International Tax and Public Finance, 11, 575–600, 2004

Tax Administration Reform and Taxpayer Compliance in India

ARINDAM DAS-GUPTA oldmonk87@yahoo.com
Goa Institute of Management, Ribander, Goa, India

SHANTO GHOSH SGhosh@lecg.com
Law and Economics Consulting Group LLC, 2000 Powell St, Suite 600, Emeryville, CA 94608

DILIP MOOKHERJEE dilipm@bu.edu
Boston University, Department of Economics, 270 Bay State Rd, Boston, MA 02215

Abstract

This paper evaluates effects on tax compliance of simple reforms in personnel policy in the Indian income tax administration. Taxpayers voluntarily disclosing higher incomes are currently assigned to special assessment units. To avoid this, high income taxpayers have an added incentive to understate their incomes. Empirical evidence consistent with this hypothesis is found. It explains spillover effects of enforcement efforts across assessment units. We incorporate these spillovers in estimating revenue effects of increased support staff. The results imply significant compliance gains would accrue from expanded staff employment and changes in assignment procedures for staff and taxpayers.

Keywords: tax compliance, law enforcement, tax administration, tax evasion

JEL Code: H26, K42, O17

1. Introduction

Low tax compliance is a matter of serious concern in many developing countries, limiting the capacity of their governments to raise revenues for developmental purposes.[1] It is commonly acknowledged that many factors contribute to this weakness: corruption, a large informal sector, weak legal systems, ambiguity in tax laws, high marginal tax rates, paucity of adequate information and accounting systems, a culture of noncompliance, and ineffective tax administration. While it is common to think of corruption as the biggest stumbling block, the paucity of relevant data on corruption and difficulty of enacting strict anticorruption reforms limit the scope for studying and designing such reforms.[2]

In contrast, reforming certain aspects of tax administration—such as increased employment of audit officers and their support staff, or altering procedures for grouping taxpayers into assessment units—represent an alternative approach that is relatively costless, unlikely to meet with resistance either from taxpayers or from employees of the tax administration. This paper focuses on the compliance effect of such reforms in the system for income tax administration in India. We develop a model of taxpayer compliance decisions within the Indian setting and subsequently apply it to data from a sample of Indian assessment units comprising self-employed taxpayers. We find empirical patterns consistent with the theoretical predictions of the model, which suggest that simple reforms in personnel policy and organization of taxpayer units would have significant effects on tax compliance. An added advantage of our approach is that it utilizes only data aggregated at the level of assessment units, instead of expensive surveys of compliance of individual taxpayers rarely available to tax administrators in developing countries.

The model is motivated by the specific institutional practices of the Indian tax administration (spelt out in more detail in Das-Gupta and Mookherjee, 1998, 2000). The most important of these is the virtual absence of functional specialization, a phenomenon shared by many other developing country administrations. Tax audits rely very little on centralized information or automated checking of returns. The same tax officer is responsible for gathering information, receiving and storing returns, carrying out audits manually, initiating penalties and prosecution on delinquent taxpayers. So taxpayers deal principally with a single official (denoted assessing officer (AO)): the one who heads the local tax collection center (hereafter referred to as *assessment unit*) to which they are assigned. Relative frequencies of audit for different taxpayer categories are therefore determined *implicitly* by myriad institutional details of the tax administration: the number of AOs employed, the support staff assigned to them, and the way taxpayers are assigned to different assessment units. For instance, raising audit frequencies across the board will require hiring more AOs and creating more assessment units so as to reduce the number of taxpayers assigned to any given officer. Raising enforcement standards for taxpayers at certain locations will require allocating more audit personnel to the corresponding jurisdictions. Given the low levels of automation of processing filed returns, the personnel policy of the tax administration *is* effectively its audit strategy.

Economic theory suggests, therefore, that these personnel policies generate significant impact on taxpayer compliance incentives. These links are however difficult to appreciate and decipher by most tax administrations. Even in the academic literature, as Andreoni, Erard and Feinstein (1998) point out in their recent survey, the connection between compliance and enforcement has been studied only for a handful of countries outside the U.S. The paucity is particularly acute for developing countries.

Section 2 of the paper explains in more detail the institutional setup of the Indian tax administration, as well as the nature of data available. Section 3 presents the conventional Allingham-Sandmo formulation of the response of individual taxpayers to enforcement efforts, based on the assumption that taxpayers and tax administration personnel are exogenously allocated to different assessment units. We present estimates of corresponding regressions of aggregate tax returns with respect to personnel assigned and enforcement efforts of AOs in question. These regressions imply a negative marginal revenue productivity of support staff in wards (assessment units earmarked for low income taxpayers), suggesting possible misspecification of the model. We investigate the possibility that misspecification may have resulted from endogenous personnel allocations made by higher officials of the tax administration in the interests of maximizing revenues. No evidence of such a strategic assignment policy is found, consistent with what is known about stated personnel policies. We proceed thereafter on the assumption of exogenous personnel assignments.

Section 4 investigates a different possible source of misspecification, arising from the practice of grouping taxpayers into different assessment units based on the incomes they voluntarily disclose. This necessitates a modification of the Allingham-Sandmo model, since the workloads of different assessment units are then endogenously determined by taxpayer compliance decisions. The assignment practice is motivated by a philosophy of targeting more audit resources towards ‘big fish’ tax-payers. In practice the latter were identified by whether they disclose an income above a given threshold (annual income of Rs. 200,000 for the years in question), in which case they were assigned to *circles* rather than *wards*. Circles were typically staffed by more experienced AOs and had lower workloads, resulting in higher frequency and intensity of audits.

This system generates a perverse incentive for high income taxpayers to disclose an income below the Rs. 200,000 threshold, simply in order to avoid assignment to a circle. Taxpayers effectively self-select into different assessment units, and their compliance decisions become interdependent. Our theoretical model incorporates this self-selection, and its implications for observed revenue patterns. Changes in enforcement standards in any unit (e.g., ward) will generate spillover effects on workloads and compliance in other units (e.g., the corresponding circle) that taxpayers can switch to.

We find evidence that these spillover effects were significant. The negative productivity of support staff in wards is explained by positive spillover effects on the corresponding circles: increased audit intensities in wards reduced the extent of strategic self-selection of high income taxpayers into them. Hence the productivity estimates (with respect to per filer revenue) based on the regressions in Section 3 are downward biased for wards, and upward biased for circles. Despite correcting for this bias, both AOs and support staff turn out to have higher productivity in circles relative to wards. The results suggest high returns to employing more audit personnel, and to reallocating personnel from wards to circles. The significance of the taxpayer self-selection incentive also suggests significant compliance gains would arise from abolishing the ward-circle distinction, and replacing it with random assignment of taxpayers to different assessment units.

Finally, Section 5 summarizes the principal results, and discusses policy implications in more detail.

2. Organization of the Indian Tax Adminstration

Taxpayers are classified into different geographic and occupation specific jurisdictions, called *ranges*. Each range in turn consists of between five and fifteen assessment units, which are either *wards* or *circles*: see Figure 1.[3] Each ward corresponds to a specific sub-jurisdiction, and includes all taxpayers from that sub-jurisdiction who declare an annual income below Rs. 200,000 on their tax returns. Those declaring above this level are assigned to a corresponding circle. There are usually fewer circles than wards, since circles frequently club together taxpayers of different sub-jurisdictions declaring above Rs. 200,000. Despite this, circles usually end up with a workload that is substantially smaller than the corresponding wards, as is indicated by the data from our field survey, described in Tables 1 and 2. On

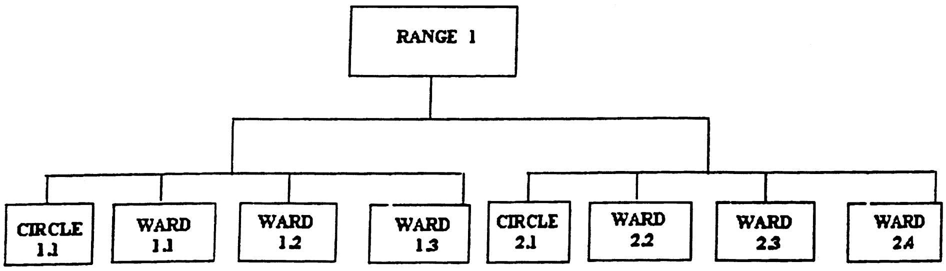

Figure 1. Organization of an income tax range.

average each ward in our sample had almost 5,000 taxpayers, twice as many as the average circle.

Each assessment unit is headed by a single assessment officer (AO), aided by a number of support staff, comprising inspectors, clerks, stenographers, tax assistants and notice servers. The system lacks functional specialization altogether, with very substantial discretion awarded to the AO, with minimal levels of hierarchical supervision.[4] For instance the AO is simultaneously responsible for (i) collecting and storing filed returns; (ii) pursuing delinquents and taxpayers in arrears; (iii) collecting information concerning potential taxpayers in the relevant jurisdiction via surveys and enquiries; (iv) carrying out a *summary assessment* of every filed return, which involves a check for arithmetical mistakes or *prima facie* errors; (v) selecting 90–120 of the filed returns for *scrutiny assessment* every year, which involves a detailed audit of the return, where the taxpayer is asked to appear before the AO and furnish supporting evidence; (vi) deciding on levels of penalties of various sorts that will be imposed on the taxpayer, or whether prosecution should be initiated, following discovery of illegitimate inaccuracies, concealment of income, or violations of various

Table 1. Assessment activity.

