International Studies Program

International Studies Program
Working Paper 03-23
December 2003

Costs and Benefits of Marginal Reallocation of Tax Agency Resources in Pursuit of the Hard-to-Tax

Dmitri Romanov

Andrew Young School of Policy Studies

Georgia State University

International Studies Program
Working Paper 03-23

Costs and Benefits of Marginal Reallocation of Tax Agency Resources in Pursuit of the Hard-to-Tax

Dmitri Romanov

December 2003

International Studies Program
Andrew Young School of Policy Studies
Georgia State University
Atlanta, Georgia 30303
United States of America

Phone: (404) 651-1144
Fax: (404) 651-4449
Email: ispaysps@gsu.edu
Internet: http://isp-aysps.gsu.edu

International Studies Program
Andrew Young School of Policy Studies

The Andrew Young School of Policy Studies was established at Georgia State University with the objective of promoting excellence in the design, implementation, and evaluation of public policy. In addition to two academic departments (economics and public administration), the Andrew Young School houses seven leading research centers and policy programs, including the International Studies Program.

The mission of the International Studies Program is to provide academic and professional training, applied research, and technical assistance in support of sound public policy and sustainable economic growth in developing and transitional economies.

The International Studies Program at the Andrew Young School of Policy Studies is recognized worldwide for its efforts in support of economic and public policy reforms through technical assistance and training around the world. This reputation has been built serving a diverse client base, including the World Bank, the U.S. Agency for International Development (USAID), the United Nations Development Programme (UNDP), finance ministries, government organizations, legislative bodies and private sector institutions.

The success of the International Studies Program reflects the breadth and depth of the in-house technical expertise that the International Studies Program can draw upon. The Andrew Young School's faculty are leading experts in economics and public policy and have authored books, published in major academic and technical journals, and have extensive experience in designing and implementing technical assistance and training programs. Andrew Young School faculty have been active in policy reform in over 40countries around the world. Our technical assistance strategy is not to merely provide technical prescriptions for policy reform, but to engage in a collaborative effort with the host government and donor agency to identify and analyze the issues at hand, arrive at policy solutions and implement reforms.

The International Studies Program specializes in four broad policy areas:

- Fiscal policy, including tax reforms, public expenditure reviews, tax administration reform
- Fiscal decentralization, including fiscal decentralization reforms, design of intergovernmental transfer systems, urban government finance
- Budgeting and fiscal management, including local government budgeting, performance-based budgeting, capital budgeting, multi-year budgeting
- Economic analysis and revenue forecasting, including micro-simulation, time series forecasting,

For more information about our technical assistance activities and training programs, please visit our website at http://isp-aysps.gsu.edu or contact us by email at ispaysps@gsu.edu.

COSTS AND BENEFITS OF MARGINAL REALLOCATION OF TAX AGENCY RESOURCES IN PURSUIT OF THE HARD-TO-TAX

by

Dmitri Romanov

Bank of Israel Research Department, Jerusalem, Israel

ABSTRACT

As a public sector project carried out by the tax administering agency, pursuit of the hard-to-tax is subject to a cost-benefit analysis. However, considerable intrinsic uncertainty of this project and endogenous dynamic relationship between costs and benefits make the standard model of cost-benefit analysis cumbersome and barely applicable. Two handy techniques may be used instead. One, encompassed by the concept of Marginal Efficiency Cost of Funds, is built around the expected return on the inspector's time invested in the audit process. As an analytical tool, this technique is potent in identifying a gradient of possible efficient reallocation of the tax agency resources; as an administrative tool, it tends to equalize the inspectors yield in the audit process. Both aspects are illustrated on the Israeli Income Tax Administration data. The second technique of cost-benefit analysis is Feinstein's econometric Detection Controlled Estimation model – amended here to include both type I and type II errors of incorrect detection of tax non-compliance.

COSTS AND BENEFITS OF MARGINAL REALLOCATION OF TAX AGENCY RESOURCES IN PURSUIT OF THE HARD-TO-TAX*

INTRODUCTION

There should be no disagreement at this conference that in a short run taxation of the hard-to-tax constitutes merely a tax administration – not tax policy – problem. This problem has numerous facets and solutions, depending on the tax in question, hard-to-tax group in question, specific legal environment and enforcement tools in hand of a tax administering agency. But whatever action is taken, one basic question has to be answered first: how much resources should be allocated to the pursuit of a given group of tax-(non)payers? As every tax administrator knows, new aims do not necessarily come with new budgets. Hence, unless the agency is provided with an earmarked budget for pursuit of the hard-to-tax, any expansion of operations toward the hard-to-tax necessarily implies a reallocation of existing resources.

As a public sector project, it seems to be the case for a standard cost-benefit analysis (as exposed, for example, by Dreze and Stern, 1987). Yet evaluating this project is not an easy job. Since neither involved administrative costs nor anticipated direct benefits are certain and well defined, modeling the world "with" the project is not straightforward. Furthermore, the world "without" the project – not to take on the hard-to-tax – is going to change endogenously because the existence of a group that effectively avoids taxation will spoil perception of justice amongst compliant taxpayers, causing imitation of the hard-to-tax avoidance practices and deterioration of future tax collection. Without accounting for dynamics of this threat (adding more uncertainty and further complicating the cost-benefit analysis model) one will likely end up with negative shadow prices prohibiting the pursuit of the hard-to-tax. But the list of *social* costs and benefits of pursuing the hard-to-tax will be still incomplete even after taking this factor into account. One has to put considerations of horizontal inequality –

* I owe to Jonathan Feinstein and Shlomo Yitzhaki for discussion and many helpful suggestions.

unaccounted for within the efficiency dimension of the cost-benefit analysis – on the benefit side to internalize negative externality associated with the existence of the hard-to-tax.

The objective of these notes is to examine a number of substantial issues related to the cost-benefit analysis of chasing the hard-to-tax, to review two convenient techniques and to illustrate their application using micro-level Israeli data.

