

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Reform, Recovery, and Growth: Latin America and the Middle East

Volume Author/Editor: Rudiger Dornbusch and Sebastian Edwards, eds.

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-15745-4

Volume URL: <http://www.nber.org/books/dorn95-1>

Conference Date: December 17-18, 1992

Publication Date: January 1995

Chapter Title: Tax Lore for Budding Reformers

Chapter Author: Arnold Harberger

Chapter URL: <http://www.nber.org/chapters/c7660>

Chapter pages in book: (p. 291 - 310)

Arnold C. Harberger

## 9.1 Introduction

When I was first invited to this conference, I was told that part of my task was to convey some of the insights and knowledge that came from long experience with both the theory and practice of public finance. This helps explain the title of this paper, and to motivate my serving up such a bouillabaisse. I like that analogy because I include some items that are very fundamental though familiar (hoping to make the soup rich and nourishing), but also some more exotic ingredients (hoping to give it a special flavor and maybe even signal the identity of the cook).

## 9.2 On Fairness, Compliance, and Corruption

More and more as I grow older, issues of tax administration and of “fairness” intrude on an economic vision that was for quite a long time dominated by considerations of allocative efficiency. I am impressed by the fact that we live in societies that are composed of many different kinds of people (I almost said “many different groups,” but thought the better of it), with widely differing objectives, values, and tastes, struggling against different constraints and obstacles to make their respective ways in life. They are all members of the broader polities that carry out the taxing and spending roles of government. They all receive benefits from the various governments that serve them, and they all recognize that somehow the bill for these services must be paid. Yet we know that governments would not get very far trying to finance themselves (as do churches, colleges and universities, and independent charitable organi-

Arnold C. Harberger is professor of economics at the University of California, Los Angeles.

zations) by voluntary contributions. Thus we come to fees, rates, and taxes—all to a greater or lesser degree involuntary contributions.

Looking around the world, one sees various kinds of patterns in the design and administration of taxes. I would distinguish four main types of relationship between taxpayers and their government: (1) those dominated by mockery and mischief, (2) those dominated by arbitrariness and fear, (3) those dominated by corruption, and (4) those dominated by a sense of fairness and voluntary compliance.

Type 1 is well represented by the relationship between taxpayers and government that prevailed until quite recently in Argentina and a number of other Latin American countries. It can be caricatured as almost a game between taxpayers and the tax authorities. The authorities raise rates, and evasion increases; they lower rates, and evasion decreases. We will see later that this sort of process occurs everywhere, but I would define type 1 as covering cases where it dominates the relations between taxpayers and government.

I recall working on a tax project in Argentina in the early 1960s, in which it fell to me to study what had happened to taxpayer compliance with the income tax during the decade of the 1950s. I did so by performing an exercise that is quite simple—one that I recommend to anybody who has access to the relevant figures (for any country, in any period). I took the official statistics on income tax receipts for 1951, classified by tax bracket. Obviously evasion was already going on in that year, and we have no easy way of knowing its extent. But we pass over that problem and simply take the level of 1951 compliance as a base. We then move to 1955 (or some other year with the relevant data) and *assume* that the relative distribution of income in the taxpaying brackets did not change between these two years. Implementing this assumption, we let  $L_1^0$  and  $U_1^0$  be the lower and upper bounds of the first tax bracket ( $B_1^0$ ) in year zero (1951) and let  $N_1^0$  be the number of taxpayers declaring income in the first bracket on that year. If prices grew by  $\dot{p}$  and real per capita income by  $\dot{y}$  between 1951 and 1955, we expect the counterpart of  $B_1^0$  in 1955 (time  $t$ ) to have a lower bound of  $L_1^0(1 + \dot{p} + \dot{y})$  and an upper bound of  $U_1^0(1 + \dot{p} + \dot{y})$ . Similarly, if the relevant population had grown by  $n\dot{n}$  between 1951 and 1955, we would expect the number of people in this bracket to have grown to  $N_1^0(1 + \dot{n})$ . We now look in the 1955 tax table to find how much tax a typical family with income of  $L_1^0(1 + \dot{p} + \dot{y})$  would have to pay (or alternatively look at the 1951 data to find out how much a typical family declaring that income actually paid in 1951). We do the same for the income of  $U_1^0(1 + \dot{p} + \dot{y})$ . Interpolating sensibly, we assign a tax between these limits to the  $N_1^0(1 + \dot{n})$  taxpayers whose incomes are calculated to be between  $L_1^0(1 + \dot{p} + \dot{y})$  and  $U_1^0(1 + \dot{p} + \dot{y})$ . Obviously, the top of the first bracket  $U_1^0$  is the bottom of the second, so we can similarly assign to  $N_2^0(1 + \dot{n})$  taxpayers a tax that is derived from period  $t$  incomes lying between  $U_1^0(1 + \dot{p} + \dot{y})$  and  $U_2^0(1 + \dot{p} + \dot{y})$ .

Projecting 1955 tax revenues in this way, we see what we would collect if

(1) the distribution of taxpayers by income bracket had remained unchanged from 1951 and (2) their degree of compliance had remained unchanged. This projected tax revenue can then be compared with actual revenue for 1955. As I recall the experiment, they found that actual 1955 revenue was barely two-thirds of the projected amount. More astounding still was that, when the same exercise was repeated for the time span 1955 to 1959, the same thing occurred. Thus, by 1959 taxpayers were paying only about four-ninths of the taxes they would have paid on the basis of the level of compliance that prevailed in 1951!<sup>1</sup>

One interpretation looks upon what happened in Argentina in this period as a sort of game between the taxpayers and the tax authorities. The taxpayers reduce their compliance; the authorities raise the rates; taxpayers respond by reducing compliance still further; the authorities raise rates again, and so on. Another interpretation simply views the reduction in compliance as the predictable result of increases in rates (see section 9.8).