	Workload	Summary asst. % of workload	Net additional demand as percent of prepaid taxes resulting from summary assessments	Scrutiny assessments		Net additional demand as percent of prepaid taxes resulting from scrutiny assessments
				Number	% of workload	
Range A	4227	76	21	110	2.3	120
Range B	3559	73	3	93	2.6	132
Range C	5754	62	0	91	1.6	145
Range D	3486	64	15	106	3	71
Range E	3446	83	−5	121	3.5	95
All ranges	4149	70	7	105	2.5	158
Wards	5011	73	3	102	2	77
Circles	2558	79	0.1	124	4.8	56
Inv circles.	269	24	169	99	37	278

Table 2. Follow up penalty and prosecution activity.

	% scrutiny cases resulting in extra demand	% of extra demand cases in which income concealment penalty was initiated	% cases with concealment penalty initiated that were actually imposed	% cases with concealment penalties imposed that were appealed by taxpayer
Range A	66	25	35	95
Range B	65	26	53	20
Range C	82	19	65	75
Range D	65	18	29	67
Range E	86	25	24	92
All ranges	74	22	31	75
Wards	76	16	39	76
Circles	56	19	12	88
Inv circles.	86	53	46	85

kinds (e.g., delays in filing, or failure to deduct income at source from employees, or failure to forward withheld taxes to the tax administration); (vii) contesting subsequent taxpayer appeals at various appellate tribunals, and pursuing prosecution cases through appellate bodies and criminal courts. The nature and outcome of these follow up activities in our sample can be gleaned from Table 2.

Since the workloads (defined by the number of filed returns that need to be assessed) end up substantially lower in circles, and each AO carries out roughly the same number of scrutiny assessments, the likelihood of being audited is much higher in a circle. Moreover, circle AOs tend to have more time available to carry out audits, given that other tasks such as summary assessments take less time owing to the lower workload: see Table 3. Given this, a high income taxpayer has a potential incentive to disclose an income of less than Rs. 200,000 simply in order to avoid filing in the circle. As some taxpayers adopt such a strategy, the benefit for others to follow is increased, as wards become more congested and relative enforcement standards become even more skewed. In effect, audit frequencies and quality end up higher for those that file above Rs. 200,000, while those filing below escape detailed investigation.

Table 3. Time allocation of assessing officer.

	Wards (% time)	Circles (% time)
Summary assessments	35	17
Scruting assessments	42	58
Examination of accounts	19	26
Other related tasks	23	32
Other duties	23	26

A notable feature of the tax administration is that increasing revenues or tax compliance did not seem to be a significant priority for most senior officials (Das-Gupta and Mookherjee, 1998, Chapter 6). Performance evaluations or targets for assessing officers concentrated mainly on quantitative workload disposal rates on a quarterly basis. Some revenue targets were also laid down, but AOs interviewed expressed the opinion that these had little bite since they were easily achieved. AOs and their immediate supervisors within the range had significant discretion over audit selection and subsequent follow up penalty and prosecution activity, with minimal supervision from higher level officials or external watchdog agencies. Mandated rewards for uncovering tax evasion were rarely paid to eligible officers, undermining their credibility. Finally, interviews with senior officials revealed lack of concern regarding implications of their administrative policies on taxpayer compliance.[5]

In light of these institutional practices, it seems legitimate to pursue a modeling strategy based on exogenous personnel assignment across assessment units. It is not surprising, therefore, to find (as we do) no evidence in favor of endogenous assignment of audit personnel, and significant revenue potential from simple changes in personnel assignment policies. Our theoretical model will therefore be based on the assumption of sequential rather than simultaneous moves between the tax administration and taxpayers, i.e., where the former selects a certain audit policy first (on an ad hoc basis), to which taxpayers respond strategically.[6]

The data available concerns forty nine tax units within five ranges, in three different major cities in India located respectively in the south, west and central parts of the country. These ranges deal exclusively with small businesses and professionals, for whom tax evasion tends to be more pronounced than for other occupations. Investigation circles were dropped from the database for the regressions owing to their exclusive nature. A few other units also had to be dropped owing to missing data entries. The data was collected directly from the official records of these tax units under the auspices of a government tax committee study, and are thus unlikely to be subjected to much reporting bias. Data was collected for two successive assessment years, 1989–90 and 1990–91, for workloads, revenues from various sources, assessment and follow up activities, besides various characteristics of the assessment unit and the assessing officer. Despite repeated efforts, we have not been successful in enlarging on the size of the sample.

Our empirical analysis focuses mainly on the determinants of prepaid taxes. This partly owes to our interest in voluntary compliance. It also turns out that prepaid taxes and total collections are very closely correlated: see Figures 2 and 3 which plots these variables across different observations in our sample.

Despite the fact that revenues fluctuate by a factor of six to eight, these are tracked closely by corresponding figures for prepaid taxes. Figure 3 plots per filer revenues and prepaid taxes, and displays the strong correspondence between the two, both in levels and fluctuations. Controlling for the range in question, besides the nature of the assessment unit (ward or circle, professional rather than business charges), variations in prepaid taxes accounted for over 80% of the variation in total revenues collected. It is not surprising therefore, that determinants of prepaid taxes are also significant determinants of total revenues, a feature which recurs consistently throughout our empirical analysis. Taxpayer compliance incentives are therefore key to revenue effectiveness of tax administration policy.

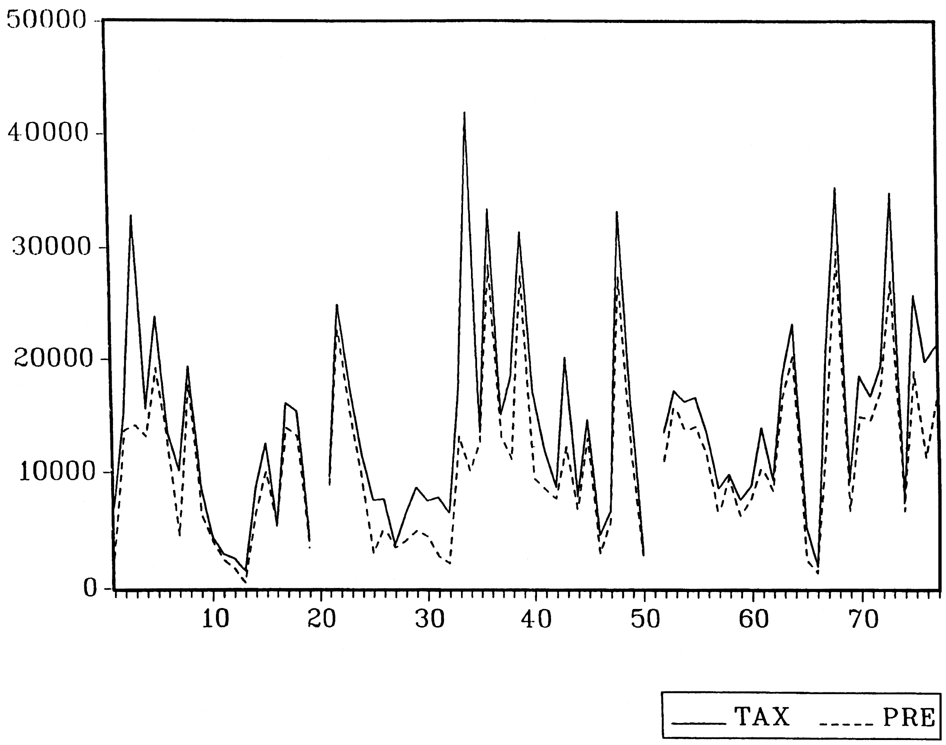

Figure 2. Total revenues and prepaid taxes.

3. Compliance Incentives with Exogenous Workloads

3.1. Compliance Incentives: The Allingham-Sandmo Equations

Let i denote the assessment unit, and t the year. In this section we assume that tax-payers are exogenously assigned to assessment units. Taxpayers differ with respect to their pretax taxable income y; each taxpayer privately knows his own income, and is assigned to a given assessment unit. The distribution of income for the set of taxpayers assigned t unit i is denoted by $F_{it}(y)$. Taxpayers are identical in all other respects; specifically they share a common constant relative risk aversion utility function defined over their aftertax income c: $u(c) = \frac{1}{\alpha}c^{\alpha}$, where $\alpha < 1, \neq 0$, with the case $\alpha = 0$ corresponding to log c. The tax law prescribes a constant tax rate τ lying between 0 and 1, and a constant penalty rate f on tax evasion established in an audit (and upheld in case the taxpayer appeals). Alternatively, if the taxpayer pays a bribe to the AO in order to avoid paying the legal penalty, f can be interpreted as the bribe rate.[7] From the taxpayer's standpoint, any payments that have to be made in the event of discovery of tax evasion by the auditor, bribes or fines, deter tax evasion, so his optimal disclosure is qualitatively similar with or without corruption.

