

**Optimal Tax Enforcement:
A Review of the Literature and Practical Implications**

by

Janet G. McCubbin*
U.S. Department of the Treasury

OTA Working Paper 90 December 2004

OTA Papers is an occasional series of reports on the research, models, and data sets developed to inform and improve Treasury's tax policy analysis. The papers are works in progress and subject to revision. Views and opinions expressed are those of the authors and do not necessarily represent official Treasury positions or policy. *OTA papers* are distributed to document OTA analytic models and data and to invite discussions and suggestions for revision and improvement. Comments are welcome and should be directed to the authors.

The author thanks Peter Brady, Janet Holtzblatt, Don Kiefer, Drew Lyon, Mark Mazur, Jim Nunns, Pamela Olson, Alan Plumley, Eric Toder, James Wetzler and many at the Internal Revenue Service and Department of the Treasury for helpful discussions and comments. All remaining errors and omissions are the author's. Comments welcome at Janet.G.McCubbin@irs.gov.

* Janet McCubbin was in the Individual Taxation Division, Office of Tax Analysis, while working on this paper. She is currently Chief, Special Studies Branch, Statistics of Income Division, Internal Revenue Service.

**Office of Tax Analysis
Department of the Treasury
Washington, DC 20220**

Abstract

Treasury and the Internal Revenue Service have recently intensified compliance research and compliance enforcement activities. This paper contributes to the effort by reviewing two classes of theoretical models of tax enforcement, identifying some practical implications that can be drawn from these models, and suggesting some ideas for future research. Both optimal tax models and game theory models suggest that close coordination between tax policymakers and the tax administrator may be necessary. Specifically, IRS officials and policymakers should think carefully about what IRS' enforcement objective should be, and how the overall objective can be implemented in various enforcement programs. The IRS must also collect the data that would allow researchers to estimate how taxpayers respond to tax and enforcement policies. Finally, Treasury, Congress and other policymakers need to take IRS constraints, incentives and likely responses into account when making tax policy and budgetary decisions, much as they take taxpayer behavior into account.

In formulating tax policy, we seek to balance a number of goals, including: economic efficiency, equity, simplicity and raising revenue. Sometimes we also aim to meet social policy goals, such as inducing low-income individuals to move from welfare to work, encouraging saving for retirement and subsidizing spending on education. Treasury analysts and officials know that tax laws will not achieve the desired policy goals if taxpayers do not understand and comply with them or the Internal Revenue Service (IRS) cannot administer and enforce them. As Roy Blough, the founder of Treasury's Office of Tax Analysis, stated so well,

“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.” (Blough, 1952, p.146, as cited in Slemrod and Yitzhaki, 2000).

Our concern about the effects of tax administration on tax policy outcomes leads us to ask what we can do to facilitate and improve tax enforcement. At the same time, there is renewed commitment on the part of the IRS to enforce the tax laws, as well as to serve taxpayers. In addition, the IRS has recently reinvigorated its compliance research program. As part of these renewed enforcement and research efforts, analysts at both Treasury and IRS are thinking about what our enforcement strategy and policies should be. My goal in this paper is to add to this effort by reviewing two classes of theoretical

models, identifying some practical implications that can be drawn from these models, and suggesting some ideas for extending the models in future research.¹

The first type of model extends the familiar optimal individual income tax problem to allow for tax evasion and enforcement. In this problem, the government policymaker chooses tax parameters and enforcement parameters such as the audit rate and penalty rate. A key implication is that the policymaker should set each tax policy parameter so that at the margin, the cost per dollar raised from each instrument (tax, audit and penalty) is the same, and this cost should equal the marginal benefit of public goods expenditure.

The second type of model recognizes that the IRS and taxpayers interact strategically. That is, taxpayers know that the probability of audit may depend on their reported income and they take this into account when filing their tax returns. In turn, the tax authority takes the taxpayers' strategic behavior into account when attempting to detect or deter noncompliance. An interesting subset of these models also incorporates interactions between tax policymakers (e.g., the Department of the Treasury or Congress) and tax enforcement agencies (e.g., the IRS). A key implication of these game theory models is that social welfare is maximized when the tax authority can commit to an announced strategy that induces compliant behavior. However, once compliant reports are made, the tax authority has an incentive to deviate from the announced audit strategy in order to avoid auditing compliant returns and save audit resources. Thus, the tax authority may be unable to make credible commitments to welfare maximizing audit strategies.

¹ A shorter version of this paper appears in McCubbin (2004b). For a similar discussion, see Cowell (2002). See also McCubbin (2004a), which reviews the theoretical and empirical models of individual taxpayer behavior that underpin the optimal enforcement models.

Optimal Tax Models

Suppose that individuals are identical and risk averse, and obtain utility from private income and public goods. They receive a fixed amount of true income of y_t that is not observed by the tax authority, and choose an amount of income to report, y_r . Individuals are subject to a constant rate of tax on reported income (τ), probability of audit (p) and penalty per dollar of underreported income (π , where π is often of the form $[1+f]\tau$, with $f > 0$).² Auditing is costly to the government, but the penalty for misreporting may be costlessly imposed once misreporting is detected. There are no other tax administrative costs. Government revenue net of enforcement costs is used to provide government goods, denoted G .³ Individuals are free-riders, in that they do not take into account the effect of their own taxes on the provision of the public good when deciding how much income to report. The taxpayer's utility function is assumed to be of the form:

$$[1-p]U(y_t - \tau y_r) + pU(y_t - \tau y_r - \pi[y_t - y_r]) + V(G).$$

The policymaker's task is to choose a tax rate, probability of detection and penalty to maximize the utility of a representative individual, subject to the government budget constraint.⁴ Let R equal expected gross tax revenue, $\tau y_r + p \pi[y_t - y_r]$, and $c(p)$ equal the

² The model can easily be extended to allow for any differentiable income tax function. It would also be useful to extend it to the choice of a tax system (see e.g. the discussion in Slemrod, 1990) and to take into account additional features of the tax code such as complexity and filing burden (see e.g., Kaplow, 1996). These considerations are beyond the scope of this paper.

³ For simplicity, it is assumed that the utility of private income and the public good is additively separable. The implications of the model are substantially unchanged when a more general functional form is used.

⁴ This model assumes that the government wants to maximize the total utility of all taxpayers. If all individuals are the same, then the problem reduces to choosing τ , p and π to maximize the utility of a single individual who represents all taxpayers. The assumption that individuals are identical is relaxed below.

cost of maintaining probability of detection p . Mathematically, the government's problem is:

$$(1) \text{ Max}_{\tau, p, \pi} [1-p]U(y_t - \tau y_r) + pU(y_t - \tau y_r - \pi[y_t - y_r]) + V(G)$$

$$\text{subject to: } G = R - c(p) \text{ and } 0 \leq p \leq 1,$$

where y_r is chosen optimally by the taxpayer.

Maximization of welfare with respect to the income tax rate requires:

$$(2) [[1-p]U'(y_t - \tau y_r) + pU'(y_t - \tau y_r - \pi[y_t - y_r])]y_r = [\partial V/\partial G][\partial G/\partial \tau].$$

That is, the policymaker should equate the expected marginal utility of foregone private consumption associated with a change in tax rate with the marginal utility of public good consumption afforded by the change in tax.

An interior solution for the optimal probability of detection requires:

$$(3) U(y_t - \tau y_r) - U(y_t - \tau y_r - \pi[y_t - y_r]) = [\partial V/\partial G][\partial G/\partial p],$$

where $\partial G/\partial p = \partial R/\partial p - c'(p)$. The left-hand side of the equation is simply individual utility when the taxpayer is not audited less utility when the taxpayer is audited, or the loss in utility resulting from an audit. The right-hand side of the equation is equal to the value of an additional unit of the public good, multiplied by the additional public good that can be provided by an increase in the audit rate. Thus, to maximize the utility of taxpayers, the policymaker should choose a level of detection that equates the marginal cost of audits (to the taxpayer, in the form of reduced private consumption) to the marginal benefit (to society, in the form of increased public goods).

Note that maximization of net revenue would require choosing p to equate the marginal cost (to the government) of audits and the marginal revenue from audits – that

is, setting $c'(p) = \partial R / \partial p$.⁵ Note further that the social welfare first order condition for the probability of audit (equation 3) can be rewritten as:

$$(4) \quad c'(p) = \partial R / \partial p - [U(y_t - \tau y_r) - U(y_t - \tau y_r - \pi[y_t - y_r])] / [\partial V / \partial G].$$

Maximizing the welfare of the representative taxpayer requires setting the marginal (government) cost of auditing equal to marginal revenue less a term that reflects the cost to the individual of being audited and the benefit of public good consumption. If the cost of each additional audit is constant or increasing as the audit rate is increased (i.e., if $c''(p) \geq 0$), this means that the probability of detection needed to maximize social welfare is less than the probability of detection required to maximize revenue. Increasing the probability of detection to the point where net revenue is maximized is not socially optimal, though it may be approximately optimal if the cost to individuals of being audited is very low or if the marginal utility of public goods is very high.

