PREVENTING TAX EVASION OF THE COMMERCIAL TAX DEPARTMENTS THROUGH STATISTICAL MODELS

*Dr. R. Venkataraman   **Ms. Seeboli Ghosh Kundu

Abstract :

Billions of dollars across the Tax administrations around the world are lost every year due to non-compliance, evasions, frauds or non-collection including both direct and indirect taxes. With access to vast quantities of data from a range of sources (e.g. financial institutions, utilities, bank transactions, social media data, etc.) both in terms of structured as well as unstructured (text, video, pdfs, etc.), tax authorities can increasingly use business rules, quantitative statistical models and advance analytics techniques to conduct audits and uncover trends and discrepancies, using new techniques such as rule-based monitoring, predictive modelling and outlier detection. This paper will showcase how the tax authorities should be moving into the predictive mode rather than post audit reactive mode for zeroing on the probable risky dealers in the indirect tax domain for highest impact of revenue recovery and collection through a statistical, scientific and information driven model for decision making.

Keywords: *Tax Evasion, Statistical Models, Logistic Regression, Misclassification, Input/Output Tax, Exempted/ Total Turnover, Predictive Model, Age of Account, Odds Ratio, Model Coefficient.*

JEL Classification : *D81, GOD, D84, H20 Taxation, F42 International policy coordination and transmission,*

Introduction:

Government's worldwide lose more than $3.1 trillion in annual revenue because of tax evasion in the indirect taxes, according to a report by World Bank. The equivalent of more than 5.1 percent of global gross domestic product never reaches the coffers of 145 national governments in the form of indirect taxes, according to the report by The Tax Justice Network, an independent group that promotes financial transparency (Rangarajan Committee, 2010). The lost taxes are mainly from the shadow economy, or cash transactions that were hidden deliberately from the tax authorities. Especially in the indirect tax domain, the different dealers working in multiple distinct sectors take the advantage of the interstate suppliers of raw materials and try to evade the regularities through differencing the input and the output taxes. These irregularities are only captured post audit after there is a substantial loss of resource and many

Dr. R. Venkataraman
Professor
Centre for Management Studies
Presidency College

Ms. Seeboli Ghosh Kundu
Senior Lecturer
ITM Business School &
Research scholar at Bharathiar University
Email : seebolighosh2011@gmail.com

a times interventions being impossible(Reserve Bank of India, 2011). Thus there is an immense requirement to create a predictive model or a scorecard predicting the likelihood of risk for different dealers. The transaction time flag has the potential to predict the probable tax evasion and thus prevent tax evasion rather than detect fraud detection post audit.

Objectives of Analysis:

✦ The Objective of this analysis it to develop a predictive risk model to predict dealers' likelihood of risk. Risk score generated can be used to scrutinize dealers for audit.

✦ The predictive model can also distinguish between the significant variables and the insignificant variables thus prioritizing the variables to be looked into.

✦ To predict the probable tax evasion with the help of transaction time flag and thus prevent tax evasion rather than detect fraud detection post audit.

Literature Review:

Kumar, Nagar and Samanta (2007) incorporated econometric model in order to identify the effectiveness of direct tax administration. Their study has given focus on the collection of personal income tax and corporation tax at pre-assessment and post assessment stage. They concluded that perceived inequity of tax system, complexity of tax laws, lack of fairness in penalty system and weak taxpayer education programmes were the factors for poor voluntary compliance. The study further pointed out the requirement of effective information system and database in order to achieve effective tax administration.

The study of Singh and Sharma (2007) was based on primary data where they tried to examine the importance and perception of tax professionals in the system of Indian Income Tax. In their study they have identified seven significant factors like reduction in tax evasion, extension of relief to taxpayers, incentives for dependents and honest taxpayers, broadening the tax base, e-filing of returns, adequacy of deductions and impact of exempt tax system that determining the effectiveness of Indian tax system.

Datar (2010) has conducted a research on direct tax code. His paper "Why the Code must be shelved" expressed his views about the importance of direct tax code. According to him people would have to waste a lot of time in understanding the new provisions of income tax law and CBDT would have to issue numerous circulars and frame several rules all over again. He also mentioned that proposed Code would neither improve efficiency nor tax collection due to deep rooted corruption.