It is interesting that cases of type 2, tax administration dominated by arbitrariness and fear, tend to come more from the history books than from the annals of recent years. To the best of my knowledge, one finds people in jail for tax evasion most often in the advanced industrial democracies, and least often in countries at the lowest income levels. But the advanced industrial democracies are examples of type 4. They make great efforts to instill in their citizens a sense of the underlying fairness of the tax system and its administration, and impose significant restrictions on their tax administrators, for the precise purpose of avoiding persecution and harassment of taxpayers by tax collectors. These safeguards were instituted as much to keep the tax collectors honest as to ensure compliance by taxpayers. Otherwise, the system can easily deteriorate into one of type 3, dominated by corruption, where taxpayers are content to buy off the collectors but in the process are subject to all sorts of capricious threats.

Modern tax administration is characterized by randomized audits and checks, with computers doing most of the selecting and with the probability of audit increasing with the potential for evasion, as well as with the likelihood that the audit will produce additional revenue. Cases are shifted from one inspector to another with some frequency, and the duties of inspectors are also changed more frequently than would be justified on strict efficiency grounds. All this is done to minimize the temptation to corruption and at the same time to standardize and routinize administrative and audit procedures.

### **9.3 On Uniform Tariffs**

Uniformity of rates of import duties can be looked upon as an administrative device to help guard against corruption, and as a compliance measure to help

1. This assumes no major change in income distribution. I believe it is correct to make this assumption because (1) income distributions tend to change relatively slowly over time, (2) exter-

instill in taxpayers a sense that they are being fairly treated, and (compared with most points of departure of trade liberalization programs) as a measure for economic efficiency. I feel that the efficiency motive would likely be dominant if one started from a tariff system that is highly protective, highly distorting, and far from uniform to begin with. But once one has corrected the most blatant disparities in rates, and once one has moved the average rate down to, say, the range between 20% and 40%, then the other two arguments take on greater weight. I doubt that there is any surer temptation to corruption than the power that a customs inspector has to classify a given import good into two or more categories, with substantially different rates. The offense need not be blatant; indeed, a careful inspector could so arrange things (by threatening to put an item in a higher-rate category and accepting a "tip" for putting it in the correct one) that the Treasury loses no money because of his misdeeds. In any case, it is easy in such circumstances for inspectors to make sure that some duty is being collected in every case in which they succumb to a bribe. Contrast this with a uniform tariff; if it is truly uniform and across the board, there is no room at all for corruption that works through classifications. Outright flouting of the law, by letting items pass without paying any duty at all, or by falsifying the documents dealing with customs valuation, becomes a virtually necessary precondition to corruption, once a uniform rate is in place.

In a sense it is obvious how tariff uniformity conduces to a sense of fair treatment on the part of taxpayers: nobody can feel he is being singled out; nobody can sense himself the victim of misunderstanding or prejudice. Everybody, simply everybody, is treated in the same way.

But the usefulness of uniformity goes beyond a simple recognition that the law, on paper, treats every importer and every import good alike. It also carries the implication that the rate of effective protection facing all actual and potential import-competing industries is the same. This in turn (so long as the policy has credibility for the future) seems to reassure potential investors in such industries that they will not be singled out for capricious or arbitrary treatment in the future. Investors may still have to worry about the rate of a general and uniform tariff being raised or lowered, but that is a different thing from a sudden withdrawal of protection from their product (and perhaps a few more) or a sudden increase in the tariffs they (and perhaps a few others) have to pay on inputs. The generality and uniformity of the tariff by themselves convey (so long as they are expected to continue) vast amounts of information and vast amounts of reassurance to economic agents. The risks that a 20% rate will move to 50% or to zero can be quite large when each tariff is treated separately; these same risks are virtually nil if they refer to moving the entire uniform structure up to a general rate of 50% or down to a general rate of zero. Eco-

---

nal evidence from Argentina in the 1950s suggests no dramatic distributional shift, and (3) even simulations based on extreme assumptions of distributional changes result in revenue reductions that are tiny in comparison with the four-ninths factor that we calculated.

conomic agents can move with more assurance, and can think in terms of longer horizons, when all they have to worry about are up and down movements of the general tariff, than when they can be placed in difficulty by all sorts of specific moves concerning their own products and their own inputs, or concerning the products of actual and potential rivals.

So uniformity of tariff rates ends up (compared with typical prereform tariff structures and with most others) by promoting a greater sense of fairness on the part of taxpayers, making tariff administration easier, making corruption much more difficult, providing an unequivocal pattern of uniform effective protection to import-competing activities, and giving clearer, potentially more reliable signals to investors (of real resources) and other economic agents. By the nature of the case, uniformity encompasses the whole range of different rates, so assertions of superiority (to preexisting tariff structures) should be conditioned on the average rate being at worst the same as, and hopefully significantly lower than, the preexisting average. And, of course, the fact remains that world welfare can generally be enhanced by lowering a uniform tariff rate all the way to zero.<sup>2</sup>

#### **9.4 Radial Reductions versus the Accordion Principle versus Sectoral Sequencing in Trade Liberalization Programs**

Having proposed uniformity of rates as a plausible and worthy goal of tariff reform, I think it reasonable to insert at this point a discussion of a technical point that is not widely appreciated. Consider a developing country that is about to embark on a major program of trade liberalization. Suppose, for sim-

2. Textbooks treat the cases of "optimum" tariffs and export taxes, often in the context of models with one import and one export good. This turns these optimum taxes into general import tariffs and general export taxes, practically by definition. The right way to apply the theory of optimum trade taxes is to recognize that optimum import tariffs are justified solely to the degree they help exploit a national monopsony position, vis-à-vis the rest of the world, while optimum export taxes are justified solely to the degree they help exploit a national monopoly position. I am prepared to concede that some countries have artificially restricted exports of particular goods so as to exploit actual and imagined monopoly positions. But I know of no country that has employed tariffs for the principal purpose of exploiting a monopsony position. Few trading entities possess enough market power on the buying side to be able to act as monopsonists. And those who do (like the United States and the European Economic Community [EEC]) never use trade policy with the aim of exploiting monopsony power. If they did, these economic giants would have high tariffs on copper, lead, zinc, tin, and other metals; and on bananas, coffee, tea, and other food items with rather inelastic total world supply, in which they account for enough of world demand so that their tariffs would significantly lower the world prices of the items in question. It is my contention that the United States and the EEC have never adopted tariffs for the purpose of pushing down world prices. They have, in a few cases, imposed trade restrictions having this effect, but in these cases (sugar in the United States, wheat and meat in the EEC are examples), the motive was patently an old-fashioned protectionist one, not a technocratic exploitation of monopsony power. Given their admirable self-restraint in this regard it is reasonable for economists to ignore monopsonistic motives where the economic giants are concerned. With respect to smaller countries, the facts themselves proclaim the absence of monopsony power.

