Chapter 22

TAX AVOIDANCE, EVASION, AND ADMINISTRATION*

JOEL SLEMROD

The University of Michigan

SHLOMO YITZHAKI

The Hebrew University of Jerusalem

Contents

* We are grateful to Wojciech Kopczuk for valuable research assistance. Helpful comments on an earlier draft were received from Jim Alm, Alan Auerbach, Len Burman, Brian Erard, Martin Feldstein, Firouz Gahvari, Roger Gordon, Jim Hines, Jonathan Kesselman, Louis Kaplow, Jim Poterba, Agnar Sandmo, Dan Shaviro, Eric Toder, Gideon Yaniv, the participants at the NYU Law School Colloquium on Tax Policy and Public Finance and the handbook conference held at the University of California at Berkeley.

Handbook of Public Economics, Volume 3, Edited by A.J. Auerbach and M. Feldstein

Abstract

Tax avoidance and evasion are pervasive in all countries, and tax structures are undoubtedly skewed by this reality. Standard models of taxation and their conclusions must reflect these realities.

This paper first presents theoretical models that integrate avoidance and evasion into the overall decision problem faced by individuals. Early models of this area focused on tax evasion, modeled as a gamble against the enforcement capability of the state. More recently, the literature has examined more general models of the technology of avoidance, with the additional risk bearing caused by tax evasion either being a special case of this technology or one aspect of the cost of changing behavior to reduce tax liability. If the cost of evasion and avoidance depends on other aspects of behavior, the choice of consumption basket and avoidance become intertwined. The paper then relates the behavior predicted by the model to what is known empirically about the extent of evasion and avoidance, and how it responds to tax enforcement policy.

The paper then turns to normative analysis, and discusses how avoidance and evasion affect the analysis of vertical and horizontal equity as well as efficiency costs; a taxonomy of efficiency costs is presented. Acknowledging the variety of behavioral responses to taxation changes the answers to traditional subjects of inquiry, such as incidence, optimal progressivity, and the optimal mix between income and consumption taxes. It also raises a whole new set of policy questions, such as the appropriate level of resources to devote to administration and enforcement, and how those resources should be deployed. Because there are a variety of policy instruments that can affect the magnitude and nature of avoidance and evasion response, the elasticity of behavioral response is itself a policy instrument, to be chosen optimally.

The paper reviews what is known about these issues, and introduces a general theory of optimal tax systems, in which tax rates and bases are chosen simultaneously with the administrative and enforcement regimes. We argue that the concept of the marginal efficiency cost of funds is a useful way to summarize the normative issues that arise, and expand the concept to include administrative costs, avoidance, and evasion.

Keywords

taxation, evasion, enforcement, avoidance

JEL classification: H2

1. Introduction

1.1. Why avoidance, evasion and administration are central, not peripheral, concepts in public finance

Most economic analysis of taxation presumes that tax liability can be ascertained and collected costlessly. As a description of reality, this is patently untrue. For example, in the U.S. the Internal Revenue Service (henceforth IRS) estimates that about 17% of income tax liability is not paid [1]; the figure for most other countries is probably higher. Furthermore, the resource cost of collecting what is paid can be large, in the U.S. probably about 10% of tax collections [2]. The tax structures themselves are undoubtedly skewed by the realities of tax evasion, avoidance, and administrative costs.

The standard models of taxation and their conclusions need to be modified in the light of these realities. Many practitioners of tax advice in developing countries believe that this change in emphasis is essential; for example, Casanegra de Jantscher (1990, p. 179) goes so far as to say that, in developing countries, "tax administration *is* tax policy" [3]. Bird (1983), Mansfield (1988), and Tanzi and Pellechio (1997) are useful summaries of the practical problems of the interaction of tax policy and tax administration in this context.

We believe that these issues are also critical in developed countries. In this setting, the issue is not the feasibility of certain taxes, but rather the optimality of alternative tax structures. For example, while in many developing countries an income tax that relies on self-reporting cannot be administered at all, in a developed country the question is to what extent optimal tax design should reflect the reality of evasion, the necessity of enforcement, and the costs of collection. In fact, tax systems do reflect these issues, although there is little systematic guidance offered by the academic public finance literature. The objective of this chapter is to collect and critique the now sizable literature that addresses these questions.

1.2. The evolution of tax structures

Scholars of the historical evolution of tax structure, notably Hinrichs (1966) and Musgrave (1969), have also stressed the importance of tax administration issues. They note that modern tax structure development has generally been characterized by a shift from excise, customs, and property taxes to corporate income and progressive

[1] Internal Revenue Service (1996).

[2] Slemrod (1996a).

[3] Others disagree. Groves (1974, p. 25) offers that: "Vetoing tax measures because of the difficulty of administering them is in most cases less compelling than doing so on the ground of their failure to conform to acceptable principles. Administration is usually amenable to improvement where violation of first principles is not. And administration of a given tax may often be improved most effectively in the process of attempting to administer it. The point is sometimes crucial in recommending taxes for so-called underdeveloped economies in our own time".

individual income taxes[4]. This shift has been made possible by the expansion of the market sector and relative decline of the rural sector, the concentration of employment in larger establishments, and the growing literacy of the population. Further changes in the technology of tax administration, including globalization and financial innovation, may now be pushing us away from progressive income taxes toward tax systems that rely more on broad-based consumption taxes such as the value-added tax (VAT), flatter rate structures for income taxation, or the "dual income tax" system recently adopted by certain Scandinavian countries, and described in Sørensen (1994).

Alt's (1983) treatment of the evolution of tax structure stresses the role of administrative and compliance costs. He argues that it has become increasingly easy to collect taxes from organized business rather than from households, and that one explanation for the widespread adoption of the VAT is that it imposes compliance costs without raising administrative costs, through incentives for self-policing. Kau and Rubin (1981) focus on changes in the cost of collecting taxes, and successfully relate growth of the U.S. federal government to reasonable correlates of collection cost, such as the literacy rate, the extent of female labor force participation, and the extent of the agricultural sector. Balke and Gardner (1991) contend that declining marginal collection costs can explain the stepwise growth in the size of government and the changes of taxation observed in the U.S. and U.K. They argue that major wars coincide with permanent improvements in tax instruments and tax collection technology, which facilitated permanent expansions in government size thereafter.

Putting aside the role of administrative issues in explaining the evolution of tax levels and tax structures, it is indisputable that these considerations are critical determinants of tax policy at a point in time. For example, an important set of generic aspects of income tax structure, such as the absence of taxation of imputed rents from consumer durables, taxation of capital gains (if at all) on a realization basis, and pre-set depreciation schedules, are undoubtedly largely driven by practical concerns of administerability. For these reasons, we believe that consideration of evasion, avoidance, and administration is essential to the positive and normative analysis of taxation. Our view corresponds closely to that of Blough (1952, p. 146):

> It is tax policy in action, not simply the wording of the statute, that determines how much the taxpayer must pay, and the effects of the payment. Knowledge of the statute is only a start in knowing a tax system. The interpretations placed on language by administrators and courts, the simplicity and understandability of tax forms, the competence and completeness of audit, the vigor and impartiality of enforcement, and the promptness and finality of action all influence the amount of revenue collected, the distribution of the tax load, and the economic effects of the tax.

[4] Although Hinrichs (1966) points out that tax structure development began with direct taxes rather than indirect taxes.

In this chapter we organize, explicate, and evaluate the modern literature that incorporates these considerations into the economics of taxation[5]. We do not claim to have put together a comprehensive survey of this literature, which is huge and multi-faceted, rather a guide to what we feel are the most important issues and contributions in this area.

1.3. Evasion, avoidance, and real substitution response

We begin with some definitions. The classic distinction between avoidance and evasion is due to Oliver Wendell Holmes, who wrote

> When the law draws a line, a case is on one side of it or the other, and if on the safe side is none the worse legally that a party has availed himself to the full of what the law permits. When an act is condemned as evasion, what is meant is that it is on the wrong side of the line ... *[Bullen v. Wisconsin (1916), 240. U.S. 625 at p. 630].*

Thus, the distinguishing characteristic of evasion is illegality[6]. In practice, of course, there are many gray areas where the dividing line is not clear, and sometimes the tax authorities may inappropriately characterize particular cases. One can draw a further distinction within the class of legal responses to taxation. At times we will refer to real substitution responses, or real responses for short, as those responses which come about because the tax law changes the relative price of different activities, and that induce taxpayers to respond by choosing a different consumption basket.

Conceptually distinct from real substitution responses are efforts to reduce one's tax liability without altering one's consumption basket, which we will refer to as avoidance. These are actions taken in response to the tax system that do not involve shifts along a given budget set. This definition covers a broad range of behaviors. One example is to pay a tax professional to alert one to the tax deductibility of activities already undertaken. Another example is to change the legal form of a given behavior, such as reorganizing a business from a C corporation to an S corporation[7], recharacterizing ordinary income as capital gain, or renaming a consumer loan as a home equity loan. A third example is tax arbitrage, when economically equivalent, but differentially-taxed, positions are held simultaneously long and short, producing tax savings. Finally, retiming a transaction to alter the tax year it falls under is an example of avoidance.

[5] Other surveys of these issues include Cowell (1990b), Andreoni, Erard and Feinstein (1998), Roth, Scholz and Witte (1989), and Alm (1999). Because of space constraints we have omitted any discussion of the role of tax practitioners, corruption in tax administration, tax amnesties, or business tax evasion.

[6] Kay (1980, p. 136) offers a different pair of definitions for evasion and avoidance: "Evasion is concerned with concealing or misrepresenting the nature of a transaction; when avoidance takes place the facts of the transaction are admitted but they have been arranged in such a way that the resulting tax treatment differs from that intended by the relevant legislation".

[7] Under U.S. tax law an S corporation retains the legal characteristics of a corporation, but is taxed as a pass-through entity such as a partnership. There are restrictions to becoming an S corporation, most notably on the maximum number of shareholders.

Fine distinction among the types of behavioral response to taxation is not possible and is for many issues not crucial. In general, changes in the tax structure will induce all the different kinds of response. Indeed, one of the goals of this chapter is to emphasize the common analytical aspects of issues that have traditionally been kept distinct.

1.4. General framework

Although there may be reasons, discussed later, for distinguishing among these categories of response to taxation, there is a common framework for analyzing these issues. Given the structure of the tax system and enforcement process, taxpayers are faced with opportunities to reduce their tax payments, or expected tax payments. There is a private cost to taking advantage of these opportunities, which may take the form of an altered consumption basket, an increasing probability of detection of, and penalty for, evasion, and/or a real resource cost of effecting avoidance or concealing evasion. This private cost depends on policies of the government that include, but are not limited to, the setting of tax rates and bases. The parameters of the tax administration and enforcement policies also matter, but these policies themselves are usually costly.

The tax system establishes the relative prices among this broad set of taxpayer activities. In the standard model, it establishes the relative price of leisure and other goods, as well as the relative price among the set of goods. In a more general framework it also sets the price of "honesty", meaning the incentives to evade, and establishes the cost and reward to legally reducing taxes via avoidance activities. The dimensions of taxpayer response interact. For example, real behavior may alter the cost of avoidance or evasion, thus changing the effective prices of real activities.

Although these are common themes, the literature to date has tended to isolate pieces of the overall problem. We follow that practice here, by beginning in Section 2 with a discussion of the now standard economic model of tax evasion[8]. Then, in Section 3, we introduce models that apply more generally to both evasion and avoidance. We then look at the empirical evidence, first in Section 4 about evasion, and then in Section 5 on avoidance. The remainder of the chapter addresses the implication for tax analysis of introducing these issues. Section 6 examines the fundamental issues of positive tax analysis, while Section 7 addresses normative issues. Section 8 concludes.

2. Theoretical models of evasion

2.1. The Allingham–Sandmo–Yitzhaki model

Suppose that the true tax base is known to the taxpayer, but is not costlessly observable by the tax collection agency. Then, under certain circumstances, the

[8] There is a vast literature which investigates non-economic perspectives on tax evasion. We do not have the space to discuss or evaluate this literature, and refer the interested reader to Roth, Scholz and Witte (1989).

taxpayer may be tempted to report a taxable income below the true value. In the seminal formulation of Allingham and Sandmo (1972) (henceforth A–S)[9], what might deter an individual from income tax evasion is a fixed probability (p) that any taxable income understatement will be detected and subjected to a proportional penalty (θ) over and above payment of the true tax liability itself. Later we introduce and discuss at length a "technology of evasion", in which evasion involves costs to the evader that might depend on the income and the amount of tax evaded.

In the A–S model, all real decisions, and therefore taxable income (y), are held fixed; only the taxpayer's report is chosen. The risk-averse taxpayer chooses a report (x), and thus an amount of unreported income y–x, in order to maximize expected utility:

$$\mathrm{EU} = (1-p)\,U(v+t(y-x))+pU(v-\theta(y-x)), \tag{1}$$

where v is true after-tax income, $y(1-t)$, t being the rate of (proportional) income tax. The von Neumann–Morgenstern utility function $U(\cdot)$ represents the individual's preferences toward risk. In this model the choice of whether and how much to evade is akin to a choice of whether and how much to gamble. Each dollar of taxable income understatement offers a payoff of t with probability $(1-p)$, along with a penalty of θ with probability p. If and only if the expected payoff to this gamble, $(1-p)\,t-p\theta$, is positive, every risk-averse taxpayer will chance some evasion, with the amount depending on the expected payoff and the taxpayer's risk preferences.

A critical issue, pointed out by Yitzhaki (1974), is whether the penalty for discovered evasion depends on the *income* understatement, as A–S assume, or on the *tax* understatement, as more accurately reflects practice in many countries. In the latter case, the maximand becomes $(1-p)\,U(v+t(y-x))+pU(v-\theta t(y-x))$, and the expected payoff per dollar of evaded income becomes $(1-p)\,t-p\theta t$. This is an important change, because it means that the tax rate has no effect on the terms of the tax evasion gamble; as t rises, the reward from a successful understatement of a dollar rises, but the cost of a detected understatement rises proportionately. The first-order condition for optimal evasion becomes

$$\frac{U'(y_{\mathrm{A}})}{U'(y_{\mathrm{U}})}=\frac{(1-p)}{p\theta}, \tag{2}$$

where y_{A} and y_{U} refer to net income in the audited and unaudited states of the world, respectively. Note that t does not appear in Equation (2), other than via an income effect in the definition of y_{A} and y_{U}. Compare this to the original A–S formulation, where t would be a multiplicative factor in the denominator of the right-hand side, implying that increases in t would proportionally increase the reward to getting away

[9] The Allingham–Sandmo paper applies to tax evasion the approach of the classic paper on the economics of crime by Becker (1968).

with understating income, but not proportionally increase the penalty, making evasion more attractive. Regardless of whether the penalty depends on the tax understatement or income understatement, more risk-averse individuals will, *ceteris paribus,* evade less. Individuals with higher income will evade more as long as absolute risk aversion is decreasing; whether higher-income individuals will evade more, as a fraction of income, depends on relative risk aversion. Evasion relative to income will decrease, increase or stay unchanged as a fraction of income depending on whether relative risk aversion is an increasing, decreasing, or constant function of income. Increases in either p or θ will decrease evasion.

