing countries.

High excise taxes on motor vehicle use, both fuel and vehicles, are also justified on the grounds of negative externalities associated with their use. Motor vehicles generate negative externalities through air, water, and noise pollution, as well as generating congestion. Some jurisdictions have applied the benefit principle to the use of revenue arising from excises related to road use, by earmarking this revenue for road repair and construction. The argument against earmarking these revenues is that it limits governmental discretion in the use of public sector resources. Nevertheless, earmarking, particularly for road-related excises, has been defended on the grounds that it may raise social welfare by revealing social preferences for public goods more effectively. Second, earmarking may enhance public support for higher taxation to finance incremental expenditure programs, which the public might not otherwise support in the absence of linkage. Legislation mandating such earmarking should contain a sunset clause requiring review of the program after a fixed period of implementation. Earmarking of narrowly based expenditure programs in the absence of externalities is generally undesirable.

**Excises as a tool for improving vertical equity**

It is generally not advisable to expand the scope of excise taxes beyond sumptuary goods, fuel, and a few luxuries. Some jurisdictions attempt to use excises as an instrument to enhance vertical equity. The desire to pursue such a strategy is understandable if direct taxation is not progressive owing to weak tax administration. Employing indirect taxes as instruments to achieve greater progressivity may lead to a proliferation of excises on a great variety of "luxury" commodities, resulting in large administrative costs and arbitrary decisions.

The case for the use of a system of excise taxes to achieve greater progressivity in indirect taxation has also been argued. There are five conditions under which this may hold. First, the cross-section income elasticity of demand for goods and services to be taxed should exceed unity and the own-price elasticity of demand should be low, so that reduction of the share of taxed commodities in household expenditure patterns following imposition will be small. Second, expenditures on excisable commodities should be a large fraction of the household income of middle- and upper-income families and such expenditures should account for a much smaller fraction of the incomes of lower-income households. Third, if excise taxation must be extended to cover commodities purchased by lower-income households, then the excise tax system should use differentiated rates applied to subgroups of commodities on the basis of quality or price. Fourth, an excise system of progressive design must be administratively feasible so that disputes and arbitrary assessments can be minimized and the system enforced as the lawmakers intend. And fifth, the system should be perceived by the public as progressive.

It may be difficult to apply these requirements in a rigorous manner in developing countries. Income elasticities of demand for the relevant commodities and up-to-date data on expenditure patterns broken down by income classes are often not available. In the absence of such estimates, generalizations about what is and is not a luxury good are apt to be based on conjecture or the arbitrary application of research findings from other jurisdictions. Differentiating excise rates within commodity groups by price and quality, with higher rates of tax applied to the higher grade and more expensive grades in the expectation that the rich will favor these, will greatly complicate administration. The designers of such "fine-tuned" measures should be aware of the technological possibilities for substitut-

---

ability. For example, kerosene, which might be tax preferred as a nonluxury good, can be mixed with diesel fuel to produce an effective motor fuel. If this practice is widespread, it may be difficult to target goods primarily used by the poor for low excise tax rates. The case for excises on luxury goods is weaker than that for sumptuary and road-use related commodities because the own price elasticities are likely to be much higher, leading to much greater distortions in consumption decisions. Consumers are deprived of satisfaction while the government gains little revenue.

Empirical evidence suggests that excises on tobacco products and beer are regressive. The empirical results for liquor are mixed, with selected empirical studies using Jamaican data finding income elasticities in excess of one. Sugar taxes are typically highly regressive and for this reason should be avoided. A recent study found that taxes on kerosene are regressive in Indonesia, Thailand, and Tunisia, while excises on gasoline are progressive and contribute to economic efficiency.

Trade-offs among goals in the use of excise taxes

It has been argued that high rates of taxation should not be applied in developing countries on those sumptuary goods with low income elasticities of demand because the resulting tax burden is highly regressive and because the low exhibited price elasticities defeat the putative intent of deterring consumption. In rebuttal, it may be noted that poverty per se should not necessarily entitle one to create negative externalities. Moreover, as noted above, with higher tax rates, demand might be more price elastic and these elasticities may be higher for groups that have not yet formed strong or addictive habits.

Taken to the extreme, if the tax on a particular commodity is set high enough, the revenue yield of the tax declines because sales of the commodity decline. At some intermediate rate between this rate and a zero rate, there is a revenue-maximizing tax rate.

At a rate below the revenue-maximizing tax rate for sumptuary goods, increasing the rate both enhances revenue and depresses consumption. Once the revenue-maximizing tax rate has been exceeded, a trade-off between the two goals emerges and policymakers must identify the relative importance they attach to generating additional revenue and reducing consumption.