plicity, that its end objective is a uniform 20% tariff on all imports. How should it proceed? Most analysts and observers agree that a useful starting point is (1) to convert quotas, licensing schemes, and other quantitative restrictions into their rough “tariff equivalents”; and (2) to take the “water” out of tariffs that are at present prohibitive (i.e., bring them close to the point where trade in fact would occur). It is from here on that schemes differ.

Those whose instincts and training lead them to think in terms of industrial groupings are likely to think of a priority ordering in those terms, liberalizing imports first in one group, then another, then another. Those who in observing actual tariff structures have tended to focus most strongly on the highest rates tend to gravitate to the “accordion principle.” This principle squeezes the rate structure from above, first reducing, say, to 100% all rates above that figure, then moving the top rate from 100% to 80%, then moving it from 80% to 60%, and so on, until the target level has been reached. (This principle may run into a sort of logical box when the target level is other than zero, but let me simply append to it the notion that, while the maximum rate goes 100% to 80%, the minimum goes from 0% to 5%, that at the next step the minimum moves up to 10%, then to 15%, and finally, say, to 20%.)

The third candidate for our scrutiny is the principle of radial reductions. This works as follows. Let the target tariff in item  $i$  be  $\tau_i^*$ . We start out with a vector of actual tariffs  $\tau_1^0$ ,  $\tau_2^0$ , and so forth, which we intend by the end of the process to bring to the levels  $\tau_1^*$ ,  $\tau_2^*$ , and so forth. The principle of radial reductions simply says that at each step we close each gap ( $\tau_i^0 - \tau_i^*$ ) by a given percentage. Thus at every step of the tariff adjustment, every single tariff moves in the correct direction by a specified fraction of the gap ( $\tau_i^0 - \tau_i^*$ ). If we are headed for an across-the-board tariff of 20% in four equal steps, then the 100% tariff goes 100, 80, 60, 40, 20; the 40% tariff goes 40, 35, 30, 25, 20; the 12% tariff goes 12, 14, 16, 18, 20; and the 0% tariff goes 0, 5, 10, 15, 20.

The radial reduction method is to be preferred because it never consciously sends a wrong-direction signal to any activity or sector. The other two methods unfortunately have such perverse signalling built in. The problem is that at every step in the process the price signals go in the wrong direction for “most” of the activities in question. This is easiest to show with sectoral sequencing. If we are moving in four steps, we will presumably have four “sectors.” When we liberalize imports in sector A, resources are being expelled from that sector. But liberalization in A causes the real exchange rate (defined as the real price of foreign currency, e.g., the real peso price of the dollar) to rise. This partially offsets the effect of tariff reductions on the activities included in A, but it gives an unequivocal signal to draw resources into the activities of sectors B, C, and D. Unfortunately, the likely end result of all the steps will be to expel resources from all four protected sectors. If this is the case, then each of the four will be given the “right” signal at only one of the steps—that step in which its particular tariffs are being adjusted. In all the other steps, it will be given the “wrong”

signals, stemming from the real exchange rate effects of the liberalizations taking place in the other sectors.<sup>3</sup>

Under the accordion principle, the effects are essentially the same. To make the exposition easy, let me suppose that all imports are initially subject to tariffs, and that the tariffs lie in the four bands: (1) 80% to 100%, (2) 60% to 80%, (3) 40% to 60%, and (4) 20% to 40%. It is easy to see that when all tariffs in band 1 are lowered to 80%, imports in categories 2–4 receive (through the real exchange rate effect of the tariff reduction in 1) a perverse price signal. Similarly when at the second step all tariffs in 1 and 2 are lowered to 60%, all imports in categories 3 and 4 receive perverse price signals. And finally, when all tariffs in 1–3 are lowered to 40%, activities in category 4 receive yet another perverse signal.

Why should we choose liberalization schemes that have the attribute of sending conflicting price signals to all or nearly all activities as the liberalization process passes from stage to stage? In particular, why should we do so when it is so easy to avoid sending conflicting signals? The obvious answer is simply to choose the radial reduction method of achieving the liberalization targets. This approach guarantees that each activity will receive a signal in the correct direction at each step of the liberalization process. This will be true for activities that in the end will be called upon to release resources, and also for activities that in the end will be called upon to increase their level of resource use. It will hold for normal cases and for anomalous ones, too.

Obviously, there are political pressures and debts, marriages of convenience, sacred cows, and taboos within most real-world settings in which liberalization actually occurs, and real-world governments must make their own judgments as to how genuine and how forceful are these considerations in any given setting. My position is as follows. Most governments will cede something to the above considerations, but those considerations should influence economists only after they bring to the bargaining table the most sensible, most natural solution that economics has to offer. So far as I can see, the principle of radial reduction wins hands down over its two principal rivals and should be the “prescription of choice” for economists as they are called upon for advice and counsel on liberalization programs.

## 9.5 On Setting the “Boundaries” of Tax and Other Policies

One of the most tricky aspects of economic policy making is the setting of the “boundaries” to be covered by a given policy. If we are to stimulate small

3. Real life is usually more complicated than simple examples, so let me mention a couple of complications, neither of which affects the essential point being made. In the first place, within any of the sectors there may be activities whose tariffs are being raised (i.e., for which  $\tau'_i > \tau_i^0$ ). Such sectors will get signals in the correct direction at each of the four steps, assuming the net effect of each step is an increase in imports. Second, there exist anomalous or perverse cases in



businesses, how do we define “small”? If in a developing country we are to favor nontraditional exports, how do we draw that line? If we have a value-added tax (VAT) that purports to be more or less general, how should we set the limits to its coverage?