THE HARD-TO-TAX: A CLASSIFICATION

To make the concept of cost-benefit analysis operational in the context of the hard-to-tax, one has to start with their classification. The following classification – one among many others possible and definitely not a comprehensive one – is build around the extent of risk borne by different hard-to-tax groups. This criterion arises from Slemrod and Yitzhaki (1987) who proved that, at the margin, the administrative cost of extra enforcement should equal the saving of excess burden[1] due to the decline in the exposure to risk. Then one has to make a distinction between evasion and avoidance (Slemrod and Yitzhaki, 2002): the former is illegal and, given a non-zero probability of detection, entails excess burden; the latter is legal and risk-free, therefore it does not bring about any excess burden – which naturally does not mean that there is no real substitution response or "sheltering" efforts (Mayshar, 1991).

By this rule, there are three groups of hard-to-tax evaders:

- The hard-to-catch (a.k.a. nomads, ghosts, non-filers, moonlighters, etc.), who evade all (relevant) taxable income by escaping the tax agency net. Representing a "corner solution" in the Allingham-Sandmo model, this group was largely sidelined by the mainstream economics-of-evasion literature. Recently, however, the hard-to-catch have attracted attention in the models of Yaniv (2003) and Polinsky (2002), and in the empirical study by Erard and Ho (1999).

[1] Yitzhaki (1987) defines excess burden of tax evasion, in utility terms, as the difference in expected utility between the case of paying the duties (taxes, fines and other applicable charges) with certainty and the alternative of paying them with uncertainty.

- The hard-to-detect, who, being in the tax agency net, succeed in underreporting taxable income and/or overreporting deductible expenses or allowances because of imperfect detection. This group also includes those who exploit tax law gray areas since they bear some risk (and associated excess burden) while gambling on whether the agency will take their case to the court, or will give up – being unable to prove the case – for fear of establishing a precedent.
- The hard-to-collect, whose tax liability is known and agreed upon, but who are unable to pay due taxes and fines. Andreoni (1992) motivates this type of evasion behavior by inter-temporal income smoothing on the part of taxpayers with borrowing constraints. Arguably, the tax collector can do nothing about this debt other than to reschedule it in anticipation of better times for the debtors. By its nature, this form of taxpayers' debt is counter-cyclical, temporal and mostly independent of the agency's actions. However, there may be other reasons for mounting arrears, such as an inadequate tax enforcement system. First, when the interest charged by the agency on late tax payments is lower than the market rate, everyone may prefer to "borrow" from the tax collector[2]. Second, when taxpayers are fined at skyrocketing rates, their aggregate debt grows exponentially, eventually becoming unmanageable and – by feeding anticipation of a tax amnesty – further worsening tax collection.

Probably the largest group of the hard-to-tax are the hard-to-levy, who escape taxation through legitimate avoidance schemes, or are exempt for any reason, *inter alia*, because the law is unenforceable. Unlike in the case of evasion, there is no *contemporary* excess burden of avoidance, but if there is some risk of taxing the hard-to-levy in the future, and they are engaged in costly avoidance activity (e.g., hiring legal or tax practitioner's advice, PR,

[2] In Israel, since 1993 this rate stands at 4 percent, as compared with 10.6 percent, the average real expected

lobbying against the change of law, etc.), then this hard-to-tax group has to be accounted for in cost-benefit analysis as well – perhaps by a social planner, not by the tax administering agency.

Henceforth I focus on the techniques of cost-benefit analysis generally applicable for the hard-to-catch and the hard-to-detect – whose treatment constitutes the lion's share of a tax administrator's routine. Having said this, one should not overlook the strategic nature of problems associated with the accumulating debt of the hard-to-collect and widening population of the hard-to-levy, the solution of which seems to be in the hands of legislators.

COST-BENEFIT ANALYSIS OF THE HARD-TO-TAX

It is helpful to consider the cost-benefit analysis of the hard-to-tax in the general normative framework of marginal efficiency cost of funds (MECF) in the presence of evasion and avoidance, as developed by Slemrod and Yitzhaki (1996, 2002). This concept can be used with respect to marginal changes in any parameter of the tax system, including tax enforcement strategies. Moreover, MECF does not assume optimality as a benchmark state of nature; that is why this concept can be used to identify incremental changes in tax (administration) policy that would increase social welfare.

Ignoring the possibility of taxpayers being at corner solutions (as are, for example, the hard-to-collect), MECF associated with tax (administration) instrument *i* is:

$$MECF_i = \frac{X_i + C_i}{MR_i - A_i} \qquad (1)$$

where MR_i is the marginal collection, A_i is the marginal administrative cost, X_i is the potential tax collection associated with this instrument, and C_i is the marginal private compliance cost, excluding expenses spent on tax avoidance.

When it comes to the marginal reallocation of the agency resources in its pursuit of tax

short-term interest in the same period.

evasion – chasing the hard-to-catch or boosting compliance of the hard-to-detect – expression (1) can be further simplified. First, all enforcement strategies aimed at the same group of taxpayers have the same X_i which is determined by the statutory tax code. Second, as long as the agency stays on the budget constraint – just reallocating its resources from one task to another – there is no marginal administrative cost. Hence, alternative tax enforcement instruments may differ only in the associated private compliance cost and generated marginal collection:

$$MECF_i = \frac{C_i}{MR_i} \tag{2}$$

Among alternative feasible tax instruments one should use one with the lowest MECF. This simple rule of choice says that the socially preferable tax enforcement strategy has to minimize private compliance cost per dollar of marginal revenue (decreased evasion). For example, consider a choice among alternative enforcement policies aimed at the same hard-to-tax group. It turns out that the rule of thumb for picking an option that maximizes the expected tax revenue is nothing but a private case of the MECF-criteria, under the assumption that the alternatives do not entail marginal compliance costs (differentiated among the taxpayers). If the choice is between searching for the ghosts and improving detection among the filers, provided both moves bring an equal expected revenue, then generally the latter option should be preferred because the marginal compliance cost borne by the ghosts tends to be much higher than that of the filing taxpayers[3].