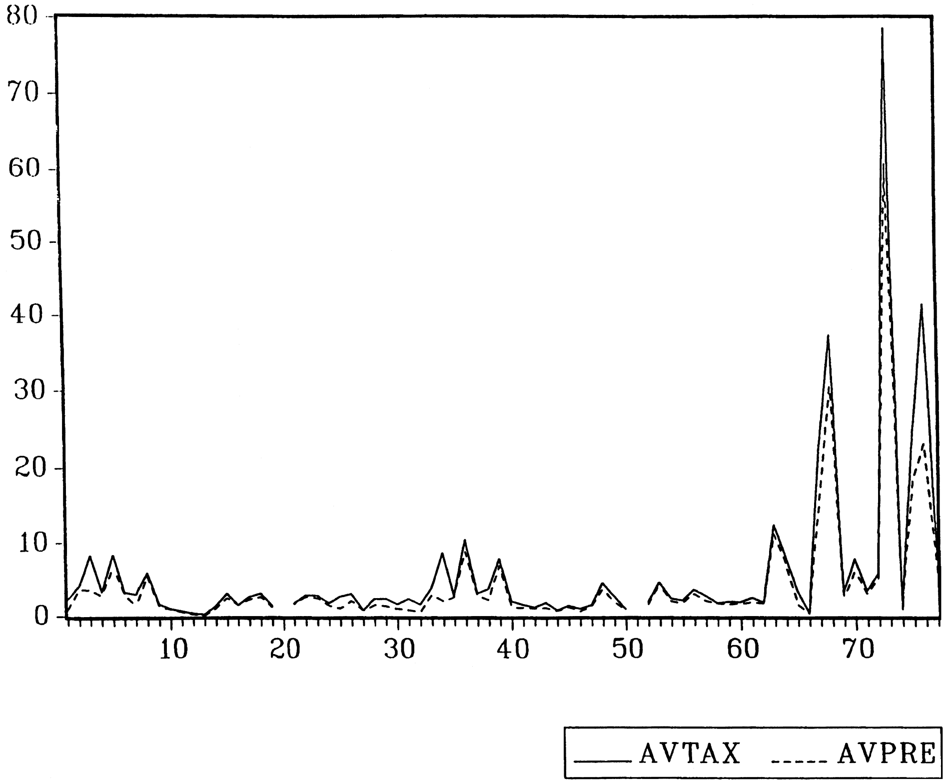

Figure 3. Per-filer revenues and per-filer prepaid taxes.

Following an audit, the AO will discover all income underreported with probability k^1_{it}; otherwise no evasion will be discovered. If evasion is discovered, the penalty at rate f will be levied with probability k^2_{it}; otherwise the taxpayer will go scot-free. Let the audit probability be denoted π_{it}.[8]

Each taxpayer knows the enforcement variables $k^1_{it}, k^2_{it}, \pi_{it}$ that characterizes this assessment unit. This assumption is reasonable in light of the fact that most businesses and professionals in India rely on tax accountants to prepare and file their tax returns. These accountants are well networked with the tax authorities and their main function is to provide detailed information concerning the relevant enforcement characteristics of the assessment unit in which the taxpayer will file. In the manner of Allingham and Sandmo (1972), then, each taxpayer confronts an audit lottery. He selects a level of income to disclose y^d not exceeding his true income y, in order to maximize expected utility

$$W(y^d, y, p_{it}) = p_{it}u((1-\tau)y - f\tau(y - y^d)) + [1 - p_{it}]u(y - \tau y^d) \quad (1)$$

where $p_{it} \equiv \pi_{it}k^1_{it}k^2_{it}$ denotes the effective probability of detection and punishment.

Given constant relative risk aversion, it is easily checked that every taxpayer decides to disclose a constant fraction r of his true income, which depends on tax and enforcement parameters p_{it}, τ, f, Hence total prepaid taxes in unit i in year t equals

$$R_{it} = r(p_{it}, \tau, f)Y_{it} \tag{2}$$

where Y_{it} denotes aggregate pretax taxable income of the taxpayer population for unit i in year t. It is well-known that r is increasing in enforcement variables p_{it} and f.

Of key interest are the determinants of p_{it}. This depends on (a) the fraction of returns that will be subjected to summary and scrutiny assessment, (b) on the quality of these assessments, which depends on the workload W_{it}, the level of support staff, as well as characteristics of the AO, including (c) penalty and prosecution activities of the assessment unit. Consequently, we specify the following equation for prepaid taxes:

$$R_{it} = K_i \cdot L_t \cdot \frac{SCR_{it}^{\alpha_0}}{W_{it}} \frac{SUM_{it}^{\alpha_1}}{W_{it}} \frac{S_{it}^{\alpha_2}}{W_{it}} PP_{it}^{\alpha_3} W_{it}^{-\alpha_4} \cdot Y_{it} \tag{3}$$

where *SCR* and *SUM* denote the number of scrutiny and summary assessments carried out, *S* is the support staff available, and *PP* is a vector of intensity of follow-up penalty and prosecution activities of different kinds. Unit-specific and year-specific effects are represented by K_i and L_t respectively. We expect that all the elasticities α_k, $k = 0, \ldots, 4$ are positive: i.e., that enforcement effectiveness is positively related to the fraction of returns assessed, the support staff available, and the effectiveness of penalty and prosecution activities, and is negatively related to the workload.

Use *PSCR*, *PSUM* and *PSS* to denote the ratios of *SCR*, *SUM* and *S* to the workload. Then equation (3) corresponds to the regression equation

$$R_{it} = G_i \cdot H_t \cdot PSCR_{it}^{\alpha_0} PSUM_{it}^{\alpha_1} PSS_{it}^{\alpha_2} W_{it}^{1-\alpha_4} PP_{it}^{\alpha_3} \cdot \eta_{it} \tag{4}$$

where G_i denotes a vector of dummies for the range to which the unit belongs, and the nature of the assessee population (e.g. whether it pertains to businesses or professionals), H_t is a year dummy representing effects of macroeconomic shifts in income and tax policy, and η_{it} is a disturbance term picking up the effects of location specific shocks in per capita income, and unmeasured characteristics of the AO, the assessment unit and the assessee population.

The first two columns of Table 4 presents the results of the regression equation (4), in both levels and first differences. In the regression we estimate elasticities with respect to PSCR, PSUM, PSS and W separately for wards and circles, by interacting these variables with respect to a circle dummy. The regression includes the following measures of penalty and prosecution effectiveness: PPNL, the fraction of summary assessment cases that involved penalties; PIMP, the fraction of penalty cases in which penalties for concealment of income were actually imposed; and APPLSUC, a measure of success with respect to contesting taxpayer appeals in past penalty cases.[9] It also includes a dummy for business (rather than professional) assessees, for circles and for the year. AO experience and various other measures of follow up penalty and prosecution activities had insignificant effects and were thus not included in the regression.

The most surprising feature of the first two columns of Table 4 is that the marginal revenue productivity of support staff in wards is negative, and is significant at 10% in the level regression. This suggests that the regression is mis-specified in some way.

Table 4. Compliance regressions.

Dependent variable: Prepaid tax revenues				
	Structural		Reduced form	
	Coeff.	First difference coeff.	Coeff.	First difference coeff.
Wards				
Workload	0.38 (0.48)	1.29	0.48 (0.3) ?	0.98*
Scrutiny probability	0.37 (0.38)	0.33		
Summary probability	1.57 (0.55)***	2.62***		
Support staff	−0.81 (0.47)*	−0.47	−0.6 (0.35) ?	−1.17
Circles				
Workload	1.58 (0.82)*	3.23*	−0.14 (0.24)	−0.08
Scrutiny probability	1.19 (0.52)**	2.18**		
Summary probability	−0.39 (0.37)	−1.31 ?		
Support staff	0.29 (0.76)	0.28	0.51 (0.72)	0.90
Per-filer penalty	0.19 (0.09)**	0.24 ?	0.17 (0.1)*	0.32 ?
Penalties imposed	0.35 (0.12)***	0.19	0.38 (0.14)***	0.45**
Appeal success rate	0.28 (0.16)*	−0.56	0.27 (0.18) ?	−0.24
Business	−0.28 (0.26)		−0.23 (0.26)	
Circle	−0.23 (4.29)		3.33 (3.44)	
Year	0.11 (0.18)		0.32 (0.18)*	
Range 1	0.70 (0.28)**		0.58 (0.25)**	
Range 2	0.69 (0.30)**		0.46 (0.27)*	
Range 14	0.51 (0.32)*		0.21 (0.29)	
Range 15	−0.02 (0.33)		−0.33 (0.29)	
N	62	27	62	27
Rsq	0.67	0.75	0.57	0.36
Adj Rsq	0.53	0.58	0.44	0.14
Root MSE	0.55		0.6	

Support staff is measured as total head-count in reduced form, and as per-filer staff in structural equation.
Numbers in parentheses are White corrected standard errors.
***: Significant at 1%, **: Significant at 5%, *: Significant at 10%, ?: Significant at 20%.