Design Issues

Scope of excise taxes

In offering advice on excise taxation, the IMF has encouraged limiting excises to five principal groups: alcohol, tobacco, automobiles, petroleum, and automobile spare parts. It has also recommended the removal of vexatious minor excises and regressive excises in favor of the general sales tax. Recent work in the applied theory of optimal taxation suggests that indirect taxes, including excises, should be imposed as close to the final sale as possible because of the potential for excises to have unexpected distributional and efficiency effects when imposed on intermediate goods.

If the excise does not apply equally to both domestic production and imports, a given excise tax rate change will require a corresponding change in the tax rate on imports, to maintain a given degree of protection. In India, for example, changes are made simultaneously in excises and in levies on imports. In Latin America, excises typically apply to both imports and domestically produced commodities.

Specific versus ad valorem rates

Conventional policy generally advocates that all excises be levied on an ad valorem basis rather than on a specific basis, because this protects the base of the tax from inflation. This is particularly important in developing countries because they are more likely to have high rates of inflation. Nevertheless, the real value of revenues may be maintained under a specific-rate excise tax if there are regular adjustments of the rate to reflect inflation.

Specific rates may be better than ad valorem rates if tax administrative capacities are limited so that undervaluation of domestic goods or imports is a common problem. Under a specific-rate excise tax, disputes over valuation do not arise, since the tax is not based on value but on physical units of the commodity. In addition, specific rates are the appropriate form of tax if the tax is intended to be externality correcting, that is, an ounce of alcohol should be taxed the same whether it is contained in high-quality spirits or not. In some
cases, therefore, it may be useful to levy the excise tax with a combination of specific and ad valorem rates or to maintain specific rates on some items.

**Ease of administration**

Excise tax liability is typically incurred as the commodity leaves the factory gate on domestic goods. Revenue officers are typically stationed inside the production facilities, particularly in the case of breweries, distilleries, and cigarette factories, to maintain on-site checks of production and shipments. Revenue authorities may also require that excisable commodities sold to the public contain an excise stamp or banderole to prevent the sales of nontaxed substitutes.
User Charges and Environmental Taxes

User Charges

GERWIN BELL

• What is the difference between user charges and taxes?
• Where are user charges optimally used?
• What are the necessary conditions for the application of user charges?
• How are user charges to be set?

With the levy of user charges for certain goods and services, fiscal authorities in effect mimic the behavior of enterprises in price setting for private goods and services. User charges are, therefore, conceptually different from taxes that are typically levied without a specific link to goods and services provided. Examples of user charges include passport and entrance fees and tolls on road use.69 While the seemingly more straightforward administration of a price system may favor the adoption of user charges especially in a context of limited institutional capacity, the theoretical and practical case for user charges is subject to certain caveats. In addition, the conceptual distinction between user charges and taxes is frequently much harder to draw in practical applications, resulting in a wide array of fiscal instruments classified as user charges.

User Charges in Principle

User charges are levied on the use of a specific good or service and thereby on the accruing benefits that economic agents receive from such use. In this way, the application of user charges allows the costs of the good or service to be distributed according to the benefits received. User charges, thus, emulate the benefit principle in public finance which states that the payment of a tax ought to correspond to the benefits received from tax-financed goods and services. Advocates of user charges stress that the allocation of government-provided goods and services through the price mechanism may help to achieve superior allocative outcomes owing to two effects: first, user charges incorporate the rationing mechanism of the price system in that the good or service is only provided to consumers who value it at least at its cost (user charge) such that demand will thereby be limited to a level below saturation demand; second, user charges also provide the information generation of the price system because the revenue obtained from user charges may be easily compared to the costs of providing goods and services and can help in the decision about the future allocation of resources. In addition, that the beneficiaries are also the payers is sometimes stressed as contributing to fairness.

Critics, on the other hand, doubt the general capability of a government to price its goods or services correctly; in the absence of correct prices, the better allocative outcome due to rationing and improved information would not necessarily be forthcoming. Critics also stress that letting all beneficiaries pay the same user charge for a good or service may not be a fair solution in the presence of unequal income distribution—at least if the good for which a charge is being levied is not a luxury good.70

On a theoretical level, the overall use of user charges is subject to two questions: first, which goods and services should be subject to user charges? Second, on which goods and services can user charges be levied? Turning to answer the first, normative, question, user charges should be applied in circumstances where they increase economic efficiency. The second, positive, question may be answered by drawing parallels to the private sector where the application of the price system for rationing is commonplace because the users of private goods and services are easily identifiable and may be excluded from the use of the goods and services if they fail to pay the market price. Therefore, to provide the answer to both questions, user charges are efficiently levied on publicly provided goods or services that enhance economic efficiency and have the characteristic that their users may be easily identified and excluded from their consumption. There are two well-defined theoretical cases where the use of user charges may support the achievement of a first-best efficient equilibrium.

