सत्यमेव जयते

Tax Administration Reform in India
Spirit, Purpose and Empowerment

(Fourth Report of TARC)

Tax Administration Reform Commission

Ministry of Finance, Government of India

February 2015

सत्यमेव जयते

Tax Administration Reform in India
Spirit, Purpose and Empowerment

Fourth Report of the
Tax Administration Reform Commission
Ministry of Finance, Government of India

New Delhi
February 2015

सत्यमेव जयते

Govt. of India
Ministry of Finance
Tax Administration Reform Commission
NBCC Plaza, 3rd Floor, Pushp Vihar, Saket, New Delhi-110017

F. No. TARC/Report/36/2014-15

Date 20/02/2015

To

Shri Arun Jaitley
Hon'ble Minister of Finance
Government of India

Sir,

We submit herewith the final, Fourth Report of the Tax Administration Reform Commission (TARC).

Dr. Parthasarathi Shome
Chairman

Y. G. Parande
Member

Sunita Kaila
Member

M. K. Zutshi
Member

S. S. N. Moorthy
Member

S. Mahalingam
Member

M. R. Diwakar
Member

Table of Contents

Preface

This volume comprises the TARC's fourth and last report. It covers the last three of its twelve terms of reference. They are:

- *To review the existing mechanism and recommend appropriate means including staff resources for forecasting, analyzing and monitoring of revenue targets.*
- *To review the existing policy and recommend measures for research inputs to tax governance.*
- *To review the existing mechanism and recommend measures to enhance predictive analysis to detect and prevent tax/economic offences.*

Perhaps these final chapters comprise the most frontier areas that are being addressed by modernising tax administrations. Many of them use models of revenue forecasting on a regular basis and it is important for the Indian tax administration benchmark itself against international practice and emulate best practices. The report first addresses revenue forecasting models in all their possibilities in Chapter XIII. Second, more esoteric perhaps is the use of predictive analysis that enables advanced tax administrations to narrow in on tax evaders while assuring the good tax payer that they would not be unnecessarily touched by the tax administration for routine examination. Indeed, predictive analysis under prevailing use has progressed to the use of complex models that are applied in fields such as prevention of frauds, more effective tax debt collection and improved customer service. Its use has resulted in what appears to be precision outcomes in selecting scrutiny and audit cases. Indeed, use of extensive analytics and modelling now underpins all key tax administartions' actions. These models are highlighted in some detail in Chapter XIV. Third, following its final term of reference, in Chapter XV the report undertakes a *tour de force* of many research areas that the Indian tax administration should take up in line with what tax administrations in advanced as well as emerging economies are already undertaking.

Apart from the TARC itself, a number of persons contributed to the successful completion of the chapters that have attempted to cover the most modern areas of tax administration analysis. They included members of the Focus Groups listed in Appendix II as well as private sector ICT based research organisations that came forward to inform the TARC with their methodologies and engagements in those areas with tax administrations across the world. The TARC deeply appreciates such rich contributions. The TARC appreciates the support it received in assistance to contribute to selected fields by its staff and research consultants.

Dr. Parthasarathi Shome

Chairman

Tax Administration Reform Commission

New Delhi

20th February 2015

Tax Administration Reform Commission

Chairman

Dr. Parthasarathi Shome	Level of Minister of State

Members - full time

Y. G. Parande	Ex-Member, CBEC
Sunita Kaila	Ex-Member, CBDT

Members - part time

M.K. Zutshi	Ex-Chairman, CBEC
S.S.N. Moorthy	Ex-Chairman, CBDT
S. Mahalingam	Ex-Chief Financial Officer and Executive Director, TCS
M.R. Diwakar	Ex-Vice President, Taxation, Murugappa Group

Secretariat

Sanjay Kumar	Secretary
Sukalyan Banerjee	Under Secretary

Research Consultants

Md. Tarique Hasan Khan

Rohitash Chaudhary

Administrative Support

Tejpal	Consultant (Administration)
Deepak Chauhan	Consultant (Administration)
Jagdish	Assistant
Ankur Verma	Steno/Assistant