One possibility is that the regression controls for the frequency of scrutiny and summary assessments. It is possible that the main task of support staff was to free up AO time by taking care of more mundane duties, enabling AOs to carry out more assessments. To capture this possibility we extend the model to allow the number of assessments of either kind to depend on support staff, controlling for workload:

$$SCR_{it} = A_i \cdot S_{it}^{\delta_1} \cdot W_{it}^{\delta_2} \cdot c_{it} \tag{5}$$

$$SUM_{it} = B_i \cdot S_{it}^{\psi_1} \cdot W_{it}^{\psi_2} \cdot d_{it} \tag{6}$$

where c_{it} and d_{it} are unit mean stochastic disturbances. These imply the following regression for the relevant assessment ratios:

$$PSCR_{it} = A_i \cdot S_{it}^{\delta_1} \cdot W_{it}^{\delta_2 - 1} \cdot c_{it} \tag{7}$$

$$PSUM_{it} = B_i \cdot S_{it}^{\psi_1} \cdot W_{it}^{\psi_2 - 1} \cdot d_{it} \tag{8}$$

Since penalty and prosecution activities proxy for unit-specific characteristics that may be correlated with assessment disposal—more efficient or conscientious officers that do better in one dimension also do better in the other—we include these variables in the estimated version of these regressions, reported in Tables 5 and 6.

Higher support staff did allow more assessments to be carried out. However this effect was significant only for assessments that were ineffective in promoting compliance: scrutiny assessments in wards and summary assessments in circles. Hence it is unlikely that the

Table 5. Scrutiny assessment.

Dependent variable: Scrutiny assessment probablity	
	Coeff.
WRD	
Workload	−0.55 (0.14)***
Support staff	0.66 (0.25)**
CRC	
Workload	−1.13 (0.14)***
Support Staff	0.05 (0.26)
Per-filer penalty	0.07 (0.04)*
Penalties imposed	0.08 (0.05)*
Appeal success rate	−0.04 (0.08)
Business	0.00 (0.09)
Circle	6.16 (1.66)***
Year	0.13 (0.07)*
Range 1	−0.35 (0.14)**
Range 2	−0.36 (0.15)**
Range 14	−0.16 (0.1) ?
Range 15	−0.48 (0.15)***
N	64
Rsq	0.92
Adj Rsq	0.9
Root MSE	0.25

Number in parentheses are White corrected standard errors. ***: Significant at 1%, **: Significant at 5%, *: Significant at 10%, ?: Significant at 20%.

Table 6. Summary assessment.

Dependent variable: Summary assessment probablity	
	Coeff.
WRD	
Workload	−0.05 (0.14)
Support staff	−0.14 (0.25)
CRC	
Workload	1.11 (0.08) ?
Support staff	0.16 (0.18)
Per-filer penalty	−0.06 (0.03) ?
Penalties imposed	0.00 (0.04)
Appeal success rate	0.05 (0.04)
Business	0.04 (0.07)
Circle	−1.57 (1.24)
Year	−0.04 (0.07)
Range 1	−0.09 (0.07)
Range 2	−0.18 (0.11)*
Range 14	−0.23 (0.08)***
Range 15	−0.36 (0.13)***
N	64
Rsq	0.53
Adj Rsq	0.4
Root MSE	0.19

Number in parentheses are White corrected standard errors.
***: Significant at 1%, **: Significant at 5%, *: Significant at 10%, ?: Significant at 20%.

puzzle of negative productivity of ward support staff can be explained in this fashion. This conclusion is reinforced by the reduced form compliance regression implied by (7) and (8), reported in the last two columns of Table 4. Here ward support staff continues to have a significant negative productivity.

One other possible source of misspecification is the assumption that support staff were exogenously assigned to different units. Suppose instead that they were allocated by higher officials in order to maximize total revenues of the range, based on information available to these officials at the beginning of the year. Let the predicted workload of unit i at year t be ω_{it}, and the prediction error be denoted v_{it}, so the actual workload was

$$W_{it} = \omega_{it} v_{it}. \tag{9}$$

Then the prediction errors will be serially uncorrelated, and uncorrelated with local income shocks, shocks in the scrutiny or summary assessment ratios for unit i. The *ex ante* expected revenue of unit i in year t, conditional on the information available to the administration, and on staff allocation of S_{it}, then equals

$$ER_{it} = g_i H_t S_{it}^a \omega_{it}^b PP_{it}^{\alpha_3} u_{it} \qquad (10)$$

where u_{it} denotes the compliance-relevant characteristics known to the administration but unobserved by us. Let Z_{it} denote $g_i H_t PP_{it}^{\alpha_3} u_{it}$, the complete set of observed and unobserved characteristics, excluding staff and workload, so the expected revenue can be written as $Z_{it} S_{it}^a \omega_{it}^b$.

Suppose the aggregate staff available to the tax administration in year t was M_t. The optimal staff allocation maximizes $\Sigma_i ER_{it}$ subject to $\Sigma_i S_{it} = M_t$. The solution to this is

$$S_{it}^* = M_t \cdot \left[\omega_{it}^b Z_{it}\right]^{\frac{1}{1-a}} \qquad (11)$$

Hence, as long as a the elasticity of reduced form revenue with respect to staff is less than one, the optimal staff allocation to unit i correlates positively with revenue enhancing characteristics of the unit, as well as with predicted workloads (assuming b is positive). Finally suppose that actual staff allocations were determined according to

$$S_{it} = S_{it}^{*\beta} s_{it} \qquad (12)$$

where $\beta \in (0, 1)$ is the responsiveness of actual allocations to revenue-optimality considerations, with s_{it} a random residual.

Under this scenario, $\beta > 0$ implies staff allocations are at least partly guided by revenue considerations. Then they end up being correlated with the disturbance term in the revenue equation, implying that the estimated support staff elasticity in Table 4 is biased. But given a is less than one, i.e., there are diminishing marginal returns to additional staff, *the estimated staff elasticity is biased upwards*. The reason is that support staff proxies for missing variables that enhance enforcement.[10] So strategic staff assignments cannot explain the finding of negative productivity in wards either.

One can also directly check for endogeneity of staff assignments, by estimating the regression predicting support staff levels corresponding to equations (11) and (12). The problem here is that the *ex ante* workload predictions ω_{it} are unobservable. The best proxy for this is the actual workload but then the regression is subject to biases resulting from measurement error:

$$S_{it} = M_t^\beta W_{it}^{\frac{b\beta}{1-a}} Z_{it}^{\frac{\beta}{1-a}} \epsilon_{it} \qquad (13)$$

where the disturbance term $\epsilon_{it} = s_{it} v_{it}^{-\frac{b\beta}{1-a}}$ is correlated with the workload, as long as $\beta > 0$. One way of overcoming this bias is to use the workload for a different year as an instrument for workload in year t. This is a valid instrument since workload prediction errors are serially uncorrelated. Table 7 presents both OLS and instrumental variable estimates of regression (13). None of the coefficients of the instrumental variable regression turn out to be statistically significant even at the 20% level.[11] This motivates our interest in exploring alternative sources of misspecification.

Table 7. Staff allocation.

Dependent variable: Support staff		
	OLS estimates (Coeff)	IV estimates (Coeff)
Workload (Wards)	0.16 (0.13)	−0.19 (1.43)
Workload (Circles)	−0.05 (0.05)	−0.2 (1.45)
Per-Filer penalty	0.01 (0.03)	−0.03 (0.06)
Penalties imposed	0.05 (0.03)$^?$	0.03 (0.04)
Appeal success rate	0.03 (0.08)	0.04 (0.09)
Business	−0.02 (0.07)	−0.02 (0.12)
Circle	1.79 (1.28)$^?$	0.16 (12.23)
Year	−0.08 (0.07)$^?$	−0.00 (0.31)
Range 1	0.09 (0.12)	0.13 (0.4)
Range 2	−0.28 (0.12)***	−0.28 (0.27)
Range 14	−0.07 (0.08)	−0.02 (0.1)
Range 15	−0.22 (0.1)**	−0.06 (0.7)
N	64	64
Rsq	0.4	
Adj Rsq	0.26	
Root MSE	0.2	

Number in parentheses are White corrected standard errors. ***: significant at 1%, **: significant at 5%, *: significant at 10%, $^?$: significant at 20%.

4. Filer Self Selection and Endogenous Workloads

The Allingham-Sandmo model is based on the assumption that the audit probability is independent of the income disclosed. This assumption is not generally valid, and is especially invalid in the Indian setting, owing to the way that taxpayers are actually assigned to wards and circles. We have noted above that wards have substantially higher workloads than circles, and conduct a similar or smaller number of scrutinies on average. Hence the scrutiny probability tends to be substantially lower in wards. Since the basis for deciding whether a taxpayer should be assigned to a ward rather than a circle is whether the declared income falls below Rs. 200,000, this provides the taxpayer with an added incentive to declare an income below Rs. 200,000, simply in order to take advantage of the lower probability that the return will be scrutinized.