Research Methodology

The research methodology includes incorporating a secondary research of 20,000 dealer samples assessed from CTD Karnataka for the financial year 2014- 2015 solely for the research purpose. The source of the data includes Dealer master, Registration data, Returns data & Audit risk output.

Hypothesis:

Hypothesis 1

H0: A statistical model used to predict the probability of riskiness of a dealer is as good as a baseline model (no model) and provides no benefit compared to random choice.

H1: A statistical model used to predict the probability of riskiness of a dealer is better than the baseline model (no model) and provides significant benefit compared to random choice.

Hypothesis 2

H0: All the factors including demographics, tax ratios and transactional variables are equally important in discriminating between a probable evader and non-evader of tax.

H1: Some of the factors are more significant than the other in discriminating a dealer between a probable evader and a non-evader.

Data Collection & the Variable characteristics :

The source of the data includes Dealer master, Registration data, Returns data & Audit risk output. The 20,000 samples comprising of 10000 confirmed risky dealers & 10,000 confirmed non risky dealers. The risky and non-risky dealers are confirmed by the audit and vigilance team on the historical data via audit reports. These 20,000 samples are randomly chosen from the dealer database of approximately 2.5 lakh dealer base. Thus the sampling technique incorporated is the stratified random sampling maintaining equal proportion of evaders and non-evaders.

Dependent Variable: A response variable was created as indicator of risk with value 1 and 0. 1 indicates dealer is risky whereas 0 indicates dealer is not risky.

Independent Variables : Following independent parameter were considered for the analysis

1. Age of Account

2. Output Input ratio

3. Total Tax Paid

4. Business Status

5. Business Type

Model Development & the Statistical Techniques Used:

The ratio that has been maintained for the training and validation data, the two parts in which the entire sample is divided is 70:30. It has been taken care that the data is divided randomly by giving every observation the equal chance of being picked removing bias. The Logistic Regression Model is created on the training data and the model validation is performed on the validation data. The full fit model, forward and backward stepwise regression are the three candidate models which are compared against each other in terms of misclassification.

Variable Selection:

Data was imported into SPSS for analysis. Logistic regression model technique was used and significant variables were selected after running multiple model iteration. Variable were selected based on Chi-square test statistics and best model was selected as mathematical equation with combination of all significant variables. The different candidate models used are forward, backward and stepwise. The best model with the least misclassification is zeroed in for scoring.

Analysis & Discussions:

Variable Selection Summary:

Variables Considered for Model	Categories	Significant variables	Definition	Model Coefficient	Odds Ratio	Impact/ Interpretation
Age of Account	NA	yes	Total number of years in the system	0.245	1.27	one unit increase in age of account the odds of being a risky dealer increase a factor of 1.27
Output Input ratio	0	yes	Output tax to input tax credit ratio	-		
	less than 1	yes	Output tax to input tax credit ratio	2.24	9.44	Dealers have higher chances of being risky if they have OI ratio less than 1
	greater than 1	yes	Output tax to input tax credit ratio	2.05	7.8	Dealers have less chances of being risky if they have OI ratio greater than 1 as compared to dealer having OI ratio less than 1
Revenue/ Tax Paid	1	yes	1 if paid tax is greater than 1 crore	1.39	4.02	Dealers with tax paid greater than 1 crore has higher chances of being risky

Source: Secondary Data

The two variables which were identified as significant are age of account and Out Input ratio. Age of account contributed significantly in the model and has positive impact on the risk variable i.e. chances of risk of dealer increase with increase in number of years in the system. Out tax to input tax credit ration was grouped into 3 groups (OI ratio equal to zero, OI ratio less than 1 and OI ratio greater than 1). OI ratio category of where OI ratio less 1 has positive impact on the risk variable i.e. dealer having OI less than 1 has higher chances of being at risk as compared to dealer with OI ratio greater than 1. That is the dealers who take more input tax credits for a period of time has the highest propensity of evading tax.

This propels us to reject the null hypothesis of hypothesis 2 and accept the alternate hypothesis that some of the factors are more significant than the other in discriminating a dealer between a probable evader and a non-evader.