Slemrod and Yitzhaki (1987) examine a similar model, in which public goods do not enter into the individual's utility function, but the government must still raise a fixed amount of revenue. The penalty is fixed, and the policymaker must choose a combination of audit and tax rates that maximizes individual utility of expected after-tax income. Given that individuals are risk averse, they prefer to satisfy the government's revenue constraint via a more certain payment (requiring a low tax rate and high audit rate) rather than a riskier payment (consisting of a higher tax rate but lower probability of audit).⁶ However, increasing the certainty of the tax by increasing the audit rate is costly.

⁵ Marginal gross revenue, $\partial R / \partial p$, is equal to $[\tau - p\pi]\partial y_r / \partial p + \pi[y_t - y_r]$. This is positive if the taxpayer underreports any income.

⁶ Recall that in the model above, expected revenue is $[1-p]\tau y_r + p[\tau y_r + \pi[y_t - y_r]] = \tau y_r + p\pi[y_t - y_r]$. A risk averse taxpayer would rather pay this expected amount with certainty, than to face a $[1-p]$ chance of a liability of τy_r and a p chance of a liability of $\tau y_r + \pi[y_t - y_r]$.

The government's objective of maximizing social welfare is equivalent to minimizing the total social cost of evasion, where the total cost of evasion is equal to the excess burden associated with an uncertain tax liability plus the administrative cost of auditing taxpayers. Slemrod and Yitzhaki demonstrate that the marginal cost of increasing the audit rate should equal the reduced excess burden attributable to the reduced uncertainty of the tax liability.⁷ The increased revenue to be gained from increasing the enforcement of the law is irrelevant in this model, because public goods do not increase individual utility and tax revenue is merely a transfer from the individual to the government. This stark formulation of the problem emphasizes the tradeoff between tolerating uncertainty and spending resources to audit.

Models that allow both the penalty and audit rates to be chosen typically result in high optimal penalty rates and low audit rates.⁸ This is because the models usually assume that audits are costly to conduct whereas penalties are costless to impose.⁹ In addition, holding the expected penalty ($p\pi$) constant, an increase in the penalty is expected to have a larger impact on compliance than an increase in the probability of

⁷ Specifically, the marginal reduction in excess burden should equal the real cost of an additional audit, or λ *marginal dollar cost, where λ is the value in terms of utility of relaxing the government budget constraint by one dollar.

⁸ See for example the groundbreaking work of Becker (1968).

⁹ Exceptions include Polinsky and Shavell (1990, 2000), who demonstrate that if it is costly to impose fines, then those administrative costs should be added to the fine. Polinsky and Shavell (1984, 2000) consider the case of fines that can be imposed without cost and imprisonment that can be imposed at some cost. Because imprisonment is costly, fines should be used to the maximum extent (up to the individual's wealth in the risk neutral case) before prison terms are imposed. Chu and Jiang (1993) show that if risk-averse individuals with differing levels of wealth face a choice of crimes of varying severity, then a combination of imprisonment and sub-maximal fines may be more efficient than maximal fines. The model presented in this section implicitly assumes that all penalties can be reduced to monetary terms. For example, a prison term can be represented in the model by the amount of money a taxpayer would pay to avoid prison. The model also assumes that the penalty is equal to a fixed percentage of the unreported

detection. Thus, the policymaker can save resources and achieve the same level of compliance by increasing penalties and reducing the probability of detection. Slemrod and Yitzhaki note that such a policy prescription is unrealistic if policy presumes that the penalty for tax evasion should be commensurate to the severity of the violation.¹⁰ Therefore, they assume that the penalty is fixed, and outside the purview of their model. However, as noted by Sandmo (1981) and others, the severity of the penalty indicated by social welfare maximization models will tend to be limited when taxpayers are risk averse, so that the policymaker is concerned with inequality. Kaplow (1989) emphasizes that while one way to reduce total risk-bearing costs is to reduce the number of individuals who bear any risk by increasing penalties to the point that taxpayers are deterred from undertaking any evasion, another way is to reduce the risk borne by those who are not deterred by reducing the penalty.¹¹

The solution to the problem at hand with respect to the penalty rate requires:

$$(5) pU'(y_t - \tau y_t - \pi[y_t - y_r])[y_t - y_r] = [\partial V/\partial G][\partial G/\partial \pi].$$

income, which is very unlikely if there are both civil and criminal sanctions. See McCubbin (2000) for a discussion of nonlinearities in the actual penalties provided for in the Internal Revenue Code.

¹⁰ Andreoni (1991) formalizes this idea and demonstrates that when guilt must be determined “beyond a reasonable doubt,” a juror’s determination of what is reasonable is likely to depend on the severity of the penalty. Thus, increasing the penalty will reduce the probability of conviction so that very high fines might encourage rather than deter criminal activity. As a result, the penalties should “fit the crime,” rather than be uniformly maximal.

¹¹ Kaplow also notes that if some tax underreporting is inadvertent, high fines add little to deterrence but impose risks on violators, and it might be efficient to expend more resources on detection and reduce penalties. Bebchuck and Kaplow (1993) show that if some individuals are harder to apprehend than others, then maximal penalties for all risk neutral offenders will result in hard-to-apprehend individuals being underdeterred and easy-to-apprehend individuals being overdeterred. (That is, the expected penalty will exceed the expected harm from the illegal act for the easy to apprehend individuals.) In that case, it is optimal to penalize hard-to-apprehend offenders more severely, perhaps by levying additional fines for destroying evidence. If easy- and hard-to-apprehend individuals cannot be distinguished, then it may be optimal to reduce penalties for all offenders and increase the intensity of efforts to apprehend criminals.

The cost to the individual of the expected change in penalty should equal the marginal benefit of increasing public goods by increasing the penalty. Because there are no costs of imposing the penalty, the change in public goods from increasing the penalty ($\partial G/\partial \pi$) is just the change in gross revenue ($\partial R/\partial \pi$).¹² But penalties may in fact be costly to impose. Moreover, the higher the penalty, the more diligent the IRS must be to ensure that it is not applied inappropriately, and the more vigorous taxpayers will be in defending themselves against the penalty.¹³

Suppose therefore that total enforcement costs are a function of both the audit rate and the penalty rate, so that $G = R - c(p, \pi)$. Equation (5) still characterizes the interior solution for the optimal penalty rate. But the public goods spending generated by an increase in the penalty, $\partial G/\partial \pi$, is reduced from $\partial R/\partial \pi$ to $\partial R/\partial \pi - \partial c/\partial \pi$. As a result, the optimal penalty rate will be lower. The optimal audit rate could be lower or higher than it would be if penalties could be costlessly imposed. When the cost of imposing penalties is increasing in the size of the penalty, it may be optimal to set a lower penalty but impose it more often by increasing the audit rate. But if that enforcement does not deter all evasion, increasing the audit rate increases the number of times that the penalty must be imposed, so that penalty costs limit the optimal audit rate.

Note that the first order conditions can be written as:

$$(6) \quad [(1-p)U'(y_t - \tau y_r) + pU'(y_t - \tau y_r - \pi[y_t - y_r])]y_r / [\partial G/\partial \tau] = [\partial V/\partial G],$$

¹² $\partial R/\partial \pi = [\tau - p\pi][\partial y_r/\partial \pi] + p[y_t - y_r]$. This is positive if the individual is underreporting any income.

¹³ Becker (1968, p. 184) notes that judges and juries may be unwilling to convict offenders when penalties are very high and that, as a result, the cost of apprehending and convicting an offender will depend on the penalty as well as the detection rate. Andreoni (1995) finds that penalties have a direct effect on reducing crime. However, he also finds an inverse relationship between penalties and conviction rates, which entirely offsets the direct effect of penalties on deterrence. Higher penalties might lead to lower conviction

$$(7) [U(y_t - \tau y_r) - U(y_t - \tau y_r - \pi[y_t - y_r])] / [\partial G / \partial p] = [\partial V / \partial G],$$

and

$$(8) pU'(y_t - \tau y_r - \pi[y_t - y_r])[y_t - y_r] / [\partial G / \partial \pi] = [\partial V / \partial G].$$

This highlights that the marginal cost to taxpayers per dollar of public expenditure financed by the tax, by audits and by the penalty should be the same, and this marginal social cost should be equal to the marginal social benefit of public goods expenditure.

Thus far I have assumed that individuals are identical. Let me now assume that there are two types of individuals: high-income individuals and low-income individuals. Suppose further that the cost of auditing a high-income person is higher than the cost of auditing a low-income person, perhaps because their returns are more complex. To focus attention on the choice of audit rates, I assume that the probability of audit may vary across the two groups, but that the tax rate and penalty rate are the same for both groups, and that penalties may be imposed without cost. The policymaker now intends to maximize the sum of individual utilities, by solving:

$$(9) \text{Max}_{\tau, p_l, p_h, \pi} \alpha EU(Y_{\text{low}}) + [1-\alpha]EU(Y_{\text{high}}) + V(G)$$

$$\text{subject to: } G = R - c(p_l, p_h), 0 \leq p_l \leq 1 \text{ and } 0 \leq p_h \leq 1,$$

where α is the fraction of individuals with low incomes and $[1-\alpha]$ is the fraction of individuals with high incomes.¹⁴ $EU(Y_{\text{low}})$ is the expected utility of income for low income taxpayers (i.e., $EU(Y_{\text{low}}) = [1-p_l]U(y_{tl} - \tau y_{rl}) + p_l U(y_{tl} - \tau y_{rl} - \pi[y_{tl} - y_{rl}])$ and

rates because defendants are willing to spend more on legal defense or because judges and juries become reluctant to convict defendants when penalties are harsher.