Increasing t has both an income effect and, possibly, a substitution effect. If the taxpayer has decreasing absolute risk aversion, the income decline makes a less risky position optimal. An increase in t has a substitution effect, increasing the relative price of consumption in the audited state of the world, and thereby encouraging evasion, if the penalty is related to income, rather than tax avoided. In the latter case, if the penalty is related to the tax evaded, a tax increase has no substitution effect, so that an increase in t *reduces* evasion as long as there is decreasing relative risk aversion[10].

This simple version of the A–S model has been criticized on the grounds that it fails a simple reality check. If p is the fraction of returns audited in the U.S., about 0.015, and θ is the statutory penalty for non-criminal evasion, about 0.2, then based on the degree of risk aversion exhibited in other situations people should be evading a lot more than they apparently do. The intriguing question becomes why people *pay* taxes rather than why people *evade.* Much subsequent research, some of it surveyed below, has been addressed to reconciling the facts with the theory[11].

In the A–S model what limits the amount of evasion attempted is the taxpayer's risk aversion. At some point further evasion becomes just too big a gamble, so that at the chosen amount of evasion the marginal gain in expected tax savings is exactly offset by the marginal disutility of the extra risk taken on[12]. The model also predicts that a

[10] Note the similarity to the standard model of the effect of taxation on the optimal portfolio, in which a tax increase can increase the demand for the risky asset [Domar and Musgrave (1944)]. One difference is that, in a portfolio model, it is arguably inappropriate to ignore the effect of the tax scheme on the variability of government revenues [Gordon and Wilson (1989)]. This issue can be sidestepped in the context of a tax evasion model, because the "risks" are independent and therefore there is no social risk involved. It is important to distinguish the effect of a change in the environment on evaded income $(y-x)$ versus the impact on evaded tax liability, $t(y-x)$. With respect to changes in p and θ, there will be no interesting distinction. However, when t increases it is certainly possible that $(y-x)$ may decline at the same time $t(y-x)$ increases.

[11] One problem with this argument is that, for many types of evasion, the effective probablity of detection is much higher than the fraction of returns audited would suggest. For example, the p for non-reporting of wage and salary income subject to information reporting by employers is probably close to 1.0. Moreover, as long as several years of returns may be audited at once, the effective p may be several times higher than a one-year probabilty of audit would indicate.

[12] In the language of Kolm (1973), the evasion is accomplished by "the mere stroke of a pen". We consider below where the evasion is facilitated by supplying labor to an "underground" sector which offers better concealment possibilities.

risk-neutral individual would either remit no tax at all, or do no evasion, depending on whether evasion has a positive expected payoff. This "either–or" prediction is eliminated if the probability of detection is an increasing function of the amount of evasion, which is likely to characterize most tax systems. The implications of introducing an endogenous p depend on the precise relationship between p and evasion. For example, consider the case [discussed in Yitzhaki (1987)] where p is an increasing function of evaded income $(y-x)$. The risk-neutral taxpayer chooses x to maximize expected income,

$$\mathrm{EY} = ((1-p[y-x])(\eta+s)+p[y-x](\eta-\theta s)), \tag{3}$$

where $s \equiv t(y-x)$ is understated tax. If $p' \equiv \partial p/\partial(y-x)$ is positive, the first-order condition becomes

$$1-p-p\theta = p'(\theta+1)(s/t). \tag{4}$$

In this case, evasion will be constrained by the fact that p increases to offset what would otherwise be an increase in expected income.

The either–or prediction in the case of a risk-neutral taxpayer is also eliminated if there are distinct sources of income, each of which is subject to its own p. For example, employee labor income has a high p (due to information reporting by employers and computer matching), while "moonlighting" income has a much lower p. Faced with this situation, a risk-neutral individual reports all or none of each of the several sources of income, but may certainly report a fraction of total income [13].

The endogenous probability of detection can of course be applied to the case of a risk-averse taxpayer. In this case, at the margin the gain in expected value is offset by a combination of increased risk-bearing and an increased probability of detection. Cremer and Gahvari (1994) generalize this notion by introducing what they call a "concealment technology", which in our notation takes the form $p(y-x,((y-x)/y),m)$, where m represents taxpayer expenditure on concealment. The notion that the probability of detection can be increased by the taxpayer's expenditure is also present in Usher (1986), Kaplow (1990), Cowell (1990a), and Mayshar (1991).

2.2. *Jointness with labor supply*

Of particular interest is the relationship between the tax report decision and other consumer decisions. Most attention has been paid to labor supply, where the individual chooses how much labor to supply and how much labor income to report. The decision about how much income to report is made simultaneously with the decision of how much to work, so that it is impossible to adjust labor supply based on whether one

[13] As discussed in the next section, differential detection rates could also affect the sectoral supply of labor.

is caught evading. This problem may be posed as how much of a homogeneous labor income to report, which is equivalent to simultaneously choosing one's consumption basket and exposure to risk[14]. Models that belong to this group are based on extensions of the A–S model. Alternatively, the problem may be posed in the context of a model of the underground economy, in which there are two sectors with possibly different equilibrium wage rates and other different circumstances. The latter class of models allows for wage adjustment in response to policy changes, and thus are general equilibrium in nature.

In the extensions of the A–S model, the first-order condition for labor supply differs from that in a model without tax evasion only in that it contains mean marginal, instead of marginal, utilities. Whether mean marginal utility is bigger or smaller than the marginal utility depends on the sign of the second derivative of marginal utility, which is the sign of the third derivative of the utility function. On top of that, if utility is non-separable, the marginal utility functions depend on the sign of cross-derivatives, which further complicates the problem[15]. Baldry (1979) and Pencavel (1979) stress the difficulty of reaching any clear-cut comparative statics conclusions from such a model; the response of reported income to changes in tax rates, penalties, and fines becomes ambiguous. Thus, most models are based on particular restrictive assumptions about the utility function. For example, if the utility function is separable in consumption and leisure, then the marginal utility of leisure is independent of consumption. If, in addition, the marginal utility of consumption is linear (as in the function $U(C,L) = \alpha + \beta C + \gamma C^2 + \delta L$), the first-order condition for optimal labor supply is

$$(1-t)\,wU_1[wL+(1-p\theta)s] = U_2[L], \tag{5}$$

where s is the tax evaded and $(1-p\theta)s$ is the expected gain from evasion. Because evasion increases expected consumption for any given amount of leisure without changing the real wage, leisure would increase, and labor supply would decline. The real wage does not decline because the evasion opportunities are independent of the amount of work done. The critical importance of the relationship between the real consumption choices and the evasion or avoidance opportunities comes up again in the more general models discussed in Section 3. There we discuss cases where the avoidance opportunities do affect the real wage. In situations where labor income in the formal sector is reported by the employer to the tax enforcement agency as a matter of course, the only way to evade tax may be by "moonlighting" – working extra hours

[14] Models of this type resemble models of choice among risky occupations [e.g., Kanbur (1979)], except that in the latter the occupational choice is usually discrete, so that a "diversified" occupational portfolio is not allowed.

[15] Moreover, the conditions that the equilibrium investigated is on the increasing portion of the Laffer curve also depend on the curvature of the third derivative of the utility function, further complicating the issue.

at a different job – or by switching completely to the informal sector or "underground economy".

2.3. *Other uncertainty*

The basic model can also be extended to deal with other sources of uncertainty. Andreoni (1992) introduces a temporal nature to the tax evasion decision, recognizing the fact that the penalty for tax evasion, if detected, is assessed later than the tax saving. Andreoni deviates from the majority of the literature which assumes efficient market environments, and instead assumes that the taxpayer is constrained by credit rationing. Due to uncertainty, the income of the taxpayer fluctuates, as does the shadow price of income. Provided that non-monetary punishments are high enough to deter one from non-repayment of penalties and tax evaded, evasion may be viewed as a way of "borrowing" from the IRS. A constrained taxpayer may find it optimal to borrow when the shadow price of money is high enough during evasion and relatively low during repayment[16]. Andreoni models a situation where, in bad times, individuals evade as a way to smooth income streams; thus the IRS is a "loan shark". The conditional repayment of the loan occurs in a better state of the world.

Another aspect of uncertainty concerns the unpredictability of the tax liability itself, which arises when the "correct" tax liability is not clearly defined[17]. Uncertainty of true tax liability can be modeled by extending the Allingham and Sandmo framework. Scotchmer and Slemrod (1989) construct a model where, upon audit, the assessed tax liability is symmetrically centered around a known value with an equal probability of one-half. In this case the very concept of income understatement becomes problematic because the taxpayer is uncertain whether any given income declaration is correct or not.

There are now three possible outcomes that the taxpayer must consider. If the return is not audited (with probability $1-p$), true taxable income is irrelevant – the taxpayer merely pays the tax due on his declared taxable income. If the return is audited, there are two possible outcomes, depending on what the assessed tax liability turns out to be. Scotchmer and Slemrod (1989) show that increasing the dispersion of possible assessed taxable incomes induces increased compliance, given weak conditions about the taxpayer's attitudes toward risk. The intuition is that, for a given reported income, more dispersion lowers income in the least desirable state of the world, when the taxpayer is audited and his taxable income is determined to be the highest possible

[16] In this case the government may find it optimal to encourage tax evasion. The optimality of such a policy depends crucially on the non-existence of alternative methods of borrowing, including negotiated payment terms with the IRS, which can in some situations be arranged in the U.S.

[17] Long (1981) argues that the IRS exploits the unpredictability of tax liability to enhance its powers by using it as a license to decide cases in whatever way serves the government's interest at the time. She also notes that unpredictability makes the IRS's burden in providing criminal intent (rather than inadvertent errors) more difficult.

value. This increases the marginal utility of income in that state of the world, which is accomplished by increasing reported income and thus subjecting oneself to a lower penalty in the event this state of the world occurs. As long as the taxpayer exhibits declining absolute risk aversion, increasing the report is the optimal response.

Beck and Jung (1987) show that this conclusion may not hold when there is a continuous range of possible taxable income assessments. In this case one marginal benefit of increasing the income report is that it reduces the probability that a fine will be assessed. For a taxpayer reporting income below the mean of possible assessment, an increased dispersion of possible assessed incomes decreases the likelihood that the income report will be declared insufficient and a fine assessed, so that this component of marginal benefit is reduced. Thus, it is theoretically possible that increased dispersion will cause a lower report.

Note that uncertainty does not reduce tax evasion by as much as it reduces aggregate noncompliance in the sense of *true* aggregate tax liability minus tax paid. This is because one effect of uncertainty is to induce some taxpayers to pay *more* tax than they are legally obligated to pay, which reduces aggregate noncompliance but not the amount of individual tax evasion.

Scotchmer (1989) allows for the possibility that, by expending resources, the taxpayer can reduce the uncertainty of tax liability. The resources can be in the form of research by the taxpayer himself, or in the form of professional assistance hired. In this case the cost of unpredictability includes not only the disutility caused by uncertain tax liability but also the resources expended to reduce the uncertainty.

2.4. General equilibrium considerations

The A–S model and its direct descendants address only the demand for tax evasion by (potential) taxpayers. One might also consider the "supply" of evasion, and ponder the general equilibrium considerations of demand having to equal supply.

One context for this extension is the underground economy. Kesselman (1989) develops a set of models in which there are two sectors – above-ground and underground – which produce two distinct goods. Workers are homogeneous in their gross productivity in each sector of the economy (and in their consumption preferences), but must work only in one sector or the other. The workers, though, have differential distaste and risk aversion for tax evasion, and differential efficiency in concealment and other skills needed to operate successfully in the underground economy.

Although the precise results are model-dependent, three general conclusions obtain: (i) much of the gain from evasion may be shifted from the evaders to the consumers of output through lower prices, and the "marginal" evader gains nothing; (ii) relative price effects tend to dampen the impact of tax rate changes on the extent of evasion, and (iii) the effects of evasion on the marginal revenue response to tax rate changes will depend on consumers' elasticity of substitution between the sectoral outputs.

A key aspect of the foregoing model is that the act of tax evasion is tightly tied to the production of a distinct good. This need not be true, as is indicated by the simultaneous presence of above-ground and underground housepainters, repair people, and so on. Still, there is certainly evidence that evasion is concentrated in particular sectors, such as those that supply services directly to homeowners, because of the small scale of production that can aid concealment and the lesser need for receipts compared to services provided to businesses.

3. General models of avoidance and evasion

Because Allingham and Sandmo addressed tax evasion as a gamble, much of the subsequent literature focused on models in which taxpayers' risk aversion, and therefore higher-order characteristics of utility functions, play an important role. This focus has to some extent obscured other important aspects of the issue, such as the tax concealment technology, and also obscured the common aspects of evasion and what we have called avoidance. To highlight these issues we turn now to more general models of behavioral response to taxation.

Mayshar (1991) poses the taxpayer's problem as

$$\max_{X,S,L,Y} U(Y,L) \quad \text{subject to} \quad X = w[L - S - m(E)], \quad Y = X - T(X,S,E), \tag{6}$$

where X is output, S is sheltering effort, L is total labor effort, and Y is consumption. Mayshar labels $T(\,)$ the "tax technology"; it specifies the maximal taxes, T, collectible from a base X, when the tax authority selects a vector E of policy instruments, while the taxpayer devotes S in labor units to sheltering activity. It is reasonable to assume that $T_X > 0$ and $T_S < 0$ and, by construction $T_E > 0$. The function $m(E)$ represents unavoidable compliance costs associated with taxpaying, measured in labor units.

Although evasion as a gamble is not explicitly treated in this model, Mayshar argues that it can be presented in this framework; to do so S is defined as that certain payment which causes the same expected utility loss as the extra risk an evader takes on, for given expected tax payments. This forms the link between the A–S models of tax evasion and the models discussed in this section[18]. From the perspective of an A–S evasion model, $T_S < 0$ means that more evasion can lower expected tax payments, at a cost of more uncertainty.

Consider the first-order conditions with respect to L and S, respectively, where asterisks indicate an optimal value:

$$-U_L(Y^*,L^*)/U_Y(Y^*,L^*) = w[1 - T_X(X^*,S^*,E)], \tag{7}$$

$$w[1 - T_X(X^*,S^*,E)] \geqslant -T_S(X^*,S^*,E), \tag{8}$$

where Equation (8) holds as an equality if $S^* > 0$.