It is worth emphasizing that being instrumental for a short-run analysis of tax agency efficiency, the MECF concept does not expose the shadow price of tax enforcement instruments in terms of horizontal or vertical inequality, whereas the two criteria may contradict each other (Wertz, 1979). The existence of the hard-to-tax causes loss of social

[3] That may explain, among other motives, why tax administrators usually opt to "fix the holes" in the existing network instead of widening it in search of ghosts. Erard and Ho (1999) point to high compliance costs as a reason for an individual to become a ghost.

welfare, which extent is positively related to the scope of tax evasion (Lambert and Yitzhaki, 1995). Notably, this negative externality is only partially accounted for in the cost-benefit analysis when the future conversion of candid taxpayers into the hard-to-tax is taken into consideration. As long as there are taxpayers choosing not to join the hard-to-tax group – because they are "pathologically honest" (Erard and Ho) or the costs of avoidance/evasion exceed the anticipated gains – horizontal inequality in enforcement will persist. When pursuing a group of the hard-to-tax with the highest evasion does not pay off to the tax agency, according to the MECF criteria, because of relatively higher administrative costs, the excessive loss of social welfare due to horizontal and vertical inequality should be put on the scales to overcome the cost-benefit analysis outcome.

Two micro-level techniques are useful for evaluating marginal expected return per dollar of administrative costs[4]. Considering the "tax-compliance game" as a set of decisions made by a taxpayer and a tax inspector along their interleaving decision trees, Yitzhaki and Vakneen (1989) suggested that the shadow prices of the tax inspector's time invested in enforcement, conditional on its duration and eventual return be computed. Feinstein (1990, 1991) developed an econometric model of Detection Controlled Estimation (DCE), yielding differentiated evasion (and aggregated tax gap) estimates controlled for imperfect detection of evasion by the tax enforcement agency.

In the two following sections I discuss these techniques and demonstrate their use in a simulation of feasible tax enforcement strategies of the Israeli Income Tax Administration, if it were interested in (counterfactual) pursuing some specific hard-to-tax groups. Since this work makes use of the dataset extensively documented in Romanov (2001), description of data is skipped here.

[4] The IRS purportedly performs this evaluation by comparing the evasion detected in scrupulous TCMP audits with that revealed in the ordinary audits (Steuerle, 1986).

SHADOW PRICE OF AUDITING RESOURCES

While such extremely efficient tools as withholding at source, information reporting and matching have become a standard practice of tax enforcement in developed and developing countries, effectively zeroing tax evasion from wages and salaries, tax administering agencies inevitably find themselves focusing on harder-to-monitor types of income. Regarding these incomes, auditing is the ultimate enforcement and deterrence tool. The expected return per unit of inspector's time invested in auditing, proposed by Yitzhaki and Vakneen (1989) as an aggregate measure of efficiency of tax enforcement among different groups of taxpayers, accounts for different quantity of labor input (administrative costs) invested in various enforcement tasks on all stages of auditing, and for different expected return – in terms of marginal tax collection – on this input. Moreover, taking into consideration the duration of the audit process, this measure implicitly accounts for marginal compliance costs. Hence, ranking of alternative tax enforcement strategies by the criteria of maximum expected return per unit of inspector's time invested in auditing will be identical to their ranking by the criteria of minimum MECF, simply because the former criterion is the inverse of the latter.

I illustrate the use of this method for evaluating different strategies of pursuing the hard-to-tax by using the example of income tax enforcement among the self-employed in Israel. Unlike the vast majority of wage-earners, who are exempt from filing because their tax liability is completely settled through withholding at source, Israel’s 275,000 self-employed taxpayers (analogous to US taxpayers filing federal Schedule C or Schedule F) have many opportunities for tax evasion, putting them at the core of the enforcement activity. They are obliged to file an annual tax return and, at least once every five years, a statement of wealth. On average, about 45-55% of the returns filed by the self-employed are audited, indicating that the average auditing rate in Israel is about ten times higher than for the comparable group of taxpayers in the United States.

The flowchart of the auditing process statute is presented on Figure 1. Depending on the taxpayer's type[5], the inspector is allocated a given time for the audit. These norms are given in column (1) of Table 1. Auditing taxpayers of type 52 and type 53, who run relatively large and established enterprises, takes more time.

A comparison of columns (2) and (3) shows that these taxpayers also "enjoy" a higher density of audits than their peers, notably those of type 30. If the audit does not terminate by agreement (outcomes 01, 02) and the taxpayer protests the auditor's assessment, asking for another audit – usually performed by a different inspector – and is granted the second hearing, it takes the agency twice the norm to process this return (not including additional resources to prepare the case for the court's hearing in outcome 09). There is wide variation in the probability that the audit will reach Stage 2: for taxpayers of type 52 it happens almost three times less frequently than for taxpayers of type 53 (column 4).

The expected return per audit day (column 5) is calculated as:

$$ER_i = \frac{\sum_j \pi_{ij} MR_{ij}}{T_i \sum_j \pi_{ij} \mu_j} \qquad (3)$$

where i indicates different types/groups of taxpayers; j indicates all possible audit outcomes (01, 02, 03, 07, 09 on Figure 1); π is the probability of terminating the audit with a given outcome; MR is the average return per audit (in terms of tax- or income-correction); T is the auditor's time invested in the process; and μ is the number of audits performed (one for outcomes 01, 02, 03, or two for outcomes 07, 09).

Columns (5) and (6) feature two measures of the tax agency return on conducted audits: the expected return per day of audit and the expected return per audit. The difference between

[5] Five types (41-53 as detailed in notes to Table 1) are defined according to the types of bookkeeping, while the latter is determined, among other factors, by the size of the taxpayer's business and its nature. Type 30 indicates a taxpayer who is employed as a manager in the enterprise that is at least partially owned/controlled by him/her.

them is clear: the former accounts for labor intensity and the chance that the audit process will extend to Stage 2, costing the agency more[6], while the latter values the audit at its face value.