For all of these decisions and a myriad of others, technical economic analysis has a lot to contribute, but so do considerations of administrative costs, of fairness, of controlling corruption and evasion, and so forth.<sup>4</sup>

This section really has two purposes. The first is to assert that almost never is an “industry” as we know it a sensible criterion on which to define the limits of a law. The exceptions of this statement are clear: pure food and drug legislation naturally encompasses the pharmaceutical and food-producing industries. Safety on railroads, buses, and airlines is appropriately dealt with by legislation specifically oriented to these activities. Protecting the public against irresponsible (as well as fraudulent) behavior by banks and insurance companies motivates legislation focusing on those industries.

Having said that, I must add that I have never been able to find anybody who could give me a single good reason motivating a tax (or subsidy) on the activities or products of the textile industry (which covers canvas tents, denim jeans, lace panties, nurses’ uniforms, and designer clothes, among many others) or of the shoe industry (which covers a range from baby booties to hunting boots to evening slippers), or of the electrical generating industry (whose alleged externalities are like mirages in the desert; they vanish before one gets close enough to touch them).

In this vein, it is easy for me to motivate a tax on noxious emissions, regardless of their source, and then end up exempting some sources because it would be too costly to try to administer a tax striking them. Likewise, it is easy to motivate a general VAT and then end up accepting one of less than full generality for a variety of administrative and pragmatic reasons. But it is very difficult

---

which a rise in the real exchange rate leads to an increase in imports of certain goods. Suppose woolen suits are both imported and made at home; assume, too, that the wool used in the domestically made suits is also imported, and that this is the principal purpose for which imported wool is used. Now let the real exchange rate rise. Less suits are imported and more made at home; because of this, imports of wool for use in making suits will also increase.

4. Note that these additional goals are quite different from “political pressures and debts, marriages of convenience, sacred cows, and taboos,” precisely because they can claim a certain generality across time and space, and a purpose of sufficient merit to justify their inclusion in our professional literature and curricula. I for one find it absurd to think of teaching young economists how to help Ferdinand Marcos raid the Philippine treasury, or to aid Anastasio Somoza as he used economic policy to distribute largesse to his friends. By the same token, but to a lesser degree, I do not feel we are at the stage where it is appropriate to teach in economics courses, or to debate in our literature, how we can join with a political candidate as he makes rash promises during a campaign, or with an incumbent as he rewards his friends and neglects (if he does not outright punish) his enemies. I am old-fashioned enough to feel that what we should write about, and teach students about, is how to assess, measure, and articulate the economic costs that such actions typically entail.

to find reasons for taxation that lead to tax boundaries that are coterminous with those of one or of a few standard industrial classifications.

A good rule to follow is that, where public safety or the protection of the public is concerned, it is reasonable to key in those motives in defining the boundaries of legislation. Where externalities are thought to exist, it is reasonable to try to measure or approximate those externalities and to implement legislation that seeks to recognize and reflect the relevant magnitudes (e.g., pollution). Where subsidies or tax incentives are justified in terms of externalities that are deemed to promote economic development, it is particularly important to try to pinpoint the assets or activities involved. I have never seen an argument that would justify subsidizing, say, the purchase or rental of buildings or vehicles on grounds of some supposed externality in, say, the drug or computer-chip industry. It behooves those who propose subsidies of this kind to demonstrate carefully and convincingly the approximate size and approximate location of the externalities in question. In brief, an “industry” label is no better at defining the appropriate boundaries of a subsidy or incentive scheme based on externalities than it is for most other policy purposes.

What is involved in the tax area is, quite generally, the weighing of the goal of efficiency against other objectives with some reasonable claim to legitimacy. For example, a VAT at a uniform rate will rarely be progressive, in the sense of falling with a higher average rate on the more well-to-do. A VAT can be made somewhat progressive, however, by applying higher rates to certain commodities bought principally by middle- and upper-income groups. Here one faces a very interesting and challenging problem—of creating, say, three large composite goods, A, B, and C, where A represents basic commodities taxed at the rate of 10%, and C represents “luxury” goods taxed at 30%, while B contains the rest of eligible commodities and is taxed at 20%. Ideally we would like these three composite goods to have relatively low elasticities of demand, so as to keep the corresponding triangles of efficiency cost relatively small. But it would be absurd to approach this problem with only that aim in view, because it would dictate lumping together, say, all refrigerators, all automobiles, all jewelry into category C, and all clothing and food into category A, so as to minimize intragroup substitution. A more suitable grouping, given the objective at hand, would be to put expensive cars, refrigerators, TV sets, suits, dresses, jewelry, and restaurant meals into category C, and to put cheap items of the same types of goods in category A, with the rest going into category B. The trick is to try to see to it that most of the relevant substitutes for items in each group are also in the same group, and that between-group substitution mainly takes place between neighboring groups.

The result of this process does not mathematically minimize efficiency cost, but certainly gives important weight to efficiency as an objective, while striving for a degree of progressivity in the indirect tax structure.

**9.6 Arguments against Ramsey Taxation**

There has been an enormous revival in the last two decades of interest in Frank Ramsey's famous problem: minimizing the efficiency cost of raising a given amount of revenue via proportional excise taxes on a subset of, say,  $k$  out of a total number  $n$  of goods and services. In the field of tax analysis the Ramsey problem is something of a bombshell because it decisively demolishes any claim that a uniform tax on the  $k$  commodities in the subset is likely to be the efficiency-cost-minimizing solution.