[18] Note that interpreting S, or C in the model of Slemrod (2001) discussed below, as the risk bearing cost of evasion, will impose restrictions on the $T(\,)$ or $C(\,)$ functions.

Expression (7) looks familiar: the marginal rate of substitution between consumption and leisure equals the net wage. But note that the effective marginal tax rate, $T_X(X^*, S^*, E)$, permits more complex marginal tax rates than the standard linear tax model, where $T(X^*, S^*, E)$ would equal tX^*, and so T_X would equal t. In Expression (7), the effective marginal tax rate may depend on the sheltering activity of the taxpayer and/or the policy instruments of the government, interpreted more broadly than simply announcing a tax schedule. Expression (8) states that, because sheltering is accomplished by using labor, at an interior optimum its opportunity cost $w(1 - T_X(\ ))$ will be equal to its marginal private gain, which is the marginal tax saving, $-T_S$.

Slemrod (2001) investigates a related model in which the private cost of achieving reductions in taxable income (denoted A, for income avoidance) is $C(wL, A)$, where wL is true labor income; he argues that, in general, $C_1 < 0$ and $C_2 > 0$[19]. If we imbed this avoidance technology into the taxpayer choice under a linear income tax, the maximization problem becomes

$$\max_{L,A} U(Y, L), \quad \text{subject to} \quad Y = w(1-L) - t(w(1-L) - A) - C(wL, A). \tag{9}$$

Before pondering the general implications of this formulation, first consider the special case where $C(wL, A) = C(A)$. In this case the first-order condition for labor supply is identical to the standard model without avoidance. The first-order condition for A is simple and straightforward, $C' = t$, implying that avoidance ought to be pursued until its marginal cost equals its marginal saving in tax liability, equal to t. In this situation a tax rate hike unambiguously increases A. Furthermore, its effect on L is no different than in the standard model, except to the extent that the income effect is altered by the possibility of avoidance.

The story is enriched when the avoidance, or tax, technology becomes $C(wL, A)$. The effective marginal return to working becomes $w(1 - t - C_1)$, where $-wC_1$ is a subsidy to working that Slemrod (2001) dubs the "avoidance-facilitating" effect; for example, a given level of allegedly work-related deductions looks more plausible if it is taken against a larger gross income. The term $(t - C_1)$ is analogous to T_X in Mayshar's model, and makes explicit how the avoidance technology influences the incentive to supply labor.

Several insights emerge from this modeling of the tax environment. First of all, the substitution effect of labor supply does not respond identically to the two components of the statutory after-tax wage rate, w and $(1 - t)$. Changes in $(1 - t)$ trigger avoidance responses which are not triggered by changes in w. While both labor supply and avoidance respond to both w and $(1 - t)$, they do not do so symmetrically. This

[19] Slemrod (2001) is less general than Mayshar in that it presumes a flat-rate statutory tax system; it does not presume that tax sheltering or avoidance must be "produced" with the taxpayer's own time. One superficial difference is the adoption by Slemrod of a cost function approach to avoidance, compared to Mayshar's production function for tax receipts.

implies that econometric studies of labor supply (and avoidance) ought to differentiate responses to w and $(1-t)$. Furthermore, one should not conclude, as does Rosen (1976), that a differential response to w and $(1-t)$ necessarily represents "taxpayer illusion"[20]; instead it could be reflecting the avoidance technology.

Mayshar and Slemrod addressed the possibility that changes in the tax system will induce from taxpayers all three types of behavioral response. For example, an increase in the rate of a proportional income tax will provide an incentive to substitute leisure for goods, to (depending on the penalty structure) increase evasion, and increase avoidance. Other interactions among the three types of behavioral response have been investigated, as well. Cowell (1990a) develops a model in which the taxpayer can evade, but can also legally shelter income for a fixed cost Γ and a constant marginal cost γ, where $\gamma < t$. These cost assumptions generate the result that if an honest (or highly risk-averse) person shelters any of his income (Y), then he will automatically shelter all of it, and will do the latter if $\Gamma + \gamma Y < tY$. Cowell then investigates whether sheltering will co-exist with evasion, and asserts that the optimum is *not* characterized by an equality between the marginal cost of avoidance and evasion. This is because sheltering reveals to the tax authority that the taxpayer's true income must be at least $\Gamma/(t-\gamma)$. He argues that there may be a class of shelterers who would also have been evaders, had it not been for the attention drawn by sheltering, and that in some cases there may be a complete polarization between evaders and avoiders.

In Cross and Shaw (1982), taxpayers must make expenditures to learn about and (in the case of avoidance) document both avoidance and evasion activities[21]. Two avenues of interaction arise. First, in a progressive tax system, expenditure on avoidance or evasion reduces the marginal tax rate, thus reducing the return to engaging in the other[22]. Second, investment in avoidance may reduce the marginal cost of evasion, or vice versa. For example, while investigating an illegal but undetectable "tax shelter", a (barely) legal tax shelter arrangement may be uncovered without much additional investment of time.

4. Descriptive analysis of evasion and enforcement

4.1. The extent of tax evasion

4.1.1. Data problems

Ascertaining the extent and characteristics of evasion immediately runs into two problems – one conceptual and one empirical. The conceptual problem is that, although

[20] Although, note that in his empirical analysis Rosen (1976) does not detect a significant differential response.

[21] In some situations, more evasion may be associated with less cost. For example, not bothering to trace a miscellaneous source of income is less costly than tracking down whatever receipt or Form 1099 would document the income. Not filing a return at all happens to minimize compliance cost.

[22] Alm (1988) also examines the simultaneous choices of evasion, avoidance and reported income, and investigates the effects of other fiscal variables on these choices.

one can assert that legality is the dividing line between evasion and avoidance, in practice the line is often blurry. Sometimes the law itself is unclear, sometimes it is clear but not known to the taxpayer, sometimes the law is clear but the administration effectively ignores a particular transaction or activity. The importance of these factors certainly differs across situations.

The other difficulty is that, by its nature, tax evasion is not easy to measure – merely asking just won't do. Several different approaches have been attempted. One approach relies on inferring the level or trends in noncompliance from data on measurable quantities, such as currency demand or national income and product accounts. The monetary indirect estimates are based on the presumption that most unreported economic activity takes place in cash, and that some time in the past the underground economy was small. In Gutmann (1977), increases in the ratio of currency to demand deposits since 1937–41 measure the underground economy; in Feige (1979), changes since 1939 in the ratio of total dollar transactions to official GNP since 1939 measure it. Tanzi (1980) estimates regressions explaining the ratio of currency to money defined as M2, and interprets the portion explained by changes in the tax level as an indication of changes in the size of the underground economy. None of these approaches is likely to be reliable, however, as their accuracy depends either on unverifiable assumptions or on how well the demand for currency is estimated. The indirect noncompliance estimates based on discrepancies between national accounts measures of income and income reported to the tax authority are also problematic. For one thing, national income estimates of several key forms of income are based on tax return data. Second, there are many inconsistencies between how income is defined for tax purposes and for national accounts. However, Engel and Hines (1999), in a study of tax evasion dynamics which focuses on the possibility of retrospective examination of previous-years' returns, study this measure of evasion in the U.S. for the years 1947 to 1993 and find that it responds as their model predicts. For example, annual fines and penalties imposed by the IRS subsequent to audits are correlated with contemporaneous and several lags of tax evasion as calculated from national income statistics.

The most reliable source of information about tax compliance concerns the U.S. federal income tax, and exists because of the IRS's Taxpayer Compliance Measurement Program, or TCMP. Under this program, approximately every three years from 1965 until 1988 the IRS conducted a program of intensive audits on a large stratified random sample of tax returns, using the results to develop a formula used to inform the selection of returns for regular audits. The TCMP data consist of line-by-line information about what the taxpayer reported, and what the examiner concluded was correct. This data formed the basis for the IRS estimates of the aggregate "tax gap", and provides much useful information about the patterns of noncompliance with respect to such variables as income, occupation, line item, region of the country, age, and marital status. While informative, it is widely recognized that even the intensive TCMP audits imperfectly reveal particular kinds of noncompliance, such as income from the underground economy.

4.1.2. Patterns of noncompliance

We cannot adequately review here what is known about the extent and nature of tax evasion for all taxes in all countries at all times. Rather, in what follows we offer a few salient facts about the recent U.S. income tax, mostly gleaned from the TCMP data just discussed.

(1) With audit coverage hovering at about 1% and an extensive information reporting and matching program, evasion is estimated to be 17% of true tax liability[23].
(2) The extent of evasion varies widely across types of gross income and deductions; for example, the 1988 TCMP reports that the voluntary reporting percentage was 99.5% for wages and salaries, but only 41.4% for self-employment income (Schedule C). These percentages clearly correlate positively with the likelihood of income understatement being detected.
(3) Evasion (as measured by underreported income, not tax liability), rises with income but at a less than proportionate rate. Christian (1994) reports that in 1988, taxpayers with (auditor-adjusted) incomes over $100 000 on average reported 96.6 percent of their true incomes to the IRS, compared to just 85.9 percent for those with incomes under $25 000[24].
(4) Within any group defined by income, age, or other demographic category, there are some who evade, some who do not, and even some who overstate tax liability[25]. For example, of middle-income (auditor-adjusted income between $50 000 and $100 000) taxpayers in 1988, 60% understated tax, 26% reported correctly, and 14% overstated tax [Christian (1994, p. 39)].

4.2. Determinants of evasion

Empirical attempts to more systematically establish how compliance responds to aspects of the tax environment have met with limited success, primarily due to the data problems discussed in Section 4.1.1[26]. Three approaches dominate the literature[27].

[23] It is probably higher in most other countries. For example, Alm, Bahl and Murray (1991) put the figure (for avoidance and evasion) at 46% for the Jamaican income tax of 1983. Richupan (1984) cites studies of tax evasion in developing countries indicating that it is not uncommon for half or more of potential income tax to be uncollected.

[24] One explanation for this pattern is almost certainly that tax returns of high-income households are more likely to attract IRS attention. Another potentially important factor is that the TCMP results do not account for the noncompliance of business entities, which are more germane for higher-income individuals.

[25] Note that Erard (1997) concludes that a large fraction of noncompliant reports may be unintentional.

[26] A former colleague, Harvey Galper, once put the problem this way: "Regression analysis of tax evasion is straightforward, except for two problems: you can't measure the left-hand side variable, and you can't measure the right-hand side variables!"

[27] In addition to the econometric methodologies discussed below, laboratory experiments typically involving students engaged in a multi-period reporting game, have been employed. [See, for example,

4.2.1. Cross-sectional analysis

Clotfelter (1983) was the first attempt to make use of the TCMP data to investigate how noncompliance responded to changes in the environment. He estimated a tobit model, explaining, for each of ten audit classes, noncompliance as a function of the combined federal and state marginal tax rate, after-tax auditor-adjusted income, and a set of demographic variables available on tax returns. The most striking conclusion is that noncompliance is strongly positively related to the marginal tax rate, with the elasticity ranging from 0.5 to over 3.0. This finding is apparently consistent with the basic A–S model, but not with the extension proposed by Yitzhaki.

Beron, Tauchen and Witte (1992) investigate TCMP data aggregated by the IRS to the three-digit zip code level. They find that increasing the odds of an audit significantly increases reported AGI and tax liability for some, but not all, of the groups. In an attempt to deal with the potential endogeneity of the intensity of enforcement, they model the simultaneous determination of tax reporting and the log odds of an audit for each of the several audit classes in each zip code area. Their instrument for this is the level of IRS resources relative to the number of returns[28]. Although Beron, Tauchen and Witte argue that it is a valid instrument because the IRS has not been able to distribute its resources among districts so as to achieve its goals, this is not convincing: it is reasonable that the IRS attempts to target its resources toward areas believed to be particularly noncompliant, thus invalidating use of IRS resources as an instrument.

Subsequent studies have produced mixed results. Of particular interest is work by Feinstein (1991), who performed a pooled cross-section analysis of 1982 and 1985 TCMP data, thus mitigating the problem that in a single cross-section (other than for cross-state differences) the marginal tax rate is a (complicated, non-linear) function of income, making it difficult to separately identify the tax and income effect. Feinstein's analysis suggests a negative impact of the marginal tax rate on evasion, which contradicts Clotfelter's results but is consistent with the A–S model as adjusted by Yitzhaki.

Klepper and Nagin (1989) investigate the characteristics of evasion across line items, and find that noncompliance rates are related to proxies for the traceability, deniability,

Baldry (1987) and Alm, Jackson and McKee (1992)]. These results are subject to the canonical criticisms of laboratory studies: that the setting is artificial, and the participants are not demographically similar to those making the actual decisions, and therefore do not come to the decision problems with the same array of experiences and expectations about the environment. These criticisms may be especially salient in this context, because the experiments differ from general problems in choice under uncertainty only by the labeling of the choice as having to do with taxes, and as compliant or not rather than gambling or not.

[28] Dubin and Wilde (1988) perform a similar analysis on the zip-code aggregated data, and use the same instrument. They defend this choice by claiming that, in an analysis of the time path of state-level IRS budgets, they were found to be independent of compliance levels, and predominantly determined by the share of total returns filed.

and ambiguity of the items, which are in turn related to the probability that evasion will be detected and punished. They also find evidence of a "substitution effect" across line items, such that greater noncompliance on one item lowers the attractiveness of noncompliance on others, because the latter jeopardizes the expected return to the former by increasing the probability of detection.

4.2.2. Time-series analysis

Dubin, Graetz and Wilde (1990) make use of state-level time series cross-section data from 1977 through 1986 to investigate the impact of audit rates and tax rates on tax compliance. They do not, though, have a direct measure of noncompliance, but instead use tax collections per return filed and returns filed per capita as (inverse) measures of noncompliance. They conclude that the continual decline in the audit rate over this period caused a significant decline in IRS collections – amounting to $41 billion by 1985.

4.2.3. Controlled experiments

As discussed above, analysis of both cross-section and time-series historical data is subject to severe difficulties of measuring the parameters of the environment and in knowing the source of any variation in these parameters. Controlled experiments can avoid all of these problems, but, for cost and other implementation reasons, are rare.

One recent exception is reported by Slemrod, Blumenthal and Christian (2001), in which the State of Minnesota Department of Revenue conducted a randomized controlled experiment with respect to four aspects of the tax compliance environment: the threat of an audit, the provision of special return preparation information services, moral appeals, and a redesigned tax form. With regard to the first, they find that, for low- and middle-income taxpayers, a threat of certain audit[29] produced a small, but statistically significant, increase in reported income, which was larger for those with greater opportunities to evade[30]. However, for high-income taxpayers the audit threat was associated with on average a *lower* income report. The authors speculate that sophisticated, high-income, taxpayers view an audit as a negotiation, and view reported taxable income as the opening (low) bid in a negotiation which does not necessarily result in the determination and penalization of all noncompliance. Based on the same experiment, Blumenthal, Christian and Slemrod (2001) find no evidence that either of two written appeals to taxpayers' consciences had a significant effect on aggregate compliance.