Table 1. Auditing the self-employed in Israel

Type of taxpayer [a]	Days per audit (1)	Share of taxpayers, % (2)	Share of audits, % (3)	Share of audits on Stage 2[b], % (4)	Expected return per audit day [c] (5)	Average return per audit [c] (6)
30	1.0	23.0	18.3	7.2	6.7	7.2
41	0.53	6.6	7.3	12.8	10.2	6.1
42	1.0	43.0	44.4	13.3	6.9	7.8
43	1.0	8.4	7.7	10.4	5.7	6.2
52	1.33	15.6	17.9	5.6	3.5	4.9
53	1.67	3.4	4.5	15.2	10.0	19.3
Total		100.0	100.0			

[a] Legend: 30 — managers-shareholders; 41 — taxpayers obliged to keep the simplest one-sided form of bookkeeping; 42 — taxpayers obliged to keep one-sided form of bookkeeping with accumulated entries; 43 — agricultural businesses without two-sided bookkeeping; 52 — taxpayers obliged to keep one-sided form of bookkeeping with cash entries; 53 — taxpayers obliged to keep two-sided form of bookkeeping and taxpayers who have volunteered for this form of bookkeeping.

[b] The share of audits with outcomes 07 and 09, by type of return.

[c] NIS 000s, 1993 prices.

This difference brings about considerable disparity in the marginal evaluation of enforcement efficiency among certain groups of taxpayers, notably in types 41 and 53. It is worth increasing the density of audits among type 41 taxpayers, because investing half-day yields, on average, the same return as the inspection taking at least twice that time (types 30, 42, 43, 52). On the other hand, it turns out that the type 53 inspection is not as promising as it

[6] Taxpayers complain and ask for the second audit when an inspector's assessment at Stage 1 seems them unreasonable and their anticipated gain (from reducing this assessment) exceeds marginal compliance costs – see model below. If Stage 2 audit does indeed reduce the assessment below the level that would trigger the taxpayer's complaint, the agency would lose not only the time of second audit (valued at its shadow price), but also a part of tax collection.

seems by gross return per audit. Auditing of these "heavyweights" effectively brings into the agency's coffers not more than that of the "lightest" tax returns. Another finding points to the waste of the tax agency's resources due to over-sampling of type 52 – the least rewarding enterprise – taking more than 20 percent of the auditors' time budget.

To summarize this exercise, allocation of enforcement resources by the Israeli Income Tax Administration may be improved by a shift, at the margin, away from labor-intensive auditing of most compliant taxpayers (type 52) toward a closer inspection of more evasive taxpayers (types 30, 41, 42). This change would be economically efficient and fair at the same time, because it improves horizontal equality bringing about more uniform distribution of effective tax burden.

In addition to identifying the gradient of the marginal reallocation of the tax agency's resources, the expected return per unit of invested time is instrumental in monitoring inspectors' yield and streamlining their activity in the audit process. As the extent of a taxpayer's evasion is often uncertain – or additional resources are required to make it certain – an auditor's assessment may actually fall within a fairly broad range. The inspector's problem is, then, to define a reasonable quantity of assessment, keeping in mind that a low assessment of evasion decreases both the return on audit time and the probability of the taxpayer's dissent, whereas a high assessment, boosting the audit's return, at the same time increases the chances of the second audit (tending to regress the assessment toward the mean) triggered by the taxpayer's complaint. These considerations, embodied in the concept of the expected return per unit of tax agency resources, would soften "tough" inspectors and harden "soft" ones, asymptotically equalizing inspectors' expected return.

To put the inspector's decision formally, let X be the taxpayer's true taxable income (unknown to the inspector); X_1 and X_2 – its assessment by the first- and second-stage auditors, respectively, when $\mathrm{E}X_2 < X_1$; t – tax function, including all penalties for evasion; a_1 and a_2 – marginal administrative costs of the first- and second-stage audits; c – marginal compliance costs of the first- and second-stage audits; p – probability that the taxpayer will disagree and get the second audit; q – exogenous probability that the second audit will lower the first audit's assessment. Without loss of generality, marginal administrative and compliance costs are assumed as being independent of the inspector's assessment.

The inspector maximizes the expected net return on the tax agency's resources invested at both stages of the audit process:

$$\max_{X_1}\{t(X_1) - a_1 - p(X_1)a_2 - p(X_1)q[t(X_1) - t(X_2)]\} \tag{4}$$

First order condition is:

$$t'(1 - pq) = p'(a_2 + q\Delta t) \tag{5}$$

where $\Delta t = t(X_1) - t(X_2)$, a positive value provided t(.) is non-decreasing in X_1. That is, the inspector will raise the assessment to the point where the marginal expected tax collection will be equal to the marginal increase in expected administrative costs associated with the second audit (costing a_2 and decreasing the assessment from X_1 to X_2 with probability q). Comparing the optimal assessment X_1^* (solving (5) subject to p'*) with $\overline{X}_1^*$ – the optimum when the second audit's cost and outcome are not taken into account (maximization of net return on the first audit ignoring endogeneity of p) – it is easy to see that $X_1^* < \overline{X}_1^*$ because in the latter case f.o.c. requires pq=1. Since this condition never holds, the inspector is found at the corner solution, making the highest possible assessment.

The probability of the second audit is determined by the taxpayer who maximizes her expected utility given the exogenous assessment X_1^*:

$$\max_{p}\{(1 - p)U(X - t(X_1^*) - c_1) + pU(X - (1 - q)t(X_1^*) - qt(X_2) - c_1 - c_2)\} \tag{6}$$

The first order condition is simply $q\Delta t = c_2$, i.e. independently of the utility function curvature the taxpayer "jumps" from p*=1 if $q\Delta t > c_2$ to p*=0 if $q\Delta t < c_2$, and randomizes complaining when the f.o.c. holds as equality. This does not mean, however, that one should expect a none-or-all reaction. In the continuum of taxpayers with different compliance costs (at least because of different valuation of time expected to be spent on the second audit) and different anticipated gain from the repeated audit (when first audit assessment is correct or undisputable, the gain is zero), only some taxpayers will complain[7]. But when a taxpayer's marginal compliance costs of standing the second audit are negligible, with non-zero expected value of tax liability reduction, the tax agency may be anyway flooded with requests for a repeated audit. To prevent this situation, the tax agency could: a) raise the taxpayers' compliance costs, for example, by requiring new evidence as a prerequisite of the second audit; b) instruct the first-stage inspectors to be "cautious" in their assessments. This is the mechanism equalizing the expected return per unit of audit time across inspectors.

