
**Minimising Selection Failure and Measuring Tax Gap: An
Empirical Model**

Sudhanshu Kumar and R. Kavita Rao

Working Paper No. 2015-150

May 2015

**National Institute of Public Finance and Policy
New Delhi
<http://www.nipfp.org.in>**

Minimising Selection Failure and Measuring Tax Gap: An Empirical Model

Sudhanshu Kumar*
and
R Kavita Rao#

Abstract

This paper presents an empirical model for minimising selection failure by tax departments in selecting cases for scrutiny assessment. This model also provides a new methodology for estimating tax gap from limited information that the department collects on a regular basis through scrutiny assessments. Using a maximum-likelihood procedure that corrects for sample selection bias, and the data on the scrutiny assessment exercise carried out by the income tax department, we estimate the model which relates the probability and extent of under-reporting to various inputs provided by the tax filer. The estimated model provides a mechanism to analyse the trade-off between two types of cases of failure - wrong selection of a case and failure to take up the potential underreporter.

JEL Classification: C52, H26.

Key words: Selection Failure, Tax Gap.

* Assistant Professor, National Institute of Public Finance and Policy (NIPFP), New Delhi.
E-mail: sudhanshu.kumar@nipfp.org.in

Professor, National Institute of Public Finance and Policy (NIPFP), New Delhi.
E-mail: kavita.rao@nipfp.org.in

1. Introduction

Income tax evasion or tax gap is a widely recognized problem that costs governments throughout the world, great deal of revenue. If it were possible for the tax department to audit all filed returns and validate all the information used by all the taxpayers in the economy, it is to be expected that the tax gap would be reduced to zero.¹ However, this is neither considered feasible nor desirable – the former since it would involve substantial administrative costs and the latter since it would be perceived as too intrusive by the taxpayer. Discussions on reforms in tax administration have shifted their focus from “enforced compliance” to “voluntary compliance”. The taxpayer too is more vocal in protesting against invasive tax administration by referring to it as “tax terrorism”. In this context, tax departments have been limiting their audit exercise to a fraction of the population. In designing their audit strategies, therefore the departments will have two concerns

1. Finding a methodology for measuring the size of tax gap – this would be a good indicator of the performance of the department where small gap would indicate a successful strategy.
2. While reducing the gap is a concern, since there are compliance benefits for the tax department from being perceived as fair, it would be useful for the department to have a methodology for selection of cases for audit where it can minimize the possibility of harassment, i.e., the possibility of auditing an honest taxpayer.

The present paper is an attempt to develop a methodology to address both these issues.

The methodologies used in the literature for estimating the tax gap can be largely classified into two categories: the macro approaches and the micro approaches.² The macro approaches focus on estimating the size of the aggregate hidden economy using various approaches like the national income-expenditure discrepancy, national income-revenue discrepancy and other econometric approaches. The two prominent ways of estimation using the macroeconomic variables are: the traditional regression approach, and using the stochastic frontier analysis. Although both these approaches are appealing in the way they handle the issue of potential tax collection, the results rely heavily on the selection of appropriate macroeconomic determinants of the tax collection. Also, they provide no information with regards to probability that a filer has underreported.

¹ It is possible to argue that the auditors may not be efficient or honest which can result in a continued tax gap, but this aspect is not the focus of the present paper. Also, one part of the tax gap is always attributed to non-filing of the tax returns as well. This paper limits its analysis to the people who underreport income in filing their tax returns.

² Very broadly, the tax gap is the difference between some estimated potential tax and the actual realisation of the tax. The term 'tax gap' is increasingly popular as a means of assessing the degree of success with which a particular tax or tax system is implemented. The Internal Revenue Service (IRS) in the US defines the tax gap as 'the difference between the tax that taxpayers should pay and what they actually pay on a timely basis'. As per this definition, the tax gap has three components: non-filing, underreporting of income and also overstating of deductions, and underpayment.

The micro approaches—taxpayer surveys and random audits—are based on grossing-up from taxpayer compliance data. The methods based on compliance activity, in some cases, take years to complete³. In micro-approach category, the main methods used to estimate tax gaps for direct taxes are random enquiries, data matching, and risk registers.

As an alternative approach, we propose a methodology to estimate the tax gap based on the regular scrutiny assessment of tax returns carried out by the tax department. Since this is a regular exercise carried out by the department, using this data for measuring tax gap is not costly and will provide a ready and reliable estimate which can be improved upon with data flowing in year on year.