If we assume a reduced form system, with taxes  $T$  as policy variables and equilibrium quantities being  $X$ , we can describe the system as

$$(1) \quad X_i = X_i^0 + \sum_{j=1}^n R_{ij}T_j \quad (j = 1, 2, \dots, n).$$

The welfare or efficiency cost of the system of taxes can be written as

$$(2) \quad WC = -\frac{1}{2} \sum_i \sum_j R_{ij}T_iT_j = -\frac{1}{2} \sum_i T_i \Delta X_i,$$

where  $\Delta X_i$  is equal to  $(X_i - X_i^0)$ . This assumes the reaction coefficients  $R_{ij}$  are constant over the relevant range. The Ramsey problem can be stated as

$$(3) \quad \text{Minimize } [-\frac{1}{2} \sum_i \sum_j R_{ij}T_iT_j],$$

subject to  $\sum_{i=1}^k T_i X_i = Y$ , a constant, and with  $T_i = 0$  for  $i > k$ . This works out as follows:

$$(4) \quad \frac{\delta}{\delta T_1} \left\{ \left[ -\frac{1}{2} \sum_i \sum_j R_{ij}T_iT_j \right] - \lambda \left[ \sum_{j=1}^k T_j X_j - Y \right] \right\} = 0;$$

$$(5) \quad -\sum_{j=1}^k R_{1j} \hat{T}_j - \lambda \sum_{j=1}^k \hat{T}_j R_{j1} = \lambda X_1.$$

Here the  $\hat{T}$  denote the efficiency-cost-minimizing values of  $T_j$ . From (5) we get, using the symmetry property  $R_{ij} = R_{ji}$ ,

$$(6) \quad \lambda X_1 = -(1 + \lambda) \sum_j R_{1j} \hat{T}_j;$$

$$X_1 = -\Delta X_1 (1 + \lambda)/\lambda.$$

In general, the Ramsey result is

$$(7) \quad \Delta X_i / X_i = -\mu \quad \text{for } i = 1, 2, \dots, k.$$

If the system is linear in the relevant range, the criterion for minimum efficiency cost is that the quantities of all goods and services in the taxed subset ( $i = 1, 2, \dots, k$ ) should all shrink by the same proportion. If the goods in the subset are neither substitutes or complements to each other (in the general-

equilibrium sense), so that  $R_{ij} = 0$  for  $i$  and  $j$  within the subset, then we have the familiar textbook result that efficiency-cost minimization is achieved when the taxes on goods 1 through  $k$  are set at rates that are inversely proportional to their elasticities of response  $R_{ij}/X_i$ . This is not the Ramsey solution in the general linear case, however. For that case, the best summary characterization is the one already given—that the quantities of all goods and services in the subset should shrink by the same amount.

I have always been troubled by this attribute of the Ramsey solution. Obviously, I am not implying that it might be wrong in an analytical sense, but rather that it could be troubling to those who would thrust it upon a society on the grounds that it was the best achievable result.

I feel that the most decisive revelation concerning the Ramsey solution is that it puts the state in the role of a sort of discriminating monopolist vis-à-vis its own citizens. This can be seen by recognizing that the problem of minimizing the efficiency cost of raising a given amount of revenue from goods 1, 2, . . . ,  $k$  is essentially the same as the problem of maximizing the revenue that can be raised from this subset of goods, while limiting the efficiency cost to a given amount. Obviously a different vector of taxes will be associated with each given amount of efficiency cost. But each of these tax vectors will have the characteristic that (for the range in which the  $R_{ij}$  are constant) the resulting reduction of the equilibrium quantities of the affected  $X_i$  will all be the same.

Now we already know that for the case of linear demand curves, constant costs, and independent markets, a discriminating monopolist will find his profit-maximizing equilibrium at the point where the quantity in each such market is just half the undistorted equilibrium quantity. The monopoly markup will be greater in the markets with more inelastic demand, and lower in those with more elastic demand. Indeed, the markup will end up being inversely proportional to the elasticity of demand in each such market—exactly the result for the tax rates in the corresponding Ramsey problem.

To see the relationship most clearly, consider the fact that the discriminating monopolist does not care (presumably) about the efficiency cost that he imposes. Thus, his solution is the unconstrained revenue maximum. It is one member of the family of solutions to the problem of maximizing revenue from a subset of goods, but subject to an efficiency cost constraint.

The final answer to the relationship between the Ramsey tax solution and the discriminating monopoly solution is as follows (for the linear case): the two solutions are identical in the case where the government is asked to raise (from goods 1 through  $k$ ) the maximum maximum of revenue. Where the government is asked to keep efficiency cost below a target level that is less than the one corresponding to the maximum maximum of revenue, then the government gets less revenue than the discriminating monopolist would. Correspondingly, instead of shrinking all affected quantities ( $X_1, X_2, \dots, X_k$ ) by one half, the government shrinks them by some smaller uniform fraction. It acts just “like” a discriminating monopolist, but exercises only a specific frac-

tion of its monopoly power. The fraction is, however, the same for all affected markets. As the efficiency cost constraint facing the government is loosened, the Ramsey solution moves *pari passu* in all affected commodities toward the discriminating monopoly solution.

In my view, this connection helps take some of the magic, some of the attractiveness, out of the Ramsey solution. It is quite fair to ask whether governments (or “societies”) would or should feel at ease in so ordering their taxes as in effect to “exploit” the tastes (or supply constraints) of their citizens. The inequality of rates that emerges from the Ramsey problem turns out to derive precisely from such exploitation. When I ask, “Is this what we want?” I am not trying to demean the quest for lower efficiency costs. But at least in my view, the link to monopolistic exploitation takes some of the sheen off the Ramsey solution and once again (by indirection) brings us back to issues of fairness and nondiscrimination.

### **9.7 Ramsey Analysis Does Not “Justify” a System of Progressive Commodity Taxes**

Writing this section is a somewhat delicate task for me, since it is so easy to be misunderstood. Let me begin by setting out three beliefs that I have held for a long time: (1) a very strong case can be made for value-added taxation over a broad base at a uniform rate; (2) if we deviate from uniformity, we should not do so just to add a last touch of polish or elegance to the tax system, but only to fulfill a serious “need” for nonuniformity in order to reflect the particular society’s own sense of values or of fairness; and (3) one pattern of deviation that in some cases will meet the criterion set out in 2 is indeed the sort of moderately progressive tax system discussed in section 9.5.