Tables

Diagrams

Graphs and Boxes

List of Annexures

Glossary of Technical Terms

ABS	Australian Bureau of Statistics
ABT	Analytical Base Table
ACE	Analytical Compliance Environment
ACES	Automation of Central Excise and Service Tax
AID	Automatic Interaction Detection
AIIR	Annual Investment Income Reports
AIR	Annual Information Return
ARIMA	Autoregressive Integrated Moving Average
ARM	Additional Resource Mobilisation
ATO	Australian Taxation Office
BE	Budget Estimates
BI	Business Intelligence
BMF	Business Master File
BRC	Budget Responsibility Committee
BSA	Bank Secrecy Act
CART	Classification and Regression Trees
CASS	Computer Aided Selection System
CBDT	Central Board of Direct Taxes
CBEC	Central Board of Excise and Customs
CBI	Central Bureau of Investigation
CBO	Congressional Budget Office
CCRs	Compulsory Compliance Requirements
CENVAT	Central Value Added Tax
CGE	Computable General Equilibrium
CIB	Central Information Branch

CIBIL	Credit Information Bureau (India) Limited
CIO	Chielf Information Officer
CIT	Corporate Income Tax
CMC	Compliance Management Cell
COIN	Custom Overseas Intelligence Network
CPC	Central Processing Centre
CPC-TDS	Central Processing Centre for Tax Deducted at Source
CRA	Canada Revenue Agency
CRISP-DM	Cross-Industry Standard Process for Data Mining
CSO	Central Statistical Organisation
DEA	Department of Economic Affairs
DES	Discrete Event Simulation
DGCEI	Directorate General of Central Excise Intelligence
DGFT	Directorate General of Foreign Trade
DGoV	Directorate General of Valuation
DGRI	Director General of Revenue Intelligence
DMSC	Data Matching Steering Committee
DODM	Directorate of Data Management
DW & BI	Data Warehouse and Business Intelligence
DWP	Department of Work & Pensions
EDW	Enterprise Data Warehouse
EFDS	Electronic Fraud Detection System
EU	European Union
FDC	Fraud Detection Centers
FET-ERS	Foreign Exchange Transaction-Eclectronic Reporting System
FinCEN	Financial Crimes and Enforcement Network
FIU	Financial Intelligence Unit

FPU	Fiscal Policy Unit
FRBM	Fiscal Responsibility and Budget Management
FTA	Free Trade Agreement
GDP	Gross Domestic Product
GOV	Government Departments/Agencies
GST	Goods and Services Tax
HHS	Household Survey
HMRC	Her Majesty's Revenue and Customs
HMT	Her Majesty's Treasury
HNWI	High Net Worth Individual
HUF	Hindu Undivided Family
ICE	Integrated Compliance Environment
ICEGATE	Indian Customs and Excise Gateway
ICES	Indian Customs EDI System
ICIJ	International Consortium of Investigative Journalists
ICRIER	Indian Council for Research on International Economic Relationship
ICT	Information and Communication Technology
IEO	Independent Evaluation Office
IMF	International Monetary Fund
I-O	Input-Output Table
IRD	Inland Revenue Department
IRDA	Insurance Regulatory and Development Authority
IRS	Indian Revenue Service
IRS	Inland Revenue Service
IRS-CI	Internal Revenue Service-Criminal Investigation
IS	Industrial Survey
ISS	Indian Statistical Services

ISTAT	Italian Institute of Statistics
ITD	Income Tax Department
ITR	Income Tax Returns
ITS	Individual Transaction Statement
IVR	Interactive Voice Response
JEFG	Joint Economic Forecasting Group
KAI	Knowledge, Analysis and Intelligence
KAIC	Knowledge, Analysis and Intelligence Centre
LDC	Lead Development Centers
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MCA	Ministry of Corporate Affairs
MIMIC	Multiple Indicators and Multiple Cause
MoUs	Memoranda of Understanding
MPE	Mean Percentage Error
MSE	Mean Squared Error
MVUE	Minimum Variance Unbiased Estimator
NACEN	National Academy of Customs, Excise and Narcotics
NADT	National Academy of Direct Taxes
NCAER	National Council of Applied Economic Research
NCB	Narcotics Control Bureau
NIA	National Income Accounts
NIFM	National Institute of Financial Management
NIPFP	National Institute of Public Finance and Policy
NMS	Non-filers Monitoring System
NN	Neural Networks
NSDL	National Securities Depositories Limited