Moving on to the second issue, starting from the premise that the tax department selects a limited number of cases for audit, whatever be the mechanism by which the department selects these cases, there is a possibility of two kinds of errors or selection failures - selection of an honest filer who yields no additional revenue but generates a reputation of the tax department harassing people and, failure to select the potential under-reporter, resulting in loss of revenue to the tax department. Ideally, the department would like to minimise both kinds of selection failures.

Using a maximum-likelihood procedure that corrects for sample selection bias, and the data on the scrutiny assessment exercise carried out by the income tax department, we estimate the model which relates the probability and extent of underreporting to information provided in the tax returns of the tax-filers. Although it is not possible to exactly pin point the under-reporter through the proposed model, the prediction from the model can act as a guiding tool for minimizing the selection failure in a scenario where resources to carry out the scrutiny assessment is limited. First component of the model assigns a probability to under-report for every taxpayer whereas the second part gives an estimate for the extent of tax gap due to under-reporting by the tax filers.

Given the estimated model, we consider two alternative strategies for selecting cases for audit – one based on the probability of under-reporting alone and the second based on the “expected quantum of under-reporting”, which is the product of the probability of under-reporting and the estimated extent of under-reporting. The performance of these models are evaluated on the basis of the two parameters of interest – extent of incorrect selection indicating scrutiny of an “honest” taxpayer and extent of additional revenue mobilisation.

If the cases taken up for scrutiny assessment were to be randomly selected, the proposed model would provide a representative coverage of the entire set of returns. The estimated versions of this model can then be applied to the population data that is, the entire set of returns, both to select the cases for scrutiny assessment and to derive an estimate of the tax gap. With the arrival of the new set of data, the model can be re-estimated on the sample before applying to the population. This leaves space for

³ In the United States, the Internal Revenue Service (IRS) uses taxpayer compliance-based methods for calculating tax gaps for direct taxes. IRS gathers the tax compliance data from randomly selected individual taxpayers (with an over-sampling of high income returns). However, these approaches are highly resource intensive, both for tax agencies and for taxpayers. See Mazur and Plumley (2007) for a detailed description.

correction from the past experiences of the scrutiny assessments in order to improve the predictive ability of the model.

Rest of the paper is organised as follows. *Section 2* describes the econometric model in detail. *Section 3* gives detail about the data on which the model is run and validated. Results are discussed in *section 4*. *Section 5* concludes the paper.

2. Methodology

For the current analysis, difference between the assessed income and the returned income of companies is named as 'gap', and is explained using the available information from the dataset. This gap is not always positive. It may be zero in some of the cases.⁴ Given the way the law is defined, the liability of tax is different for those who pay regular tax and those who pay Minimum Alternate Tax (MAT). Further, in the process of assessment, the status of a tax payer can change from being liable to MAT to being liable to regular tax. To avoid dealing with these issues, we are working with non-MAT cases alone.

As discussed above, in addition to explaining the 'gap', we also need to explain the possibility of being assessed to additional income in any given case. Excluding all the cases which yield no additional income in assessments would result in biases in the estimates derived, since we are not explaining why, in some cases, there is no additional income assessed. Thus, given the structure of the dataset and the questions to be answered, we need to set up a model which utilises all the available information efficiently to answer both questions jointly: factors explaining the probability of getting assessed for additional income i.e., assessed income being higher than the returned income, and the extent of additional income determined in the assessment.

In the literature, there is a class of models which provides a framework to address such issues. However, to best of our knowledge there is no application of such models for analysing the tax underreporting within the existing literature. The Heckman Selection-Correction Model (Heckman, 1979) can be used, to assess whether selection bias is present, to identify factors contributing to the selection bias, and to control for this bias while estimating the outcome. In the present case, the bias is defined in terms of which cases are associated with additional assessed income within the assessed cases.⁵ We assume that the characteristics given in the dataset can help in explaining the selection issue to some extent.

We employ the two-step Maximum Likelihood Heckman Model by first estimating a selection, and then the outcome equation, adjusting for selection bias, if there is any. This exercise helps us in understanding the factors that determine both the possibility of being assessed for additional income, which could also be interpreted as the possibility of

⁴ Any case where the gap is negative has been excluded from the analysis since this would be an aberration. This makes our data on gap censored at zero from below.