Figure 2, showing spectacular differences in the extent of income corrections by audit outcomes – with little variation across taxpayer types, suggests that taxpayers' decisions in the course of the audit process may be more homogeneous than those of tax inspectors. At outcome 01 audits merely correct a taxpayer's minor accounting errors, ending up, on average, with a negligible discrepancy of 0.5 percent of the reported taxable income. At outcome 02 the average assessment – accepted by a taxpayer – stands at 18 percent of the reported income, with some 80 percent of corrections being between 10 and 20 percent. Strikingly, repeated audits, terminated to the taxpayers' satisfaction (outcome 07), close with a similar income correction (21 percent at average), probably indicating that the Israeli self-employed[8] perceive an assessment of evasion at 15-20 percent of their true income as a reasonable and agreeable audit's outcome. Anyway, this is not the case for unearthed large-scale evasion, supposedly present at outcome 03. With the average assessment of 96 percent of the reported taxable income (ranging from 75 percent for type 30 to 110 percent for type 41), these

[7] For these reasons knowledge of the taxpayer's decision rule is not very instructive for the inspector solving (5). The first-stage inspector takes expectation over X_2 (conditional on X and X_1) and c_2 possible for a given taxpayer. That is why p' in (5) is continuous in spite of the fact that p'* probably is not.

[8] Of course, this is not a general "sensitivity threshold of evaders". In fact, this trait should not even be generalized to the whole population of Israeli self-employed because it is sample-specific, and given a strong

taxpayers either were not eligible for the repeated audit (e.g., because of serious bungles in bookkeeping), or failed to convince the second inspector to diminish the assessment.

The concept of the expected return on time invested in auditing – built to account for the endogenous relationship between the tax agency's inputs and outputs – can easily be applied to evaluating other enforcement activities, not only those consuming auditors' time as in the example of optimal audit assessment given above.

Generally, this technique – choosing activities maximizing tax agency's net expected return – is a mirror reflection of its alternative – a technique choosing activities minimizing undetected evasion – discussed in the following section.

DETECTION CONTROLLED ESTIMATION OF EVASION

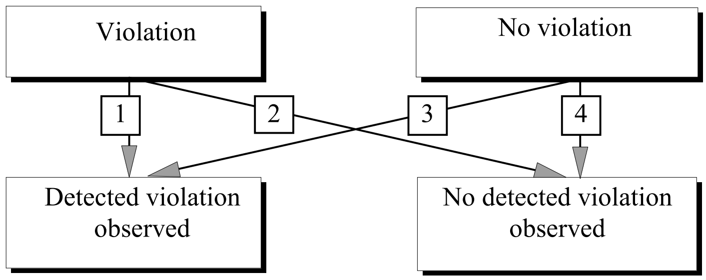
By definition, the hard-to-tax are more evasive than other taxpaying public. This feature has two closely related components – higher evasion per se and lower detection per audit. As a result, inference from the audited sample regarding the "true" extent of evasion by different groups of taxpayers may be biased. Feinstein's (1990, 1991) econometric framework of DCE, accounting for incomplete detection of tax code violations, yields statistical estimates of the tax gap; these estimates are very close to the IRS figures.

Figure 3 shows the basic problem of inferring marginal distribution of violations given a set of observed inspection outcomes. The observed case status misrepresents the true status when two "miscarriages of justice" happen: overdetection (type I error) and underdetection (type II error). The third possible inspection outcome – correct detection – occurs when inspector either detects the violation that indeed took place, or properly states that no violation happened.

---

selectivity of audit sample it is a sure bet to say that other taxpayers would not agree with a 15-20 percent addition to their reported taxable income.

Figure 3: Detection in Random Sample



Legend: arrows 1 and 4 – correct detection; arrow 2 – type II error; arrow 3 – type I error.

This problem has been tackled recently by the misclassification literature (e.g., Hausman et al., 1998, Lewbel, 2000) in the binary semi-parametric and non-parametric setting. Feinstein developed and estimated a number of parametric models for binary and continuous variables representing violation (tax non-compliance), all of them under assumption that the probability of overdetection (type I error) is zero[9]. However, this assumption seems to be too restrictive. As seen in the previous section, the number of tax returns reaching the second stage of the audit process is not negligible, leading in many cases to significant correction of the first-stage assessments. Thus, given considerable private and social costs associated with overdetection – taking also into consideration that some taxpayers choose not to complain[10] (p*=0 if $q\Delta t < c_2$) – it becomes clear that the DCE model has to be amended to incorporate this audit outcome. In the remaining part of the paper I restate the Feinstein's model in a way allowing this amendment, and present the comprehensive model of detection controlled estimation.

[9] Discussing this case as false detection, Feinstein (1990) claims that it "may well be less empirically relevant than the related problem of 'gray areas' of the law which can affect empirical analysis whenever monitors misinterpret compliant behavior as illegitimate" (pp. 241-242).

[10] There is no gain of complaining and asking for the second audit in cases of correct detection of compliance (arrow 4 on Figure 3) and underdetection (arrow 2). A sufficient condition for complaining in a case of correct detection of violation (arrow 1) is that second inspection has positive conditional probability, q, of accepting the complainant's claim of being a case of overdetection. Since producing the evidence supporting this claim is clearly an act of forgery, a claimant may be fined if caught cheating with probability 1-q. Sufficiently high level of fine prevent complaining in this case. Thus, no complaints are expected to be filed in the cases of correct detection and underdetection.