[29] The audit threat was delivered by letter in January following the tax year.

[30] The approach is a "difference-in-difference" analysis; that is, the increase in reported income over the previous year of the treatment group is compared to the increase in reported income of the control group.

5. Descriptive analysis of avoidance

5.1. Dimensions of avoidance

Stiglitz (1985) distinguishes three basic principles of tax avoidance within an income tax: postponement of taxes, tax arbitrage across individuals facing different tax brackets (or the same individuals facing different marginal tax rates at different times), and tax arbitrage across income streams facing different tax treatment. Many tax avoidance devices involve a combination of these three principles. In an example used by Stiglitz, the basic feature of an Individual Retirement Account (IRA) is the postponement of tax liability until retirement; if the individual faces a lower tax rate at retirement than at the time the income is earned, then the IRA also features tax arbitrage between different rates. Finally, if the individual can borrow to deposit funds in an IRA and the interest incurred to finance the deposit is tax deductible, then the IRA is a tax arbitrage between two forms of capital, one of which is taxed, and the other of which is not taxed[31]. Stiglitz argues that, with perfect capital markets, these three principles can be exploited to eliminate all taxes while leaving the individual's consumption and bequests unchanged relative to the zero tax case, and facing no more risk than in the original situation. But capital markets are not perfect, and therefore all tax liability is not eliminated by tax avoidance[32], and to reduce tax liabilities distorting actions (such as investment in sectors where it is easier to convert ordinary income into capital gains) are utilized. There is considerable empirical evidence testifying to the extent and tax sensitivity of these kinds of avoidance behavior.

5.1.1. Retiming

There is abundant support for the notion that the timing of certain transactions can be extraordinarily responsive to changes in tax rates. Perhaps the most striking example was the response of capital gains realizations to the tax rate increase scheduled to occur on January 1, 1987, but fully anticipated by the fall of 1986. Aggregate realizations in 1986 were *twice* what they were in any previous year or for several years thereafter. As Burman, Clausing and O'Hare (1994) document, capital gain realizations on corporate stock in December of 1986 were seven times higher than in the previous December. Another striking example of timing response is provided by Goolsbee (2000), who documents that, in advance of the expected 1993 increase in the U.S. top individual

[31] The IRA example makes clear that in certain cases (some of) the avoidance behavior is the result of a conscious tax policy choice, in this case with the intent of increasing saving. Another excellent example is capital gains, where taxation upon realization rather than accrual allows for deferral of tax liability, often into periods of lower taxation, and where gains are completely excused from taxation at death due to the step-up of tax basis.

[32] There are also policy responses to avoidance, such as limits on loss offsets and interest deductions.

tax rate, corporate executives realized a huge amount of income in 1992, primarily through exercising non-qualified[33] stock options.

Sophisticated econometric techniques using panel data have been developed for separately identifying the timing responses to tax rate changes over time from the permanent behavioral response to a changed tax rate. These new techniques have been applied to both capital gains realizations [Burman and Randolph (1994)] and charitable contributions [Randolph (1995)]. In both cases the results suggest that the retiming effect dominates the permanent effect.

5.1.2. Tax arbitrage

Tax arbitrage activity takes advantage of inconsistencies in the tax law, featuring economically offsetting positions which have asymmetric tax treatments. Examples range from sophisticated derivative financial instruments to the more mundane cases of doing tax-deductible borrowing to finance tax-deferred IRA contributions or tax-exempt bond purchases.

5.1.3. The classification of income

The classic example of income reclassification, also termed income shifting, is turning ordinary capital or labor income into preferentially-taxed capital gains. In another example, Maki (1996) and Scholz (1994) have documented that, following the Tax Reform Act of 1986, there was a large shift from no-longer-deductible consumer interest into still-deductible mortgage or home equity loans. There is anecdotal evidence that, following the introduction of the R&D credit in the United States, much business activity was "discovered" to have a significant research component. Gordon and MacKie-Mason (1990, 1997) have investigated how, when the Tax Reform Act of 1986 lowered the top individual rate below that of the corporate rate, there was a large shift from C corporations into S corporations, which are taxed like partnerships and therefore are not subject to the corporation income tax. Gordon and Slemrod (2000) discuss the shifting of income between the corporate and individual tax base via the method of compensation, and document evidence of such shifting in the United States.

5.2. The extent of avoidance

No one has attempted to calculate for avoidance a counterpart to the aggregate evasion "tax gap". There is, though, some indirect evidence that the avoidance tax gap is

[33] For non-qualified stock options, the difference between the exercise price and the issue price is taxable at ordinary income tax rates at the time of exercise, and is deductible from the employer's taxable income at the same time.

large. Gordon and Slemrod (1988) calculated that the U.S. tax system of 1983 raised approximately zero revenue from taxing capital income, due to the combination of legislated deviations from a pure income tax and tax arbitrage[34]. As to the incidence of the avoidance opportunities, Agell and Persson (1990) and Gordon and Slemrod (1988) argue that the availability of tax arbitrage opportunities will generally benefit those at the bottom and top of the tax rate distribution, to the disadvantage of those in the middle. This generally corresponds to low- and high-income individuals, respectively, but there are exceptions to that rule; high-income individuals benefit through their ownership of tax-preferred pension assets.

6. Fundamentals of tax analysis

Having completed a review of the positive, or descriptive, analysis of tax evasion and avoidance, we turn now to the normative analysis of taxation. However, before we proceed to that task, we must first reconsider the fundamental building blocks of tax analysis – the evaluative criteria of equity and efficiency – to check whether these concepts need to be revised.

6.1. Equity

6.1.1. Vertical equity

Analyses of the distributional impact of taxation, especially those based on tax return data, ought to account for the presence of evasion. The evidence cited in Section 4.1 – that noncompliance as a fraction of true income *declines* with true income – suggests that standard analyses of incidence based on the statutory rates and base may understate the progressivity of the tax burden[35]; Bishop, Chow, Formby and Ho (1994) find this for the United States using the 1985 TCMP data, although Alm, Bahl and Murray (1991) reach the opposite conclusion about Jamaica[36].

6.1.2. Horizontal equity

Horizontal equity – the idea that equals should be treated equally by the tax system, or that tax liability should not depend on any of a set of irrelevant characteristics –

[34] It is likely that the Tax Reform Act of 1986 mitigated the avoidance tax gap by reducing the dispersion of marginal tax rates and tightening up the rules about tax arbitrage behavior.

[35] This evidence also suggests that tax return data may overstate the inequality in the distribution of incomes. Because the data on tax evasion are flawed, one should keep in mind that the theoretical arguments discussed in Section 5.2 imply that data on reported incomes will *understate* the true dispersion of income.

[36] A complete incidence analysis would account for the costs borne by evaders in the form of exposure to risk and concealment expenses, neither of which is accounted for in the studies mentioned.

is central to an assessment of the impact of tax avoidance and evasion. To see this, compare two tax situations, one in which there is a linear income tax rate of 20% and everyone reports their true income, and another in which the tax rate is 40% and everyone (costlessly) reports exactly half their income. In this case the two systems are identical with respect to both horizontal and vertical equity. Now imagine that, in the second system, on average everyone reports half their income, but that the fraction differs systematically by income. In that case replicating the progressivity of the first tax system will require a more complicated, non-linear, system of rates. If, however, evasion varies *within* income classes, no revision of the tax rate schedule can compensate, and there will be horizontal inequity.

In the context of the A–S model of tax evasion, the horizontally inequitable tax burden will depend on the taxpayer's degree of risk aversion. Less risk-averse households will gain more from the availability of a gamble with given positive expected value. In contrast, common parlance would ascribe any horizontal inequity to variations in honesty, with the honest, or dutiful, citizens left holding the bag by the dishonest. In the typical economic model, though, there are no honest or dishonest individuals, only utility-maximizers; thus, this distinction can be introduced only artificially by simply positing that some individuals do not pursue tax evasion. The same kind of artificial differentiation across people can be made with regard to tax avoidance by positing that some people have an aversion to such behavior; as Steuerle (1985, p. 78) says: "Some taxpayers simply do not enjoy playing games no matter what the certainty of the return; the U.S. tax system is designed to insure that such individuals pay a greater share of the tax burden than those who are not so hesitant". Steuerle (p. 19) concludes that "taxpayers pay unnecessary taxes because of the simplicity of their filing response or their lack of knowledge of the tax laws".

6.1.3. Incidence

The theory of tax incidence – who bears the burden of a given tax structure – begins with three basic principles: (i) the burden of all taxes must be traced back to individuals; (ii) individuals with relatively elastic demand (or supply) of a taxed good tend to escape the burden of tax imposed on that good; and (iii) in the long run the incidence of a tax levy does not depend on which side of the market bears the legal responsibility for remitting the tax to the government. Introducing avoidance and evasion preserves the methodological importance of the first two principles[37], but calls the third into question. A complete analysis of the incidence of a particular tax requires specifying the remittance process and positing an avoidance technology for both the suppliers and demanders of the taxed good.

Avoidance opportunities alter the analysis of incidence for two separate reasons. First, their presence affects the behavioral response to a change in the tax system,

[37] Section 2.4 discusses some models of tracing the incidence of tax evasion.

and this alters what otherwise would be the change in equilibrium prices. Second, the presence of avoidance alters the link between tax-inclusive prices and welfare. This suggests that the incidence (not to mention the efficiency) of a tax may depend on which side of the market the responsibility for remittance falls. That is in stark contrast to the standard model, under which that is irrelevant to the long-run incidence[38].

6.1.4. Are changes in the social welfare function necessary?

In models with heterogeneous citizens, the standard objective function is a social welfare function which has as arguments the utility level of each citizen – accepting the individuals' own relative valuations of goods and services – where the shape of the social welfare function implicitly determines the social value placed on the distribution of utilities as opposed to the sum of utilities. In the presence of uncertainty, the *expected* utilities of individuals are the relevant arguments – accepting the risk preferences of consumers. Cowell (1990b) questions the appropriateness of according the same social weight to investigated and guilty taxpayers as is applied to the innocent or uninvestigated, and argues that there may be a case for putting a specific discount on the utility of those "who are known to be antisocial" (p. 136). Cowell investigates a few alternative social objective functions, including one in which any private benefit derived from the proceeds of evasion is assigned a social weight of zero, but in our opinion no convincing alternative that provides reasonable policy prescriptions has yet been presented.

6.2. A taxonomy of efficiency costs

In the standard model the efficiency cost of taxation is entirely due to the fact that, because of the change in relative prices, individuals are induced to select socially suboptimal consumption baskets – to substitute away from relatively highly-taxed goods to relatively lightly-taxed goods, such as leisure. A standard exercise in optimal taxation theory is to describe the tax system that minimizes these costs, or to describe the tradeoff between these costs and the distribution of welfare in the society.

In the presence of avoidance and evasion, a broader concept of efficiency cost is needed. In what follows, we describe and comment on three additional components of the social cost of taxation and discuss the problems that arise in introducing these costs into formal models of optimal taxation.

6.2.1. Administrative costs

Tax administrations deal, among other things, with information gathering. But this is a difficult element to model because information varies in quality. There is a qualitative

[38] There are exceptions. Consider, for example, the debate between Tanzi (1992) and Dixit (1991) over the implications of tax collection lags for the optimal amount of inflationary finance. Tanzi (1992) argues that, when consumption taxes are collected by firms in advance and held by them for the duration of the collection lag, inflationary finance implies a real redistribution of income from consumers to sellers.

difference between an auditor "knowing" that a given taxpayer is evading and having sufficient evidence to sustain a court finding to that extent. Also, the cost of gathering information depends on how accessible the information is, and whether it can be easily hidden. There are several advantages to taxing a market transaction relative to taxing an activity of the individual such as self-consumption. First, in any market transaction there are two parties with conflicting interests. Hence, any transaction has the potential of being reported to the authorities by one unsatisfied party. A second property is that the more documented the transaction, the lower is the cost of gathering information on it. For this reason it is easier to tax a transaction that involves a large company, which needs the documentation for its own purposes, than to tax a small business, which may not require the same level of documentation. Finally, market transactions establish arms-length prices, which greatly facilitate valuing the transaction. Administrative cost may also be a function of the physical size and the mobility of the tax base (it is harder to tax diamonds than windows), whether there is a registration of the tax base (e.g., owners of cars, holders of drivers' licenses), the number of taxpayer units, and information sharing with other agencies[39]. It is also an increasing function of the complexity and lack of clarity of the tax law.

Administrative costs possess two additional properties that complicate the modeling of tax administration issues: they tend to be discontinuous and to have decreasing average costs with respect to the tax rate. To see the first property, consider two commodity tax rates, denoted by t_1 and t_2. If $t_1 = t_2$, then only the total sales of the two commodities need be reported and monitored. If, however, the two rates differ even slightly, then the sales of the two commodities must be reported separately, doubling the required flow of information. There are decreasing average costs because the cost of inspecting a tax base does not depend on the tax rate (except to the extent that people are more inclined to cheat with a higher tax rate). Hence, a higher tax rate reduces the administrative cost per dollar of revenue collected [Sandford (1973)]. Administrative cost may also be a function of the combination of the taxes employed and their rates, because the collection of information concerning one tax may facilitate the collection of another tax (e.g., inspection of VAT receipts may aid the collection of income tax).

6.2.2. *Compliance costs*

Slemrod (1996a) estimates that, for the U.S. income tax, the private compliance cost is about 10 cents per dollar collected. Sandford (1995) presents estimates for a variety of taxes in several countries. Some of that cost is an unavoidable cost of complying with the law, and some of it is voluntarily undertaken in an effort to reduce one's tax bill, but in either case it approximately represents resource costs to society. In almost all cases the *private* compliance costs dwarf the *public* administrative costs of collecting taxes,

[39] A good description of the properties of administrative cost can be found in Shoup, Blough and Newcomer (1937, pp. 337–551).

which the IRS estimates at 0.6 cents per dollar collected for all the taxes it administers. Integrating compliance costs into formal models in a meaningful way is tricky. As an example of the modeling difficulties this topic poses, consider the following problem: when is it optimal to delegate to employers the authority to collect taxes and convey information about employees, thus requiring the administration to audit both the taxpayer agent and the taxpayer himself, and when is it optimal to deal only with the employee? Clearly, given that the employer already has the necessary information, it would save administrative costs to require him to pass it along to the tax administrator. This might also reduce total social costs if the cost of gathering information by the administration is higher than the increase in cost caused by imposing a two-stage information-gathering system[40].

However, the potential efficiency of involving taxpayers in the administrative process must be tempered with a practical consideration. Administrative costs must pass through a budgeting process, while compliance costs are hidden. Hence, there may be a tendency to view a policy which reduces administrative cost at the expense of an equal (or greater) increase in compliance costs as a decrease in social cost, because it results in a decrease in government expenditures. We will discuss this issue further in Section 7.