¹⁴ Many optimal tax models assume a utilitarian social welfare function (i.e., one that maximizes the sum of individual utilities). See Kaplow (1995) who argues that utilitarianism is more consistent with the Pareto principle (which holds that an option unanimously preferred by individuals should be adopted) than alternative social welfare functions. See also Schroyen (1997) who shows how the optimal response to growing evasion might be different when the objective is to guarantee a minimum standard of living than when the objective is to maximize total welfare.

$EU(Y_{high})$ is the expected utility of income for high-income taxpayers (thus, $EU(Y_{high}) = [1-p_h]U(y_{th} - \tau y_{rh}) + p_h U(y_{th} - \tau y_{rh} - \pi[y_{th} - y_{rh}])$). As before, reported income is chosen by taxpayers to maximize expected utility of private income. Note that this formulation assumes that the policymaker can distinguish between low- and high-income taxpayers. This is not the case if some high-income taxpayers report low incomes. However, it is perhaps reasonable to assume that the policymaker has sufficient information to group taxpayers into broad categories.

Interior solutions for p_l and p_h are characterized by:

$$(10) \quad \alpha [\partial EU(Y_{low}) / \partial p_l] / [\partial G / \partial p_l] = [\partial V / \partial G]$$

and

$$(11) \quad [1-\alpha] [\partial EU(Y_{high}) / \partial p_h] / [\partial G / \partial p_h] = [\partial V / \partial G].$$

The policymaker wants to set the marginal social cost of low-income audits per dollar of public good financed by low-income audits equal to the marginal social cost of high-income audits per dollar of public good financed by high-income audits, and both of these equal to the marginal benefit of the public good. Recall that $\partial G / \partial p_l = \partial R / \partial p_l - \partial c / \partial p_l$. All other things equal, the policymaker will want to use more low-income audits than high-income audits because low-income audits are assumed to cost less ($\partial c / \partial p_l$ is assumed to be less than $\partial c / \partial p_h$ for any given level of p). However, the policymaker will also take into account differences in gross revenue generated by low-income and high-income audits (i.e., differences in $\partial R / \partial p$), as well as differences in the marginal private cost of audits (as reflected in $\partial EU(Y_{low}) / \partial p_l$ and $\partial EU(Y_{high}) / \partial p_h$). The concavity of the utility function will tend to make the marginal utility cost of audits greater for lower-

income taxpayers, reducing the incentive for the policymaker to audit lower-income taxpayers.¹⁵

Suppose now that the policymaker (e.g., Congress) delegates the choice of audit rates to a tax administrator (e.g., IRS), and that the objective of the tax administrator is to maximize tax and penalty revenues. The policymaker also provides the administrator with an optimally-sized budget, $B^* = c(p_l^*, p_h^*)$, where p_l^* and p_h^* (along with the optimal tax and penalty rates) solve the social welfare maximization problem. The tax administrator's problem is:

$$(12) \text{ Max}_{p_l, p_h} R$$

$$\text{subject to: } B^* \geq c(p_l, p_h), 0 \leq p_l \leq 1 \text{ and } 0 \leq p_h \leq 1.$$

As before, R is expected gross revenue, or

$\alpha[\tau_l y_{rl} + p_l \pi_l [y_{tl} - y_{rl}]] + [1 - \alpha][\tau_h y_{rh} + p_h \pi_h [y_{th} - y_{rh}]]$. Let λ be the shadow value of relaxing the administrator's budget constraint. Then optimal interior solutions for p_l and p_h will be given by:

$$(13) \partial R / \partial p_l = \lambda * \partial c / \partial p_l$$

and

$$(14) \partial R / \partial p_h = \lambda * \partial c / \partial p_h$$

or $[\partial R / \partial p_l] / [\partial c / \partial p_l] = [\partial R / \partial p_h] / [\partial c / \partial p_h]$. The tax administrator will want to set audit rates such that at the margin, the revenue from low-income audits per dollar spent on

¹⁵ $\partial EU(Y_i) / \partial p_i$ is utility of income in the case where the taxpayer is audited less utility in the case where he is not, or $U(y_{ti} - \tau y_{ri}) - U(y_{ti} - \tau y_{ri} - \pi [y_{ti} - y_{ri}])$, where $i = l$ or h indicates low or high. $\partial R / \partial p_i = \tau \partial y_{ri} / \partial p_i + \pi [y_{ti} - y_{ri}] - p_i \pi \partial y_{ri} / \partial p_i$, times the fraction of the population that is type i . In this simple example I have focused on the choice of audit rates and assumed that the tax and penalty rates are the same for both groups. To maximize social welfare the tax and enforcement parameters are chosen simultaneously, and the marginal tax rate may vary across groups. If the tax rate is allowed to vary across the groups and the optimal tax rate is higher for higher income taxpayers, then the incentive to audit higher income taxpayers will be further increased. The social welfare maximizing policymaker will take all of these factors into account.

low-income audits is equal to the revenue from high-income audits per dollar spent on high-income audits. In other words, the tax administrator will want to set rates such that the last dollar spent on low-income audits yields the same gross revenue as the last dollar spent on high-income audits. Otherwise, revenue could be increased by reallocating audit resources between the two groups.

Notice that the revenue maximizing tax administrator does not take into account the private cost of audits, and therefore need not choose $p_l = p_l^*$ or $p_h = p_h^*$. Earlier we saw that unconstrained revenue maximization results in too much expenditure on auditing, relative to the social welfare maximizing level. Now we see that even if the tax administrator's total auditing budget is the correct size, $B^* = c(p_l^*, p_h^*)$, a revenue maximizing administrator is not likely to choose the social welfare maximizing audit rates for each audit class when taxpayers are not identical. This suggests that there must be coordination between the policymaker who sets tax rates, penalty rates and budgets, and the tax administrator who chooses audit rates, in order to achieve maximum social welfare. We will return to this topic near the end of the next section.¹⁶

To summarize the results from these social welfare maximization models: the policymaker must simultaneously solve the first order equations for τ , p and π to find their optimal values. This requires information about the individual's utility function and

¹⁶ It is reasonable to ask whether we should care about the utility of tax evaders. Clearly we do care about them to some extent, otherwise the penalty would be set as high as possible (including the use of severe criminal sanctions for even relatively minor infractions). What is more likely is that society cares little about the costs of audits imposed on evaders, as long as those costs are not too large. This is somewhat difficult to model. We can only identify evaders ex post. Since we want to discount the utility of potential evaders only when they choose to evade, we cannot simply give them lower weights, ex ante, in the social welfare function. Moreover, if audits are costly even to compliant taxpayers, if taxpayers make unintentional reporting errors, or if auditors make errors in assessing penalties, then it remains the case that there are private enforcement costs that will be taken into account by a social welfare maximizer but not a revenue maximizer.

responsiveness to enforcement parameters that are not readily available. In particular, we have no reliable estimates of $\partial y_r / \partial p$ and $\partial y_r / \partial \pi$. Nevertheless, the form of the solution provides useful insights as to what the policymaker's goals should be. If the policymaker wishes to maximize social welfare, then each policy parameter should be set so that at the margin, the cost per dollar raised from each instrument (tax, audit and penalty) is the same, and this cost should equal the marginal benefit of public goods expenditure. The cost of each instrument is the expected utility cost to taxpayers of increasing the tax, audit or penalty rate. These private costs would be ignored if revenue maximization were the objective, and therefore revenue maximizing choices are not generally social welfare maximizing.

The model presented here is very simple, and there are a number of ways in which it might be extended. A number of authors have shown that endogenous labor income might reduce the efficiency of the income tax and corresponding enforcement, or, evasion can offset the excess burden of labor taxes. See for example Sandmo (1981), Slemrod and Yitzhaki (1987), Kaplow (1990), Mayshar (1991) and Schroyen (1997). See also Kopczuk (2001), who finds that it may be optimal to allow some avoidance activities to continue when they are more available to the poor and more efficient at accomplishing redistribution than increasing the progressivity of tax rates. Interested readers may also see Chen (2003) for analysis of the effects of evasion and enforcement on economic growth in a general-equilibrium framework.