One justification for the economic efficiency of user charges is based on the core philosophical foundation

---

69See Brownlee (1961), Copeland (1992), and Bat (1993).

70See Bird and Miller (1990).
for government participation in the economy, namely on the correction of externalities. One form of such corrective action is the provision of public goods that are undersupplied by the private sector because of their positive consumption externalities. Public goods are defined to be goods that are nonrival in consumption and that allow the accommodation of an extra consumer at zero cost. The requirement, however, for well-functioning user charges that nonpaying users ought to be excluded restricts the cost recovery of public goods through user charges to such public goods for which exclusion of users can be easily enforced. This is true, for example, for national parks that can only be entered through specific gates, but this does not hold for educational radio broadcasts.

Another form of such corrective action is necessitated by the existence of negative externalities. The problem there is that the market may not price certain factors that are used as inputs into production or consumption or both. Economic agents who base their optimization on market prices will thus demand too much of these factors. Examples include the use of air to pollute as well as the recreational use of parklands. Proponents of user charges point out that the government can increase economic efficiency by identifying the nonpriced good or service and charging a fee for its use equivalent to the externality. Of course, this is the classical foundation for Pigouvian taxation, and, in fact, a Pigouvian tax can be understood as a user charge if the service for which it is being charged is defined widely enough (e.g., "pollution-carrying capacity of air" may be the service for which fiscal authorities charge through a Pigouvian excise tax on emissions). Critics of user charges for externalities argue that the existence of externalities per se does not justify user charges if the government does not know the true extent of the externality. They underscore that markets have oftentimes been successful in eliminating potential externalities such that excludable public goods (such as pay TV) are frequently provided by private enterprises.71

The second theoretical justification of user charges rests on natural monopolies for private goods. Natural monopolies are firms whose cost functions exhibit decreasing average costs over the relevant scale of production, primarily owing to large fixed costs compared to variable cost (e.g., electric power networks). Efficient pricing would require that price equals marginal cost; as average costs are declining (since marginal costs are lower than average costs), however, this pricing rule implies that the price will be lower than the average costs, that is, the firm will be making a loss. An industry with such cost functions will be dominated by a monopolist who will raise prices such that a profit is assured, leading to a suboptimal provision of the good being produced. Government production of such a good can achieve efficiency if the user charge for the product is set at the level of marginal costs. Since such pricing will lead to a loss, however, the government has to finance its production through general revenue or to subsidize private production to assure a first-best outcome. Critics argue that the deadweight loss associated with the revenue generation for the subsidy needs to be taken into account when evaluating the optimality of such a scheme. Furthermore, they point out that government production would rarely take place at minimum cost and would also reduce the competitive pressure to lower costs and improve technology through innovation. In practice, the line between excludable public goods and private goods produced in a natural monopoly is likely to be hard to draw (e.g., highways).

As a corollary, the above suggests which goods or services may not necessarily be subjected to user charges. If a certain good or service is provided because of a merit foundation, that is, the government wants services to be consumed to a higher degree than they would be demanded at market prices (even in situations where major positive consumption externalities are not present), the application of economically efficient user charges would be self-defeating. This may be a problem confronting developing countries in numerous government programs aimed at improving social sectors such as education. Charging cost prices for primary education will often risk the attainment of social targets; charging for higher education, however, may lead to a more efficient allocation of resources and increase equity. Extending the last point, user charges are also inappropriate where certain goods or services are provided with a redistributional intention, i.e., where the idea is to provide certain in-kind goods or services in order to achieve some income redistribution (e.g., food stamps); of course, such a redistribution would be offset by levying user charges on the beneficiaries. The generally adverse redistributitional effect of user charges does not hold if they are levied on goods or services with a high income elasticity (luxury goods). The prevailing view in public finance is, however, that such redistribution be better done through income subsidies than in kind.

In developing countries, situations are often found where the government provides a private good or service which is not subject to natural monopoly problems. The prevailing view is that both revenue and

expenditure objectives for the budget as well as overall economic efficiency will be better served by the privatization of the provision of the particular good or service. Proponents of the use of user charges in such circumstances stress that the resources required for successful privatization may not be at hand in many instances.