6.2.3. The risk-bearing costs of tax evasion

In the Allingham–Sandmo model, tax evasion occurs only if the taxpayer expects to increase his expected income by evading taxes, including the expected fines that he would have to pay if he were caught; it continues until, at the margin, the increased expected income is offset by the increased risk-bearing. Hence, a taxpayer who evades taxes increases both his exposure to risk and his expected income. This additional exposure to risk is a deadweight loss to society. In principle, the taxpayer could be better off under an agreement whereby the taxpayer pays at least as much as the government currently collects, while the government ceases to audit. Assuming a risk-neutral government, the risk-bearing cost of tax evasion is equal to the risk premium that the taxpayer would be ready to pay in order to eliminate the exposure to risk [Yitzhaki (1987)]. Depending on the other assumptions about the probability of detection, the penalty structure, and risk aversion, the risk-bearing costs of evasion may be a continuous function that increases with the tax rates. These costs are in addition to the compliance costs voluntarily incurred by an individual attempting to minimize the expected cost by camouflaging the evasion or shifting to an otherwise less remunerative occupation.

[40] Note that a withholding system requires two information gathering systems and might generate incentives for the withholding agent to evade the taxes it collects, or to collaborate with withholdees in withholding less than required [Yaniv (1988, 1992)]. In a period of rapid inflation, the gain to the agent from withholding may exceed the cost.

7. Normative analysis

7.1. Optimal tax administration and enforcement

Avoidance and evasion pose two challenges for the normative analysis of taxation. First, they introduce a new set of policy instruments whose optimal setting is at issue. These include the extent of audit coverage, the penalty imposed on detected evasion, and the structural integrity of the tax code itself, which determines the extent and nature of avoidance opportunities. Second, they invite a rethinking of standard taxation problems, including the optimal setting of commodity tax rates and optimal progressivity.

7.1.1. Optimal penalties

Consider the A–S model of a representative consumer whose true income is exogenous and whose only choice concerns how much of that income to report. This choice depends on two policy instruments set by the government, p, which has a resource cost due to the need for auditors and the related infrastructure, and θ, which is a fine for detected evasion, which is a transfer with no resource cost.

It has been well known since Becker (1968) that in this setting a government concerned with maximizing the *ex ante* utility of its representative citizen will want to set θ as high as possible, allowing p to be as low as possible. This policy of "hanging violators with a probability of zero" deters evasion while minimizing the resource cost of the deterrent – p represents a real resource cost but θ is simply a transfer. But this kind of model ignores, *inter alia,* the possibility of a corrupt tax administrator who abuses the system or, alternatively, harshly punishes someone who commits an honest mistake[41]. The harsher the penalty, the more damage that can be inflicted by a corrupt administrator or, in the case of an honest mistake, the more capricious the system is. Hence, the harsher the penalty, the more detailed and cautious the prosecution process should be, although this may increase its administrative costs. In the absence of modeling the interaction between the penalty rate and administrative costs, analytical models usually assume a ceiling on the penalty rate.

7.1.2. Optimal randomness

Auditing some taxpayers and not others inevitably introduces some *ex ante* uncertainty and some *ex post* horizontal inequity. This suggests a link to an earlier literature in public finance, in which Stiglitz (1982) and Weiss (1976) each argued that, even in a world of risk-averse citizens, it may be optimal for the government to introduce some randomness into its net tax (or transfer) to individuals. The argument depended on

[41] Polinsky and Shavell (2000) examine this and other issues involved in the optimal setting of penalties for crime including but not restricted to tax evasion.

the second-best nature of the problem, in which an income tax distorted the labor–leisure choice. For some utility functions, Stiglitz and Weiss argued, the introduction of random payments induced people to work harder, thus mitigating the labor market distortion; in some cases the value of the increased labor more than offset the utility loss from the randomness introduced.

This argument has clear implications for the optimal enforcement of the income tax, because it suggests that one of the presumed social benefits of greater enforcement – the reduced uncertainty of payment of a given expected value of taxes – may be mitigated by the increased labor supply distortion. Weiss uses approximations around the point of no evasion to describe the condition under which allowing some degree of evasion can both increase revenue and increase welfare. However, Yitzhaki (1987) shows that, in the examples used by Weiss, the condition that allows successful evasion is identical to the condition that the solution is on the declining portion of the Laffer curve; in this case, *any* reduction of the tax rate would increase welfare and increase revenue. This suggests that the improvement was not caused by allowing evasion. We conclude that neither the practical nor hypothetical relevance of this point has yet been demonstrated.

7.1.3. The optimal extent of enforcement

For a given penalty structure how much resources should be devoted to enforcing the tax laws? Or, in other words, what is the optimal probability of detection, p? Many widely-used textbooks and several IRS commissioners presume that the answer is to increase p until the marginal increase of revenue thus generated equals the marginal resource cost of so doing. As Slemrod and Yitzhaki (1987) show, however, this rule is incorrect. Intuitively, although the cost of increasing p (hiring more auditors, buying better computers, etc.) is a true resource cost, the revenue brought in (through assessed fines as well as higher compliance) does not represent a net gain to the economy, but rather a transfer from private citizens to the government. The correct rule equates the marginal social benefit of reduced evasion to the marginal resource cost; the social cost is not well measured by the increased revenue, but is in this model related to the reduced risk bearing that comes with reduced evasion[42]. This result implies that privatization of revenue collection will inevitably lead to a socially excessive amount of resources devoted to that purpose unless restrictions are put on the resources and behavior of the agency.

7.1.4. Optimal auditing rules

One of the key simplifying assumptions of the Allingham–Sandmo model is that the probability of evasion being detected is fixed and unrelated to any actions of the

[42] Note, though, that Baldry (1984) has shown that complete enforcement of income tax laws (p high enough to deter evasion) is inefficient.

taxpayer. In Section 2.1 we investigated the implication of p increasing with the amount of evasion, but this relationship was exogenously imposed. Other models allow the audit strategy of the tax collection agency (henceforth the IRS) to depend on the report of the taxpayer in a way that maximizes an explicit objective function; the taxpayer, in turn, forms some expectation of what the IRS' auditing rule is, and acts accordingly. In modeling the game between taxpayers and the IRS, researchers have generally assumed that the IRS attempts to maximize net revenue collected. As we discussed earlier, this is not likely to characterize the social-welfare-maximizing solution to how big the enforcement budget ought to be, although it might characterize the optimal allocation of resources for a given IRS budget. Another critical model element is whether it is assumed that the IRS can commit to an announced audit rule, or whether it cannot commit, and therefore will opportunistically audit whatever returns it wishes once the returns are filed. Finally, it is critical whether the IRS budget is assumed to be fixed.

Following Reinganum and Wilde (1985), models of this question generally assume that the probability of audit depends on reported income only. Most papers conclude that the optimal strategy in this context is to randomly audit individuals who report below some threshold level of income. In equilibrium only low-income individuals report honestly, while high-income taxpayers report exactly at the threshold level of income and are never audited. Sanchez and Sobel (1993) derive this result in the context of risk-neutral taxpayers with a continuous distribution of actual income and no labor supply decisions, and where penalties for detected evasion are bounded and exogenously set. Cremer and Gahvari (1996) reach similar conclusions when they allow for endogenous labor supply, although they consider just two types of individuals. Mookherjee and Png (1989) consider risk-averse individuals. Imposing mild restrictions on the level of risk aversion, they show that the optimal policy is characterized by random audits and finite penalties. It is still true that above some income level taxpayers are not audited, but it is no longer true that everyone reporting an income below that level is honest[43]. Scotchmer (1987) relaxes the assumption that the IRS can only observe the taxpayer's report, and instead assumes that it is possible to assign taxpayers to a number of audit classes based on observable characteristics. Although the optimal audit policy within each class is similar to that described above, this policy introduces a regressive bias to the effective tax system, because the agency will audit taxpayers with low-income reports with higher probability than high-report taxpayers, thus making it less attractive for low-income taxpayers to underreport income. This bias may be difficult to undo through the statutory tax system if the tax code cannot depend on the audit class.

This state of affairs provides an obvious temptation to the IRS to reverse its pre-announced audit rule and instead to audit only those taxpayers that report exactly the threshold level of income; those that report below the threshold are, after all, reporting

[43] Note that, in all of the papers in this literature, in the optimal policy taxpayers revealed to be honest are rewarded, a decidedly counterfactual prediction.

truthfully. Because of that temptation, an announced precommitment is not likely to be credible. Describing the equilibrium outcome in the absence of precommitment is more complex, as Andreoni, Erard and Feinstein (1998) discuss. One class of models, first investigated by Graetz, Reinganum and Wilde (1986), introduces a set of taxpayers at each income level who report truthfully regardless of their incentives to do otherwise. This enriches the model because it implies that at each level of income report there are both honest and evading taxpayers. Melumad and Mookherjee (1989) take another tack by assuming that although the government cannot commit to a particular audit policy, it can commit to the total amount spent on audits. In this context they demonstrate that the problem of commitment may be solved by delegating this task to a separate agency, and they describe the optimal contract that guarantees a unique equilibrium and provides incentives for the agency to audit optimally. Such a contract is welfare improving.

In the context of models of tax compliance in which the strategies of both the taxpayers and the IRS are objective-maximizing, the impact of a change in, say, the tax rate, depends on one's forecast of how both sets of actors respond. For example, if the tax rate increases it may become optimal for the IRS to audit more returns; in Graetz, Reinganum and Wilde (1986), with an unconstrained budget, an increase in the tax rate on the high-income taxpayers who are potential evaders decreases evasion. Whether this prediction turns out to be accurate depends on whether in practice the IRS budget increases concomitantly with the tax rate, and there is no empirical evidence that supports this.

7.1.5. Optimal allocation of enforcement resources

Administrative costs are inputs into the revenue raising process. But what should be the target of the administration, and how should economic considerations be introduced into the tax-revenue production function? To address this issue, one has to define the objectives of the tax administration and its production function – how much revenue is produced with different combinations of inputs (subject, of course, to the tax law). Then one can analyze whether the allocation of funds for administration is efficient or to check whether, as Tanzi and Pellechio (1997) put it, "personnel are often assigned to tasks that have low productivity while important functions get unattended".

Yitzhaki and Vakneen (1989) develop a model that introduces microeconomic considerations into the management of tax administration[44]. They assume that the objective of the administration is the maximization of revenue and that taxpayers can be classified into groups based on having returns of similar complexity. These assumptions allow them to present the inspection process of tax returns as a decision tree in which the "inspector" has to spend a given amount of his time to review

[44] See also Wertz (1979), the first modern treatment of the optimal allocation of the work force in a tax collection agency.

the return, and the reaction of the taxpayer (whether to appeal) is determined by the quality of the assessment. If they continue to disagree, the results are determined by the court. The solution to this decision tree problem can be determined in a dynamic programming model. Estimation of the decision tree enables one to estimate the present value of future tax revenue that is collected by each activity of the tax administration. Yitzhaki and Vakneen argue that an administration should equalize the rate of return, in terms of tax revenue, for each activity. This principle should govern sampling of tax returns for inspection, as well as which items on the return to inspect.

7.2. Optimal tax systems

The previous section addressed how to evaluate the appropriate setting of tax enforcement instruments, for a given specification of tax base and rates. The more general problem is to consider all of these aspects simultaneously, what Slemrod (1990) calls the theory of "optimal tax systems". Certainly, the ease of administering various taxes has critical implications for the optimal structure of tax systems. Tax codes which are based on unobservable and practically unmeasurable quantities (such as an ability tax) often look desirable on paper. Integrating the issue of administrative ease into normative tax theory requires a shift of emphasis away from the structure of preferences, which has been the principal focus of optimal tax theory, toward the technology of tax collection.

7.2.1. The choice of tax instruments

With some exceptions, optimal tax theory has dealt with the issue of administering a tax by making extreme assumptions about what kinds of taxes are available to the policymaker. The fundamental results of optimal tax theory depend on implicit assumptions about which taxes can be administered and which cannot. The problem of optimal commodity taxation is interesting only because the possibility of lump-sum taxation is ruled out[45], presumably because it is infeasible. Production efficiency is desirable only if all commodities can be taxed and 100 percent taxation of profits is feasible (or if no profits exist). When consumers are not identical, an ability tax dominates an income tax because it causes no distortion in behavior. The study of optimal income taxation is appropriate when ability taxes are ruled out, usually by appealing to the difficulties of measuring ability for the purpose of basing tax liability on it.

Extreme assumptions about the feasibility of tax instruments are analytically convenient, but incorrect. Ability can be measured, although with some expense and

[45] To be sure, the optimal commodity taxation literature yields insights about the less analytically tractable, but more realistic, multi-person environment. Nevertheless, in most models in which the use of lump-sum taxes is limited, this is done as an assumption rather than as a choice based on the costs and benefits of this instrument.

error. On the other hand, income cannot be measured perfectly, and the degree of accuracy in income measurement depends on the resources expended toward this goal.

Extreme assumptions about the feasibility of tax instruments may also preclude consideration of fundamental changes in policy[46]. For example, a common assumption made in optimal taxation models of developing countries is that income and consumption arising in the agricultural sector are not taxable, although marketable surplus is taxable. Much interesting analysis proceeds from this assumption, but none asks at what point it makes sense for a country to attempt to tax agricultural income, even assuming that it will have only limited success in doing so. There is clear evidence [Riezman and Slemrod (1987)] that countries with low literacy rates tend to rely on highly distorting but (relatively) easily collectable import and export taxes, and shy away from efficient but administratively difficult land taxes. Under what conditions should an imperfect land tax be tried? The answers to these questions depend on the resource cost of administering the new tax instrument relative to its effectiveness, or degree of success. This latter notion has several dimensions, including the true revenue yield and the extent and nature of the mistakes that are made in administration.

Stern (1982) models the choice between an optimal nonlinear income tax, in which income is costlessly observable, and a system of differential lump-sum taxes based on characteristics of taxpayers which can be ascertained with some error. The lump-sum tax system is superior if there are no errors in classifying individuals but, when enough mistakes are made, income taxation may be the preferred system. Stern's analysis recognizes that the two tax systems each have their own information requirements (the lump-sum system requires classifying individuals, the income tax system requires observing incomes). The two systems will also likely have different administrative costs as well, although for the sake of simplicity Stern assumes these costs are identical. Greater accuracy in the classification of individuals could be achieved with higher cost, as could more accurate measurement of income[47].