NTC	National Targeting Centre
OBR	Office for Budget Responsibility
OECD	Organisation for Economic Co-operation and Development
OLTAS	On-Line Tax Accounting System
OLTP	Online Transaction Processing
ONS	Office for National Statistics
OTA	Office of Tax Analysis
PAN	Permanent Account Number
PAYE	Pay as Your Earn
PCA	Post-clearance Audit
PHA	Policyholders' Account
PIT	Personal Income Tax
PMS	Performance Management System
PPAC	Petroleum Planning and Analysis Cell
PTM	Personal Tax Model
R&D	Research and Development
RAU	Revenue Analysis Unit
RBA	Reserve Bank of Australia
RBI	Reserve Bank of India
RDH	Risky Descriptions
RE	Revised Estimate
RGDP	Real Gross Domestic Product
RMD	Risk Management Division
RMS	Risk Management System
RMSE	Root-Mean-Square Error
RMSQ	Root Mean Square Error
ROC	Registrar of Companies

ROC	Receiver Operating Characteristic
RP	Relative Prices
RPPD	Research, Policy and Planning Department
RRP	Return Review Programme
SAS	Statistical Analysis System
SCI	ISTAT Business Survey Data
SD	System Dynamics
SEMMA	Sample, Explore, Modify, Model, & Assess
SFIO	Serious Froud Investigation Office
SHA	Shareholders' Account
SKAT	Danish Tax Administration
SME	Small and Medium Enterprises
SMS	Smart Message Service
SOPs	Standard Operating Procedures
SPV	Special Purpose Vehicle
SQL	Structured Query Language
SSI	Small Scale Industry
STA	Swedish Tax Agency
STR	Suspicious Transaction Reports
SUNAT	Superintendencia Nacional de Administración Tributaria
TAN	TDS Account Number
TAR	Tax Arrear Recovery
TARC	Tax Administration Reform Commission
TARF	Tax Analysis and Revenue Forecasting
TAX	Tax Department
TCS	Tax Collected at Source
TDS	Tax Deduction at Source

TEP	Tax Evasion Petitions
TIL	Theil Inequality Statistics
TPA	Tax Policy and Analysis
TPB	Tax Policy Branch
TPDS	Third-party data store
TPL	Tax Policy and Legislation
TR	Real Tariff Revenues
TRU	Tax Research Unit
USD	US Doller
VAR	Vector Auto-Regression
VAT	Value Added Tax
V-RAM	Valuation Risk Assessment Module
VTTL	VAT Theoretical Total Liability
WCW	Wealth Conversion Weight
Web-CBRS	Web Currency & Banking Retrieval System

Chapter XIII

Revenue Forecasting

Chapter XIII

Revenue Forecasting

List of contents

Chapter XIII

Revenue Forecasting

XIII.1 Need for revenue forecasting

Objective, accurate and timely analysis of the budget, inter-governmental finance, revenue and expenditure trends are essential for the smooth functioning of a government. Arriving at accurate estimates of tax collections and expenditure estimates is necessary from the budgetary point of view. Together, the estimates indicate the likely deficit that will have to be financed through public borrowing, seigniorage, cutting expenditures or through other deficit financing measures. Making an accurate and timely projection of the likely tax revenues through revenue forecasting is, thus, paramount in setting as well as in achieving public finance policy goals. This assists the government in managing and aligning expenditure and revenue flows. When revenue targets are not met, expenditure programmes can get undermined, which leaves the government with two options – either to contain fiscal expenditures or to allow the budget deficit to expand by the amount of revenue shortfall. Either option endangers fiscal discipline and reduces the allocative efficiency of the government budget. The imposition of expenditure controls to put a cap on the fiscal deficit weakens the planning and budgeting linkage and puts expenditure programme at the mercy of politics. The ability to project accurately future tax resources is thus critical to avoid budgetary shortfalls. Small errors in projecting revenue can result in serious budget problems such as large surpluses or deficits, making revenue forecasting fundamental to governments at all levels. The Fiscal Responsibility and Budget Management Act (FRBM Act), 2003, has brought to the centre-stage the need for revenue forecasting to ensure long-term macroeconomic stability. Section 7 of the FRBM Act enjoins the Finance Ministry to report to Parliament progress in revenue collections and the finances of the country on a quarterly basis.