Model coefficient:

The model coefficients give the quantum of impact of the significant variables on the target. Also the sign of the coefficient showcases, the direction in which the variables are impacting (i.e. in the positive direction or in the negative direction). The category of Output/ Input Ratio less than 1 has the highest coefficient of 2.24 followed by the category Output/ Input Ratio greater than 1 (2.05), Revenue/Tax Paid (1.39) and Age of Account (0.245). All the variables are in the positive direction thus showcasing that an increase of one unit for each of them contributed in the riskiness of the dealers.

Odds Ratio:

The odds ratio is a relative way of gauging the impact of the independent variables on the target variable. The odds ratio estimates indicate by what factor the odds of tax evasion increase for each unit change of associated input. Thus a dealer who is in the system more than 1 year has the odds of tax evasion increasing by 1.27 times than a dealer who is less than a year compared to the stated dealer. One unit change of O/I ratio for the category (less than 1) increases the odds ratio by 9.44 times while one unit change of O/I ratio for

the category (greater than 1) increases the odds ratio by 7.8 times. The odds ratio for Revenue/ Tax paid is 4.02.

Summary from the output of Logistic Regression

Classification Table

	Observed	Predicted		
		Risk_A		Percentage Correct
		0	1	
Step 1	Risk_A 0	248	50	83.2
	1	102	79	43.6
	Overall Percentage			68.3

The classification table showcases that 68.3% of the times the model have been able to rightly predict a probable risky dealer as a risky dealer and a probable non-risky dealer as a non- risky dealer. Thus this is significantly more than the base line model (Classification=50%). This propels us to reject the null hypothesis of hypothesis 1 and accept the alternate hypothesis that the statistical model used to predict the probability of riskiness of a dealer is better than the baseline model (no model) and provides significant benefit compared to random choice

Conclusion & Recommendation:

This paper has showcased how the tax authorities can move into the predictive mode rather than post audit reactive mode for zeroing on the probable risky dealers in the indirect tax domain for highest impact of revenue recovery and collection through a statistical, scientific and information driven model for decision making. A statistical and scientific model used to predict the probability of riskiness of a dealer is better than the baseline model (no model) and provides significant benefit compared to random choice . A figure of 18% more classification (68% for the model vs 50% randomly chosen) is a significant lift in terms

of intervening at the transaction time or pre-audit for highest impact of tax revenue collection and recovery. Also all the variables are not equally significant in terms of discriminating between tax evader and non-evader and prioritization should be incorporated for the significant variables.

The recommendation should be moving into the predictive mode rather than post audit reactive mode for zeroing on the probable risky dealers for highest impact of revenue recovery and collection and a statistical, scientific and information driven model for decision making is the right direction to it.

Bibliography:

• Bryman, A (2006) Integrating quantitative and qualitative research: how is it done?' Qualitative research, Vol.6, No. 1, pp. 97 – 113 Sage

• Creswell John W., (2003) Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 2nd Ed. Sage Publications.

• Creswell. J. W. & Miller. D. L. (2000): Determining validity in qualitative inquiry. Theory into Practice, 39(3), 124-131.

• Datar P. Arvind, "Why the Code must be shelved", Business Line , 8th May 2010, p. 9.

• Government of India (2006).Towards Faster and more Inclusive Growth - An Approach to the 11th Five Year Plan. Planning Commission.

• Kelkar, V. (2010).Financial Inclusion for Inclusive Growth. ASCI Journal of Management, 39 (1), 55-68.

• Kumar, Nagar and Samanta, "Performance of Income Tax Department: An Appraisal",

Indian Management Studies Journal , Vol. 9, 2007.

- Philip Genschel and peter schwarz ((2011),Oxford University Press.

- Rangarajan Committee (2008). Report of the Committee on Financial Inclusion, Government of India.

- Reserve Bank of India (2011).Report of Trend and Progress of Banking in India 2000-11, Mumbai.

- Saunders, M, Lewis, P Thornhill, A, 2007; Research methods for business students, 4th Edition, Prentice Hall

- Singh, Jaspal & Sharma, Poonam, "Tax Professionals Perception of the Income Tax System of India an Empirical Evidence", the ICFAI Journal of Public Finance, Volume 5, No. 1, February 2007, pp. 45-56