If some taxpayers are always honest or if audits impose costs even on compliant taxpayers, it is likely to be optimal to increase penalties and reduce audit rates, because penalties are only imposed on dishonest taxpayers and audits on honest taxpayers will be

unproductive (yielding neither increased deterrence nor increased direct audit revenue). This assumes of course that auditors do not mistakenly impose penalties on honest taxpayers. Scotchmer and Slemrod (1989) consider a model with uncertainty about the tax liability that will be assessed in an audit. They find that if reducing randomness is costly, revenue is increased by allowing or increasing uncertainty.¹⁷

The model above assumes that taxpayers are fully informed and audits uncover all misreporting. It would be useful to allow for unintentional taxpayer errors and errors in detection by auditors. Similarly, the models assume that tax evasion is always punished when it is detected by an audit (although some constrain the maximum penalty by the taxpayer's income or wealth to preclude the possibility that penalties cannot be collected). If this is not the case, the question of whether or not to audit should take into account not just the additional tax and penalty that might be assessed, but whether the tax and penalty will be sustained after appeal and whether it can be collected. As a practical matter, the problem is complicated by the fact that examinations and collections are separate functions within the IRS. Thus, the examiner deciding whether or not to pursue an audit is not the same individual who will be responsible for collecting any resulting tax and penalties.

¹⁷ Reinganum and Wilde (1988) consider a similar model in which audit costs are uncertain. As a result, taxpayers are uncertain about the amount of noncompliance that the tax authority will tolerate. Reported income and net revenue are at first increasing, but then decreasing, as uncertainty about audit costs increases. Therefore the tax authority will want to maintain a modest amount of uncertainty and secrecy about audit techniques. In this model, the tax authority seeks to maximize revenue and taxpayers are risk neutral. It is not clear that allowing uncertainty is optimal if taxpayers are risk averse and the policymaker seeks to maximize social welfare. Alm (1988) considers the consequences of tax uncertainty more generally, including uncertainty about future changes in tax rates and the tax base. He finds that tax base uncertainty is likely to generate behavioral responses that increase the size of the income tax base, while tax rate uncertainty always leads to decreases in the income tax base. If uncertainty about the tax base increases the tax base, the policymaker may be able to reduce tax rates and increase social welfare despite the fact that uncertainty, *ceteris paribus*, reduces welfare. Despite the possibility that under certain circumstances, specific policies incorporating uncertainty may be welfare enhancing, Alm finds that

Another interesting way to extend the optimal enforcement model would be to allow for tax preparers or other advisers. Given current interest in the problem of abusive tax shelters, it would be useful to allow taxpayers to, in a sense, purchase insurance against detection and penalties by purchasing a shelter rather than simply not reporting the income. Presumably, it is desirable to deter such behavior to the extent that sheltering activities result in pure deadweight loss. Corporate income taxes and evasion could also be modeled, perhaps replacing maximization of the utility of individual income by profit maximization, shareholder value maximization, or maximization of the utility of the manager's salary, where the salary depends on corporate income and evasion.

The model can also be extended to allow for an endogenous probability of audit. This is a key feature of the game theory models, discussed in the next section.

Game Theory Models

A second class of models seeks to more fully account for strategic interactions between taxpayers and the tax administrator (e.g., IRS), and in some cases, between tax policymakers (e.g., Congress or the Treasury Department) and the tax administrator. These models recognize that the probability of detection need not be constant across taxpayers and can depend on the taxpayer's reported income or other signals from the taxpayer. Taxpayers are expected to take into account the likely effect of their reports on the probability of audit when filing their returns. In turn, the tax authority takes the taxpayers' strategic behavior into account when deciding on the audit rule. These models are more realistic, because we know that the IRS does take reported information into

uncertainty often reduces revenue, welfare or both. Thus, more stable tax and tax administration policies may very well be beneficial.

account when choosing returns to audit.¹⁸ Unfortunately, these models are also more complicated and difficult to solve. Therefore most authors focus on special cases or examples to demonstrate how strategic interactions might affect the optimal audit strategy, rather than attempting to characterize general results.

Most of these models assume that taxpayers are risk neutral, rather than risk averse. Risk aversion is arguably a more appropriate assumption, as it is consistent with diminishing marginal utility of income at the individual level and a concern for equity at the social level. Many of these models also assume that the government's objective is to maximize revenue (including taxes paid voluntarily and additional taxes and penalties paid after audit, net of audit costs) rather than to maximize social welfare, thus side-stepping equity considerations altogether. Assuming risk neutrality and revenue maximization makes the problem more tractable, and also affects the implications of the models. A key distinction among these models is whether the tax authority is assumed to be able to commit to an announced audit strategy.

In perhaps the earliest accounting for strategic behavior in an optimal enforcement model, Landsberger and Meilijson (1982) consider whether more revenue can be raised when the probability of audit is based on the taxpayer's past behavior than when the probability of audit is purely random. Taxpayers are assumed to be risk neutral (though the results also hold if taxpayers are risk averse, as long as the probability of audit is sufficiently small). In the authors' proposed scheme, taxpayers are initially assigned arbitrarily to one of two states, which are characterized by different audit rates.

¹⁸ For example, it is well-known that IRS ranks returns by a discriminant function (DIF) that uses 1988 Taxpayer Compliant Measurement Survey data to estimate the likelihood that a substantial tax change will be found, and that higher scoring returns are more likely to be audited. See for example IRS Publication 556 and U.S. General Accounting Office (1999).

If a taxpayer in state 1 is audited and found to be noncompliant, he or she is moved to state 2, where the audit rate is higher, for the next reporting cycle. If a taxpayer in state 2 is audited and found to be truthful, he or she is moved to state 1, where the audit rate is lower. This scheme is announced to taxpayers, and the tax authority is able to commit to the scheme. The authors find that the state-dependent audit rule raises more revenue net of audit costs than random audits in a wide variety of cases.¹⁹

Similarly, Greenberg (1984) shows that the number of tax evaders can be made “arbitrarily small” by assigning taxpayers to three groups, based on past compliance.²⁰ In Greenberg’s scheme, taxpayers with a history of repeated misreporting are always audited. The author notes that if taxpayers could agree to misreport all of the time, then all of them would end up in this group. Since the tax authority presumably does not have sufficient resources to audit everyone, the threat to audit all repeat offenders would not be credible, and Greenberg’s scheme would fail. Greenberg assumes that the audit threat is credible because it is too difficult for taxpayers to cooperate.

¹⁹ In their seminal paper on tax compliance, Allingham and Sandmo (1972) show that if a taxpayer’s past reporting is investigated whenever the current report is audited and found to be noncompliant, then more income is reported than would otherwise be the case. This occurs because as periods of noncompliant behavior accumulate, the passage of time becomes equivalent to an increase in the penalty for noncompliance in the current period. In addition, because current noncompliance affects the expected penalty in future periods, the taxpayer will be more compliant in the current period than would otherwise be the case.

²⁰ Specifically, a fraction α of the population is initially assigned to state 1 and the rest are assigned to state 2. In state 1, taxpayers are audited at rate $\rho/2$, where ρ is a probability sufficient to prevent all myopic taxpayers from cheating. If taxpayers in state 1 are audited and found to have misreported income, they are moved to state 2. Taxpayers in state 2 are audited at rate $\alpha/[1-\alpha]*[\rho/2]$, and move to state 1 if they are audited and found to have reported truthfully, and to state 3 if they are audited and found to have misreported. Taxpayers in state 3 are always audited and always remain in state 3. Taxpayers in state 1 will find it optimal to evade tax and taxpayers in states 2 and 3 will find it optimal to report truthfully. Therefore taxpayers will oscillate between states 1 and 2 as they are audited, no one ever moves to state 3, and the proportion of the population in each group remains the same as initially allocated. No more than the fraction α of taxpayers will be noncompliant. The results do not depend on the risk neutrality of taxpayers; no restriction is made on the utility function other than it is increasing in income.

Reinganum and Wilde (1985) compare purely random audits to a cutoff rule, whereby the tax authority audits no one who reports income greater than or equal to some specified level and audits everyone who reports income less than the cutoff level. They find that if taxpayers are risk neutral, the cutoff audit policy raises more revenue net of audit costs than purely random audits.²¹ However, if taxpayers are risk averse, then random audits might raise more revenue than the cutoff rule. Under the cutoff strategy, all taxpayers with income below the cutoff report truthfully (because they are all audited) and all taxpayers with income at or above the cutoff report the cutoff amount. Only truthful taxpayers are audited.²² The purpose of audits in this scheme is not to detect evasion, but to deter evasion and ensure that higher income taxpayers report at least the cutoff amount of income. This highlights the nature of the commitment problem. Once taxpayers have been induced to report correctly, it is wasteful to audit them. But if the tax authority does not follow through on the commitment to audit taxpayers who report less than the cutoff amount, the enforcement scheme does not work.