The revenue forecasting model also serves to assess, on a continuing basis, the revenue impact of changes in tax and non-tax policies (for example, trade policies, industrial policies and business/financial regulations) and other economic and structural changes. This assessment is not only required at the ex-ante stage, i.e., at the time of making the budget, but also during mid-fiscal year evaluation. Regular evaluations contribute to a more precise and efficient policy-making process, improving the quality of government planning and policy formulation by evaluating the impact of proposed fiscal changes on the economy and of other economic changes or shocks during the year, such as lower economic growth, commodity inflation, changing trade patterns, changes in the global economic environment, drought conditions affecting agriculture, etc., which may be difficult to anticipate at the time of the budget making.

The revenue forecasts help in measuring potential tax revenue that is likely to be collected given the existing tax base and the present tax rate. The actual tax collection represents the extent to which the tax administration is able to realise this potential. The performance of the tax administration, known as the tax effort, can be evaluated on the basis of actual and potential tax collection. This evaluation can be used to establish benchmarks to monitor the performance of the tax administration and to stimulate efforts to identify the economic reasons for the inability to achieved desired tax collections. This exercise, along with detailed tax analysis, can help identify ways to expand the tax base under existing tax legislations and improve tax administration.

XIII.2 Methods of forecasting, analysing and monitoring revenue

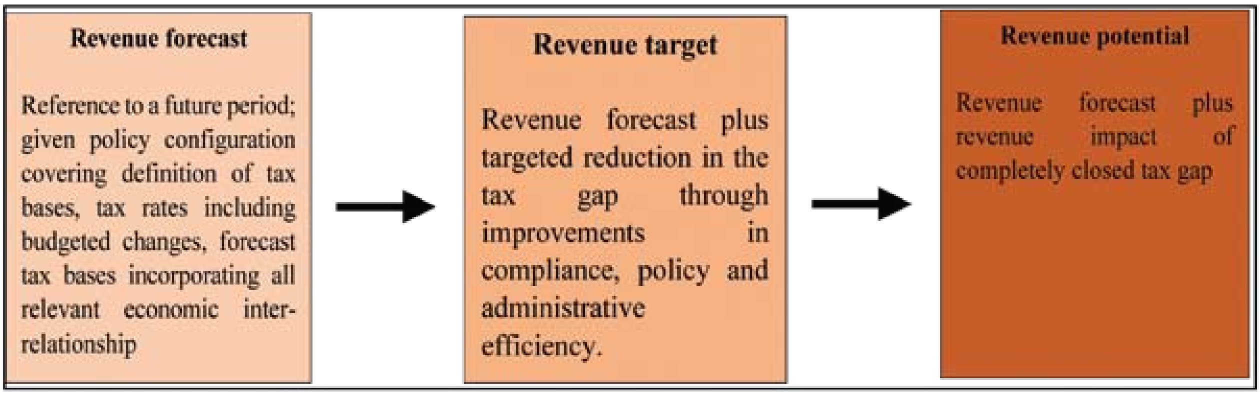
Revenue forecast, revenue potential, and revenue target are three distinct, although closely related concepts. Revenue forecast is the forecast tax revenue for a given tax policy structure. Revenue target can be equal to or more than the revenue forecast as it may aim to cover partially the tax gap, which could be due to compliance, policy, and administrative efficiency gaps. Revenue potential is usually higher than both the revenue forecast and the revenue target as it estimates potential tax revenue after fully closing the tax gap, i.e., with 100 per cent compliance, policy parameters (rates, exemptions, etc.) suitably benchmarked against other comparable systems (countries or states), and administrative efficiency kept at the highest attainable. Thus,

Revenue potential ≥ Revenue target ≥ Revenue forecast

As the excess of tax revenue potential over actual tax revenue is called the tax gap, it usually refers to a past period for which actual tax revenue data are already available. If it is estimated for a future period and the comparison is made with a reliable tax revenue forecast, we would have an 'estimated tax gap'.