⁵The selection bias problem arises because the error term in the outcome equation is correlated with the error term in the selection equation. This means that the error term in the outcome equation will not have mean zero, and will be correlated with the explanatory variables. This, in turn, leads to inconsistent estimates.

under-reported incomes, as well as the factors that explain the quantity of such additional income, that is, the dimensions of evaded income. Thus, we have two components to the model: the *selection component*, and the *outcome component*.

In sum, the selection component helps us in explaining the probability of a tax filer getting assessed positively i.e., at a figure higher than the returned income, and the outcome component explains the size of under-reporting of the income.

Let y_2^* denote the outcome of interest, here the 'gap', and this outcome is observed if the other latent variable $y_1^* > 0$. Here, y_1^* determines whether a tax filer is assessed positively, and, y_2^* determines level of the assessment.

Thus, the two-equation model comprises a selection equation for y_1 , where

$$y_1 = 1 \text{ if } y_1^* > 0 \\ = 0 \text{ otherwise}$$

and a resultant outcome equation for y_2 , where

$$y_2 = y_2^*, \quad \text{if } y_1^* > 0 \\ = --, \text{ otherwise}$$

Here, y_2 is observed only when $y_1^* > 0$.

We assume a linear model with the additive errors.

$$y_1^* = x_1' \beta_1 + \varepsilon_1 \\ y_2^* = x_2' \beta_2 + \varepsilon_2$$

where 'x' denotes the explanatory variables. The errors ε_1 and ε_2 are possibly correlated. It is assumed that the correlated errors are jointly normally distributed and homoscedastic, that is

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right]$$

where, the normalization $\sigma_1^2 = 1$ is used because only the sign of y_1^* is observed.

The dataset is truncated at zero, that is, the 'gap' is either zero or greater than zero. This makes the dataset suitable for the Heckman-selection bias model to determine whether selection bias is present. If there is any selection bias, then we look at the factors contributing to the selection bias. For the suitability of the model, the error terms in selection and outcome equations should be significantly correlated with each other. It justifies the presence of a selection bias, and hence, both the selection and the outcome models should be jointly estimated.

To estimate the tax gap from this model, ideally, the model generated from the sample of audited returns should be run on the population of tax returns so as to generate an estimate of the probability of assessing additional income and of the additional income if so assessed. The expected additional income pooled across all taxpayers would constitute an estimate of the tax gap for the population.

$$Tax\ Gap = \frac{\sum_{population} gap * probability}{\sum_{population} returned\ income}$$

The prediction based on the estimated model assigns a probability of under-reporting for each tax filer. The model also predicts the potential tax liability of each taxpayer. Based on these predictions, there are two ways to select a case for audit: based on the probability alone or, based on the expected tax gap for taxpayer. In the first approach, the taxpayers can be arranged in descending order by probability and those with the highest probability can be selected, the exercise continuing until the desired number of cases for scrutiny has been identified. This could also be viewed as a method where the department determines a cut-off probability above which the tax filer is expected to have possibly under-reported its income and hence taken up for audit.

Other possible way is to arrange taxpayers in descending order of expected gap and select cases so as to maximise the expected gap. In case the department does not have limitations on the number of cases it can audit, this method would be akin to auditing as many cases that are sufficient to achieve the expected revenue target set by the tax department.

The selections made through either of these approaches have an advantage over the random selection of the cases in terms of reduction in the cases of selection failures. To compare the performance of these two alternative approaches, the two criteria of interest would be:

- 1.) The extent of unreported incomes recovered (to be referred to as “gap covered”), and,
- 2.) The extent to which “honest” taxpayers are harassed.

3. Data

The results in this paper are based on scrutiny assessment for corporate income tax payers in India for the year 2009-10.⁶ The data was provided by the Income Tax Department (ITD) and was initially analysed for the project, ‘Study on unaccounted income/wealth both inside and outside the country’ at the National Institute of Public Finance and Policy. Information on some of the subheads in the income tax return (ITR) file was used for estimating and validating the model. Information on following variables was available to us as a part of the mentioned project: Total Income, Total Assessed Income, Total Tax liability, Brought Forward Loss, Deduction, Exemption, Depreciation, Profit Before Tax, and Sector Code. However, the returns filed by the taxpayers include information on many more variables. Incorporating all these variables might improve the performance of the model.