The optimal tax system framework has also been applied to a more immediately policy-relevant choice, that between direct and indirect taxes. It has frequently been claimed that a shift from income taxation to value added taxation can combat evasion by taxing the spending on goods from the compliant sector by individuals who evade taxes on their income. Boadway, Marchand and Pestieau (1994) consider the optimal mix between a general non-linear income tax and commodity taxes under the assumption that evasion is possible *only* for the income tax. Granting this assumption

[46] The desirability of introducing the choice among tax instruments into the optimal tax problem has been noted by, among others, Hahn (1973) and Atkinson and Stiglitz (1980, p. 363), who state that "for a complete theory of the choice of tax base, a fully articulated model is necessary of the information available to the government and cost of observing the different characteristics". Diamond (1987, p. 640) agrees that this would be ideal, but adds that the standard simplifications "may do little damage to the policy conclusions if the set of feasible policies is well chosen, although the problem of choosing well is a difficult one".

[47] This is an example of the issue of optimal "tagging" discussed in Akerlof (1978).

provides a strong case for commodity taxation to supplement an income tax. The authors recognize that the results would have to be "seriously adjusted" (p. 73, fn. 2) if there is more evasion on indirect than on direct taxes. In contrast, Kesselman (1993) concludes that changing the tax mix toward indirect taxes will have little or none of the claimed anti-evasion effects. Underlying this conclusion is a two-sector model in which the income tax is paid only by workers in the above-ground sector, and the indirect tax is paid completely by above-ground workers but incompletely in the underground sector. This is justified on the grounds that to evade the income tax successfully requires evasion of the indirect tax on output as well, since honest reporting of gross sales for the indirect tax would signal to the authorities the extent of the income tax evasion. Which analysis better captures the reality depends on the technology of tax avoidance and evasion.

7.2.2. Presumptive taxes

The general nature of the optimal tax systems problem is well illustrated by considering a class of taxes – known as presumptive taxes – which are a pervasive element in the tax systems of many developing countries. This kind of tax makes sense in cases where the otherwise desirable tax base is difficult for the tax authorities to measure, verify, and monitor. As a substitute for the desired base is the "presumed" tax base, which is derived from a formula, which itself may be simple or complex, based on more readily monitored items[48]. For example, at one time in Israel a taxi driver had a choice of a tax based on book income or a levy on the accumulated mileage of the taxicab; for shopkeepers, the alternative to a tax on book income was a tax based on the square footage of the shop and other observable characteristics of the business. The wide variety of presumptive taxes used in the developing world is nicely surveyed in Tanzi and Casanegra de Jantscher (1989) and in Rajaraman (1995).

The problem that presumptive taxes address – the difficulty of monitoring certain potential tax bases – is not confined to developing countries, and use of presumptive taxes, albeit with different names, is also widespread in developed countries. Examples include the use of fixed depreciation schedules in place of asset-specific measures of the decline in asset value (economic depreciation), taxation of capital gains on a realization basis, and floors on deductible expenses. Slemrod and Yitzhaki (1994) and Kaplow (1994) analyze the U.S. standard deduction as a presumptive tax; a higher value reduces the administrative and compliance cost of monitoring itemized deductions,

[48] There are two general categories of presumptive taxes. In the first, tax liability is based on an easily monitorable base which is presumably correlated with the ideal tax base. The tax on taxicab mileage or a tax on electricity used by a laundry are examples. In many cases, the monitorable base is a specific input, and the presumptive tax is actually a tax on an input. The second category includes (effective or *de facto*) exemptions or floors, intended to eliminate monitoring costs of "nonfruitful" populations. Examples include exempting businesses with less than a certain number of employees, or floors on deductible expenses.

but it increases horizontal inequity by increasing the range of taxpayers for which the "proper" amount of deduction is replaced by a single number[49].

Upon reflection it is clear that all taxes are presumptive, to some degree. The conceptually pure tax base – be it the flow of income, wealth, sales revenue, or something else – cannot be perfectly measured, and the tax authority is constrained to rely on some correlate of the concept. We label particular taxes as presumptive when the calculation of the tax base deviates in a substantial way from the ideal concept. But there is a pervasive tradeoff between accuracy and the costs of complexity[50].

7.2.3. Optimal commodity taxes

The characterization of optimal commodity taxes is a cornerstone of the standard theory of optimal taxation, dating back to Ramsey (1927). The standard theory, though, assumes that taxes on all commodities can be verified and collected costlessly.

Yitzhaki (1979) investigates the optimal size of the commodity tax base in a representative consumer economy when there is a resource cost, related to administration, to adding goods to the tax base. If, as he assumes, preferences over all goods are Cobb–Douglas, then uniformity of rate for all taxed goods is optimal. Expanding the tax base to cover more goods will reduce the excess burden of taxation, but it increases the administrative cost. The optimal tax system equates the marginal excess burden of raising a dollar of revenue to the marginal administrative cost, and thus minimizes the total resource cost of raising revenue. Wilson (1989) generalizes the framework to constant-elasticity-of-substitution utility functions.

The fact that changes in administrative costs are likely to be discontinuous with respect to changes in tax policy is important in more general treatments of the optimal set of tax instruments. The theory of optimal taxation tells us that, except in special cases, all goods should be taxed at different rates. However, it is likely that administrative cost depends on the number of different tax rates as well as the number of commodities taxed. This is not an issue when one assumes a utility function that implies uniform optimal taxes (e.g., Cobb–Douglas), but is very important under more general preferences; in that case there is a tradeoff between administrative and compliance costs on the one hand and the standard excess burden on the other.

Both the Yitzhaki and Wilson papers assume that a commodity is either in the tax base and taxed at the uniform rate, or out of the tax base entirely. Boadway, Marchand and Pestieau (1994), Cremer and Gahvari (1993) and Kaplow (1990) investigate general characterizations of optimal commodity taxation with evasion, administrative costs and costly enforcement[51]. In Cremer and Gahvari (1993), the optimal tax on a

[49] See also Sadka and Tanzi (1993), who argue in some situations for a presumptive tax on assets as a substitute for a income tax.

[50] Kaplow (1994, 1996) addresses the equity and efficiency issues involved in making this tradeoff.

[51] Skinner and Slemrod (1985) suggest that enforcement policy can be part of an optimal tax system in which the statutory tax rates are constrained to be suboptimal; for example, lax enforcement of a good whose statutory tax rate exceeds the optimal rate may be appropriate.

commodity is, *ceteris paribus,* lower when the elasticity of induced avoidance response to a tax increase is higher; intuitively, this increases the marginal social cost per dollar raised from taxing that commodity.

7.2.4. Optimal progressivity

In the optimal linear income tax literature, where only a demogrant and single marginal tax rate are chosen, what constrains redistribution is the marginal excess burden caused per dollar raised by the marginal tax rate, and the fact that this ratio increases with the marginal tax rate levied. Cremer and Gahvari (1994) investigate how the introduction of evasion and concealment expenses change the optimal setting of a linear income tax, when the audit probability is also optimally chosen. They characterize the optimal marginal tax rate in the presence of evasion, but conclude that one cannot hope for an unambiguous result in general about whether in a model with evasion the marginal tax rate is higher or lower compared to in a model without evasion.

If other aspects of the tax system are not set optimally, there is no presumption that the tax rate that is optimal, given the value of the other instruments, is also the global optimum. To be concrete, if enforcement instruments are set suboptimally, so that the marginal cost of raising revenue is higher than it need be, then the optimal tax rate will appear lower than if the enforcement parameters are set optimally.

The point is that the optimal level of taxes or tax progressivity can be properly assessed only simultaneously with the instruments the government uses to control avoidance and evasion. Slemrod (1994) constructs an example of this issue by modeling a two-person economy in which the only possible response to taxation is avoidance. The government must choose three instruments to maximize social welfare: a demogrant, a (single) marginal tax rate, and an avoidance-control policy denoted p, which at a cost reduces both the level of avoidance and its responsiveness to changes in the marginal tax rate. An example shows that, with p set suboptimally, the optimal policy can be to lower t; however, a superior policy is to raise both p and t. The intuition here is that the calculation of marginal excess burden of the marginal tax rate should be done assuming the other policy instruments are set optimally. Using the metaphor of Okun (1975), the "leak" in the revenue system, which limit both redistribution and the size of the public sector, can be "fixed", albeit at some cost.

Slemrod and Kopczuk (2001) expand on this notion by isolating the effect of a policy instrument on the elasticity of taxable income, which summarizes the magnitude of the behavioral response to taxes that limits optimal progressivity. They formally characterize the optimal elasticity, emphasizing that in many settings it is appropriate to think of this as a policy choice rather than an exogenous constraint. In a special case where the policy instrument is the breadth of the tax base, Slemrod and Kopczuk show that more egalitarian societies will feature lower elasticities of taxable income, as will societies with a lower marginal cost of tax administration. Thus, this research simultaneously addresses optimal progressivity and the optimal ease of collecting taxes, and focuses on a critical difference between real substitution responses on the

one hand and avoidance and evasion responses on the other. Economists nearly always assume that the former is an immutable, or primitive, parameter that is immune to policy (or any kind of) manipulation. Whatever the truth of that assumption as it applies to, say, labor supply response to taxation, it is certainly untenable as it applies to avoidance and evasion responses. Their availability is certainly a (perhaps highly constrained) policy choice. Truly optimal tax policy does not accept the current state of administration and enforcement as given, but instead chooses these aspects and the statutory tax structure together.

7.3. The marginal efficiency cost of funds

A principal theme of this chapter is that acknowledging the range of behavioral responses to taxation suggests a rich set of new empirical and conceptual issues and alters the answers to some fundamental questions of public finance. For some other questions, though, the anatomy of behavioral response may not matter. For example, Feldstein (1999) argues that, for the purpose of calculating the marginal efficiency cost of taxation, the critical parameter is the tax rate elasticity of taxable income, and the etiology of the elasticity – be it increased leisure, evasion, or increased untaxed fringe benefits, for example – is irrelevant. The intuition is that at the margin people are willing to incur a dollar's worth of cost to save a dollar of taxes, and that cost may take the form of a distorted consumption basket, a fee to an accountant, or increased exposure to the risk of punishment for evasion. However, because Feldstein derives this conclusion in a model which allows real substitution response but neither avoidance nor evasion, it begs the question of whether the taxable income elasticity is a sufficient statistic for measuring the efficiency cost of raising taxes and for comparing the relative efficiency of alternative ways to raise revenue.

This problem has been treated in the context of the concept referred to as the marginal cost of funds or marginal efficiency cost of funds, and was developed by Usher (1986), Mayshar (1990, 1991), Wildasin (1984), and Slemrod and Yitzhaki (1996). This model also allows us to place the issues raised above into a more general normative framework. We first discuss the concept in the absence of administrative costs, evasion, or avoidance, and then extend it to apply to these issues[52].

Following Mayshar (1991), assume that the government sets a level of public goods, G, and a vector E of tax policy instruments so as to maximize $V(U^*(E, w), G)$, where w is the wage rate and U^* is the utility derived from private goods. He shows that the optimum is characterized by MBF = MCF_i, where MBF is the social marginal benefit of funds (in terms of private consumption), and MCF_i is the marginal cost of

[52] Although what follows can be generalized to apply to a multi-individual framework and to public goods, here we restrict ourselves to the representative individual model of tax analysis. See Slemrod and Yitzhaki (2001) for a treatment of the more comprehensive problem.

funds of tax instrument i. Mayshar and Yitzhaki (1995) decompose the MCF_i term into:

$$MCF_i = DC_i^* MECF_i, \tag{10}$$

where DC_i is Feldstein's (1972) distributional characteristic of the tax instrument, while $MECF_i$ is the marginal efficiency cost of the tax instrument. In the absence of evasion or avoidance, $MECF_i$ is equal to X_i/MR_i, where X_i is the change in revenue assuming no behavioral response, and MR_i (marginal revenue) allows behavioral response. Thus, in the case of an income tax, X_i/MR_i equals $1/(1 + \varepsilon_i)$, where ε_i is the elasticity of taxable income with respect to tax instrument i.

Note that the above interpretation is not limited to reforms involving tax rates. One may define the marginal cost of funds with respect to marginal changes in any parameter of the tax system (e.g., income brackets, exemption levels, penalties for tax evasion, etc.). Nor does its application rely on an assumption that tax policy has been set optimally. As Slemrod and Yitzhaki (1996) show, away from the optimum the MECF concept can be used to identify incremental changes in the tax system that would increase social welfare.

To see how the MECF can be extended to evasion and avoidance, recall that the potential change in tax revenue (assuming an inelastic base) is X_i but, because of taxpayers' response, the government collects only MR_i. We can divide the potential tax X_i into two components as follows:

$$X_i = (X_i - MR_i) + MR_i, \tag{11}$$

where MR_i dollars are collected and $(X_i - MR_i)$ "leaks" outside the tax system. The critical question is how to evaluate, from a social point of view, the leaked dollars. To do this one must ask how much a taxpayer is ready to expend (on the margin) to save a dollar of taxes or, alternatively, how much utility loss he is willing to suffer to save a dollar of taxes. The answer is that a rational taxpayer will be ready to sacrifice up to, but no more than, one dollar in order to save a dollar of taxes. Hence, on the margin the private cost, which is equal to "leaked" dollars multiplied by their cost per dollar, is $X_i - MR_i$; the collection of MR_i dollars results in a loss of $(X_i - MR_i)$ to the taxpayer over and above the taxes paid. If we assume that the utility loss to the individual (private cost) of the leaked tax revenue should be accorded the same social cost as the utility loss due to the taxes paid, then the cost to society of transferring a dollar to the government is $(X_i - MR_i)/MR_i = (X_i/MR_i) - 1$. The total marginal cost to the individual taxpayer, including the taxes paid, is X_i/MR_i.

Consider now a taxpayer who also has the option to evade part of the additional tax. On the margin, he would be ready to sacrifice utility valued at one dollar (in additional risk bearing due to evasion and/or due to substitution to cheaper but less rewarding activities) in order to save a dollar of taxes. Hence, we do not have to know whether the "leak" was through evasion or real substitution to evaluate the costs to society.

The same rule applies to avoidance activity and, in fact, to any activity under taxpayer control. Therefore, all one needs to know is the potential tax (i.e., assuming an inelastic tax base) that will be collected from a change of a parameter of the tax system, and the actual change (taking into account all behavioral responses) in order to evaluate the marginal efficiency cost of raising revenue. It is in this sense that Feldstein's (1999) claim about the central importance of the elasticity of taxable income generalizes to avoidance and evasion.