The fact that under a successful deterrence scheme only truthful taxpayers are audited also suggests a problem with one stated IRS objective. The IRS seeks to reduce

²¹ To understand this result, note that the tax authority seeks to maximize revenue and is not subject to a budget constraint. In addition, in this highly stylized model, the marginal cost of audits is constant and, when taxpayers are selected randomly for audit, the expected marginal benefit of an audit is also constant. As a result, the revenue maximizing random audit strategy will involve auditing all taxpayers (if the expected benefit exceeds the cost) or auditing no taxpayers (if the expected benefit is less than the cost). Under the alternative cutoff policy, the probability of audit is conditioned on income so that the expected marginal benefit of auditing is not constant. In addition, conditioning the probability of audit on income is useful because it induces taxpayers to reveal information. In particular, high income taxpayers are induced to reveal that they have income at least as great as the cutoff amount. Some generalizations of the model are discussed below.

²² Since the game-theoretic models (especially those that allow pre-commitment to an audit strategy) often imply truthful reporting by all taxpayers or all audited taxpayers, the size of the fine is sometimes irrelevant, and may be even be zero. This is in contrast to the first class of social welfare maximization models, where a large penalty is often recommended.

the number of audits that do not produce a direct change in tax liability.²³ The IRS does not, to my knowledge, routinely estimate the indirect revenue produced by audits in the form of deterrence. Focus on the so-called no-change rate is therefore misguided, if audits do in fact induce improvements in taxpayer behavior.²⁴ Of course, if the IRS faces a tight budget constraint, has not committed to a particular strategy, and is unable to induce a high level of compliance, then auditing taxpayers who appear to be noncompliant rather than auditing taxpayers who appear to be compliant seems entirely reasonable. Nevertheless, it may be optimal for the IRS to audit some taxpayers who appear to be compliant in order to maintain a credible enforcement threat. Fortunately, matching third-party information (such as Form W-2) to taxpayer reports allows the IRS to verify at least part of nearly every taxpayer's return, at minimal cost to the IRS and minimal burden to the taxpayer. This contributes enormously to taxpayer compliance.²⁵

Scotchmer (1987) emphasizes that cutoff rules create a regressive bias. Unless the cutoff is high enough to induce truthful reporting by all taxpayers, only lower income

²³ See for example, IRS press release IR-2002-05, January 2002 which cited a reduction in "needless 'no change' audits" resulting from improved audit selection algorithms as one purpose for the National Research Program.

²⁴ Rewarding the performance of individual auditors based on detected evasion would be similarly problematic, because the auditor's reward would fall if the enforcement strategy was effective in inducing compliance. For a discussion of this problem see Sanyal (2002) and Melamud and Mookerjee (1989).

²⁵ Requiring information reporting is another tool, in addition to penalties, tax rates and audit rates, which is under the control of policymakers. To the extent that information reporting is required, compliance is typically improved and automatic verification can substitute for auditing. Requiring information reporting may also be viewed as a way to facilitate IRS' ability to commit to an enforcement strategy. In this vein, Treasury and IRS have recently taken steps to rationalize and strengthen disclosure rules for transactions associated with potentially abusive tax shelters, and have asked Congress to change rules for promoter registration and enact penalties for non-disclosure. Note however, that IRS is unable to follow up on every discrepancy identified by matching tax returns to information returns. This is because while the initial matching is nearly automatic, IRS manually screens mismatches to ensure that, for example, the third-party report appears correct and the income is actually missing from the return rather than merely reported on the wrong line. In addition, after spurious mismatches are screened out and the remaining taxpayers are contacted, IRS must expend resources to answer taxpayer questions and review any responses submitted by taxpayers in support of their initial reports. See U.S. Department of the Treasury (2003), p. 5.

taxpayers are audited, and only higher income taxpayers misreport income. In contrast, a random audit policy in Reinganum and Wilde's model would generate either all evasive reports or evasion by low income taxpayers and truthful reporting by higher income taxpayers. Scotchmer demonstrates that if taxpayers can be grouped into classes by factors (such as reported income or income source) that are correlated with true income, then it will still be revenue maximizing to audit only lower income taxpayers within each class, but the regressive bias can be offset overall by auditing higher income classes at higher rates (by setting higher cutoffs in those classes).

Border and Sobel (1987) generalize the work of Reinganum and Wilde by relaxing assumptions about the tax scheme. Reinganum and Wilde assume that taxpayers must pay the minimum of a fixed amount of tax or all of their income, or, in an alternative model, that taxes are proportional. Border and Sobel assume only that fines and taxes cannot exceed the taxpayer's wealth. Border and Sobel retain the assumption that taxpayers are risk neutral and the tax authority maximizes revenue net of audit costs. They find that optimal taxes are monotonically increasing in reported wealth and optimal audit rates are monotonically decreasing in reported wealth. Rewards for truthful reporting by taxpayers who are subsequently audited are also optimal. In the revenue maximizing scheme, taxpayers are divided into at most three groups. Taxpayers in the lowest reported-wealth group (which might contain no taxpayers in equilibrium) are always audited. Taxpayers in the highest reported-wealth group are never audited. Taxpayers in the intermediate group are audited with some probability between 0 and 1. The audit scheme remains revenue maximizing when taxpayers are risk averse. All taxpayers report truthfully.

Mookherjee and Png (1989) also assume that the audit authority can commit to an announced audit strategy. However, Mookherjee and Png assume that the authority seeks to maximize social welfare rather than revenue, and that individuals are risk averse. Mookherjee and Png focus on an insurance problem, but their results also apply to the tax problem. Like Border and Sobel, they find that truthful reports should be rewarded. Unlike Reinganum and Wilde (who find that a simple cutoff rule dominates) and Border and Sobel (who find that a cutoff rule combined with some stochastic audits dominates), Mookherjee and Png find that rates should be between 0 and 1 for all classes rather than determined by cutoff rules. The audit rate may be 0 only for individuals whose reported payment is equal to the maximum possible payment. The optimal probability of audit is still decreasing with income and the optimal tax is increasing with income, when income and tax are limited to two values (high and low). It is not clear whether this result holds when there are three or more possible income levels.

Cremer and Gahvari (1995) extend the model to allow for risk averse individuals with two different levels of ability (and therefore different wage rates) and endogenous labor supply. They obtain similar results: high-wage individuals are never audited, low-wage individual are audited at some probability less than one, and social welfare is maximized when honest reporting is rewarded. Wane (2002) considers a more general tax and enforcement problem, and obtains very different results. In his model, taxpayers can be induced to generate a subset of the Pareto frontier through the use of pre-announced audit rules for each taxpayer that depend on the reports of all taxpayers. In this outcome, all individuals except the most able evade some amount and none is

audited. Evasion by less able taxpayers assists the social planner in obtaining the preferred outcome. (This is similar in spirit to the result obtained by Kopczuk (2001).)

Chu (1990) and Ueng and Yang (2001) consider an optimal tax problem with a feature similar to a cutoff rule. In their models, taxpayers may choose between paying a fixed amount of tax with certainty and paying what they claim they would owe under the regular tax system subject to a possible audit. If taxpayers choosing the regular tax are audited with certainty, the scheme devolves to a simple cutoff rule. Chu and Ueng and Yang show that the introduction of the alternative fixed tax can be Pareto improving. Because some risk averse taxpayers will choose to pay a slightly higher tax with certainty than be subject to potential audit, revenue can be increased without making any taxpayer worse off. Like the cutoff rule, the alternative fixed tax induces taxpayers with true (regular) liabilities above a certain amount to reveal themselves (by choosing the alternative tax).

Graetz, Reinganum and Wilde (1986) and Reinganum and Wilde (1986, 1991) argue that it is unrealistic to assume that the tax authority can credibly commit to an audit strategy. Even if pre-commitment strategies are optimal, the tax authority might not be able to make a credible commitment because audit resources are scarce and taxpayers know that the tax authority faces an incentive to conserve resources by not auditing accurate reports. In addition, Reinganum and Wilde (1986) argue that individual taxpayers cannot verify that the IRS has followed through on its announced strategy and that there is no “higher power” to force the government to follow through on auditing commitments. It is arguable that taxpayers could verify whether the IRS has used some strategies. For example, if the IRS pledges to audit some group with certainty, at least

some taxpayers in that group would know if the IRS did not follow through on that pledge. In addition, IRS does publish average audit rates for various types of taxpayers. Nevertheless, it is clear that the IRS faces an incentive to conserve resources by not auditing taxpayers who appear to be compliant, once tax returns have been filed.