Once this is recognized, the analysis of the previous Section needs to be modified. Workloads in wards and circles are no longer exogenous, and represent a possible source of downward bias of the estimated productivity of ward support staff. Increased staff might enhance enforcement in wards, causing high income taxpayers to stop filing in the ward and switch to the corresponding circle. The effect on measured revenues of the ward would tend to be negative.

4.1. *Self Selection Model*

The possibility of self-selection complicates the Allingham-Sandmo model of individual disclosure decisions for a number of different reasons. First, the probability of detection now depends on the amount of income disclosed: if less than Rs. 200,000 is disclosed, the return will be filed in a ward, where the scrutiny rates tend to be lower. Indeed, the scrutiny probability falls discontinuously just as the amount disclosed crosses the Rs. 200,000 threshold. Second, the disparity between the ward and circle scrutiny probabilities depends on the filing decisions of *other* taxpayers, since these affect their relative workloads. Hence each taxpayer will have to form conjectures concerning the filing decisions of other taxpayers in the same jurisdiction: one cannot identify the decision of each taxpayer in isolation.

Moreover, the fact that scrutiny rates are so much lower in wards compared to circles make it difficult to explain why anybody files in a circle at all. We believe the reason is that the probability of evasion being discovered is an increasing function of the extent of evasion, unlike the assumption of a constant probability made by the Allingham-Sandmo model. Those with incomes sufficiently in excess of Rs. 200,000 could then prefer to file in a circle, since the alternative of filing in the corresponding ward involves too great a risk of discovery.

Let a taxpayer's true income be denoted by y, and the disclosed income y^d. Fix the jurisdiction and the year, so filers can file either in a given ward or the corresponding circle. The distribution over true income in the given jurisdiction is denoted the distribution function $F(y)$, which we assume is continuous and strictly increasing. The tax rate is $\tau \in (0, 1)$. If the taxpayer discloses income y^d above Rs. 200,000, the return will be filed in the circle corresponding to his jurisdiction, for which the taxpayer conjectures a probability β of scrutiny. If he discloses below Rs. 200,000, the return will be filed in the corresponding ward, for which the scrutiny probability is conjectured to be ρ. These scrutiny probabilities will be endogeneously determined.

If the return is scrutinised, whether in a ward or circle, the AO will find evidence of concealment with probability $k_1(1 - \frac{y^d}{y})$, where k_1 is a parameter lying between 0 and 1. Hence the probability of detecting concealment increases with the amount concealed. In the event of discovery of concealed income, penalties at the constant rate f on the extent of taxes evaded will be imposed with some probability $k_2 \in (0, 1)$, which like k_1 is specific to the assessment unit in question.

Let the conjectured scrutiny probability be denoted

$$p(y^d) = \begin{cases} \rho & \text{if } y^d < 2 \\ \beta & \text{otherwise} \end{cases} \tag{14}$$

where we normalize units of income to Rs. 100,000 each. Then the taxpayer's expected utility conditional on a given scrutiny probability π is given by

$$\begin{aligned} W(y^d, y, \pi) = {} & k_1 k_2 \pi \left(1 - \frac{y^d}{y}\right) u((1-\tau)y - \tau f(y - y^d)) \\ & + \left[1 - k_1 k_2 \pi \left(1 - \frac{y^d}{y}\right)\right] u(y - \tau y^d). \end{aligned} \tag{15}$$

and optimal disclosure y^d of a taxpayer with true income y maximizes:

$$EU(y^d, y) = W(y^d, y, p(y^d)) \tag{16}$$

Let the optimal disclosure be denoted $y^d(y; \beta, \rho)$, as a function of true income and the conjectured scrutiny probabilities. This in turn determines the workloads of the corresponding ward and circle: all those declaring $y^d < 2$ will file in the ward, and the rest in the circle. An *equilibrium* arises when the resulting scrutiny probabilities are exactly as conjectured by all the taxpayers.

It is easily checked that for given scrutiny probability π, the expected utility W is strictly concave in y^d. However, the scrutiny probability is discontinuous at a disclosure of $y^d = 2$, causing the payoff functions to be nonconcave and discontinuous. The key step in the analysis is provided by the following proposition, whose proof is provided in the Appendix.

Proposition 1. *Given conjectured scrutiny probabilities ρ and β, there exists a threshold income $y^*(\beta, \rho)$ exceeding Rs.* 200,000, *with the property that a taxpayer files in the circle if and only if his true income exceeds y^*. The function $y^*(., .)$ is continuous, increasing in β and decreasing in ρ.*

All those with true incomes below Rs. 200,000 will file in the ward. Those with incomes slightly in excess of Rs. 200,000 will also file in a ward, as they would not like to declare their entire income, and an added incentive arises if $\rho < \beta$. So the only people filing in circles are those with an income sufficiently in excess of Rs. 200,000 that the likelihood of being discovered if they file in a ward is too high.

Figure 4 illustrates the argument in the case where $\beta > \rho$. As shown in Appendix A, we can normalize the payoff function per unit income of the taxpayer, and represent the disclosure problem as selecting the fraction disclosed r to maximize $V(r, \pi) \equiv \frac{1}{\alpha} \{\pi(1-r)[(1-\tau-\tau f(1-r)]^\alpha + [1-\pi](1-\tau r)\}^\alpha$. Figure 4 graphs V as a function of r. Disclosure ratios below $\frac{2}{y}$ correspond to filing in the ward, and above correspond to filing in the circle. Within any region, payoffs are concave in r, but there is a discontinuity at the switchpoint $\frac{2}{y}$. The true income of the taxpayer only affects the switchpoint between the two payoff functions. For a taxpayer with true income y_L the payoff jumps downward at the switchpoint: such a taxpayer will file in a ward and disclose an income at the upper endpoint of Rs. 200,000. The threshold y^* is the income level at which a taxpayer is indifferent between filing in the ward and the circle. Those with higher incomes (such as y_H) will prefer to file in a circle. The ultimate result of the self-selection effect is that taxpayers with true incomes between Rs. 200,000 and y^* disclose a lower fraction of their income ($\frac{2}{y}$, rather than $r(\beta)$).

Proposition 1 implies that the allocation of workload between ward and circle will depend on conjectured scrutiny probabilities. The fraction of taxpayers in the jurisdiction that file in the ward is given by $F(y^*(\beta, \rho))$. A higher scrutiny probability in the ward or a lower scrutiny probability in the circle will reduce this fraction, as fewer high income taxpayers find it advantageous to self-select into a ward.

To close the model, we have to specify how taxpayers form conjectures concerning the scrutiny probabilities. Assume (for reasons explained in the previous Section) that staff allocations across the ward and circle are exogenously given. And suppose that the number

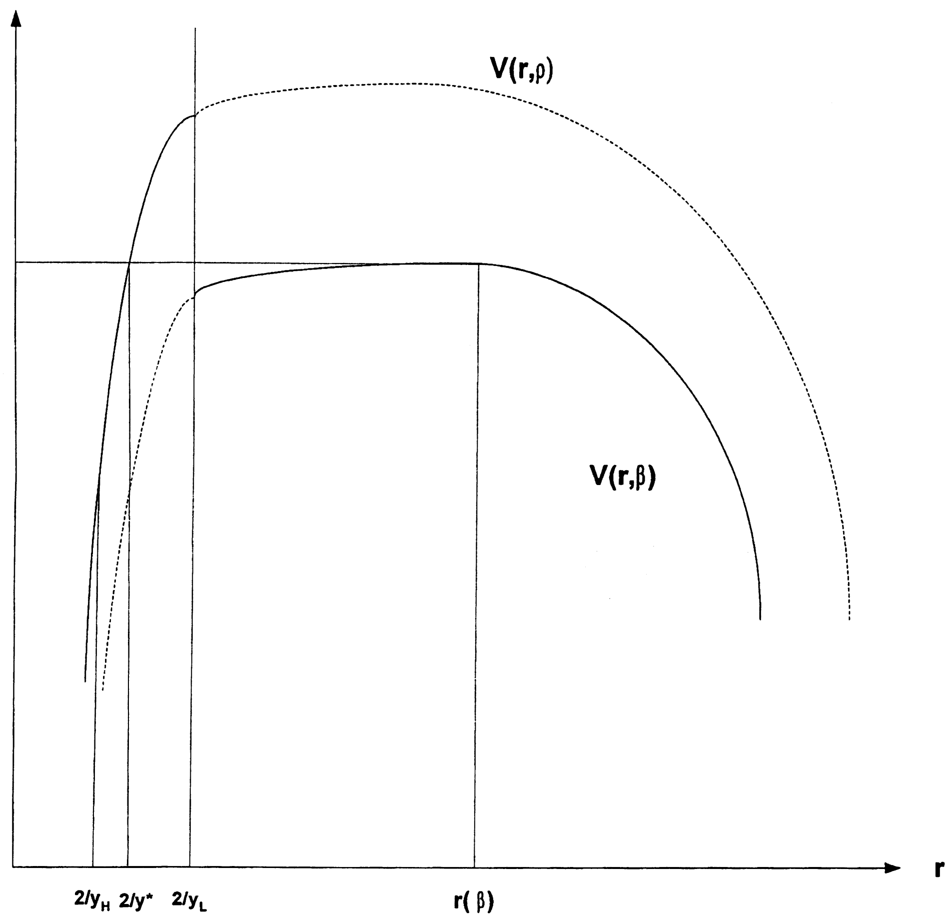

Figure 4. Payoff function when $\beta > \rho$.

of assessments V_w, V_c are decided by the AO depending on the staff available and the workload, as given by (5). Then $V_w = V_w(S_w, W_w)$ and $V_c = V_c(S_c, W_c)$, where S_w, S_c denote the staff levels and W_w, W_c the corresponding workloads. Assume that these functions are continuous in the workloads.