Ideally, the model developed using this data should be run on the population of taxpayers, to get an estimate of the tax gap. However, since we did not have access to

⁶ Scrutiny assessment refers to the process where the return is taken up for audit by the tax department. In such cases, the income assessed can vary from income returned.

full dataset, we have adopted an alternative approach – the available sample is being treated as the population. This population has been divided into two subgroups through random sampling, a ‘developmental sample’ and a ‘validation sample’. The developmental sample consists of 70 per cent of randomly selected observations from the newly defined population and, is used to develop the model. The estimated model is validated by applying it to the remaining observations which is being referred to as the validation sample. In order to establish the robustness of the results so obtained, the process has been repeated 100 times over. In what follows, we report the summary results obtained from these 100 simulation rounds.

4. Results and Analysis

The model is first run on the developmental sample using all the information available with us for the tax filers, and then validated with the remaining sample. *Table 1* presents the estimated result using the developmental sample in one such simulation round as an illustration. The summary statistics validate usefulness of the two-step estimation model. The variables used are from the list of information that was provided for each tax filer. This estimated model is used to predict the probability of tax under-reporting and the tax gap for the validation sample.

Variables (in logarithm except the sectoral dummies)	Selection Model			Outcome Model		
	Coefficient	z-stat	p-value	Coefficient	z-stat	p-value
Tax liability	0.045	6.160	0.000	2.400E+07	2.10	0.036
Brought forward loss	0.005	1.610	0.107	6.215E+06	2.07	0.038
Deduction	0.029	9.670	0.000	9.454E+06	2.60	0.009
Exempt	0.019	2.290	0.022	4.740E+06	0.70	0.486
Depreciation	-0.007	-3.280	0.001	-7.043E+06	-2.72	0.007
Total income	0.049	7.430	0.000	2.370E+07	2.28	0.023
Profit before Tax	0.000	1.910	0.057	2.575E-01	109.37	0.000
Manufacturing	-0.749	-18.360	0.000	-1.090E+09	-3.04	0.002
Agent	-0.824	-5.390	0.000	-1.100E+09	-2.83	0.005
Builder	-0.907	-18.520	0.000	-1.130E+09	-2.95	0.003
Contractor	-0.491	-7.220	0.000	-9.680E+08	-2.94	0.003
Professionals	-0.734	-5.140	0.000	-1.080E+09	-2.87	0.004
Service	-0.840	-18.970	0.000	-1.060E+09	-2.85	0.004
Financial service	-0.739	-16.760	0.000	-1.060E+09	-2.98	0.003
Trade	-0.706	-16.800	0.000	-1.050E+09	-3.01	0.003
Entertainment	-0.927	-4.400	0.000	-1.130E+09	-2.56	0.010

Number of observations = 11947
Censored observations = 4802
Uncensored observations = 7145
Wald chi2(16) = 12419.79 Prob> chi2 = 0.000
Rho = 0.85; p-value for Mills-lambda = 0.001

Tax Gap

While ideally the estimate of tax gap should be based on the population of tax returns, given that we had access to limited set of data, as discussed in section 2, we have split our sample into a developmental sample and a validation sample. The model estimated on the developmental sample is run on the validation sample. In order to validate the effectiveness of the estimate, we compare the predicted tax gap with the actual tax gap for the validation sample. It may be recalled that in the validation sample, we have information on assessments and the corresponding increases in income if any. Results from this validation exercise are summarised in *Table 2*.

Table 2: Predicted Gap and the Actual GAP		
	Predicted Tax Gap	Actual Tax Gap
Mean	0.33	0.32
Std. Dev.	0.10	0.04
t- test for equality of mean = 0.93 (The means are not statistically different from each other)		