Reinganum and Wilde (1986) therefore assume that the IRS and taxpayers play a sequential game, in which taxpayers first file their returns and then IRS audits taxpayers based on the reported information. Risk neutral taxpayers understand how their reports will influence the probability of audit and report optimally to maximize net income. When IRS receives a report, it asks, what true income would make this report an optimal report on the part of the taxpayer, and chooses returns to audit based on that information. As under the pre-commitment scenario, it is revenue maximizing to audit taxpayers with lower reported incomes at a higher rate than other taxpayers in the same audit class. Unlike in the pre-commitment case, taxpayers with greater incomes underreport less than taxpayers with lower incomes in equilibrium. This is because optimality for the IRS after returns are filed requires that underreporting decline with true income. If the absolute amount of underreporting rose with true income, IRS would want to increase the probability of audit for higher income taxpayers.²⁶

Graetz, Reinganum and Wilde focus on differences between the standard tax evasion models that focus only on the taxpayer's decision given fixed tax and enforcement parameters, and more realistic game-theoretic models that account for institutional constraints and interactions between the IRS and taxpayers. For example, in their model, an increase in a proportional tax rate or in tax progressivity increases the

²⁶ There are multiple equilibria in this model. The authors focus on separating equilibria, that is, cases in which types are separated out by their reports.

incentive for underreporting but also increases the gain to auditing. The audit effect tends to dominate, so that noncompliance falls when tax rates rise, regardless of taxpayers' risk preferences. The authors also note that if IRS faces a binding budget constraint, then the probability that any taxpayer is audited depends on the behavior of other taxpayers. The probability that a particular taxpayer is audited is lower when more taxpayers misreport their income. This may in part explain the finding that taxpayers are more likely to be noncompliant if they believe that many other taxpayers are evading taxes. The perception that evasion is widespread may encourage noncompliance because it reduces the perceived probability of detection, rather than or in addition to inducing change in taxpayer preferences.²⁷

In an extension to their model, Graetz, Reinganum and Wilde (1989) examine taxpayer incentives and revenue maximizing audit and penalty policies when taxpayers can obtain expert opinion letters to avoid penalties when claiming questionable deductions. In a similar vein, Reinganum and Wilde (1991) consider the effect of introducing tax practitioners who can reduce a taxpayer's filing costs and costs associated with an audit (but not the result of the audit, fine or probability of an audit). Play is sequential, in that tax practitioners set a fee, then risk neutral taxpayers decide whether or not to use a practitioner and how much income to report, and finally the tax authority decides which returns to audit. As usual, the revenue maximizing audit rate is decreasing in reported income. Compliance may be higher or lower on practitioner prepared returns

²⁷ Similarly, Erard and Feinstein (1994) extend the model of Reinganum and Wilde (1986) to allow for the presence of honest taxpayers and a binding budget constraint. They find that an increase in the fraction of taxpayers who are strategic changes the distribution of reported income and, given the fixed budget, the probability of audit conditioned on reported income. Thus while the total number of audits remains constant, strategic taxpayers cheat more and face a lower probability of audit.

than on self-prepared returns. This is because using a practitioner reduces the costs to the taxpayer of complying with an audit, but practitioner returns are audited at a higher rate (if practitioners, as well as taxpayers, are subject to penalties). Net revenue may be higher when practitioners are used even if compliance is lower, when practitioners are subject to penalties. In this case, the revenue maximizing tax authority, somewhat perversely, would prefer lower compliance. The authors find that while there are a number of possible equilibrium solutions to the model, the outcome that fits stylized facts suggests that practitioner efficiencies are large so that taxpayers prefer using practitioners, that compliance is lower when practitioners are used, and that the increased noncompliance is not offset by preparer penalties so that revenue is lower when practitioners are used. The authors therefore conclude that practitioner penalties should be increased.

Boadway et al. (2002) develop a model in which taxpayers (in their specific case, buyers and sellers) must act in concert to evade taxes. They find that increasing sanctions increases incentives for cooperation between taxpayers. This can increase the ability of taxpayers to commit to cooperative behavior, thus increasing evasion. As usual, the ability of the players to commit to a course of action is key to the analysis. In addition to examining interactions between taxpayers, the paper makes a unique contribution in that it focuses on the effect of increasing penalties, whereas most of the game theory models emphasize audit selection schemes. Additional research on the effect of penalties would be useful.

As noted by Melumad and Mookherjee (1989), we immediately suspect that the tax authority will not be able to achieve the optimal welfare level without pre-

commitment because the no-commitment scenario imposes the additional constraint that the strategy be revenue maximizing *ex post* (i.e., after returns are filed and the taxpayers' reports are known). This indeed turns out to be the case: expected revenue is lower when the tax authority cannot pre-commit to the optimal strategy. This suggests that IRS should focus on effecting deterrence rather than on maximizing *ex post* revenue, and should perhaps consider making more (or more specific) announcements about its audit plans.²⁸

Slemrod, Blumenthal and Christian (2001) found that some high-income taxpayers reported less income and tax when told that their Minnesota tax returns would be reviewed with certainty. Slemrod et al. hypothesize that this is because when the probability of audit is uncertain, taxpayers may have an incentive to report more income and tax in the hope of avoiding being selected for an audit. If an audit is certain, this benefit of increased income and tax reporting is eliminated. In addition, when faced with a certain audit, some taxpayers might choose to take a more aggressive reporting stance, viewing their initial report as the starting point for negotiations with the tax agency. Other authors (e.g., Sheffrin and Triest, 1992) point out that perceptions that the tax system is unfair or that others are dishonest may be associated with increased noncompliance. Therefore announcements must be carefully crafted, to avoid giving impression that large numbers of taxpayers are evading tax, thereby increasing noncompliance. In light of these problems it would be wise to limit pre-commitment

²⁸ IRS does make some fairly general announcements. See for example, IRS press release FS-2002-12 issued September 2002, which states "The Internal Revenue Service is realigning its audit resources to focus on key areas of non-compliance with the tax laws." The areas identified included offshore credit card users, high-risk, high-income taxpayers, abusive schemes and promoters of schemes, high-income non-filers, unreported income and the National Research Program.

announcements to certain groups of taxpayers, and to study the effects as was done in the Minnesota experiments.

A third type of game-theoretic model recognizes that in the U.S. and in other countries, policymakers set tax rates, penalties and an auditing budget, but delegate the development of an audit strategy to a separate agency. Cremer, Marchand and Pestieau (1990) consider the case where a social welfare maximizing government chooses a proportional tax rate, and a revenue maximizing audit authority chooses an audit cutoff level and audit rate. The government takes the likely revenue maximizing audit strategy into account when choosing the tax rate. (In game theory terminology, the government is a Stackelberg leader.) Taxpayers are risk neutral, but government nonetheless cares about equity and therefore the social welfare function is assumed to be concave. Pre-commitment is possible, and in equilibrium only taxpayers who report income under the cutoff level are audited. Not surprisingly, the authors find that the social welfare maximizing government would choose a lower cutoff level (i.e., would audit fewer returns) than the revenue maximizing audit authority chooses. Therefore the government may choose a lower tax rate in order to induce the audit authority to choose a lower audit threshold.

Sanchez and Sobel (1993) analyze a similar model, but do not restrict the audit strategy to a simple cutoff regime. They find revenue is maximized when risk neutral taxpayers are separated into at most three groups. The lowest income group is audited at a rate just sufficient to induce truthful reporting (i.e., at rate $p = 1/[1+f]$ where taxpayers pay $[1+f]$ times the unreported tax when they are audited). The intermediate group is audited at a smaller but positive rate, and all individuals underreport income. The highest

income group is not audited and therefore underreports to the greatest extent. As expected, the social welfare maximizing government chooses a lower auditing budget than the audit authority prefers.²⁹

In an interesting modification to this model, Macho-Stadler and Perez-Castrillo (1997) consider the case of two different types of income – easy to detect and hard to detect. To maximize revenue, taxpayers with income detected with probability greater than or equal to $1/[1+f]$ when they are audited will be audited at a rate sufficient to induce truthful reporting if income is below a cutoff value. Taxpayers with income detected with probability less than $1/[1+f]$ will always be audited, but the certainty of audit is not sufficient to induce truthful reporting given the uncertainty of detection. Therefore they all report the lowest possible income, and audits generate some direct tax and penalty revenue (which is consistent with stylized facts). Higher income taxpayers with either type of income are not audited, and report the cutoff values. The revenue maximizing audit authority will tend to devote more resources towards the group with easy to detect income, because those audits are more effective at inducing compliance. An increase in auditing resources and thereby the cutoff level for the hard-to-detect group has the perverse effect of inducing some taxpayers to report less income. Taxpayers with income above the new cutoff level will increase reported income to the cutoff amount, but taxpayers with income now falling below the cutoff can no longer avoid an audit by reporting the cutoff amount. Therefore they report the minimum amount. Similarly, the analysis of Cowell and Gordon (1995) suggests that intense enforcement with audit rates

²⁹ Sanchez and Sobel also provide a general characterization of the revenue maximization problem. They emphasize that a simple cutoff strategy can be shown to be revenue maximizing only if taxpayers are risk neutral and the function $t'(y)[1-F(y)]/f(y)$ is monotonically decreasing in true income y , where t' is the

conditioned on taxpayer reports can be detrimental if it leads strategic taxpayers to stop filing returns altogether. As a result, a combination of random and deterministic audits is likely to be most effective.