So I come here not as an advocate, *peremptorily* arguing against the idea of a progressive system of indirect taxes, but rather as a teacher, pleading with my readers not to try to defend such a system with the wrong line of argument.

That is what I feel has been widely done; even more, it is a conclusion to which economists might quite naturally jump, after a moderate exposure to the idea of Ramsey taxation. Let me take you down what I see as the primrose path that leads to error, and then identify the source of the error. The starting point is the vision of a completely general tax that in principle would have zero efficiency costs. A very simple model reflecting this vision would have labor as the only factor of production, and would go on to postulate a zero-elastic supply of labor.<sup>5</sup>

5. Such a model is not in principle as far from reality as it may at first look. If the scheme used to implement the resulting tax or taxes is a VAT system, one can design the scheme so as effectively to exempt the capital factor from taxation. This can be done by employing a VAT of consumption type, which can be accomplished most easily through the credit method of administration. Under this method each firm pays tax (at the requisite rate or rates) on its total sales, then deducts the VATs that it paid on all its purchases. The consumption type of tax entails allowing such deductions

In such a system there are two equivalent taxes that are completely neutral (i.e., have zero efficiency cost). One is a flat-rate tax on all labor (which strikes the income of people as they earn it), the other a flat-rate tax on all purchases of consumption goods and services (which strikes the income of people as they spend it). The latter tax can be implemented via a VAT of the consumption type.

Now we modify the original problem by releasing the assumption of zero-elastic labor supply. Our uniform tax now covers only  $n - 1$  goods and services, the  $n^{\text{th}}$  (and untaxed) one being leisure. We also know that this uniform tax will not typically be neutral, since it distorts the choice between labor and leisure and, more generally, the choices between leisure on the one hand and each of the remaining  $n - 1$  goods and services on the other.

If we are precluded from taxing leisure, the Ramsey solution (for a linear system) entails imposing a set of  $\hat{T}_i$  such that the quantities of the  $n - 1$  goods and services other than leisure all contract by the same proportion. This entails placing higher taxes on goods that are complementary to leisure, or that are less-than-average substitutes for leisure, and placing lower taxes on those that are better-than-average substitutes for leisure.

This result can be seen intuitively. Suppose there existed one good, say  $X_4$ , which was used only in fixed proportions with leisure. Then the fact that we are precluded from taxing leisure would be no constraint at all. The pair of goods ( $X_4$  and leisure) would be like left and right shoes, and we could work on the principle that a tax of \$2 per pair of shoes can be replicated either by a tax of \$2 per right shoe or by a tax of \$2 per left shoe. Applied to our case, leisure would be like the right shoe,  $X_4$  like the left one. We know that a uniform tax on all goods including leisure would be neutral. We hypothetically create such a tax, but then, being precluded from directly taxing leisure, we load the total tax on the pair ( $X_4$  plus  $X_n$ ) onto  $X_4$  alone. Because of the rigidity of proportions between  $X_4$  and  $X_n$ , this modified tax system is equivalent to the uniform one, and hence is also neutral.

If we can't find a target with completely fixed proportions to leisure, then things with close-to-fixed proportions can serve as surrogates. The Ramsey

---

for capital as well as current outlays, so that no distinction need be made between the two; for this reason it is easier to administer than a VAT of the income type (in which capital assets purchases are not immediately deductible but must be capitalized and then depreciated over time, for tax purposes) or one of the product type (in which capital assets are neither expensed nor depreciated, but in which they nonetheless must be distinguished from current inputs, which *are* expensed). When capital assets are expensed (as occurs with the consumption type of tax), one can say that the capital factor is thus freed of tax. This is self-evident for a constant rate ( $t$ ) of VAT. The firm pays  $t$  on the service yield ( $S$ ) of capital assets as it accrues over time, but pays  $-t$  on the acquisition cost ( $C$ ). Considering that at the firm's own rate of discount the present value of the service yield (PVS) will tend to equal the present value of acquisition cost (PVC), it becomes clear that no net tax payment is involved. The government in effect "invests"  $tC$  and takes a return of  $tS$ ; it thus ends up as a  $t\%$  partner in each capital asset so treated, rather than as a taxing agent.

solution says that goods whose quantities increase least when the price of leisure rises should be taxed at the highest rates, and vice versa. This is the source of the notion that the Ramsey solution tells us to tax more heavily those goods that are complements to, or less-than-average substitutes for, leisure.

So far we have traveled a considerable distance without getting into trouble. The trouble appears right at this point, as we face the question of which goods are complementary to leisure, which are good substitutes, and which are poor substitutes. For something close to twenty years now, I have performed an experiment with my public finance classes and with occasional seminar audiences. Let us make a list, I suggest, of goods that are complements to leisure, and I then receive suggestions from the floor. "Television sets," says one. "Movies," says another. "Vacation trips," says a third. Then comes a whole barrage: sporting goods, restaurant meals, concerts, night clubs, summer houses, leisure-time clothes, and so forth.

When the list gets long enough, I tell my students that they really should do better than that—they are not really thinking like economists. To think like an economist about this problem one should pose a scenario that will automatically reveal the answer.

I have two such scenarios. They both start with a full equilibrium, where everybody is working his or her desired number of hours at the going wage. In the first scenario, this equilibrium is disturbed by a government decree, requiring each of them to work one or two hours less, at the same wage as before. This move entails no first-order income effect. Each person has less money income, counterbalanced by more leisure. Their money income may go down from \$500 to \$475, and we have to ask, which goods will see an increase in demand (these will be complements to leisure), which goods will see a small decrease (these will be poor substitutes), and which goods will see a big decrease (these will be good substitutes for leisure).