Calculating the MECF involves two critical assumptions that deserve further attention. The first of these is that at the margin the taxpayer sacrifices exactly one dollar (instead of *up to* one dollar) to reduce tax liability by one dollar. However, it may be that the taxpayer is at a corner solution with respect to behavioral response, so that the marginal utility loss may be less than a dollar. For an example of a taxpayer at a corner, consider the case of Individual Retirement Accounts (IRAs). An employee can contribute up to $2000 per year into an IRA, deduct the contribution from taxable income, pay no tax on accrued earnings in the account, and pay tax on the principal when withdrawn. Although IRAs were designed to increase saving, there is nothing to prevent an individual who in the absence of taxes would have invested $4000 in a similar account from diverting $2000 into the IRA. There may be a cost to this, as IRAs have early withdrawal penalties which in some cases limit the flexibility of using these funds. Thus, contributing to an IRA can save taxes, does not require a change in one's consumption basket, but may entail some cost. However, it cannot be presumed that, at the margin of an IRA contribution, the private value of the sacrifice is equal to the tax saving; the IRA contribution is limited to $2000 only because of the statutory limit on contributions. As another example, consider the MECF of raising the tax rate on labor income in a situation where, in an economy with two taxpayers, one taxpayer reports no labor income at all and, at that corner, is bearing risk valued at 20 cents (rather than a dollar, as would be true at an interior solution) to evade, including penalties, an expected value of one dollar. The other taxpayer, with identical labor income, reports all of it. Assuming no labor supply or avoidance response, the MECF with respect to an increased tax rate is 1.2.

To take account of the possibility of the taxpayer being at such corner solutions, one can generalize the expression for the MECF by introducing a parameter γ, $0 < \gamma \leqslant 1$, which is a weighted average of the marginal value to the taxpayers of the leaked revenue, $X_i - \mathrm{MR}_i$. Introducing γ reduces the simplicity of the MECF expression because its value varies depending on the situation under study.

The second critical assumption is that the cost borne by taxpayers in the process of reducing tax liability is equivalent to the social cost. This is certainly true in many situations, such as when the private cost takes the form of a distorted consumption basket. But in some cases the private cost is not identical to the social cost. An example is when the act of the taxpayer causes some externality. Consider the case where being caught evading imposes a stigma on the taxpayer, as in Benjamini and Maital (1985) or Gordon (1989), and assume that the larger the number of evaders the lower the stigma attached to each act. In this case the social cost of evading taxes diverges from

the private cost because the potential evader does not take into account the impact of his action on other members of the society.

Fines for tax evasion present another example of the potential divergence between the private and social costs of tax-reducing activities. The possibility of a fine for detected tax evasion is certainly viewed as a cost by the taxpayer, but from society's point of view it reduces the amount of revenue that would otherwise have to be collected. (This is in contrast to imprisonment, unless the prisoner is forced to produce socially valuable products while imprisoned.) Thus, the MR term should include fine collections. Note that, if the fine itself is the policy instrument, this argument implies that its MECF could be close to zero, and almost certainly less than one, making an increase in fines look like an attractive policy option indeed. As discussed in Section 7.1.1, there are reasons unrelated to efficiency cost minimization which render undesirable increasing fines for tax evasion without limit.

Applying the MECF rule to administrative and compliance issues clarifies the common thread running through models of optimal tax systems. In the generic problem, there are two ways to raise revenue: to increase a set of tax rates, and by so doing to increase excess burden, or via an alternative which involves increasing administrative costs [e.g., by broadening the tax base as in Yitzhaki (1979), or by increasing the probability of a tax audit, as in Slemrod and Yitzhaki (1987)]. On the margin, it is optimal to equalize the marginal costs of raising revenue under the two alternatives. If one defines the costs of taxation as deadweight loss plus administrative costs, at an optimum the MECF of each tax rate should be equal to the MECF of administrative improvements that raise revenue. In calculating the MECF of administrative improvements, it is important to account for the fact that these expenses come out of funds that were presumably raised with tax instruments that have an MECF in excess of one. In other words, administrative improvements that raise net revenue decrease the excess burden; hence, on the margin and for given revenue, the saving in excess burden should be equal to the increase in administrative costs. In this way, the MECF criterion can be applied to tax administration, too[53].

Compliance costs are additional costs imposed on the taxpayer. Therefore, they should be added to the burden imposed on the taxpayer. They serve as a substitute to administrative costs, but the expenses are borne directly by the taxpayer rather than through the government budget.

The revised MECF that includes all these factors, derived and discussed in Slemrod and Yitzhaki (1996), is

$$\mathrm{MECF}_i = \frac{\gamma(X_i - \mathrm{MR}_i) + C_i + \mathrm{MR}_i}{\mathrm{MR}_i - A_i}, \tag{12}$$

[53] Yitzhaki and Vakneen (1989) use the term "the shadow price of a tax inspector", which is the revenue collected by adding another tax inspector. Note that the MECF is actually the reciprocal of the shadow price of a tax inspector.

where γ is the social value of the utility the taxpayer is sacrificing at the margin in order to save a dollar of tax. C_i is the marginal private compliance cost associated with the ith instrument, A_i is the marginal administrative cost, and $MR_i - A_i$ is the net revenue collected at the margin. The intuitive interpretation of the expression is the same as before, with some qualifications. The potential tax is X_i. $X_i - MR_i$ is leaked at a social cost of γ per dollar, MR_i is collected by the government, and C_i is the additional involuntary compliance cost. Hence, the total burden on society is the sum of those components. Of the MR_i collected by the government, A_i is spent on administration, leaving $MR_i - A_i$ in the coffers. The MECF is the burden on society divided by what is collected after subtracting the cost of doing business. This yields the marginal costs of a dollar collected.

Because in Equation (12), C_i is added in the numerator and A_i is subtracted in the denominator, the key conceptual difference between the two is explicit – only the latter uses revenue raised from taxpayers. To illustrate this difference, consider that a tax for which $C_i = MR_i$ (with A_i and $X_i - MR_i = 0$) might conceivably be part of an optimal tax regime (if the MECFs of other instruments exceed two), but it would never be optimal to have $A_i = MR_i$, for at the margin this instrument has social cost but raises no revenue.

As emphasized in Slemrod (1998), applying this notion using empirical estimates of the taxable income elasticity must be done with care. Foremost is the need to consider the elasticity of the *present value* of tax revenues. Recall that a class of avoidance responses involves the retiming of taxable-income-generating events. If a tax policy change causes retiming, focusing only on the revenues in a subset of periods will bias the findings. For example, the taxable income response to an anticipated future decrease in tax rates must consider the lost revenue in the period before the tax rate changes[54]. Similarly, if a tax change causes an increase in deferred compensation, the increased future tax liability must be netted against any decline in current tax payments. Furthermore, any change in taxable income in one tax base must be netted against changes in taxable income in other bases. For example, if a decline in personal tax rates causes a shift from C corporation status to S corporation status, the increased personal taxable income must be netted against decreases in corporate taxable income.

8. Conclusion

The possibilities for evasion and the difficulties of administration have always shaped tax systems. Until recently, formal analysis of taxation largely ignored these realities. After a quarter of a century of research on the topic, it is time to put to rest the claim

[54] Feldstein (1999) recognizes the timing effect, but whether the empirical analysis of the 1986 tax changes from which he derives a taxable income elasticity [Feldstein (1995)] does or does not is a controversial question, with Slemrod (1996b) arguing that timing effects dominate the results.

that this is an understudied area. Instead, it is a vibrant area of research that has clarified the positive and normative analysis of taxation.

The research has clarified that when the tax structure changes, people may alter their consumption basket, but they also may call and give new instructions to their accountant, change their reports to the IRS, change the timing of transactions, and effect a set of other actions that do not directly involve a change in their consumption basket. In many cases, particularly for high-income taxpayers, this latter set of responses has larger revenue and welfare implications than the real substitution responses, such as labor supply, that tax analysis has traditionally focused on.

Early models of this area focused on tax evasion, modeled as a gamble against the enforcement capability of the state. More recently, the literature has examined more general models of the technology of avoidance, with the additional risk bearing caused by tax evasion either being a special case of this technology or one aspect of the cost of changing behavior to reduce tax liability. A critical aspect of this technology is whether the avoidance is inframarginal, in which case only income effects are involved, or whether its cost depends on other aspects of behavior. If the latter is true, the choice of consumption basket and avoidance become intertwined because certain activities may facilitate avoidance, which alters their effective relative price or return.

Acknowledging the variety of behavioral responses to taxation greatly enriches the normative analysis of taxation. It changes the answers to traditional subjects of inquiry, such as incidence, optimal progressivity, optimal commodity taxation, and the optimal mix between income and consumption taxes. It also raises a whole new set of policy questions, such as the appropriate level of resources to devote to administration and enforcement, and how those resources should be deployed. A recurring question that runs throughout this chapter is whether the standard toolkit of positive and empirical analysis can be applied to avoidance, evasion, and administration. The answer is a qualified yes, as this chapter hopefully demonstrates.

In one respect, though, the policy perspective does change in an important way. The magnitude of real substitution response, such as labor supply, to taxes is presumed to be an immutable function of preferences, and not susceptible to policy manipulation in a free society. With respect to avoidance and evasion, though, this hands-off approach is not appropriate. On the contrary, there are a variety of policy instruments that can affect the magnitude and nature of avoidance and evasion response, ranging from the activities of the enforcement agency to how tightly drawn are rules and regulations. The same kind of cost–benefit calculus applies to the choice of these instruments, implying that the elasticity of behavioral response is itself a policy instrument, to be chosen optimally.

A key challenge for the future is to add more empirical content to the theoretical models of taxpayer and tax agency behavior. This will require, *inter alia,* addressing the technology of raising and avoiding taxes. This is the analogue to the critical role for traditional taxation theory of the empirical investigation of the structure of individuals' preferences. Although by their nature the appropriate data are often difficult to come

by, new approaches such as controlled field experiments and analysis of changes in tax administration are promising.

It would also be fruitful to incorporate public choice considerations into the analysis. In some case administrative difficulties as well as widespread avoidance and evasion are caused by the inability of compromise-seeking legislators to agree upon a well-defined law. Furthermore, there is apparently no political constituency for tax simplicity and facilitated adminstration. Combining analysis of the public choice mechanisms that produce tax systems with the kind of normative analyses discussed in this chapter may lead to a more complete understanding of the reality of taxation.

References

Agell, J., and M. Persson (1990), "Tax arbitrage and the redistributive properties of progressive income taxation", Economic Letters 34(4):357–361.

Akerlof, G. (1978), "The economics of 'tagging' as applied to the optimal income tax, welfare programs, and manpower planning", American Economic Review 68(1):8–19.

Allingham, M.G., and A. Sandmo (1972), "Income tax evasion: a theoretical analysis", Journal of Public Economics 1(3–4):323–338.

Alm, J. (1988), "Compliance costs and the tax avoidance-tax evasion decision", Public Finance Quarterly 16(1):31–66.

Alm, J. (1999), "Tax compliance and administration", in: W. Bartley Hildreth and James A. Richardson, eds., Handbook on Taxation (Marcel Dekker, New York) pp. 741–768.

Alm, J., R. Bahl and M.N. Murray (1991), "Tax base erosion in developing countries", Economic Development and Cultural Change 39(4):849–872.

Alm, J., B.R. Jackson and M. McKee (1992), "Estimating the determinants of taxpayer compliance with experimental data", National Tax Journal 45(1):107–114.

Alt, J. (1983), "The evolution of tax structures", Public Choice 41(1):181–222.

Andreoni, J. (1992), "IRS as loan shark: tax compliance with borrowing constraints", Journal of Public Economics 49(1):35–46.

Andreoni, J., B. Erard and J. Feinstein (1998), "Tax compliance", Journal of Economic Literature 36(2):818–860.

Atkinson, A.B., and J.E. Stiglitz (1980), Lectures in Public Economics (McGraw-Hill, New York).

Baldry, J.C. (1979), "Tax evasion and labor supply", Economics Letters 3(1):53–56.

Baldry, J.C. (1984), "The enforcement of income tax laws: efficiency implications", Economic Record 60(169):156–159.

Baldry, J.C. (1987), "Income tax evasion and the tax schedule: some experimental results", Public Finance 42(3):357–383.

Balke, N.S., and G.W. Gardner (1991), "Tax collection costs and the size of government", Mimeo (Southern Methodist University).

Beck, P., and W.O. Jung (1987), "An economic model of taxpayer compliance under complexity and uncertainty", Journal of Accounting and Public Policy 8:1–27.

Becker, G.S. (1968), "Crime and punishment: an economic approach", Journal of Political Economy 76(2):169–217.

Benjamini, Y., and S. Maital (1985), "Optimal tax evasion and optimal tax policy", in: W. Gartner and A. Wenig, eds., The Economics of the Shadow Economy (Springer, Berlin) pp. 245–264.

Beron, K.J., H.V. Tauchen and A.D. Witte (1992), "The effect of audits and socioeconomic variables on compliance", in: Joel Slemrod, ed., Why People Pay Taxes (University of Michigan Press, Ann Arbor) pp. 67–89.

Bird, R.M. (1983), "Income tax reform in developing countries; the administrative dimension", Bulletin for International Fiscal Documentation 37:3–14.
Bishop, J.A., K.V. Chow, J.P. Formby and C.-C. Ho (1994), "The redistributive effects of noncompliance and tax evasion in the U.S", in: John Creedy, ed., Taxation, Poverty, and Income Distribution (Edward Elgar, Aldershot) pp. 28–47.
Blough, R. (1952), The Federal Taxing Process (Prentice-Hall, Washington, D.C.)
Blumenthal, M., C.W. Christian and J. Slemrod (2001), "Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota", National Tax Journal 54(1):125–138.
Boadway, R., M. Marchand and P.M. Pestieau (1994), "Towards a theory of the direct-indirect tax mix", Journal of Public Economics 55(1):71–88.
Burman, L., and W. Randolph (1994), "Measuring permanent responses to capital gains taxes in panel data", American Economic Review 84(4):794–809.
Burman, L., K. Clausing and J. O'Hare (1994), "Tax reform and realizations of capital gains in 1986", National Tax Journal 45(1):1–18.
Casanegra de Jantscher, M. (1990), "Administering the VAT", in: Malcolm Gillis, Carl S. Shoup and Gerardo Sicat, eds., Value-Added Taxation in Developing Countries (The World Bank, Washington, D.C.) pp. 171–179.
Christian, C.W. (1994), "Voluntary compliance with the individual income tax: results from the 1988 TCMP study", in: The IRS Research Bulletin, 1993/1994, Publication 1500 (Rev. 9-94) (Internal Revenue Service, Washington, D.C.) pp. 35–42.
Clotfelter, C.T. (1983), "Tax evasion and tax rates: an analysis of individual returns", Review of Economics and Statistics 65(3):363–373.
Cowell, F.A. (1990a), "Tax sheltering and the cost of evasion", Oxford Economic Papers 42(1):231–243.
Cowell, F.A. (1990b), Cheating the Government: The Economics of Evasion (MIT Press, Cambridge).
Cremer, H., and F. Gahvari (1993), "Tax evasion and optimal commodity taxation", Journal of Public Economics 50(2):261–275.
Cremer, H., and F. Gahvari (1994), "Tax evasion, concealment, and the optimal linear income tax", Scandinavian Journal of Economics 96(2):219–239.
Cremer, H., and F. Gahvari (1996), "Tax evasion and the optimum general income tax", Journal of Public Economics 60(2):235–249.
Cross, R., and G.K. Shaw (1982), "On the economics of tax aversion", Public Finance 37(1):36–47.
Diamond, P.A. (1987), "Optimal tax theory and development policy: directions for future research", in: David Newbery and Nicholas H. Stern, eds., The Theory of Taxation for Developing Countries (Oxford University Press, New York) pp. 638–647.
Dixit, A.K. (1991), "The optimal mix of inflationary finance and commodity taxation with collection lags", International Monetary Fund Staff Papers 38(3):643–654.
Domar, E.D., and R.A. Musgrave (1944), "Proportional income taxation and risk-taking", Quarterly Journal of Economics 58:388–423.
Dubin, J.A., and L.L. Wilde (1988), "The empirical analysis of federal income tax auditing and compliance", National Tax Journal 41(1):61–74.
Dubin, J.A., M.J. Graetz and L.L. Wilde (1990), "The effect of audit rates on the federal individual income tax, 1977–1986", National Tax Journal 43(4):395–409.
Engel, E., and J.R. Hines Jr (1999), "Understanding tax evasion dynamics", Working Paper 6903 (National Bureau of Economic Research).
Erard, B. (1997), "Self-selection with measurement errors: a microeconometric analysis of the decision to seek tax assistance and its implications for tax compliance", Journal of Econometrics 81(2):319–356.
Feige, E.L. (1979), "How big is the irregular economy?" Challenge 22(5):5–13.
Feinstein, J. (1991), "An econometric analysis of income tax evasion and its detection", RAND Journal of Economics 22(1):14–35.
Feldstein, M. (1972), "Distributional equity and the optimal structure of public prices", American Economic Review 62(1):32–36.