User Charges and Taxes

As previously indicated, the actual distinction between taxes and user charges is sometimes hard to define as was demonstrated for Pigouvian excise taxes, if the "service" on which the tax is levied is defined on a wide enough scope. Nevertheless, there are substantial differences between user charges and taxes. On a conceptual level, some analysts have pointed out that user charges are voluntary in that economic agents can affect the payment of the user charge through their behavior, but that taxes are involuntary. This distinction, however, is usually not of practical relevance since, for example, also the income tax may be considered voluntary, as economic agents can affect the amount of their income tax through their labor-leisure choice while, on the contrary, the user charge for a driving license may not be voluntary in circumstances where the operation of a car is essential. Another legal distinction defines the difference according to intent: if the intent at the introduction of a fiscal measure, on the one hand, was general revenue collection, that measure is deemed a tax; if the intent, on the other hand, was to ration the provision of a particular good or service among competing users, that measure is deemed a user charge.

In more practical terms, a country deciding between the imposition of a tax or the levy of a user charge can do so by considering a catalog of features. First, the allocative impact of user charges may not always be desirable if user charges cannot be imposed uniformly. Toll roads provide a good example of this point: if not all roads can be made toll roads, the imposition of tolls on certain roads may lead to economically suboptimal substitution by economic agents to roads which do not charge tolls even if this implies longer distances and slower travel. In such a context, the imposition of an excise tax on gasoline would go further towards recuperating the costs associated with road traffic. A similar case can be made for the collection of waste disposal costs through taxation if one wants to avoid the adverse incentives under user charges to dispose of waste illegally.

Second, the user charge may enhance the more economic use of a good or service in a way that a tax cannot. For example, the tuition payment for public schooling would probably lead to a more efficient use of schooling than if schooling were financed from general revenue sources. As was indicated before, however, such incentives will have to be weighted against merit and income redistribution objectives.

Third, and extending the last point, different concepts of fairness have different bearing on user charges and taxes. If fairness dictates that the beneficiaries of certain government-provided goods or services ought to pay for them, a user charge is the more appropriate instrument. Conversely, if the ability to pay is considered as the overriding fairness principle, the application of a tax is indicated.

Fourth, if the administration of user charges is too complex or resource intensive, the alternative would be a tax that is as closely related as possible to the good or service the government is providing. For example, the manpower requirement for toll collection on roads may be too large so that the indirect collection of user fees through gasoline taxation may be more efficient.

Fifth, if the taxation of certain economic activities is already high, the imposition of user charges on goods or services which are used as input into these activities may worsen the overall economic allocation rather than improve it. This may often be a problem in developing countries where the tax base is limited.

Sixth, a government may find it useful to determine the composition of its budgetary expenditure to a certain extent outside the political process. This could be a second-best approach to help prevent certain essential budgetary outlays from excessive cutbacks in periods of fiscal adjustment. It may impose more (long-run quasi constitutional) discipline on the budgetary process that would otherwise be heavily influenced by short-run organized interest groups, politicians, and bureaucracy. Such arguments normally favor user charges over taxes. Within the budget, user charges generally fund specific programs or agencies, while taxes other than social security generally flow into general revenues. This differentiation, however, gets blurred in the case of earmarking taxes if the taxed good or service is very close to the earmarked use (e.g., gasoline taxes that finance highway expansion and upkeep). But if the taxed service is very different from the earmarked use (state lotteries for education) the similarity with user charges disappears while the

---

72See Richardson (1993).

overall budgetary structure still prevails. Arguments against earmarking include the political process as no longer determining the budget and earmarked resources providing incentives for the budgetary beneficiaries not to reduce their costs.

User Charges in Practice

Pricing

From the perspective of economic efficiency, user charges should reflect the marginal cost of the service. The easiest case where this would be possible is the case of zero marginal cost as, for example, for one more car's use of an uncongested highway. In this case, however, the cost of operation would have to be either financed through general revenues which are usually raised in a distortionary fashion or through certain—in general—second-best qualifications to the marginal cost pricing rule. Such qualifications can be in the form of price discrimination according to the relative elasticity of groups of demand (Ramsey pricing) or price discrimination in accordance with the relative capacity use (peak-load pricing). Other qualifications of the marginal cost pricing rule consist of two-part tariffs (where a fixed sum is charged plus a fee per unit of service, e.g., telephone services) or block rates. There are cases, however, where the deviation from marginal cost pricing is efficient even when the marginal costs are zero—that is, in the case of externalities. The marginal cost of pumping one more ounce of effluent into a river may in fact be zero, yet the cost to society is not. Similarly, an admission fee of zero to national parks may not appropriately reflect the scarcity value of this resource.