Let N denote the total size of the taxpayer population in this jurisdiction; this is also exogenous and common knowledge among taxpayers. Also suppose that they all conjecture that a fraction a_w of them will decide to file in the ward. Then the corresponding conjectured scrutiny probabilities are

$$\beta(a_w \mid S_c, N) = \frac{V_c(S_c, (1 - a_w)N)}{N(1 - a_w)}$$
$$\rho(a_w \mid S_w, N) = \frac{V_w(S_w, a_w N)}{a_w N} \tag{17}$$

Inserting these into the optimal filing decision, the fraction of these taxpayers that will subsequently decide to file in the ward is denoted $F(y^*(\beta(a_w \mid S_c, N), \rho(a_w \mid S_w, N); X_c, X_w))$, where X_w, X_c denote ward and circle characteristics that bear on their respective enforcement intensities (e.g., penalty effort). An equilibrium is characterized by the condition that taxpayers' conjectures concerning each other's filing decisions are correct:

$$a_w = F(y^*(\beta(a_w \mid S_c, N), \rho(a_w \mid S_w, N); X_c, X_w)). \tag{18}$$

It is evident that the right hand side of (18) is continuous in a_w, and maps the unit interval to itself. So there must exist at least one equilibrium. It is possible that there is more than one equilibrium: for instance if taxpayers conjecture a large fraction of them will file in a ward, the resulting congestion effect will lower the scrutiny probability in the ward, and raise it in the circle. This in turn will encourage a large number of them to file in the ward.

Nevertheless equilibria will typically be locally unique, with local properties described by a function of the form

$$a_w = A_w(S_w, X_w; S_c, X_c; N). \tag{19}$$

From Proposition 1 it is evident that A_w is locally increasing in X_c and locally decreasing in X_w: this is just the self-selection effect. The same is true for support staff levels, provided that they result in higher enforcement intensities.

Equilibrium workloads are then given by

$$\begin{aligned} W_w &= N A_w(S_w, X_w; S_c, X_w; N) \\ W_c &= N[1 - A_w(S_w, X_w; S_c, X_c; N)]. \end{aligned} \tag{20}$$

Increased enforcement intensity in any ward will then cause a reallocation of workloads in favor of the corresponding circle. This is the main prediction of the self-selection model, which we shall test below. Note, however, that the effect of an increase in the population size N on workloads cannot be signed, since it reduces per capita enforcement levels in both wards and circles.

Implications for compliance are as follows. Per filer prepaid taxes AR_w and AR_c are given by

$$\begin{aligned} AR_w &= \tau\{r(\rho^*, X_w)\mathcal{E}[y \mid y < 2] + 2[F(y^*(\beta^*, \rho^*; X_w, X_c)) - F(2)]\} \\ AR_c &= \tau r(\beta^*, X_c)\mathcal{E}[y \mid y > y^*(\beta^*, \rho^*; X_w, X_c)] \end{aligned} \tag{21}$$

where $r(\rho, X_w)$, $r(\beta, X_c)$ denote the optimal fraction of income disclosed while filing in a ward or circle respectively, ρ^* and β^* denote equilibrium scrutiny probabilities

$$\begin{aligned} \rho^* &= \frac{V_w(S_w, A_w \cdot N)}{A_w N} \\ \beta^* &= \frac{V_c(S_c, (1 - A_w)N)}{(1 - A_w)N} \end{aligned} \tag{22}$$

and $\mathcal{E}$ denotes conditional expectation. The reduced form expressions for per filer prepaid taxes are as follows:

$$\begin{aligned} AR_w &= AR_w(S_w, X_w, S_c, X_c, N) \\ AR_c &= AR_c(S_w, X_w, S_c, X_c, N). \end{aligned} \tag{23}$$

Finally, using the fact that total prepaid taxes equals the product of average prepaid taxes and workload, we obtain the reduced form equation for total prepaid taxes:

$$\begin{aligned} R_w &= R_w(S_w, X_w, S_c, X_c, N) \\ R_c &= R_c(S_w, X_w, S_c, X_c, N) \end{aligned} \tag{24}$$

Terms involving y^* represent the self-selection effect, while r represents the conventional Allingham-Sandmo effect. Since these typically operate in different directions, effects of increased levels of enforcement, either in the same unit, or in the corresponding unit, cannot be predicted in general. An increase in X_w causes some relatively high income filers to switch back to filing in the circle, bringing down the per filer average in the ward. Running against this is the fact that lower income individuals who continue to file in the ward, now disclose a higher proportion of their income, thus raising the ward average. The model is now consistent with negative *measured* effects of certain variables affecting the quality or quantity of enforcement, such as support staff. However one can corroborate such an explanation based on induced self-selection effects directly from the corresponding workload regression.

If enforcement in the *associated* ward or circle affects prepaid taxes in any given unit, the model rationalizes this as the effect of induced self-selection. Does the theory impose any restrictions on the nature of this dependence? For wards, an increase in enforcement X_c in the associated circle would cause y^* to increase, and filers to switch from the circle to this ward. Those who switch have higher incomes than all those previously filing in the ward: hence this will raise per filer prepaid taxes in the ward, as well as in the circle in question. However, the induced switch in workloads causes the average scrutiny probability to decline in the ward, and increase in the circle: which affects disclosure decisions of the non-switchers in the opposite direction. If this congestion effect is negligible relative to the pure self selection effect, we would expect per filer revenues to *increase* in wards following an increase in X_c. For circles this would be reversed: an increase in X_w should lower per filer compliance.

The model clarifies the biases in estimates of the overall revenue effectiveness of increasing enforcement in any given assessment unit. For instance, if the self-selection effect is significant, and increased support staff in a ward enhances its enforcement level X_w substantially, then some high income filers switch from filing in the ward to the corresponding circle. Note that this will tend to bring down the per filer revenue average in both the ward and the circle, since those who migrate are the highest income types previously filing in the ward, and the lowest income types currently filing in the circle! The revenue effects would *appear* to be negative in both the ward and the circle, whereas the actual effect on total revenues would be substantially positive (as the migratory taxpayers now disclose a larger income). To assess the effectiveness of any enforcement variable, one should therefore look at total rather than per filer revenue regressions, and examine effects not only on revenues within the same unit, but also the spillover effects on the corresponding associated ward or circle.

4.2. Empirical Estimates

The main problem with estimating these regression equations is that they were derived on the basis of the assumption that there was only one ward and one circle within any range.

However, every range actually contains more than one ward, and some of them have more than one circle. And often a given circle clubs together high-income disclosing filers from disparate jurisdictions that correspond to different wards. Indeed, some ranges have a single circle and multiple wards.

It is straightforward to extend the theory to accommodate these complexities, at the cost of increasing the set of relevant independent variables. The workload of any given assessment unit will depend on staff allocations and other enforcement-relevant characteristics in every other unit in the same range. In a ward, for instance, the staff level of other wards is also relevant, since they affect incentives of taxpayers in those jurisdictions to file in the common circle. From a practical point of view, however, these feedback effects could be negligible, and not worth including in the estimated regression equations. Indeed, aggregate indices of enforcement in other wards within the same range turned out to be statistically insignificant (at 10%) when included in the list of independent variables for the ward estimates. So it makes sense to exclude them in order to conserve on degrees of freedom.

But even if we seek to include only the enforcement variables in the same unit and the corresponding ward/circle that filers of any jurisdiction can select between, there is a problem of measuring these enforcement variables. For instance, in a range with a single circle and many wards, we cannot identify a single ward that ‘corresponds’ to the circle. The circle clubs together the taxpayers from *all* jurisdictions that disclose incomes exceeding Rs. 200,000. So there is a need to construct a suitable average of the enforcement variables of all wards associated with that single circle. Appendix B describes the procedure we followed for this.

Table 8 presents the estimates of workload and compliance regressions (19), (20), (24) and (23). These provide evidence in favor of the self-selection hypothesis. Workloads and prepaid taxes exhibit significant spillovers with respect to enforcement intensities in adjacent wards or circles. The direction of these spillover effects match the theoretical predictions in most respects.