The best course at this point is to follow the advice of the Austrian school, and simply look inside ourselves for the answer (introspection, they call it). My answer is that if I were faced with such an arbitrary cut in my working hours and in my money income, I would react much the same as if I had the same cut in income (say from an increase in taxes), without the drop in working hours. The result is not exactly the same because the "time constraint" behaves differently in the two cases. But I certainly would allow my behavior to be guided more by a 5% cut in money income than by two extra hours of leisure. Following this line of reasoning I conclude that the cross-elasticities of demand for different goods and services with respect to changes in the price of leisure ( $\eta_{lm}$ ) are not exactly proportional but are close to proportional to the income elasticities of demand ( $\sigma_i$ ) of those items. Exact proportionality prevails if people pay no attention at all to the change in the time constraint.

Once this example has been presented, my audiences over the years have overwhelmingly agreed with me that the real complements to leisure are likely

to be the inferior goods (those with  $\sigma_i < 0$ ), that the strong substitutes for leisure are strong luxury goods (those with  $\sigma_i > 1$ ), and that the weak substitutes for leisure are probably "ordinary necessities" (with  $1 > \sigma_i > 0$ ). These audiences have certainly recognized the absurdity of their original line of thinking about the problem, which departs from the premise that anything you use in your leisure time is a complement to leisure. That premise, taken seriously, leads to the conclusion that everything is a complement to leisure except for overalls and lunch pails (the things you actually use while at work).

The second scenario leads to the same conclusion. In that scenario everybody starts in equilibrium at the prevailing wage, and is now offered the opportunity to work overtime at a premium wage. Everybody will respond by working more (or at worst the same), because, once again, no first-order income effect is involved. Do you think they will take *fewer* vacation trips, go to *fewer* movies, buy *fewer* restaurant meals, own *fewer* summer houses, and so forth, as a consequence of working a few more hours at overtime rates? Much more plausible is the result that they spend the extra income in much the same way as they would spend the proceeds of a reduction in taxes or of an annuity received as an inheritance.

One final shot on the point being made here. I am asserting that the  $\eta_{in}$  (cross-elasticities of demand for goods with respect to the price of leisure) are approximately proportional to the  $\sigma_i$  (income elasticities of demand for the respective goods). My parting shot is that we had better all hope it is that way, because the things we measure and label as income elasticities are in reality complex jumbles in which the  $\sigma_i$  and the  $\eta_{in}$  are inextricably mixed. Consider a typical time series demand analysis. Most of the changes in real income recorded over, say, a twenty- or thirty-year period will stem from changes in real wages; another part will come from changes in income from capital. Well, the part coming from wage rate changes affects consumption of good  $i$  by both an income effect ( $\sigma_i$ ) and a substitution effect ( $\eta_{in}$ ) with respect to the price of leisure. Only the part coming from changes in capital income gives us a pure income effect.

Consider now a typical cross-section analysis of demand. Once again the main differences in income among the people (or households) represented in the analysis are likely to stem from differences in their hourly wage rates. Once again the response of demand for good  $i$  with respect to these income differences incorporates effects from both  $\sigma_i$  and  $\eta_{in}$ . Only the response of demand to differences in the capital income of the different households reflects a pure income effect.

So, if we want to *really* measure income elasticity of different goods and services, we have to treat as separate variables in our demand functions the real income stemming from labor on the one hand and the real income stemming from nonlabor income in the other. Only the coefficient of the latter gives us a pure income effect. But if the two coefficients are close together, if we



cannot distinguish very clearly between the ways small increments of the two types of income (starting from a given base) would be spent, then the data would be corroborating the point that I am making here.

The end of this long lesson is that the Ramsey analysis would lead us to tax most highly the inferior goods, next most highly the ordinary necessities, and least highly of all, the luxuries. As I survey my students and colleagues, looking for supporters of this package as serious, real-world legislation, I find an empty set. Moral of the story: please do not propose a progressive structure of excise taxes, and invoke the Ramsey rule as your reason!

## 9.8 More on Drawing Boundaries and on Evasion

The theme of this section is welfare triangles. The general-equilibrium version of the textbook triangle ( $-1/2T_1\Delta X_1$ ) is the “generalized triangle”  $-1/2\sum_i T_i\Delta X_i$ , where  $\Delta X_i$  is measured from the undistorted equilibrium ( $X_1^0, X_2^0, \dots, X_n^0$ ) to the full equilibrium ( $X_i$ ) with all the  $T_i$  in place. As indicated in section 9.5, the trick in setting boundaries is to put close substitutes together within a classification, so as to tax them all at a single rate. By treating packages of close substitutes as composite commodities, we eliminate tax-induced substitution among them; that is, we eliminate what we can call within-group substitution. What we are left with, if we have several different groupings of commodities, is between-group substitution.

Consider that we have a tax  $T_3$  on four-door sedans,  $X_3$  only. The welfare triangle  $-1/2T_3\Delta X_3$  will have an enormous base, compared with what is feasible through good boundary drawing. The base  $\Delta X_3$  would include substitution between four-door sedans and two-door sedans, between four-door sedans and coupes, convertibles, station wagons, and so forth, as well as between four-door sedans and motorcycles, between four-door sedans and panel trucks, between four-door sedans and other trucks, and finally between four-door sedans and everything else.

The tax  $T_3$  thus defined qualifies soundly as an utterly stupid tax. It is slightly improved by setting  $T_4$  on two-door sedans at the same rate, and putting  $T_5, T_6$ , and  $T_7$  on coupes, convertibles, and station wagons, also at the same rate. At the end of this process we have a tax  $T_a$  on all automobiles, which if at the same rate raises much more money than did  $T_3$ , and which if set to raise the same amount of money as  $T_3$  can do so at a much lower efficiency cost (because of a greatly lower rate and because the elasticity of demand for automobiles is very much less than for sedans).