Feldstein, M. (1995), "The effect of marginal tax rates on taxable income: a panel study of the 1986 Tax Reform Act", Journal of Political Economy 103(3):551–572.
Feldstein, M. (1999), "Tax avoidance and the deadweight loss of the income tax", Review of Economics and Statistics 81(4):674–680.
Goolsbee, A. (2000), "What happens when you tax the rich: evidence from executive compensation", Journal of Political Economy 108(2):352–378.
Gordon, J. (1989), "Individual morality and reputation costs as deterrents to tax evasion", European Economic Review 33(4):797–805.
Gordon, R.H., and J.K. MacKie-Mason (1990), "Effects of the Tax Reform Act of 1986 on corporate financial policy and organizational form", in: Joel Slemrod, ed., Do Taxes Matter? The Impact of the Tax Reform Act of 1986 (MIT Press, Cambridge, MA) pp. 91–131.
Gordon, R.H., and J.K. MacKie-Mason (1997), "How much do high taxes discourage incorporation?" Journal of Finance 52(2):477–505.
Gordon, R.H., and J. Slemrod (1988), "Do we collect any revenue from taxing capital income?" in: Lawrence Summers, ed., Tax Policy and the Economy (MIT Press and National Bureau of Economic Research, Cambridge, MA) pp. 89–130.
Gordon, R.H., and J. Slemrod (2000), "Are 'real' responses to taxation simply shifting between corporate and personal tax bases?" in: Joel Slemrod, ed., Does Atlas Shrug? The Economic Consequences of Taxing the Rich (Harvard University Press, Cambridge; the Russell Sage Foundation, New York) pp. 240–280.
Gordon, R.H., and J.D. Wilson (1989), "Measuring the efficiency cost of taxing risky capital income", American Economic Review 79(3):427–439.
Graetz, M.J., J.F. Reinganum and L.L. Wilde (1986), "The tax compliance game: toward an interactive theory of tax enforcement", Journal of Law, Economics, and Organization 2(1):1–32.
Groves, H.M. (1974), Tax Philosophers: Two Hundred Years of Thought in Great Britain and the United States (The University of Wisconsin Press, Madison). Published posthumously; edited by Donald J. Curran.
Gutmann, P.M. (1977), "The subterranean economy", Financial Analysts Journal 33(6):26–34.
Hahn, F. (1973), "On optimum taxation", Journal of Economic Theory 6(1):96–106.
Hinrichs, H.H. (1966), A General Theory of Tax Structure Change During Economic Development (The Law School of Harvard University, Cambridge, MA).
Internal Revenue Service (1996), "Federal tax compliance research: individual income tax gap estimates for 1985, 1988, and 1992", Publication 1415 (Internal Revenue Service, Washington, D.C.)
Kanbur, S.M. (1979), "On risk taking and the personal distribution of income", Journal of Political Economy 87(4):769–797.
Kaplow, L. (1990), "Optimal taxation with costly enforcement and evasion", Journal of Public Economics 43(2):221–236.
Kaplow, L. (1994), "The standard deduction and floors in the income tax", Tax Law Review 50:1–31.
Kaplow, L. (1996), "How tax complexity and enforcement affect the equity and efficiency of the income tax", National Tax Journal 45(1):135–150.
Kau, J.B., and P.H. Rubin (1981), "The size of government", Public Choice 37(2):261–274.
Kay, J.A. (1980), "The anatomy of tax avoidance", in: Income Distribution: The Limits to Redistribution; Proceedings of the 31st Symposium of the Colston Research Society, University of Bristol (John Wright) pp. 135–148.
Kesselman, J.R. (1989), "Income tax evasion: an intersectoral analysis" Journal of Public Economics 38(2):137–182.
Kesselman, J.R. (1993), "Evasion effects of changing the tax mix", The Economic Record 69(205): 131–148.
Klepper, S., and D. Nagin (1989), "The anatomy of tax evasion", Journal of Law, Economics, and Organization 5(1):1–24.
Kolm, S.-C. (1973), "A note on optimum tax evasion", Journal of Public Economics 2(3):265–270.

Long, S. (1981), "Social control in the civil law: the case of income tax enforcement", in: H.L. Ross, ed., Law and Deviance (Sage Publications, Beverly Hills) pp. 185–214.
Maki, D.M. (1996), "Portfolio shuffling and tax reform", National Tax Journal 49(3):317–329.
Mansfield, C. (1988), "Tax administration in developing countries", International Monetary Fund Staff Papers 35(1):181–197.
Mayshar, J. (1990), "On measures of excess burden and their application", Journal of Public Economics 43(3):263–289.
Mayshar, J. (1991), "Taxation with costly administration", Scandinavian Journal of Economics 93(1): 75–88.
Mayshar, J., and S. Yitzhaki (1995), "Dalton-improving tax reform", American Economic Review 85(4):793–807.
Melumad, N., and D. Mookherjee (1989), "Delegation as commitment: the case of income tax audits", RAND Journal of Economics 104(2):139–163.
Mookherjee, D., and I.P.L. Png (1989), "Optimal auditing, insurance, and redistribution", Quarterly Journal of Economics 104(2):399–415.
Musgrave, R.A. (1969), Fiscal Systems (Yale University Press, New Haven and London).
Okun, A.M. (1975), Equality and Efficiency: The Big Tradeoff (The Brookings Institution, Washington, D.C.).
Pencavel, J. (1979), "A note on income tax evasion, labor supply and nonlinear tax schedules", Journal of Public Economics 12(1):115–124.
Polinsky, A.M., and S. Shavell (2000), "The economic theory of public enforcement of law", Journal of Economic Literature 38(1):45–76.
Rajaraman, I. (1995), "Presumptive direct taxation: lessons from experience in developing countries", Economic and Political Weekly (Mombai, India), May 6-13.
Ramsey, F.P. (1927), "A contribution to the theory of taxation", Economic Journal 37:47–61.
Randolph, W. (1995), "Dynamic income, progressive taxes, and the timing of charitable contributions", Journal of Political Economy 103(4):70–138.
Reinganum, J.F., and L.L. Wilde (1985), "Income tax compliance in a principal agent framework", Journal of Public Economics 26(1):1–18.
Richupan, S. (1984), "Measuring tax evasion", Finance and Development 21(4):38–40.
Riezman, R., and J. Slemrod (1987), "Tariffs and collection costs", Weltwirtschaftliches Archiv 123(3): 545–549.
Rosen, H.S. (1976), "Tax illusion and the labor supply of married women", Review of Economics and Statistics 58(2):485–507.
Roth, J.A., J.T. Scholz and A.D. Witte (1989), Taxpayer Compliance, Vol. 1, An Agenda for Research; Vol. 2, Social Science Perspectives (University of Pennsylvania Press, Philadelphia).
Sadka, E., and V. Tanzi (1993), "A tax on gross assets of enterprises as a form of presumptive taxation", Bulletin for International Fiscal Documentation 47(2):66–73.
Sanchez, I., and J. Sobel (1993), "Hierarchical design and enforcement of income tax policies", Journal of Public Economics 50(3):345–369.
Sandford, C. (1973), The Hidden Costs of Taxation (Institute for Fiscal Studies, London).
Sandford, C., ed. (1995), Tax Compliance Costs: Measurement and Policy (Fiscal Publications, Bath).
Scholz, J.K. (1994), "Tax progressivity and household portfolios: descriptive evidence from the surveys of consumer finances", in: Joel Slemrod, ed., Tax Progressivity and Income Inequality (Cambridge University Press, Cambridge) pp. 219–267.
Scotchmer, S. (1987), "Audit classes and tax enforcement policy", American Economic Review 77(2): 229–233.
Scotchmer, S. (1989), "Who profits from taxpayer confusion?" Economics Letters 29(1):49–55.
Scotchmer, S., and J. Slemrod (1989), "Randomness in tax enforcement", Journal of Public Economics 38(1):17–32.

Shoup, C., R. Blough and M. Newcomer, eds (1937), Facing the Tax Problem (Twentieth Century Fund, New York).
Skinner, J., and J. Slemrod (1985), "An economic perspective on tax evasion", National Tax Journal 38(3):345–353.
Slemrod, J. (1990), "Optimal taxation and optimal tax systems", Journal of Economic Perspectives 4(1):157–178.
Slemrod, J. (1994), "Fixing the leak in Okun's bucket: optimal tax progressivity when avoidance can be controlled", Journal of Public Economics 55(1):41–51.
Slemrod, J. (1996a), "Which is the simplest tax system of them all?" in: Henry Aaron and William Gale, eds., Economic Effects of Fundamental Tax Reform (The Brookings Institution, Washington, D.C.) pp. 355–391.
Slemrod, J. (1996b), "High-income families and the tax changes of the 1980s: the anatomy of behavioral response", in: Martin Feldstein and James Poterba, eds., Empirical Foundations of Household Taxation (University of Chicago Press, Chicago; and NBER) pp. 169–192.
Slemrod, J. (1998), "Methodological issues in measuring and interpreting taxable income elasticities", National Tax Journal 51(4):773–788.
Slemrod, J. (2001), "A general model of the behavioral response to taxation", International Tax and Public Finance 8(2):119–128.
Slemrod, J., and W. Kopczuk (2001), "The optimal elasticity of taxable income", Journal of Public Economics, forthcoming.
Slemrod, J., and S. Yitzhaki (1987), "The optimal size of a tax collection agency", Scandinavian Journal of Economics 89(2):183–192.
Slemrod, J., and S. Yitzhaki (1994), "Analyzing the standard deduction as a presumptive tax", International Tax and Public Finance 1(1):25–34.
Slemrod, J., and S. Yitzhaki (1996), "The cost of taxation and the marginal efficiency cost of funds", International Monetary Fund Staff Papers 43(1):172–198.
Slemrod, J., and S. Yitzhaki (2001), "Integrating expenditure and tax decisions: the marginal cost of funds and the marginal benefit of projects", National Tax Journal 54(2):189–201.
Slemrod, J., M. Blumenthal and C.W. Christian (2001), "Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota", Journal of Public Economics 79(3): 455–483.
Sørensen, P.B. (1994), "From the global income tax to the dual income tax: recent tax reforms in the Nordic countries", International Tax and Public Finance 1(1):57–79.
Stern, N.H. (1982), "Optimum taxation with errors in administration", Journal of Public Economics 17(2):181–211.
Steuerle, C.E. (1985), Taxes, Loans, and Inflation: How the Nation's Wealth Becomes Misallocated (The Brookings Institution, Washington, D.C.).
Stiglitz, J.E. (1982), "Utilitarianism and horizontal equity: the case for random taxes", Journal of Public Economics 18(1):1–33.
Stiglitz, J.E. (1985), "The general theory of tax avoidance", National Tax Journal 38(3):325–337.
Tanzi, V. (1980), "The underground economy in the United States: estimates and implications", Banco Nazionale del Lavoro Quarterly Review 135:427–453.
Tanzi, V. (1992), "Theory and policy: a comment on Dixit and on current theory", International Monetary Fund Staff Papers 39(4):957–966.
Tanzi, V., and M. Casanegra de Jantscher (1989), "The use of presumptive income in modern tax systems", in: Aldo Chiancone and Ken Messere, eds., Changes in Revenue Structures; Proceedings of the 42nd Congress of the International Institute of Public Finance (Wayne University Press, Detroit) pp. 37–51.
Tanzi, V., and A. Pellechio (1997), "The reform of tax administration", in: Christopher Clague, ed., Institutions and Economic Development: Growth and Governance in Less-Developed and Post-Socialist Countries (Johns Hopkins Press, Baltimore) pp. 273–292.

Usher, D. (1986), "Tax evasion and the marginal cost of public funds", Economic Inquiry 24(4):563–586.
Weiss, L. (1976), "The desirability of cheating incentives and randomness in the optimal income tax", Journal of Political Economy 84(6):1343–1352.
Wertz, K.L. (1979), "Allocation by and output of a tax-administering agency", National Tax Journal 32(2):143–156.
Wildasin, D.E. (1984), "On public good provision with distortionary taxation", Economic Inquiry 22(2):227–243.
Wilson, J.D. (1989), "On the optimal tax base for commodity taxation", American Economic Review 79(5):1196–1206.
Yaniv, G. (1988), "Withholding and non-withheld tax evasion", Journal of Public Economics 35(2): 183–204.
Yaniv, G. (1992), "Collaborated employee–employer tax evasion", Public Finance 47(2):312–321.
Yitzhaki, S. (1974), "A note on 'income tax evasion: a theoretical analysis", Journal of Public Economics 3(2):201–202.
Yitzhaki, S. (1979), "A note on optimal taxation and administrative costs", American Economic Review 69(2):475–480.
Yitzhaki, S. (1987), "On the excess burden of tax evasion", Public Finance Quarterly 15(2):123–137.
Yitzhaki, S., and Y. Vakneen (1989), "On the shadow price of a tax inspector", Public Finance 44(3):492–505.