

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:

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

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.

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.

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.

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.