We can then conceptually understand the relationship between these three concepts more closely from the following flow chart.

Diagram 13.1: Revenue potential, revenue target, and revenue forecast



The estimation of one or more of these concepts may be, and often are, undertaken by tax departments, the policy wing of the broader government (Ministry of Finance/Department of

Economic Affairs), and bodies of independent researchers, particularly those concerned with fiscal and macroeconomic policy, to serve different purposes.

A revenue forecast takes into account all available information at the time of making the forecast including proposed budget changes, changes outside the budget regarding the tax base, and statutory provisions including those relating to tax rates, exemptions and coverage of tax bases. A revenue forecast would refer to a specific period or periods and estimate the revenues that are likely to be raised given the current macroeconomic conditions and compliance levels. It also takes into account the revenue impact of any policy changes being undertaken with reference to the forecast period. Revenue forecasts are ideally derived from a consideration of variables affecting the tax base for a given policy configuration including changes applicable in the forecast period. Revenue forecasts need to take into account effect of policy changes on both revenue and the tax base.

Revenue targets

Tax authorities or governments are often interested in determining revenue targets, which could be based on a desire to improve compliance and administrative efficiency. The first relates to the behaviour of taxpayers and the second to that of the tax authorities. In fixing the revenue target for a given period, it is likely that the authorities may assume that compliance may be improved but not up to hundred per cent. In such a case, the revenue target would be higher than the revenue forecast and lower than the revenue potential. Revenue targeting has to have an underlying revenue forecast on which the revenue impact of administrative and compliance improvements may be added.

Alternatively, revenue forecasts as well targets can be derived from exogenous considerations like deriving the tax revenue target by starting with the fiscal deficit target, plugging in forecasts or estimates of expenditures and non-tax sources (non-tax revenues, borrowing, and non-debt capital receipts) of financing that expenditure. In such a case, the revenue forecast is actually a residual and may not be linked directly to the movement of tax bases.

In a developing economy that starts out with a low tax-GDP ratio, revenue targeting may serve as a policy tool to increase the tax-GDP ratio to a desired long-term target. Such a target may itself be benchmarked against tax-GDP ratios of other similarly placed countries or derived from a view, based on ideological concerns, on the role of the government/public sector in the economy.

In India, for example, when the era of planning began, the tax-GDP ratio was quite low. The idea of ‘additional resource mobilisation’ was used to increase the tax-GDP ratio by widening the tax base, increasing tax rates, and reducing compliance gaps.

Revenue targets may be fixed above revenue forecasts to reduce the compliance gap. After increasing the tax-GDP ratio during the period 1950-51 to 1989-90 from about 6 per cent to close to 16 per cent, the tax-GDP ratio in India, considering the central and state governments together, has stagnated, with some variations, in the range of 16-17 per cent for nearly 25 years implying that there has be no effective additional resource mobilisation (ARM) from the tax side during the period.

Since the focus of revenue targets is mainly on administrative efficiency, the aggregate revenue target should be decomposed into the relevant jurisdictions, sector-wise or area-wise. In this process, a considerable amount of disaggregated data relating to sectoral, state-wise, area-wise, and commodity-wise growth of real activities and prices may be fruitfully employed.

If the disaggregated targets are independently derived, they must maintain consistency conditions with the aggregate target. In practice, disaggregated targets hardly ever maintain consistency with the aggregate revenue target and are usually based either on rules of thumb such as a 5 per cent or 10 per cent growth, or are based on a trend growth rate. These disaggregated targets can go way out of line whenever the economy is going through strong cyclical movements. If the disaggregated revenue targets have no link with the aggregate target and are not adjusted in response to changes in economic conditions, it can lead to highly distorted assessments of administrative performance by revenue officials.