Let N_i^* be the latent propensity of individual *i* to commit violation. The *extent* of committed violation, n_i, follows tobit specification:

$$\begin{cases} n_i = x_i\beta + \varepsilon_i \\ n_i = 0 \end{cases} \text{if} \begin{cases} N_i^* > 0 \\ N_i^* \leq 0 \end{cases} \qquad (7)$$

where x_i is an array of individual's traits and opportunities of violation, ε_i is the random disturbance i.i.d. $N(0, \sigma_\varepsilon^2)$, by assumption.

Detection is modeled as a two-phase process. The first phase deals with the probability of correct detection, while the second phase – conditional on the failure to correctly detect violation or compliance at the first phase – specifies the type of error (type I or type II). In fact, this setup rests on the assumption that inspection is a fair trial where the inspector makes her best to reveal the truth, and errors happen only when the case is unclear. This assumption naturally holds on a court's floor in the presence of a benevolent judge and the procedural balance between the disputing sides. In an administrative inspection process, however, the balance is usually shifted towards the inspector wearing two hats – the prosecutor's, and the judge's. A fair trial assumption implies, therefore, that the inspector's stance is not *a-priori* biased either toward implicating individuals in imaginary violations or toward overlooking non-compliance.

The first phase of detection process – modeling the likelihood of correct detection – is specified by Equation (8). The latent indicator variable y_i equals 1 if the inspector correctly detects the extent of non-compliance and 0 otherwise:

$$y_i^* = z_{1i}\gamma_1 + v_i \text{ and } y_i = 1(y_i^* > 0) \qquad (8)$$

Here, 1(Q) denotes indicator function of the event Q; vector z_{1i} represents the factors easing or hindering correct detection in the case *i*, including inspector's, individual's and inspection process attributes[11]; and v_i is assumed to be a standard normal random disturbance

[11] Among these factors one may mention the case complexity, the inspector's caseload and the inspector's

(normalized because its scale is unidentified).

Let the extent of detected non-compliance in case *i* be:

$$m_i = z_{2i}\gamma_2 + \omega_i \tag{9}$$

where the disturbance term ω_i is i.i.d. $N(0, \sigma_\omega^2)$, by assumption.

When inspection terminates with the assessment of non-compliance $m_i>0$, it must be one of three mutually exclusive events: correct detection ($m_i=n_i$); overdetection ($m_i>n_i$), with the private case of $n_i=0$ (i.e. accusation of a compliant individual in cheating); and partial detection ($m_i<n_i$). Observing no detected non-compliance ($m_i=0$) may follow either from correct detection of individual's compliance ($m_i=0|n_i=0$), or from the failure to detect any violation whatsoever ($m_i=0|n_i>0$).

Then, the probability of type I error (overdetection) is:

$$\alpha_{1i} \equiv \Pr(m_i > n_i \mid n_i \geq 0,\, y_i = 0) \tag{10}$$

and the probability of type II error (underdetection) is:

$$\alpha_{2i} \equiv \Pr(0 < m_i < n_i \mid n_i > 0,\, y_i = 0) + \Pr(m_i = 0 \mid n_i > 0, y_i = 0) \tag{11}$$

The first component of (11) is the likelihood of partial detection, whereas its second component is the probability of failure to uncover any non-compliance.

Provided disturbance terms ε_i, v_i and unconditional ω_i are independent,[12] likelihood of set of the cases with observed detected non-compliance is:

experience. The first two factors are negatively related to the probability of correct decision, when the inspector's experience improves the chances of detecting the truth in a case of the given complexity. While affecting the chances of correct detection, these factors do not bend the inspection's outcome to error of either type.

[12] The assumption of independence is not crucial here. Generalization for the correlated disturbances is straightforward.

$$L1= \begin{aligned}&\prod_{m_i>0}\{\Pr(m_i=n_i\cap y_i=1)+\Pr(m_i<n_i\mid y_i=0)+\Pr(m_i>n_i\mid y_i=0)\}=\\
&\prod_{m_i>0}\{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{m_i-x_i\beta}{\sigma_\varepsilon}\right)\Phi(z_{1i}\gamma_1)+\int_0^{m_i}\frac{1}{\sigma_\varepsilon}\phi\left(\frac{n_i-x_i\beta}{\sigma_\varepsilon}\right)\Phi(-z_{1i}\gamma_1)\frac{1}{\sigma_\omega}\phi\left(\frac{m_i-z_{2i}\gamma_2}{\sigma_\omega}\right)dn_i+\\
&\int_{m_i}^{\infty}\frac{1}{\sigma_\varepsilon}\phi\left(\frac{n_i-x_i\beta}{\sigma_\varepsilon}\right)\Phi(-z_{1i}\gamma_1)\frac{1}{\sigma_\omega}\phi\left(\frac{m_i-z_{2i}\gamma_2}{\sigma_\omega}\right)dn_i\}=\\
&\prod_{m_i>0}\{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{m_i-x_i\beta}{\sigma_\varepsilon}\right)\Phi(z_{1i}\gamma_1)+(1-\Phi(z_{1i}\gamma_1))\frac{1}{\sigma_\omega}\phi\left(\frac{m_i-z_{2i}\gamma_2}{\sigma_\omega}\right)\int_0^{\infty}\frac{1}{\sigma_\varepsilon}\phi\left(\frac{n_i-x_i\beta}{\sigma_\varepsilon}\right)dn_i\}=\\
&\prod_{m_i>0}\{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{m_i-x_i\beta}{\sigma_\varepsilon}\right)\Phi(z_{1i}\gamma_1)+(1-\Phi(z_{1i}\gamma_1))\frac{1}{\sigma_\omega}\phi\left(\frac{m_i-z_{2i}\gamma_2}{\sigma_\omega}\right)\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\}\end{aligned} \qquad (12.1)$$