The workload regressions appear in the first two columns of Table 8 (corresponding to equations (19) and (20) respectively). We present both regressions because each is subject to a distinct set of estimation problems. The one predicting absolute workloads (20) shown in the second column suffers the problem that data on the size of the taxpayer population in the jurisdiction (N in the model of the preceding section) is not available. This problem does not afflict regression (19) shown in the first column which predicts relative workloads of wards and circles. But this is subject to the measurement problems we just described above.

Both regressions contain evidence in favor of self-selection. It is most compelling when we consider the effect of penalty and prosecution activity, especially for circles. An increase in per filer penalty rate in a circle causes a significant reduction in the workload of the circle, and a significant increase in the workload of associated wards. Conversely a *ceteris paribus* increase in penalty rates in the ward induces a significant switch in the opposite direction. Support staff variations exhibit qualitatively similar effects on relative workloads, but these effects are not statistically significant.

The equations for prepaid taxes indicate strong spillover effects, both with respect to staff levels and penalty rates. Increased support staff in wards is associated with a revenue

Table 8. Regressions for the self selection model.

Dependent variable	Fraction filing	Workload	Avg. prepaid taxes	Total prepaid taxes
Wards				
Staff	0.01 (0.06)	0.16 (0.22)	−0.65 (0.45) ?	−0.65 (0.53)
Assoc. circle staff	2.30 (1.82)	−3.64 (2.69) ?	7.86 (3.02)**	7.25 (2.04)***
Per-filer penalty	−0.01 (0.01)	0.0 (0.05)	0.13 (0.12)	0.15 (0.11) ?
Ass. circle per-filer penalty	0.06 (0.03)**	0.06 (0.08)	0.31 (0.16)*	0.29 (0.18) ?
Penalties imposed	−0.02 (0.01) ?	0.01 (0.05)	0.33 (0.15)**	0.35 (0.14)**
Ass. circle penalties imposed	0.05 (0.06)	0.15 (0.11) ?	0.39 (0.3)	0.43 (0.29) ?
Circles				
Staff	−0.48 (0.53)	−1.18 (0.54)**	2.73 (0.72)***	1.48 (0.36)***
Assoc. ward staff	1.8 (2.5)	−2.85 (3.04)	5.9 (3.98) ?	7.38 (2.73)***
Per-filer penalty	−0.26 (0.14)*	−0.75 (0.25)***	1.33 (0.19)***	0.64 (0.11)***
Ass. ward per-filer penalty	0.77 (0.28)***	−0.07 (0.35)	0.81 (0.46)*	1.32 (0.3)***
Penalties imposed	−0.12 (0.15)	0.09 (0.19)	0.68 (0.21)***	0.18 (0.15)
Ass. ward penalties imposed	0.5 (0.28)*	0.53 (0.26)*	−0.93 (0.45)**	−0.45 (0.35)
Range 1	−0.8 (0.65)	1.77 (0.95)*	−2.65 (1.07)**	−1.95 (0.73)***
Range 2	0.32 (0.29)	−0.3 (0.47)	1.4 (0.69)**	1.63 (0.65)**
Range 14	−0.2 (0.13) ?	0.27 (0.2) ?	−0.29 (0.36)	−0.26 (0.35)
Range 15	0.45 (0.32) ?	−0.04 (0.52)	0.72 (0.57)	1.17 (0.41)***
Circle	2.58 (2.18)	−0.8 (2.53)	0.23 (4.13)	−2.77 (3.09)
Year	−0.02 (0.06)	0.16 (0.1)*	−0.11 (0.19)	0.14 (0.19)
Business	−0.01 (0.03)	−0.03 (0.08)	−0.28 (0.28)	−0.32 (0.29)
N	72	72	70	70
Rsq	0.97	0.82	0.77	0.61
Root MSE	0.17	0.3	0.62	0.58

loss in the ward concerned, and a positive and large spillover effect into the corresponding circle (statistically significant at 15%). The point estimates imply an aggregate elasticity (incorporating both own and cross effects) between 1 and 2. In other words, the spillover effects dominate the own effects, resulting in a significant positive productivity estimate.

With regard to increased support staff in circles, both own and cross revenue effects are positive and significant (both statistically and quantitatively). The aggregate productivity of support staff continues to be higher in circles than in wards, even after incorporating spillover effects. This suggests the value of reallocating staff towards circles from wards.

Finally, higher per filer penalty rates in either wards or circles generate positive own and cross effects on revenues and per filer compliance. The revenue effects of penalty imposition rates are less significant in the case of circles, but the direct own effects are positive and significant for wards.

5. Conclusion

We have found evidence consistent with the model of taxpayers strategically self-selecting into wards or circles. This arises owing to the phenomenon of assigning taxpayers to different assessment units on the basis of the incomes they voluntarily disclose. It implies that variations in support staff or enforcement effort in any given unit generates spillover effects on the workload and compliance in related units. These effects explain why the measured revenue productivity of ward support staff on the ward's own revenues were negative. Once the spillover effects are incorporated the estimated revenue productivity of ward staff exceeds 1, suggesting the scope for expanding the scale of staff employed in these units. The estimated productivity of staff in the circles is even higher.

What are the principal policy implications? One general implication is that taxpayer incentives for voluntary compliance matter significantly, and are substantially affected by enforcement efforts, especially in circles. The determinants of these incentives as described above, suggest useful directions for reform. First, consideration should be given to removing the ward/circle distinction, replacing it with random assignment rules. This will remove the strategic underfiling incentive, with beneficial compliance effects. Second, the estimates indicate significant revenue productivity with respect to expansion of support staff and assessing officers. Third, revenue gains may be achieved by reallocating support staff from wards to circles, where they appear to be more productive. Finally, penalty and prosecution effort appears to have significant effects on compliance. It would be worthwhile to encourage such efforts by improving the quality of information available to AOs, and including measures of penalty effort in their performance evaluations.

All of these reforms are simple and unlikely to meet with much opposition from employees of the tax administration. Other more sophisticated and ambitious reforms could also be considered. These include reducing discretion of assessing officers with regard to selection and conduct of audits, increasing competition across assessing officers, closer supervision of audits, reforming performance evaluation and personnel allocation procedures, centralized audit selection procedures based on income disclosures and information generated by third parties. Many of these have been discussed in Das-Gupta and Mookherjee (1998, 2000). However they are less easy to achieve in the light of the scale and expense of the reform efforts, and opposition from tax administration employees.

Appendix A: Proof of Proposition 1

Note that

$$W(y^d, y, \pi) \equiv \left(1 - \frac{y^d}{y}\right)\pi \frac{1}{\alpha}[y(1-\tau) - \tau f(y - y^d)]^\alpha + \left[1 - \left(1 - \frac{y^d}{y}\right)\pi\right]\frac{1}{\alpha}(y - \tau y^d)^\alpha$$

is strictly concave in y^d. Moreover, maximization of W is equivalent to selection of a fraction r of true income to be disclosed, to maximize

$$V(r, \pi) \equiv \frac{1}{\alpha}\{\pi(1-r)[(1-\tau-\tau f(1-r)]^\alpha + [1-\pi](1-\tau r)^\alpha\}.$$

Let $r(\pi)$ denote the optimal disclosure fraction for fixed π; this is easily verified to be strictly less than one, and a strictly increasing function of π. Moreover, the maximized value of V is strictly decreasing in π.

However, the value of π actually depends on whether the disclosed income is below or above 2. Hence there is a function $\pi = \pi(r)$, which takes the value ρ if r is smaller than $\frac{2}{y}$, and β otherwise. The problem faced by a taxpayer with true income y is to select r to maximize $V(r, \pi(r))$, subject to the constraint $\pi(r) = \rho$ if $r < \frac{2}{y}$, and β otherwise.

Consider first the case where β is bigger than ρ. Then this problem is solved as follows. First check if $V(0, \rho) \geq V(r(\beta), \beta)$. If this is the case, then every taxpayer is better off filing in the ward and disclosing a infinitesimal income (since $V(., \rho)$ must be strictly increasing at zero disclosure). In this case $y^* = \infty$.

Now consider the opposite case: $V(0, \rho) \geq V(r(\beta), \beta)$. Then there is unique $r^* \in (0, 1)$ such that $V(r^*, \rho) = V(r(\beta), \beta)$. This follows from the continuity of V in r, and the fact that $V(r, \rho) > V(r, \beta)$ for all r, so $V(r(\beta), \rho) > V(r(\beta), \beta)$.

Define $y^* = \frac{2}{r^*}$. Then we claim that the taxpayer will file in the ward, i.e., the optimal r will fall below $\frac{2}{y}$, if and only if $y < y^*$.

To show the 'if' part, note that $y < y^*$ corresponds to the condition that $r^* < \frac{2}{y}$, so it is feasible for the taxpayer to disclose r^* and be subject to the scrutiny probability ρ, i.e., file in the ward and do at least as well as what she could get by filing in the circle $V(r(\beta), \beta)$. She can do even better by filing in the ward and disclosing r in the neighborhood of r^*, since $V(., \rho)$ cannot be maximized at r^* (if it were, the maximized value of $V(., \rho)$ would be the same as the maximized value of $V(., \beta)$).