Linking administrative efficiency to meeting or missing revenue targets can lead to highly distorted decisions regarding penalising or rewarding tax officials. At the macro level, missing budgeted revenue forecasts can lead to missing out on the FRBMA targets of fiscal deficit, revenue deficit and debt. On the one hand, this may lead to unduly aggressive behaviour of tax officials and on the other, it may lead to undesirable short-term changes in tax policies such as hiking tax rates and withdrawing otherwise desirable exemptions. When the realisation of revenue shortfall comes in the fiscal year, the government often resorts to expenditure cuts, which has deleterious effects on growth and, therefore, the tax base, creating a vicious circle.

The key improvement in prevailing practices relating to revenue targeting in India must relate to ensuring (a) consistency of aggregate revenue target and the sum of disaggregated revenue targets and (b) creating a mechanism that ensures that if a revision of the aggregate revenue target is undertaken within a financial year because of changed economic conditions or policy changes, corresponding changes are incorporated in the disaggregated revenue targets.

Determining revenue potential

Revenue potential for the same period is likely to be higher than the revenue forecast. As long as compliance levels are not hundred per cent, there will always be a positive tax gap. Usually, the tax gap is estimated for past periods, in which case it is defined as the excess of revenue potential over actual revenue. If the revenue potential is estimated for a future period, it will be the forecast revenue potential, and the comparison would be with the estimated revenue potential and forecast revenue.

Potential tax revenue is the theoretical tax liability as per the tax law, the rate structure, and full compliance. Tax gaps can be decomposed between policy gap, compliance gap, and administrative efficiency gap. The compliance gap refers to the behaviour of taxpayers and administrative efficiency gap refers to tax collectors. The policy gap is the result of using lower than standard rate for some categories, compared to suitable benchmarks (other countries, other states depending on the context), giving exemptions, etc. The second gap arises due to non-compliance. The third gap arises due to administrative inefficiency. Usually, all the three gaps are highly interdependent, and it is often difficult to precisely decompose a tax gap into its constituent gaps.

Two broad strategies for estimating the tax gap are the (a) top-down approach and (b) bottom-up approach. In the former, the tax base is constructed using macro data, such as national income data, supplemented by other data compiled independently of the tax departments. In the bottom-up approach, micro-level data from tax departments and field offices and specific surveys are used to obtain direct estimates of revenue gaps. Benchmarking summary measures of tax gaps with comparable countries offers valuable insights into whether comparatively high tax gaps exist because of design deficiencies, lax administrative implementation, taxpayer behaviour or policy inadequacies.

In India, in spite of the existence of large compliance and administrative gaps, hardly any attempt has been made to measure the tax gap either at the central level or in the states for any of the major direct and indirect taxes. Such an exercise, undertaken on a regular basis by the Ministry of Finance at the centre and finance departments in states, can be a valuable guide to augmenting revenues, designing tax reforms, simplifying tax structures, reducing compliance costs, and improving the efficiency of tax administration.

In all three cases, aggregate and disaggregated approaches and bottom-up and top-down approaches are possible. The level of aggregation may refer both to the variable that is to be forecast and the methodology that is to be used. For example, the variable to be forecast may be the aggregate of all tax revenues under the control of the central government. Or it may be forecast using a highly macro approach such as using buoyancy and GDP at current prices or using different levels of disaggregation by dividing the overall tax base into components and sectors. Disaggregation is always a matter of degree. These are complementary and not mutually exclusive approaches. The complementarity of the two approaches can be used to the government's and tax authorities' advantage since following the two approaches simultaneously increases the information base of both tax policy and fiscal policy. The aggregate approach gives a handle to tax and fiscal authorities to endogenise the effects of cyclical changes in the economy on tax revenues and provides a mechanism to incorporate the effect of changes in tax revenues on important macro variables in the economy like consumption and investment.

The difference between the top-down and bottom-up approaches is similar to the differences between the aggregated approach and disaggregated approach. It also may refer to both the variable under study and the methodology.