The fundamental problem is, however, that in most cases, marginal cost is not observable so that the cost has to be estimated. This may be relatively easy when markets for comparable goods or services exist (e.g., parking spaces in the inner city). Critics, however, point out that—in general—this task is usually infeasible for a government that is subject to uncertainty and ignorance and also to certain strategic behavior which is aimed at distorting the cost of government-produced goods or services. In the presence of such problems, the government will also not be able to determine the optimal capacity to supply, and the fees charged will lead to unknown cross subsidization. Such cross subsidization would then face the possibility of becoming subject to the political process. In a realistic macroeconomic environment where inflation is very often present to a nonnegligible extent, even user charges whose prices had once been set optimally will have to be adjusted to keep pace with the general increase of the price level.

Identification of the user and intensity of use

A user is defined as an economic agent who is to any extent responsible for the depletion of a particular good or service. In cases where nonexcludable public goods are targeted for user charges, the question of how to deal with free riders arises. Identifying the actual users may oftentimes be difficult as is illustrated by the example of road users. The licensing of cars does not charge road users such as pedestrians, horses, and bicyclists. The imperfect identification of intensity of use may be illustrated by sewer use. In the absence of metering, a user charge on a sewer connection does not identify users according to their intensity of use. In cases where users or their intensity of use are imperfectly identified, cross subsidization among users will be the rule.

Types of user charges

The following typology of user charges is usually drawn:

- **User fees**: These are payments on services consumed to yield a direct benefit to the user. Examples include royalties on natural resources, canal, bridge, and highway tolls; lease and rental payments; and charges for recurring sales of resources (e.g., water, minerals, and timber) and products (e.g., power); use of land (e.g., for grazing livestock); use of facilities (e.g., natural parks); services (e.g., mail delivery, waste disposal); and permits and licenses.

- **Regulatory fees**: These are very much like a tax in that they are solely based on the government's sovereign power to regulate particular economic agents or activities. Examples of regulatory fees include passport and judicial fees, customs service user fees, patent and copyright fees. In contrast to user fees, regulatory fees may or may not benefit the user (e.g., fees for patents may be beneficial but fees for inspections of certain set standards may not). In the latter case, the economic rationale for the regulation is the compensation for externalities imposed by unregulated behavior. This holds for external costs as well as benefits (e.g., drug certification charges may be set lower than marginal cost to recoup the positive externality involved when curing patients infected with contagious diseases).

---

74 See Katz (1987).
76 See Richardson (1993).
Conclusions

• The difference between user charges and excise taxes is not clear-cut, and user charges may often be substituted for taxation and vice versa.

• User charges may augment efficiency if they reduce externalities or price the goods or services of a natural monopoly.

• The efficiency of user charges depends on the pricing of the goods or services concerned.

• In inflationary environments, user charges will have to be reset frequently.

• The fairness of user charges has to be judged according to the prevalence of the benefit principle over the ability-to-pay principle and the income elasticity of the goods or services concerned.

Environmental Taxes

David C. L. Nellor

• What is the rationale for environmental taxes?

• What are the different types of environmental taxes?

• What factors influence the extent to which environmental taxes might be used?

The increasing use of environmental taxes, albeit mainly in industrial countries, is moving in the opposite direction from the more general thrust of tax reform which has been toward neutral broad-based tax systems that permit lower tax rates. Contrary to this trend in tax reform, environmental taxes are directed toward changing the relative prices of productive inputs and consumer goods that are linked to environmental damage. Most notable is the prominence now given to the introduction of carbon/energy taxes as a vehicle to meet both environmental and revenue objectives.

This note has three main parts: first, it looks at the rationale for environmental taxes; second, it examines the nature of environmental taxes; and third, it considers the factors that influence the scope for and design of environmental taxes.

The Rationale for Environmental Taxes

Environmental objectives

Environmental damage arises when the social costs of an activity that uses an environmental resource exceed the social benefits of that activity. This divergence between social benefits and costs can arise owing to one or both of two factors. The first factor is market failure—the failure of markets to bring about a solution where marginal social cost and benefits are equated—which could arise, for example, when there are poorly defined property rights over forest use or there are emissions of waste into rivers or the atmosphere. The second factor is policy failure—the use of a microeconomic policy that changes relative prices so as to encourage excessive use of a natural resource—for example, the use of chemical pesticides or water resources may be subsidized by the government.

A variety of policy instruments are available to address these environmental problems. In the case of market failure, structural measures (such as land reform that include the definition of property rights), regulatory policies (such as those involved in forest management), or tradable permits (in the case of waste emissions), could be appropriate. In the case of policy-induced environment problems, removal of the policy or its modification may be appropriate. In yet other cases, taxes may provide a solution to environmental problems whether caused by market or policy failure. Economic instruments—policy instruments that use the price mechanism such as taxes and tradable permits—are often preferred by economists because they do not require a costly regulatory apparatus and encourage reduction of environmental damage by the least costly means.