Melumad and Mookherjee consider the case where a social welfare maximizing government cannot commit to an audit strategy, but can commit to an audit budget and a payment structure for the tax authority. They demonstrate that the delegation of the audit authority can achieve truthful reporting and the full-commitment level of social welfare if the audit authority's *ex post* incentives are structured to be consistent with the policymaker's *ex ante* preferences. Direct limited commitment (where the policymaker commits to an overall audit budget but does not delegate auditing authority) yields a higher level of social welfare than no commitment, but does not achieve the full-commitment level of social welfare.³⁰

To summarize the results from this section, models where pre-commitment is possible imply that taxpayers who are induced to report truthfully (and often only taxpayers who are known to be truthful) will be audited. In these models, the purpose of the audit is to deter evasion rather than to collect additional tax and penalties. These models suggest that minimizing the number of unproductive audits might not be an

marginal tax function, $F(y)$ is the cumulative distribution of income and $f(y)$ is the associated probability density function.

³⁰ It is reasonable to ask why the policymaker would be able to commit to a particular payment structure for the audit authority but unable to commit directly to an audit strategy. In the authors' primary model, the audit authority is given a salary of $\alpha - \beta[A - A^*]$ if collected fines are 0 and $\alpha + \beta F$ if collected fines are greater than 0, where α and β are positive constants, A is realized aggregate audit expenditures, A^* is the commitment level of audit costs and F is aggregate fines collected. The auditor's salary is maximized (at α) when the auditor uses $A = A^*$ to induce truthful reports and collects no fines. But once truthful reports are received, the policymaker has an incentive to change the game, by for example giving the auditor a salary of α if he conducts no audits at all, thereby saving the audit expenses A^* . The authors argue that the policymaker will be unable to change the incentives after reports are received because it would be hard for legislators to manipulate bureaucrats outside of the budgetary process. Even if this argument is unconvincing, the finding that policymakers can increase social welfare by creating appropriate incentives for bureaucrats is an important one.

appropriate objective for the IRS in selecting returns for audit, especially if measured audit productivity includes only the direct effect of audits and not the deterrence effect. In addition, the models suggest that the IRS consider pre-committing to some portions of its audit plans. Pre-commitments could initially be limited to certain taxpayers or issues, so that the effect could be studied as in Slemrod et al. Partly because these models usually assume that the objective is revenue maximization, they imply that lower-income taxpayers should be audited at a higher rate. The resulting regressive bias can be offset by dividing taxpayers into different audit classes, and auditing taxpayers reporting lower incomes at higher rates than other taxpayers within the same class.

It would be useful to extend these models to account for even richer interactions between the IRS and taxpayers. For example, game theory models may be particularly useful for cases where negotiation between the taxpayer and the IRS is possible. Presumably, the possibility that negotiation will occur results in more extreme initial positions (i.e., more aggressive initial reports and more aggressive assessments) by both taxpayers and the IRS. The effect on revenue, enforcement costs, and social welfare is not immediately clear.

It is also important to recognize that IRS relies on a variety of enforcement tools (such as automatic SSN verification and the matching of third-party data including Forms W-2 and 1099 to income tax returns) as well as traditional audits. In addition, an audit may be more or less intense, focusing on many items or only a few. An audit may end very quickly, with the taxpayer agreeing to IRS' recommended assessment soon after receiving an initial notice, or it may proceed through a number of steps including litigation and collection procedures. IRS must decide what resources to commit to a

particular case at each stage, and taxpayers must decide how to act at each stage.

Therefore it might be fruitful to model the taxpayer-IRS interaction as a multi-stage game. While many of the more general findings cited in this paper are likely to hold even after additional enforcement tools are taken into account, it is important to recognize that the IRS' optimization problem is much more complicated than the models discussed here, and to think about how those complexities might change the optimal policy choices.³¹

Game-theoretic models make an important contribution to the optimal enforcement literature by recognizing that data reported by taxpayers convey signals to the tax authority that can be used in selecting returns for audit. An overarching question still to be answered is how we should balance the use of information reported by the taxpayers and by third parties (as well as DIF formulae and other information obtained from compliance research projects) with the benefits of making (and carrying out) pre-commitments that might result in truthful reports. In other words, we should ask whether there are pre-commitment strategies that can encourage compliance while at the same time making use of information received by the IRS and minimize audits of taxpayers who are likely to be compliant. This question is particularly interesting given that the IRS faces a binding budget constraint as well as a mandate to provide service to taxpayers in addition to enforcement.

Conclusion

As discussed by Plumley and Steuerle (2002), tax authorities must have clear objectives for balancing the myriad aspects of tax administration, in order make coherent

³¹ The tax administrative problem is also complicated because the IRS is composed of tens of thousands of employees. It is difficult align their individual incentives and performance goals with the overall objective. I thank Eric Toder for this observation.

resource allocation decisions. Plumley and Steuerle argue that the objective should be to collect the right amount of tax at minimum cost to taxpayers, subject to a budget constraint. In other words, both tax underpayments and tax overpayments should be minimized, and individual costs of understanding and complying with the tax code should be considered. Congleton (2002) suggests that the tax administrator might try to maximize tax revenue, minimize errors, or punish evaders. Maximizing revenue is the same as minimizing errors if there are no tax overpayments. Punishing evaders is consistent with minimizing the (direct) no-change rate. Shoup (1969) states that a fixed amount of enforcement resources can be allocated to maximize revenue, maximize accuracy, or equalize the degree of accuracy among taxpayers. None of these objectives is generally equivalent to social welfare maximization.

The finding from both the optimal tax literature and the game theory literature that maximizing revenue does not maximize social welfare even when the auditing budget is the optimal size suggests that coordination between the tax administrator and tax policymakers may be necessary. Policymakers can and do mandate changes in IRS's audit strategy, or otherwise signal to IRS that they want a change. For example, Congressional appropriators mandated a study of audit rates and "the fairness with which IRS compliance efforts are being implemented."³² Congressman Charles Rangel introduced a bill (H.R. 1661) which would, in part, require IRS to audit all taxpayers

³² See FY 2003 Treasury and Postal Appropriations Bill, House report and *Report to Congress: IRS Tax Compliance Activities*, Department of the Treasury, June 2003. The report language begins: "The Committee remains very concerned about the decline in IRS compliance actions and the degree to which IRS compliance resources are appropriately allocated to meet the most pressing needs." Fairness was not explicitly defined in the committee report. IRS interpreted it to mean that returns are not selected arbitrarily, and stated that "the IRS broadly meets that standard by selecting returns with a view towards identifying the least compliant taxpayers and promoting voluntary compliance. Ideally,... the potential burden on compliant taxpayers should be factored into the selection process."

“likely to have an unpaid Federal income tax liability of more than \$1,000,000,” and to audit “high income taxpayers likely to owe taxes” at as least as high a rate as it audits “low income taxpayers likely to owe taxes.”³³ Representatives Rosa DeLauro, Jim Cooper and Carolyn Kilpatrick offered an amendment to the FY 2005 appropriations bill to remove \$75 million from an EITC compliance program and redirect it to compliance efforts against large and midsize corporations.³⁴

The models of tax enforcement reviewed in this paper are highly stylized, but they provide useful insights regarding the nature of an optimal enforcement strategy. They suggest that IRS and policymakers should continue to think about what IRS’ objective function should be, and how the overall objective can be implemented in various enforcement programs.³⁵ IRS should also collect the data that would allow researchers to estimate the parameters of the taxpayer, IRS and policymaker first order conditions, and thereby choose the best allocation of IRS resources. Fortunately, IRS has taken a critical step in this direction by implementing a new National Research Program. It is crucial that this research effort be expanded beyond the current study of individual income tax returns, so that it can provide information on the efficient allocation of IRS resources across different types of tax entities. Finally, Treasury, Congress and other policymakers might need to consider IRS’ incentives and likely responses when making tax policy or budgetary decisions, much as they take account of taxpayer incentives and behavior.

³³ H.R. 1661, Taxpayer and Fairness Protection of 2003 (introduced April 9, 2003) Section 340. In addition, IRS would be required to audit all taxpayers likely “to have unreported income or structured transactions which are considered by the Secretary [of the Treasury] to be high risk.”

³⁴ See press release from the office of Congresswoman Rosa DeLauro, September 4, 2003.

³⁵ Dissent among members of Congress and between members of Congress and Executive Branch officials begs the question, if IRS is to maximize social welfare, what social welfare function should be used?

References

- Allingham, Michael G., and Agnar Sandmo (1972). "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1 (November): 323-338.
- Alm, James (1988). "Uncertain Tax Policies, Individual Behavior, and Welfare," *The American Economic Review*, 78(1): 237-245.
- Andreoni, James (1991). "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?" *Rand Journal of Economics*, 22(3): 385-395.
- Andreoni, James (1995). "Criminal Deterrence in the Reduced Form: A New Perspective on Ehrlich's Seminal Study," *Economic Inquiry*, 33: 476-483.
- Bebchuck, Lucian Arye and Louis Kaplow (1993). "Optimal Sanctions and Differences in Individuals' Likelihood of Avoiding Detection," *International Review of Law and Economics*, 13: 217-224.
- Becker, Gary S. (1968). "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76: 169-217.
- Blough, Roy (1952). *The Federal Taxing Process*, Washington, DC: Prentice Hall.
- Boadway, Robin, Nicolas Marceau and Steeve Mongrain (2002). "Joint Tax Evasion," *Canadian Journal of Economics*, 35(3): 417-435.
- Border, Kim C. and Joel Sobel (1987). "Samurai Accountant: A Theory of Auditing and Plunder," *Review of Economic Studies*, 54: 525-540.
- Chen, Been-Lon (2003). "Tax Evasion in a Model of Endogenous Growth," *Review of Economic Dynamics*, 6: 381-403.
- Chu, C. Y. Cyrus (1990). "Plea Bargaining with the IRS," *Journal of Public Economics*, 41: 319-333.
- Chu, C. Y. Cyrus and Neville Jiang (1993). "Are Fines More Efficient Than Imprisonment?" *Journal of Public Economics*, 51: 391-413.
- Congleton, Roger D. (2002). "Risk-Averse Taxpayers and the Allocation of Tax Enforcement Effort: Law Enforcement or Leviathan? Some Empirical Evidence," *Public Finance Review*, 30(5): 456-476.