Disaggregated and sectorally detailed approaches facilitate incorporation of institutional factors and sector and area details, both administrative and concerning the tax base, which may not be possible for a macro approach to incorporate. A tax authority that develops ways in which the complementarity of the two approaches is used is likely to be more successful in designing and running an effective tax system.

Linking revenue targeting to revenue forecasting

Revenue targets must be closely linked to revenue forecasts and should be revised as the forecasts are revised. The frequency of such revisions can be decided but often such revisions are important within a financial year if the original forecasts go out of line because of unanticipated exogenous events such as crude oil price shocks. Revenue forecasts may also need to be revised due to policy interventions within a financial year.

Revenue forecasting should be couched in model frameworks that provide responses to changes in the macroeconomic environment and are detailed enough to enable estimation of the revenue impact of fiscal policy actions. Fiscal policy actions can be at the macro-level, including variations in the fiscal deficit, and tax-specific such as increasing a tax rate or exemption limit.

Since macro-econometric modelling is a highly specialised exercise, it is best for the tax department or the Ministry of Finance to develop institutional collaboration in this matter, possibly with more than one institution, to obtain a broad-based understanding of the macroeconomic environment and the impact of contemplated fiscal changes, including the revenue impact of changes in government expenditures.

Supplementing model based revenue forecasts with tax specific revenue models

For important central taxes such as personal income tax, corporate income tax, and domestic indirect taxes and import duties, tax specific models that utilise the vast amount of information on taxpayers, transactions, and tax bases should be developed. These models could use more detailed and micro-level information relating to individuals, firms, and transactions relating to sale and purchase in the domestic and international markets.

The personal income tax model should be designed to pick up the links between revenues and distribution of personal incomes, age, sex, marital status and any other relevant features in respect of which differential tax treatment may be given, including exemptions and deductible allowances. This model would also be able to highlight summary measures like the overall progressivity of the personal income tax system and the revenue implication of exemptions, etc., i.e., tax expenditures.

The corporate income tax micro model should be designed to reflect specific provisions regarding depreciation allowances, provisions for net operating loss carry-overs, rules for computing and claiming tax credits, minimum corporate income tax provisions, interest deductions, and tax credit limitations. This model should also facilitate computation of tax expenditure in the case of corporate income tax.

The domestic indirect taxes micro model should be designed to capture estimation of input taxes that are to be credited in subsequent stages. It can provide an estimate of expenditures according to different categories including private final consumption expenditure, government final consumption expenditure, investment expenditure and exports at a suitably disaggregated level consistent with available data, including input-output tables. The model would also estimate the revenue cost of zero-rating exports. While the macro approach can provide one estimate, the micro approach would provide comparable and consistent estimates from a sectorally disaggregated perspective.

In this section, different methods for tax revenue forecasting have been explained. While forecasting tax revenues, tax policy units, i.e., Tax Policy and Legislation (TPL) in the CBDT (Central Board of Direct Taxes) and Tax Research Unit (TRU) in the CBEC (Central Board of Excise and Customs), are also called upon to analyse the impact of tax policies on different entities in the economy – taxpayers, businesses and government – and to estimate the revenue implications of tax measures on these entities to ensure a healthy fiscal situation within the economy. Methods to analyse tax policy in the context of forecasting, such as the distributional impact on taxpayers, impact on different businesses, change in investment climate, has been discussed in the second part of the section. Methods to monitor tax collections, a crucial part of the budget exercise, have been discussed in the following section. Monitoring tax revenues is particularly important for ensuring regular flow of revenues as this affects treasury management.