The choice of policy instruments—regulatory or tradable permits versus tax—may depend on the source of uncertainty. If there is better information concerning efficient levels of consumption or production of a good but uncertainty in knowing how to set tax rates to achieve that level of production or consumption, a solution of tradable permits or regulation could be preferable. From an economic perspective, the issue may be to determine which will create the greatest cost if the policymaker is wrong—setting the quantity incorrectly or the price via a tax.

Tax policy objectives

Environmental taxes permit a given level of revenue to be raised with lower efficiency costs than otherwise possible. It is sometimes noted that environmental taxes provide a “double dividend” because they (1) reduce the social cost of environmental damage as well as (2) permit a reduction in the rates of other socially

\footnote{This note uses this economic definition although it is recognized that the term “environmental damage” has taken on a variety of meanings.}
costly taxes thereby contributing to an overall reduction in the social cost of the tax system. Environmental objectives are met by raising the tax rate on emissions until the marginal social cost of using environmental resources is equal to the marginal social benefit of using the resources. From a revenue perspective, the rate of tax should be set such that the marginal social cost of setting the tax rate is equal across alternative tax instruments. The attractiveness of garnering a double dividend, of reduced environmental damage and revenue without social cost from environmental taxes, should not be underestimated. Many estimates for the United States suggest that the marginal welfare cost of taxation is about one third or higher. Thus taxes, such as carbon taxes which have relatively large tax bases, could dramatically improve the efficiency with which revenues are raised.

**Environment-Related Taxes**

A variety of taxes may be related to the environment—four categories are noted here. The first two categories of taxes—Pigouvian taxes or indirect environment-targeted taxes—can resolve or ameliorate environmental problems by changing the relative prices of activities—thereby discouraging various activities—that result in environmental damage. The third category of taxes are taxes that have unintended—favorable or unfavorable—environmental implications. These taxes, such as income taxes, which have incentives encouraging certain forms of economic activity, have played an important role in environmental damage in some developing countries. The fourth category of taxes are those often labeled as environmental taxes, but which are revenue sources earmarked for expenditures on the environment, such as the feedstock taxes in the United States which are earmarked for the “Superfund” to clean up hazardous waste dumps.

**Pigouvian taxes**

Pigouvian taxes are the classic textbook solution to resolving environmental problems. These taxes are specific rate taxes levied on units of emissions or on the units of damage generated by an activity. The rate of tax per unit is set to equal the marginal social cost of an activity such that the total marginal cost of the activity is equated to its marginal benefit. The tax rate is specific, not ad valorem, because the rate of tax is set according to the damage generated by an activity rather than the price of that activity.

Pigouvian taxes are viewed as efficient solutions to many environmental problems because they use the price mechanism to encourage modification to economic activities which, in an ideal world, is less costly to administer, and encourage an efficient reduction in damage compared with regulatory mechanisms. Taxes induce efficient allocation of pollution reduction across consumers and in the composition of that adjustment, across scale of the activity, technological change, and factor substitution.

There are, of course, a number of well-known practical difficulties in implementing Pigouvian taxes. The difficulty of determining the marginal social cost to set the correct tax rate is a major impediment to the ideal use of these taxes. There are also issues of monitoring, measurement, and compliance.

**Indirect environmental taxes**

In view of the practical difficulties in levying Pigouvian taxes, other taxes targeted to achieving environmental objectives can be employed. These taxes—termed indirect environmental taxes—are levied on the use of productive inputs or consumption goods where the use of those goods is related to emissions or environmental damage. Indirect environmental taxes change relative prices but are not levied directly on the damage or emission.

The case for indirect environmental taxes is thus made on one or both of two grounds.

- Where an input to production or consumption activities is and can only be used in fixed proportion to the emission. These cases arise when there is no “end-of-pipe” technology permitting modification of input usage that will influence emissions. Even in these cases, however, the danger is that taxing inputs may remove the incentive for innovation of such end-of-pipe technology that could reduce emissions.
- When taxes on inputs are the most cost-efficient policy. Alternative policies are evaluated recognizing the administrative costs of their implementation—this process may eliminate policy options that are ideal from a narrow efficiency perspective. This is probably the most common environmental rationale for input taxation and arises because the administrative costs of taxing emissions are prohibitive.

The efficiency of indirect environmental taxes is dependent on the nature of the link between the tax base and environmental damage. As noted above, a Pigouvian tax is levied directly on the environmental damage. An indirect environmental tax will be as equally efficient as the Pigouvian tax if it is levied on a tax base—such as inputs or final goods—that is linked by a definable functional relationship—ideally a fixed proportional relationship—to that damage. The less well defined the relationship between the use or consumption of the taxed goods and the environmental
damage, the less effective are these taxes as indirect environmental taxes. The relationship between the use of taxes and their effectiveness in securing environmental objectives can be illustrated by two indirect environmental taxes: a carbon tax and a petroleum tax.