Should it be a function consistent with the enacted tax code, or should it be altered to reflect, for example, the preferences of the current president or Treasury officials? I thank James Wetzler for this observation.

- Cowell, Frank (2002). "Sticks and Carrots," presented at the conference *The Crisis in Tax Administration*, sponsored by the Brookings Institution and the University of Michigan's Office of Tax Policy Research, November 7 and 8, Washington, DC.
- Cowell, F.A. and J.P.F. Gordon (1995). "Auditing with Ghosts," in G. Fiorentini and S. Peltzman, eds., *The Economics of Organised Crime*, London, Cambridge University Press: 185-196.
- Cremer, Helmuth and Firouz Gahvari (1995). "Tax Evasion and the Optimum General Income Tax," *Journal of Public Economics*, 60: 235-249.
- Cremer, H., M. Marchand and P. Pestieau (1990). "Evading, Auditing and Taxing," *Journal of Public Economics*, 43: 67-92.
- Erard, Brian and Jonathan S. Feinstein (1994). "Honesty and Evasion in the Tax Compliance Game," *RAND Journal of Economics*, 25(1): 1-19.
- Graetz, Michael J., Jennifer F. Reinganum and Louis L. Wilde (1986). "The Tax Compliance Game: Toward and Interactive Theory of Law Enforcement," *Journal of Law, Economics, and Organization*, 2(1): 1-32.
- Graetz, Michael, Jennifer Reinganum and Louis Wilde (1989). "Expert Opinions and Taxpayer Compliance: A Strategic Analysis," *California Institute of Technology Social Science Working Paper 710*.
- Greenberg, Joseph (1984). "Avoiding Tax Avoidance: A (Repeated) Game-Theoretic Approach," *Journal of Economic Theory*, 32: 1-13.
- Kaplow, Louis (1989). "The Optimal Probability and Magnitude of Fines For Acts That Are Definitely Undesirable," *NBER Working Paper No. 3008*.
- Kaplow, Louis (1990). "Optimal Taxation with Costly Enforcement and Evasion," *Journal of Public Economics*, 43: 221-236.
- Kaplow, Louis (1995). "A Fundamental Objection to Tax Equity Norms: A Call for Utilitarianism," *National Tax Journal*, 48(4): 497-514.
- Kaplow, Louis (1996). "How Tax Complexity and Enforcement Affect the Equity and Efficiency of the Income Tax," *National Tax Journal*, 49(1): 135-150.
- Kopczuk, Wojciech (2001). "Redistribution When Avoidance Behavior is Heterogeneous," *Journal of Public Economics*, 81: 51-71.
- Landsberger, Michael and Isaac Meilijson (1982). "Incentive Generating State Dependent Penalty System," *Journal of Public Economics*, 19: 332-352.

- Macho-Stadler, Ines and J. David Perez-Castrillo (1997). "Optimal Auditing with Heterogeneous Income Sources," *International Economic Review*, 38(4): 951-968.
- Mayshar, Joram (1991). "Taxation with Costly Administration," *Scandinavian Journal of Economics*, 93(1): 75-88.
- McCubbin, Janet (2000). "EITC Noncompliance: The Determinants of the Misreporting of Children," *National Tax Journal*, 53(4): 1135-1164.
- McCubbin, Janet (2004a). "Individual Income Tax Compliance: A Literature Review," unpublished working paper.
- McCubbin, Janet (2004b). "Optimal Tax Enforcement: A Review of the Literature and Practical Implications," *Proceedings of the 96th Annual Conference of the National Tax Association*, forthcoming.
- Melumad, Nahum and Dilip Mookherjee (1989). "Delegation as Commitment: the Case of Income Tax Audits," *Rand Journal of Economics*, 20(2): 139-163.
- Mookherjee Dilip and Ivan Png (1989). "Optimal Auditing, Insurance and Redistribution," *Quarterly Journal of Economics*, 103: 399-415.
- Plumley, Alan H. and C. Eugene Steuerle (2002). "What Should the Ultimate Objective of the Internal Revenue Service Be? A Fresh Look From an Historical Perspective," presented at the conference *The Crisis in Tax Administration*, sponsored by the Brookings Institution and the University of Michigan's Office of Tax Policy Research, November 7 and 8, Washington, DC.
- Polinsky, A. Mitchell and Steven Shavel (1984). "The Optimal Use of Fines and Imprisonment," *Journal of Public Economics*, 24: 89-99.
- Polinsky, A. Mitchell and Steven Shavel (1990). "Enforcement Costs and the Optimal Magnitude and Probability of Fines," *NBER Working Paper No. 3429*.
- Polinsky, A. Mitchell and Steven Shavel (2000). "The Economic Theory of Public Enforcement of Law," *Journal of Economic Literature*, 38(1): 45-76.
- Reinganum, Jennifer F. and Louis L. Wilde (1985). "Income Tax Compliance in a Principal-Agent Framework," *Journal of Public Economics*, 26: 1-18.
- Reinganum, Jennifer F. and Louis L. Wilde (1986). "Equilibrium Verification and Reporting Policies in a Model of Tax Compliance," *International Economic Review*, 27(3): 739-760.

- Reinganum, Jennifer F. and Louis L. Wilde (1988). "A Note on Enforcement Uncertainty and Taxpayer Compliance," *Quarterly Journal of Economics*, 103(4): 793-98.
- Reinganum, Jennifer F. and Louis L. Wilde (1991). "Equilibrium Enforcement and Compliance in the Presence of Tax Practitioners," *Journal of Law, Economics, and Organization*, 7(1): 163-181.
- Sanchez, Isabel and Joel Sobel (1993). "Hierarchical Design and Enforcement of Income Tax Policies," *Journal of Public Economics*, 50: 345-369.
- Sandmo, Agnar (1981). "Income Tax Evasion, Labour Supply, and the Equity-Efficiency Tradeoff," *Journal of Public Economics*, 16: 265-288.
- Sanyal, Amal (2002). "Audit Hierarchy in a Corrupt Tax Administration: A Note with Qualifications and Extensions," *Journal of Comparative Economics*, 30: 317-324.
- Scotchmer, Suzanne (1987). "Audit Classes and Tax Enforcement Policy," *American Economic Association Papers and Proceedings*, May, 229-233.
- Scotchmer, Suzanne and Joel Slemrod (1989). "Randomness in Tax Enforcement," *Journal of Public Economics*, 38(1): 17-32.
- Schroyen, Fred (1997). "Pareto Efficient Income Taxation Under Costly Monitoring," *Journal of Public Economics*, 65: 343-366.
- Sheffrin, Steven M. and Robert K. Triest (1992). "Can Brute Deterrence Backfire? Perceptions and Attitudes in Taxpayer Compliance, in Joel Slemrod, ed., *Why People Pay Taxes -- Tax Compliance and Enforcement*, University of Michigan Press, Ann Arbor: 193-218.
- Shoup, Carl S. (1969). *Public Finance*, Chicago: Aldine Publishing Company.
- Slemrod, Joel (1990). "Optimal Taxation and Optimal Tax Systems," *Journal of Economic Perspectives*, 4(1): 157-178.
- Slemrod, Joel, Marsha Blumenthal and Charles Christian (2001). "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota," *Journal of Public Economics*, 79: 455-483.
- Slemrod, Joel and Shlomo Yitzhaki (1987). "The Optimal Size of a Tax Collection Agency," *Scandinavian Journal of Economics*, 89(2): 183-192.
- Slemrod, Joel and Shlomo Yitzhaki (2000). "Tax Avoidance, Evasion, and Administration," *NBER Working Paper No. 7473*.

Ueng, K. L. Glen and C. C. Yang (2001). "Plea Bargaining with the IRS: Extensions and Further Results," *Journal of Public Economics*, 81: 83-98.

U.S. General Accounting Office (1999). *Tax Administration: IRS' Return Selection Process*, GAO/GGD-99-30, February.

U.S. Department of the Treasury, Internal Revenue Service (2003). *Report to Congress: IRS Tax Compliance Activities*, July 15.

Wane, Waly (2002). "Income Taxation and Tax Evasion in a Finite Economy," *Annals of Economics and Finance*, 3: 407-431.