The likelihood of the complementary set containing the cases with no detected non-compliance is:

$$L2= \begin{aligned}&\prod_{m_i=0}\{\Pr(n_i=0\cap y_i=1)+\Pr(m_i=0\mid n_i>0,y_i=0)=\\
&\prod_{m_i=0}\{\left(1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right)\Phi(z_{1i}\gamma_1)+\int_0^{\infty}\frac{1}{\sigma_\varepsilon}\phi\left(\frac{n_i-x_i\beta}{\sigma_\varepsilon}\right)\Phi(-z_{1i}\gamma_1)\frac{1}{\sigma_\omega}\phi\left(\frac{-z_{2i}\gamma_2}{\sigma_\omega}\right)dn_i\}=\\
&\prod_{m_i=0}\{\left(1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right)\Phi(z_{1i}\gamma_1)+(1-\Phi(z_{1i}\gamma_1))\frac{1}{\sigma_\omega}\phi\left(\frac{-z_{2i}\gamma_2}{\sigma_\omega}\right)\int_0^{\infty}\frac{1}{\sigma_\varepsilon}\phi\left(\frac{n_i-x_i\beta}{\sigma_\varepsilon}\right)dn_i\}=\\
&\prod_{m_i=0}\{\left(1-\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\right)\Phi(z_{1i}\gamma_1)+(1-\Phi(z_{1i}\gamma_1))\frac{1}{\sigma_\omega}\phi\left(\frac{-z_{2i}\gamma_2}{\sigma_\omega}\right)\Phi\left(\frac{x_i\beta}{\sigma_\varepsilon}\right)\}\end{aligned} \qquad (12.2)$$

There are five parameters to estimate in this model: β, γ_1, γ_2, and two standard errors, σ_ε and σ_ω. This parametric model is identified by non-linearity of the probability distribution functions meeting the standard regularity conditions[13]. Given that the monotonicity condition (Hausman et al., 1998) $\alpha_{1i}+\alpha_{2i}=\Pr(y_i=0)<1$ is held by construction, a semi-parametric model is identified if each of x_i, z_{1i} and z_{2i} meets the exclusion condition – includes at least one variable with an unbounded support excluded from two other arrays (see Feinstein, 1990, Theorem A2).

[13] The proof follows Feinstein's (1990) Theorem A1.

REFERENCES

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

Dreze, J., and Stern, N., 1987. The theory of cost-benefit analysis. In A. J. Auerbach, and M. Feldstein (eds.) *Handbook of Public Economics, volume 2.* Elsevier Science B.V., 909-990.

Erard, B., and Ho, C.-C., 1999. Searching for ghosts: who are the nonfilers and how much tax do they owe? Working Paper 99-11, Department of Economics, Carleton University.

Feinstein, J. S., 1990. Detection controlled estimation. *Journal of Law and Economics* 33, 233–276.

Feinstein, J. S., 1991. An econometric analysis of income tax evasion and its detection. *RAND Journal of Economics* 22 (1), 14–35.

Hausman, J. A., Abrevaya, J., and Scott-Morton, F. M., 1998. Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics* 87 (2), 239-269.

Lambert, P. J., and Yitzhaki, S., 1995. Equity, equality and welfare. *European Economic Review* 39 (3-4): 674-682.

Lewbel, A., 2000. Identification of the binary choice model with misclassification. *Econometric Theory* 16 (4), 603-609.

Mayshar, J., 1991. Taxation with costly administration. *Scandinavian Journal of Economics* 93 (1), 75-88.

Polinsky, A. M., 2002. Optimal sanctions when wealth can be audited. The paper presented at the 2002 NBER Summer Institute.

Romanov, D., 2001. Analysis of organizational problems of income tax enforcement. PhD dissertation. The Hebrew University of Jerusalem.

Slemrod, J., and Yitzhaki, S., 1987. The optimal size of a tax collection agency. *Scandinavian*

Journal of Economics 89 (2), 183-192.

Slemrod, J., and Yitzhaki, S., 1996. The cost of taxation and the marginal efficiency cost of funds. *International Monetary Fund Staff Papers* 43 (1), 172-198.

Slemrod, J., and Yitzhaki, S., 2002. Tax avoidance, evasion, and administration. In A. J. Auerbach, and M. Feldstein (eds.) *Handbook of Public Economics, volume 3*. Elsevier Science B.V., 1423-1470.

Steuerle, C. E., 1986. *Who should pay for collecting taxes?* American Enterprise Institute for Public Policy Research, AEI study 448.

Wertz, K. L., 1979. Allocation by and output of a tax-administering agency. *National Tax Journal* 32 (2), 143–155.

Yitzhaki, S., 1987. On the excess burden of tax evasion. *Pubic Finance Quarterly* 15 (2),123-137.

Yitzhaki, S., and Vakneen, Y., 1989. On the shadow price of a tax inspector. *Public Finance* 3, 492-505.

Yaniv, G., 2003. *Auditing ghosts by prosperity signals*. Unpublished manuscript.