Conversely, if $y > y^*$, then $r^* y > 2$, so it is not feasible for the taxpayer to file in the ward and disclose r^* fraction of her true income. She must disclose less in order to file in the ward. Now $V(., \rho)$ must be strictly increasing at r^* and hence (by concavity) at all $r < r^*$. So for any $r < r^*$, the taxpayer is worse off filing in the ward: $V(r, \rho) < V(r^*, \rho) = V(r(\beta), \beta)$. Hence it is optimal for such a taxpayer to file in the ward.

Hence in the case where $\beta > \rho$ we can find $y^* > 2$ which depends on β and ρ, such that the set of taxpayers filing in the ward is exactly those with true incomes below y^*. It is evident from the definition of y^* that it is increasing in β and decreasing in ρ.

The argument is similar in the case where $\beta < \rho$. □

Appendix B: Construction of Variables for Regressions in Table 8

Consider first the case where a range contains a number of wards and a single circle. We assume that the relative proportions of representatives of these different jurisdictions in the circle exactly equal the relative workload sizes of the different wards. This is valid provided that different jurisdictions are characterized by the same equilibrium proportion filing in wards. This assumption permits us to estimate the population sizes of the different jurisdictions. Morever, we assume that filers in any unit are selected randomly for scrutiny, so that the number of scrutinies applicable to circle-filers of any given jurisdiction equals the corresponding proportion of this jurisdiction in the circle. This gives us the appropriate measures of enforcement to be used in the ward regressions.

We also need measures of enforcement for the single circle. For a taxpayer of a given jurisdiction this is the measure corresponding to the appropriate ward; we then have to aggregate across the different jurisdictions (wards), again using their relative workloads as weights. Given our log-linear formulation of the regression equations for an individual taxpayer, this procedure is exactly correct for predicting circle level revenues provided that the geometric and arithmetic averages of revenues within every assessment unit is the same. Given constant relative risk aversion, this in turn requires the distribution of true income amongst filers in any unit to not be skewed. Since this assumption is unlikely to be valid, this introduces measurement bias.

In the case with multiple circles as well as wards, we extend this method of proportional allocation of workload across populations of different jurisdictions. A jurisdiction is defined as a particular ward-circle pair, and different jurisdictions are characterized by a similar proportions filing in wards. We also need to assume, as above, that enforcement efforts within any unit get uniformly allocated across different jurisdictions falling within that unit, and that income distributions within any jurisdiction are not excessively skewed. These assumptions enable us to construct suitable measures of enforcement for each assessment unit in the sample. It ought to be evident from this discussion that the regressions are prone to measurement error, a problem which is inherent in the nature of the data, and can be overcome only if different jurisdictions can be separately identified (e.g., with micro-level data).

Finally, in order to test equation (19) for allocation of workloads between wards and circles, we also need to measure the relevant dependent variable. For this we follow the same procedure as above. For example, in the case of a single circle and multiple wards, we allocate the workload of the circle to the different jurisdictions according to the relative workloads of the respective wards. This generates an estimate of the total taxpayer population size for each jurisdiction, as well as the allocation of that population between the ward and the circle.

Acknowledgment

We are grateful to D.P. Panta of the Indian Income Tax Department for his collaboration in designing the questionnaire for assessment units and subsequently collecting the data, and to him as well as Arbind Modi for enlightening us on myriad details of the functioning of the Indian tax administration. This paper is based on a project carried out by Das-Gupta and Mookherjee for the National Institute of Public Finance and Policy, New Delhi. For useful comments we are grateful to John Wilson, an anonymous referee, and participants at the 1998 NEUDC conference at Yale, and the Columbia University-World Bank conference on Tax Design and Institutional Reforms in Developing Countries held in February 2000.

Notes

1. It also causes developing countries to rely excessively on regressive production and trade taxes that generate cascading deadweight losses. Low income countries with per capita income below $360 raise tax revenues

that are one-seventh of GDP on average, compared with one-third of GDP raised by countries with income above $6000. Income and trade taxes account respectively for 3.5% and 5.3% of GDP in poor countries, as compared with 11.0% and 0.7% for countries above $6000. See Burgess and Stern (1992) for further details.

2. For theoretical analyses of anti-corruption policies, see Besley and McLaren (1993), Mookherjee and Png (1995) and Mookherjee (1998). Kahn, de Silva and Ziliak (2001) is one of the few papers providing empirical analysis of an anti-corruption reform in tax administration.
3. In addition, there are a few elite *investigation circles*, which include a few high profile taxpayers concerning whom the tax administration has some special incriminating evidence, for example from raids or searches. We exclude these special units from our empirical analysis because of their exceptional nature.
4. The AO reports to the Deputy Commissioner (DC) heading the relevant range. For a variety of reasons described in detail in Das-Gupta and Mookherjee (1998, Ch. 6), the range DC and the AO are in turn subjected to minimal supervision. There is much evidence that audits, both internal and external, focus excessively on technical correctness of assessment duties, and ignore errors of omission. Moreover, the tax administration pays little attention to these audit reports in appraising the performance record of individual officers, while deciding on promotions and transfers: they focus instead almost exclusively on achievement of quantitative targets concerning disposal of workload. Moreover, centralized instructions concerning audit strategy, or penalty and prosecution effort are conspicuous by their near absence. AO compliance with the few loose audit guidelines is not effectively monitored.
5. For instance, the progressive expansion of summary assessments at the expense of scrutiny assessments since the early 197s was rationalized by them as the obvious means of ‘disposing’ an ever-increasing assessment workload.
6. The alternative modelling approach assumes that the tax authorities and taxpayers move simultaneously in selecting their respective strategies, and that the former’s objective is to maximize expected net revenues: see the models of Graetz, Reinganum and Wilde (1986) and Reinganum and Wilde (1986). See Andreoni, Erard and Feinstein (1998) and Mookherjee (1997) for further elaboration of the contrast between the two modeling approaches.
7. See Mookherjee and Png (1995) and Mookherjee (1997) for models of bribery where the bribes turn out to be proportional to the extent of evasion discovered.
8. In the event of bribery, p_{it} is simply the probability that tax evasion will be discovered by the auditor, and $k_{it}^2 = 1$.
9. This was constructed by computing a weighted average of the proportion of penalties that were actually upheld by appeals courts, following a taxpayer appeal.
10. Moreover, the estimate of a is itself biased upward, and our estimate of a was indeed significantly less than one, so the hypothesis of diminishing returns to staff is upheld.
11. Moreover, the estimated coefficients with respect to workload or dummies for ranges 1, 2 and 14 have signs exactly opposite to that predicted by (13) and the structural or reduced form estimates of the compliance equation. Elasticities with respect to compliance-relevant observed characteristics are quantitatively and statistically insignificant. The results are also similar for the OLS equation, with the single exception of the workload elasticity. But even this elasticity is fairly small and not statistically significant at 10%.

References

Allingham, M. and A. Sandmo. (1972). “Income Tax Evasion: A Theoretical Analysis,” *Journal of Public Economics* 1, 323–338.

Andreoni, J., B. Erard and J. Feinstein. (1998). “Tax Compliance,” *Journal of Economic Literature* 36, 818–860.

Besley, T. and J. McLaren. (1993). “Taxes and Bribery: The Role of Wage Incentives,” *Economic Journal* 103, 119–141.

Burgess, R. and N. Stern. (1992). “Taxation and Development,” STICERD Development Economics Discussion Paper No. 42 (London School of Economics).

Das-Gupta, A. and D. Mookherjee. (1998). *Incentives and Institutional Reform in Tax Enforcement: An Analysis of Developing Country Experience*. New Delhi: Oxford University Press.

Das-Gupta, A. and D. Mookherjee. (2000). "Reforming Indian Income Tax Enforcement." In S. Kahkonen and A. Lanyi (eds.), *Institutions, Incentives and Economic Reforms in India*. New Delhi, Thousand Oaks, London: Sage Publications.

Graetz, M., J. Reinganum and L. Wilde. (1986). "The Tax Compliance Game: Toward an Interactive Theory of Tax Enforcement," *Journal of Law, Economics and Organization* 2(1), 1–32.

Kahn, C., E. de Silva and J. Ziliak. (2001). "Performance-Based Wages in Tax Collection: The Brazilian Tax Collection Reform and its Effects," *Economic Journal* 111, 188–205.

Mookherjee, D. (1997). "The Economics of Enforcement." In A. Bose, M. Rakshit and A. Sinha (eds.), *Issues in Economic Theory and Policy: Essays in Honour of Tapas Majumdar*. New Delhi: Oxford University Press.

Mookherjee, D. (1998). "Incentive Reforms in Developing Country Bureaucracies: Lessons from Tax Administration." In B. Pleskovic and J. Stiglitz (eds.), *Annual World Bank Conference on Development Economics 1997*. Washington D.C.: The World Bank.

Mookherjee, D. and I. Png. (1995). "Corruptible Law Enforcers: How Should They be Compensated?" *Economic Journal* 105, 145–159.

Reinganum, J. and L. Wilde. (1986). "Equilibrium Verification and Reporting Policies in a Model of Tax Compliance," *International Economic Review* 27(3), 739–760.