A carbon tax is equally efficient as a Pigouvian tax because there is a fixed relationship between the tax base—fossil fuels—and the environmental damage. A carbon tax is intended to address global warming concerns arising from the emission of carbon dioxide that results from the combustion of fossil fuels. A Pigouvian tax would be levied on the emission of units of carbon dioxide. However, because there is no end-of-pipe technology that permits a change in the relationship between the combustion of fossil fuels and emissions of carbon dioxide, a tax on the fossil fuels is equivalent in its effect to a Pigouvian tax—there is a fixed proportional relationship between the use of the fuel and the emissions.

Petroleum taxes intended to address local air quality concerns are not as efficient as Pigouvian taxes on emissions or damage; there is no fixed relationship between the use of petroleum and environmental damage. A tax on petroleum is not related to the emission of pollutants because the pollutants are also a function of the vehicle's operating efficiency, driving speed, road quality, etc. In this case, although a tax on petroleum may reduce pollution, it is not an efficient policy from a theoretical perspective. An efficient tax on vehicle emissions would likely result in a reduction in emissions from a variety of reactions to the increased cost of emissions depending on the least-cost means of reducing emissions. For example, emissions could be reduced by lower mileage traveled, improved vehicle maintenance, and technological adjustment, such as the use of catalytic converters. Indirect environmental taxes such as petroleum taxes do not encourage adjustment on all of these margins. In fact, they may discourage the use of some methods of emissions adjustment, such as changes in the use of catalytic converters.

**Taxes with unintended environmental implications**

Some taxes have unintended environmental implications. Excise on various energy products are often viewed as environmentally favorable when they discourage environmentally damaging activities. Taxes could, however, unintentionally have environmentally harmful consequences by creating distortions that cause environmental damage. For example, an income tax may favorably treat agricultural investment that encourages capital intensive activities in areas unsuited to that form of agriculture.

**Earmarked taxes**

Some taxes are labeled as environmental taxes but are taxes that raise revenue earmarked for environmental purposes rather than taxes that change the relative price of using environmental resources. There are sound fiscal arguments against the use of earmarked taxes; they imply a coincidence between the level of revenue raised and the expenditure requirement, and they discourage effective evaluation of expenditure alternatives. In some countries, however, tax measures have been "legitimized" by identifying the importance of the environmental expenditure to which they are put.

**Factors Influencing the Scope for and Design of Environmental Taxes**

In considering tax reforms that may meet environmental objectives, the policymaker has two general considerations. First, in reviewing all taxes, consideration can be given to whether or not there are special incentives or tax rate preferences which might unintentionally result in environmental damage. Second, the policymaker can consider what role Pigouvian and indirect environment-targeted taxes can play. In the following text, the focus is on the scope and design of environment-targeted taxes.

Pigouvian taxes are difficult to define and administer. Efficient setting of a Pigouvian tax requires determining the tax base of emissions or damage and setting a tax rate such that the marginal social cost of environmental damage is equal to the marginal cost of abatement. Determining the shape and nature of these damage and cost functions, even within reasonable degrees of accuracy, is a considerable challenge. Many of these difficulties—specific to environmental taxes such as ecological valuation and damage measurement—apply to varying degrees to indirect environment-related taxes.

Many environmental concerns have a spatial or temporal dimension, and it can be difficult to structure a tax related to the damage or emission when this is the case. For example, the costs of air pollution from automobiles are highest in urban areas and often at particular times of the day. Taxes on gasoline, for example, are blunt instruments to deal with these aspects of environmental damage. Work at the World Bank has suggested using a combination of taxes and regulatory devices as the best method to address this difficulty.

Other aspects of environmental damage may be exacerbated by the imposition of an environmental tax even if a link between the tax and a source of environ-
mental damage is established. A commonly mentioned example is the use of a tax on petroleum fuels in a poor country that exacerbates the costs of market failure in forests. In Thailand, for example, a tax on chemical fertilizers, which in most circumstances is environmentally beneficial, resulted in environmental damage because it encouraged extensive farming, rather than intensive farming, by forest clearing on land where property rights were poorly defined. The extent of market and policy distortion is probably greater in many developing countries.

Environmental taxes are likely to be ineffective if there is macroeconomic instability. Setting a specific environmental tax is wasted effort when rates of inflation are high. Thus, the type of fine tuning of environmental taxes discussed in industrial countries may not be appropriate in all countries.