But that is not the whole story by any means. I mentioned panel trucks and motorcycles for a reason, for we know from experience that very interesting, very curious things can happen to these vehicles if automobiles are very heavily taxed. I recall some decades ago when Chile had very heavy taxes (they were actually tariffs) on cars, but much lower ones on panel trucks. The result was a vast increase in the importation of panel trucks, which was spawned

by an entirely new Chilean industry—that of “converting” panel trucks into something that looked like and served as a car. The panels were invariably knocked out and replaced by glass windows. Upholstered seats were installed in the back. Sometimes new doors were added, sometimes not.

In the case of motorcycles, my example comes from Indonesia. Again the story starts with a heavy tax on imports of cars, with a much lower (maybe even zero) tax on motorcycles. But here it was the motorcycles that underwent conversion. Three-wheel cycles were converted, by artful additions, into virtual buses, or at least taxis. Sometimes a single bench was added, with the passenger looking backward. Other times the cycle was stretched at the back, with two benches going down each side, and maybe even with an extra little running board cutting laterally across the rear (where the rear bumper of a car would be). I must say I was truly astounded when I saw my first eight-passenger motorcycle, but I came to relish the experience.

I hope that readers can see the close analogy between these newly spawned conversion industries and what we have come to regard as rent-seeking activity. The difference between the tax rates on cars and trucks, or between those on cars and motorcycles, generates the use of real resources (1) to take advantage of a potential economic rent generated by a misguided law, (2) to avoid paying the tax on cars, and (3) to reflect the same kind of tax-induced substitution that always takes place when given items are taxed and actual or potential substitutes are taxed either at lower rates or not at all. If the above were a multiple-choice question, my answer would be “all of the above.” Answers 1, 2, and 3 are different descriptions of the same phenomenon. The curious behavior of the panel-truck converters in Chile and of the motorcycle converters in Indonesia is in principle just as natural an economic consequence of taxation as it is for people to drink more beer when the tax on wine is raised. But surely resources were in some sense wasted in the process; efficiency costs would be dramatically reduced if the boundaries of the automobile tax had been so drawn as to include converted panel trucks and converted motorcycles, if not all trucks and all motorcycles as well.

This brings me to the topic of tax evasion. This is just another way in which the bases  $\Delta X_i$  of the generalized triangle  $-1/2 \sum_i T_i \Delta X_i$  end up being made bigger. And it is quite analogous to the reconversion activities referred to above. They would be called *tax-evasion activities* if they were illegal; as it is they can be fairly labeled *tax-avoidance activities*. But is there any real difference between converting a motorcycle into a “taxi” using real resources so as to *avoid* the tax on cars, compared with using real resources to smuggle TV sets or whiskey into a country so as to *evade* its tax on these items? And, going one step further, is there any real difference between using real resources to smuggle whiskey on the one hand, and using real resources to hide certain receipts from the income tax collector? I for one think that in none of these cases is there any real difference.

Just as the welfare triangle associated with the tariff on cars is made bigger

by the existence of the “conversion” activities (whether of the Chilean or of the Indonesian type), so too is the welfare triangle for imported whiskey made larger by the fact of smuggling. If in the base of the whiskey triangle we have substitution toward gin, toward rum, toward wine, toward beer, and toward all other things, we also have to insert substitution away from taxed whiskey and toward untaxed (i.e., smuggled) whiskey.  $\Delta X_{13}$  (if that is whiskey) is *not* the difference between whiskey consumption in the undistorted equilibrium and whiskey consumption in the presence of the tax. No, it is instead the difference between whiskey consumption in the undistorted equilibrium and “taxed-whiskey consumption” in the presence of the tax.

Out of this comes an insight that I think of as reasonably penetrating: when we are analyzing taxes, the definition of a commodity is what the tax law itself, *plus its administration*, determines it to be. Just as the size of the triangle’s base can be greatly affected *via* bad or good drawing of boundaries (i.e., definition of what is being taxed), so too it can be greatly affected by the ease of administration and by its effectiveness. Ease of administration largely depends on how the law is written; effectiveness of administration depends on how much effort and ingenuity the administrators put in and, of course, on how honest they are.

One final point: once we recognize that evasion is “just another substitute commodity,” we can see that it is absolutely natural and normal that it should increase with every notch by which the tax rate itself is raised. It is just the same as people drinking more beer as the tax in whiskey goes from 40% to 60%, and drinking still more beer when it goes from 60% to 80%. At the margin, evasion efforts will be pursued until the extra cost of evading the tax on another unit of the commodity is itself just equal to the tax rate.

This helps explain the behavior of Argentine taxpayers described in section 9.2—how with every increase of the tax rate imposed by the authorities there was a new burst of additional evasion. That is something that is perfectly natural, perfectly predictable, perfectly understandable in economic terms.

How then can tax authorities cope with this type of natural phenomenon? I think the answer is overkill. If the penalties for evasion are high enough, most people will be “honest” even if the tax does go up from 40% to 60% to 80%. By raising the costs of evasion high enough, we may be able to drive people into that corner solution called “honesty.” This is an important and serious message, and one that lies at the heart of needed tax and administrative reforms in many developing countries.

## References

- Atkinson, Anthony B., and Joseph E. Stiglitz. 1980. *Lectures on public economics*. New York: McGraw-Hill.

- Boskin, Michael J., and Charles McLure, eds. 1990. *World tax reform: Case studies of developed and developing countries*. San Francisco: Institute for Contemporary Studies.
- Due, John F. 1990. Some unresolved issues in design and implementation of value-added taxes. *National Tax Journal* 42 (December).
- Gillis, Malcolm, ed. 1989. *Tax reform in developing countries*. Durham, NC: Duke University Press.
- Ito, Takatoshi, and Anne O. Krueger. 1992. *The political economy of tax reform*. Chicago: University of Chicago Press.
- Newbery, D., and Nicholas Stern, eds. 1989. *The theory of taxation for developing countries*. London: Oxford University Press.
- Tait, Alan A. 1988. *Value-added tax*. Washington, DC: International Monetary Fund.
- Tanzi, Vito, ed. 1990. *Public finance, trade, and development*. Detroit: Wayne State University Press.
- World Bank. 1988. *World development report, 1988*. Washington, DC: World Bank.

This Page Intentionally Left Blank