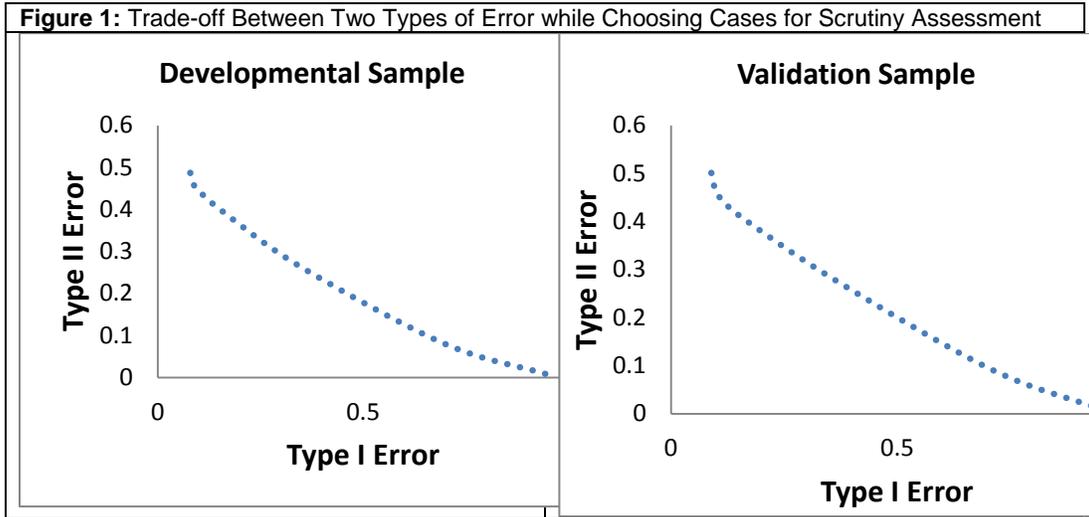
The results indicate that the average predicted gap is statistically not significantly different from the average actual gap. In other words, this approach provides a good mechanism for estimating the tax gap. In other words, if the model estimated on the scrutiny assessment data can be run on the population of returns filed in the country, the results suggest that it would provide a reliable estimate of the tax gap in the country.

Minimising Selection Failures

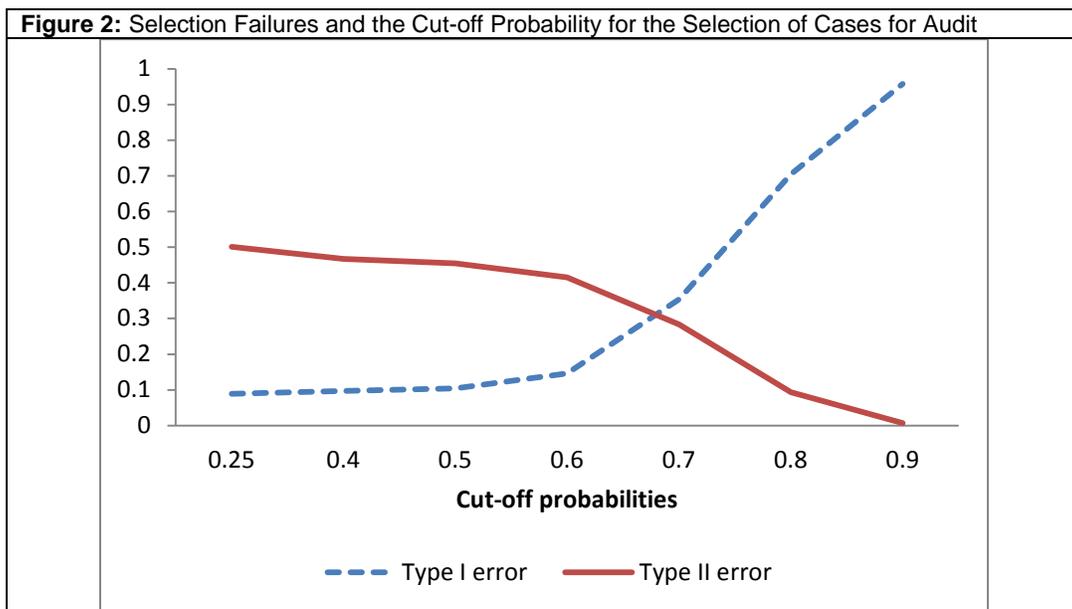
One of the objectives of the exercise is to select the cases for scrutiny assessment (in order to catch the under-reporter) from the list of tax filers based on prediction of the model. Thus, the selection exercise has the null hypothesis of 'a selected filer being an under-reporter'. Such selection exercise has possibility of committing both types of errors: Type I error- incorrect rejection of true null and the Type II error- failure to reject a false null. For the given null in current exercise, Type I error means failure to catch an under-reporter and Type II error corresponds to the scrutiny assessment of an honest tax filer. Type I indicates letting go away the culprits whereas Type II signals about the harassment of a genuine filer. In absence of a perfect way to catch the culprit, a balance between the two types of errors is desirable depending on the preference of the tax authority.