Environmental taxes may not have the desired incentive effects of modifying production methods that reduce environmental damage. Rather, companies such as state enterprises that are not profit maximizing pass on environmental taxes without seeking to reduce the tax burden by modifying their production methods. This is an important concern in the case of economies in transition but also in many developing countries where public utilities and enterprises may play an important role in the economy.

There are also political economy concerns. First, there is a danger that the use of environmental taxes, by intentionally using the tax system to modify incentives, will encourage numerous other demands to use the tax system for special treatment of specific activities. The experience with special tax provisions is not encouraging. Second, a political impediment to the introduction of environmental taxes is the argument that they harm international competitiveness. Partly as a result of concerns regarding international competitiveness, many proposals for environmental taxes have been made at the international level. For example, the European Community has proposed that a carbon tax be introduced in its member countries, but its implementation is dependent on other major countries introducing measures with comparable effect. These international agreements are inevitably difficult to complete.

There is a potential conflict between revenue objectives and environmental targeting. One tax policy instrument is unlikely to efficiently achieve both objectives. The nature of the possible trade-off between revenue and environmental objectives will vary from case to case. For example, in the presence of severe environmental damage, the tax rate may need to be set so high to meet environmental objectives that there is an inverse relationship between meeting revenue and environmental objectives. Revenue objectives that call for a tax rate that exceeds the rate justified by environmental objectives, however, is perhaps the most likely case. For a given revenue objective, the tax rate should be higher than that justified by environmental considerations if the tax elasticity is greater than zero at the environmental optimum based on the presumption that other taxes have social costs. In these circumstances, an additional dollar of revenue can be raised by increasing the rate of environmental tax and reducing other tax rates with a net reduction in social costs of the tax system.

Raising environmental taxes may conflict with equity objectives. For example, the empirical literature qualifies, although it does not refute, the common perception that a carbon tax will be regressive. Raw data for the United States support the conclusion that a carbon tax is regressive; the percent of income going to fossil fuel consumption in the lowest decile of income is 10.1 percent compared with 1.5 percent of the highest decile of income. European data, however, do not unambiguously support this conclusion. Other factors suggest that the tax may not be as regressive. For example, using permanent income to assess the progression of carbon taxes shows that the tax is considerably less regressive and, in a number of European countries, it is no longer regressive at all. In developing countries, institutional factors may mean that a carbon tax would not be as regressive as it would appear on the surface. The prevalence of price controls that pass forward of taxes is sometimes illegal and the likely price inelasticity of supply all reduce the prospect of forward shifting of the tax, making it more likely that it will be borne by capital owners and thus more progressive.

Conclusion

A variety of policy instruments are available to address environmental problems that arise when the social costs of an activity that uses an environmental resource exceed the social benefits of that activity. Economic instruments—policy instruments that use the price mechanism, such as taxes and tradable permits—are often preferred because they do not require a costly regulatory apparatus and encourage reduction of environmental damage by the least costly means. Fiscal policy objectives can also be enhanced by environmental taxes because they permit a given level of revenue to be raised with lower efficiency costs than otherwise possible. Environmental taxes are thus said to provide a "double dividend" because they (1) re-
duce the social cost of environmental damage as well as (2) permit a reduction in the rates of other socially costly taxes, thereby contributing to an overall reduction in the social cost of the tax system.

A variety of taxes may be related to the environment—four categories are noted here. The first two categories—Pigouvian taxes or indirect environment-targeted taxes—can resolve or ameliorate environmental problems by changing the relative prices of activities—thereby discouraging various activities—that result in environmental damage. The third category is that of taxes having unintended—favorable or unfavorable—environmental implications. The fourth category is that of taxes often labeled as environmental taxes, but which are revenue sources earmarked for expenditures on the environment.

In considering tax reforms that may meet environmental objectives, the policymaker has two general considerations. First, in reviewing all taxes, consideration can be given to whether or not there are special incentives or tax rate preferences which might unintentionally result in environmental damage. Second, the policymaker can consider what role Pigouvian and indirect environment-targeted taxes can play.

In view of the practical difficulties in levying Pigouvian taxes, other taxes targeted to achieving environmental objectives can be employed. These taxes—termed indirect environmental taxes—are levied on the use of productive inputs or consumption goods where the use of those goods is related to emissions or environmental damage. The scope for these taxes and their design is influenced by a number of factors, inter alia, administrative factors, the spatial or temporal dimension of the environmental damage, other aspects of environmental damage, macroeconomic conditions, industrial structure of the economy, political economy issues, potential conflicts between revenue and environmental objectives, and equity objectives.
References