Figure 1. The Audit Process as a Sequential Game

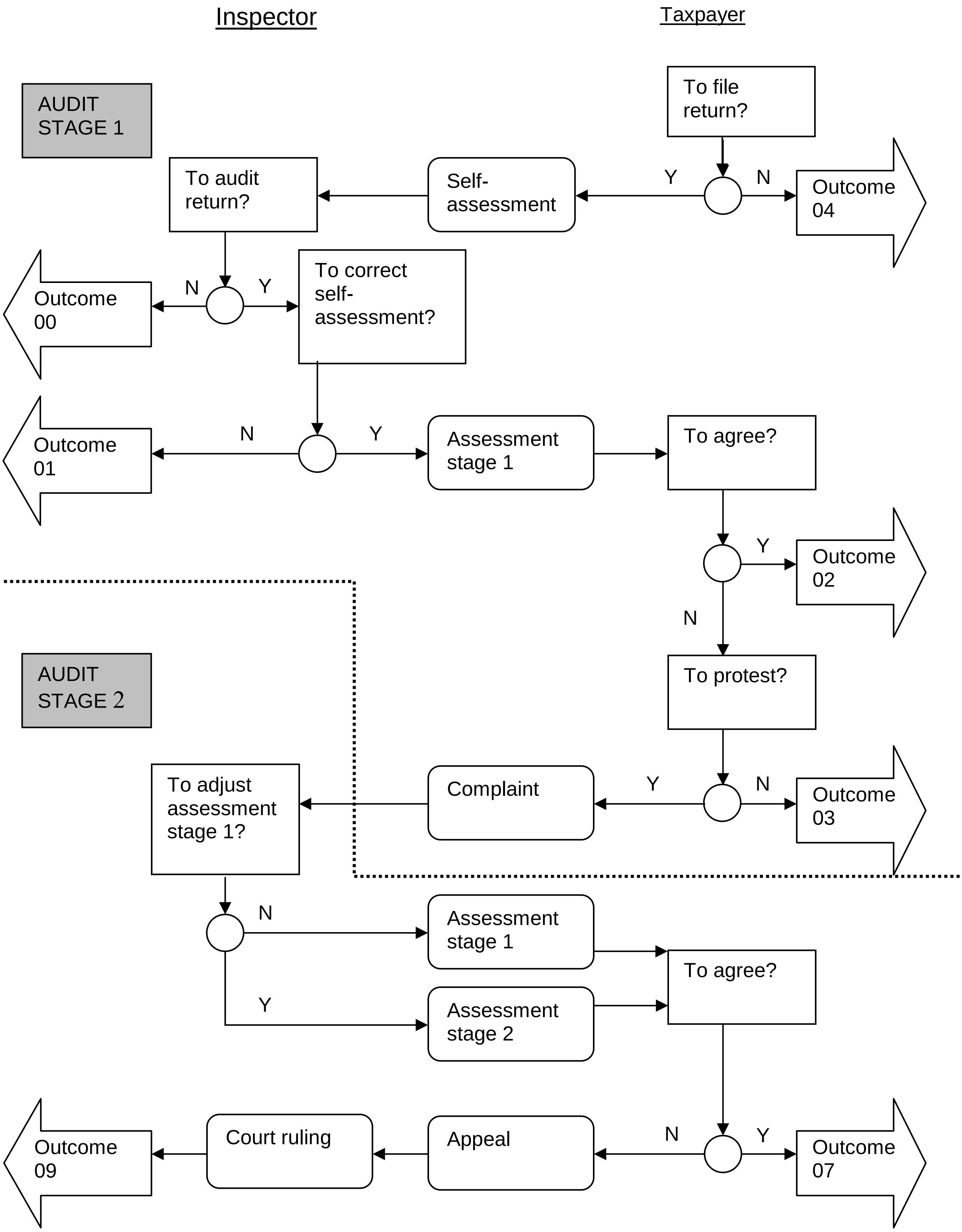
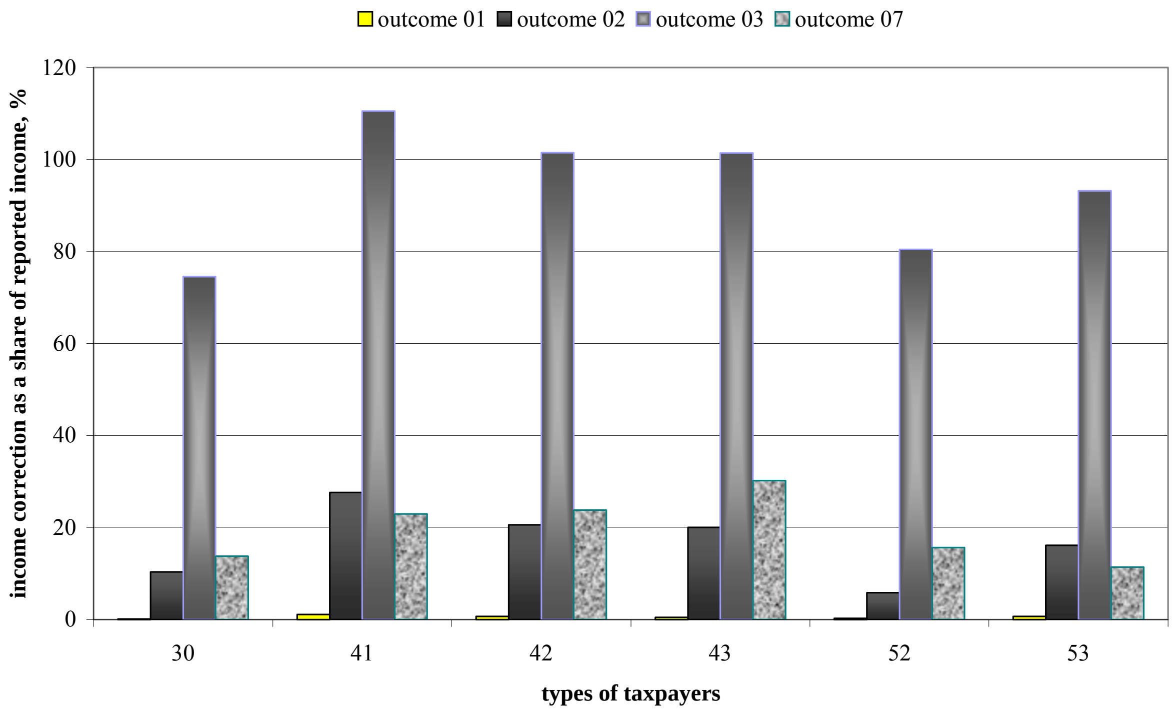


Figure 2: Income correction in different audit outcomes, by types of taxpayers



Legend of outcomes

Outcome 01: taxpayer's self-assessment accepted with minor technical corrections.

Outcome 02: inspector's assessment in agreement with taxpayer.

Outcome 03: inspector's assessment without taxpayer's agreement.

Outcome 07: inspector's assessment in the repeated audit (after complaint), in agreement with taxpayer.

Legend of taxpayers' types

30: managers-shareholders;

41: taxpayers obliged to keep the simplest one-sided form of bookkeeping;

42: taxpayers obliged to keep one-sided form of bookkeeping with accumulated entries;

43: agricultural businesses without two-sided bookkeeping;

52: taxpayers obliged to keep one-sided form of bookkeeping with cash entries;

53: taxpayers obliged to keep two-sided form of bookkeeping and taxpayers who volunteered for this form of bookkeeping.