The trade-off between these two is better presented in *figure 1* based on the sample selection for scrutiny assessment through the estimated model. *Figure 1* suggests that there is an inverse relation between two types of the errors i.e. as the tax

authority increases sample size for the audit in order to catch the under-reporter the possibility of the cases of harassment of an honest filer rises. Type II error for the scrutiny assessment based on the existing approach of the tax authority in India for the corporate income tax filer is 0.40 for the year 2009. The proposed model provides a way to minimize the type II error in the selection of the cases.



The tax authority can use a cut-off probability by looking at the *figure 1*, in order to select the cases for audit. The cut-off probability may be chosen to give a set number of samples depending on the administrative capability of the department to carry out detailed audit of the filers. Also the tax authority can use the information contained in *figure 2* to supplement the information set before deciding about the cut-off probability. *Figure 2* presents the extent of Type I and Type II error as we increase a cut-off probability for selecting a case for scrutiny assessment. This gives a tool to assess the extent of associated selection failure for a chosen cut-off probability.



Alternatively, if the tax authority aims to take up a fixed number of cases for the audit, it may like to take up the cases which maximise its expected revenue gain from the scrutiny assessment. This can be done from the predicted outcome or the tax gap by the estimated model for each of the filers. To do so, a fixed number of samples from the descending order of predicted tax gap are selected. This way, cases which are suspected to give more revenue are selected for the assessment exercise.

Given that tax departments are constrained by finances, administrative capacity and sometimes by law, about number of cases can be taken up for scrutiny assessment. We assume the number to be 10 percent of the population. As discussed earlier, it is possible to select cases based on two approaches. Since the information in the development sample has already been used for developing the model, the present exercise focuses on selecting a sample out of the validation sample based on the two alternative approaches. To assess the performance of these two alternative methods, *Table 3* provides two summary numbers.

One point of comparison is based on estimates of type II error as a percentage of the total number of cases taken up for scrutiny assessment. This measure captures the extent to which “honest taxpayers” have been subject to scrutiny – a lower number here is clearly preferable. The second point of comparison is - of the total gap assessed by the department for the validation sample, what is the gap that is captured by the selected sample. In other words, what is the effectiveness of the selected sample in covering the tax gap within the population. This will be referred to as “gap covered”.

Since in the exercise for validating the model for measuring tax gap, we have developed 100 simulations, the present effort also works with these many simulations. In each case, 10 percent sample is drawn based on the two alternative approaches. *Table 3* presents the average for these criteria across the 100 simulations. In *table 3*, the “gap covered” of value 1 would represent the model where scrutiny assessment captures all of

the under-reported incomes. From the results, we observe that the selection based on the cut-off probability has lower type II error. While selection based on expected revenue gain shows a marginally better performance in the “gap covered”, the differences are not large.

	Based on cut-off Probability		Based on Revenue Gain	
	Mean	Std. dev.	Mean	Std. dev.
Type 2 error	0.12	0.01	0.24	0.06
Gap Covered	0.63	0.13	0.67	0.12

5. Conclusion

Tax authorities around the world carry out scrutiny assessment exercise in order to catch the under-reporter on a sample selected from the total available tax filers. In absence of complete information, in any of the scrutiny assessment exercises, there is always a possibility of carrying out a scrutiny assessment of an honest filer or there are chances that a potential underreporter stays outside the purview of the tax authorities. Both the cases of failure - wrong selection of a case and failure to take up the potential under-reporter - have a cost. Thus, any exercise where a sample of tax filers is to be selected from the population needs to make a balance between these two types of the possible selection errors.

In this paper, we develop an empirical model that is useful in reducing the selection failures in taking up the cases for scrutiny assessment. The empirical model helps to detect the tax filers who cheat by filing fraudulent income tax returns by assigning a probability of under-reporting, and then estimates the shortfall in corporate income tax revenue due to the under-reporting. Based on a cut-off probably that takes care of the trade-off between the two types of selection failures, the sample is selected for the audit. Other alternative to this is to select the cases in order to maximize the expected revenue gain. We find that the selection based on the cut-off probability has lower Type II error compared to the selection based on maximizing the revenue gain.

The two step estimation procedure provides an estimate for the size of under-reporting by the tax filers or the tax gap. To the tax authority, the expected tax gap tells the estimated potential revenue that can be collected through the scrutiny assessment of the tax filers.

References

- Heckman, J. J., 1979. "Sample Selection Bias as a Specification Error", *Econometrica*, 47(1);153-161 (January).
- Mazur, M. and Plumley, A., 2007. "Understanding the Tax Gap". *National Tax Journal* 60; 569–576.