XIII.2.a Revenue forecasting methods

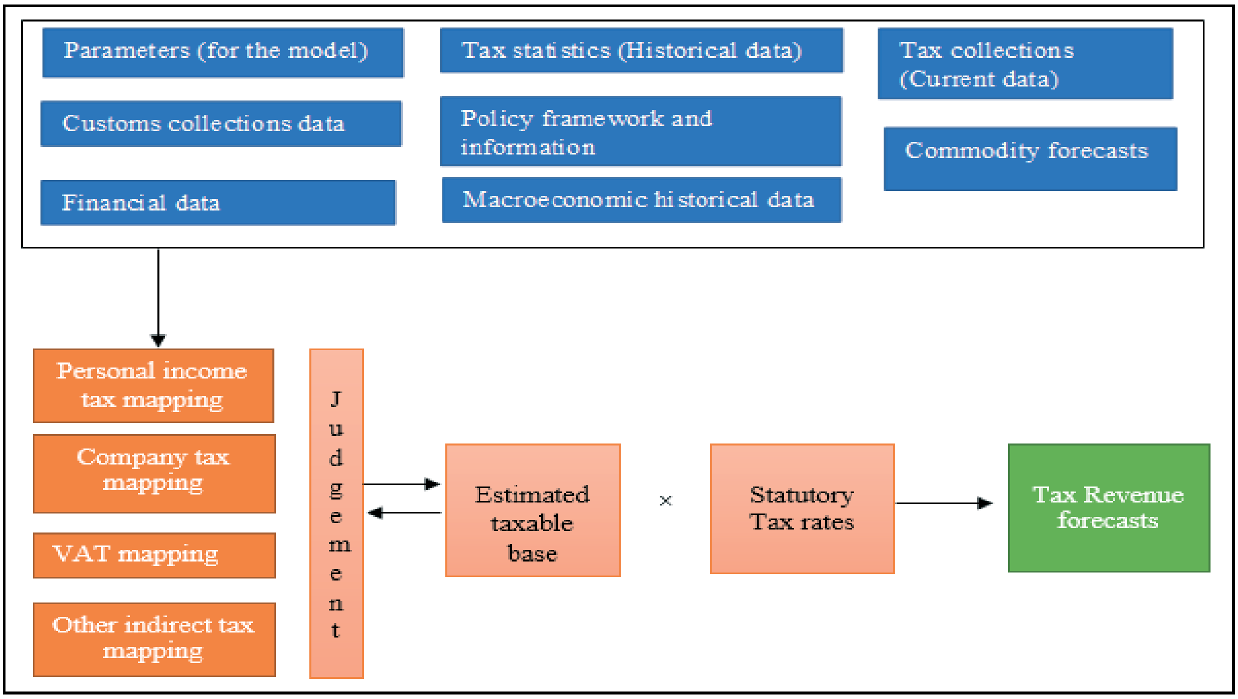
Different tax forecasting models are employed by tax administrations, depending on data availability, ambition level in terms of the accuracy of projections, and resources employed. Any model employed to forecast tax revenue needs to take into account the underlying interdependencies of the sectors of the national economy. These sectors – primary, secondary and tertiary sectors – contribute in an inter-woven manner to sale, income and taxes. For example, a decrease in the tariff on goods can increase the consumption of goods. This will spur consumer demand for the product and, as a result, the base for taxes on domestic production and sales can expand. The unmet demand for the product will have to be supplemented by imports, expanding the import tax base and increasing revenue from customs. An increase in domestic production and sales will increase companies' profit (the base for corporate income tax). The increase in companies' profits may have a positive impact on wages and salaries, which in turn will increase the demand for consumption and imported goods. But this will also result in an increase in interest rates, bringing all the sectors of the economy into play.

While attempting to model each of these activities in a comprehensive forecast model may be complicated and cumbersome, even though desirable,[326] tax administrations often use the basic understanding that tax revenues are a function of the tax base, tax rate and compliance ratio, expressed as $T \equiv \phi(x, y, \eta)$, where T is tax revenues, x be the tax base, y the tax rate and η the tax compliance ratio. This equation is at the base of forecasting methods. But, the techniques for forecasting revenues can range from subjective judgment/expert opinion to asking taxpayers, who are better informed about their taxpaying capacity, to methods that are sophisticated, requiring econometric or computer-based

326 The Computable General Equilibrium (CGE) model does that. This has been described later in the chapter.

micro-simulation analysis; more sophisticated models may lead to gains in policy analysis. Sophisticated models also require detailed data from sources such as national accounts, tax returns, or household/industrial surveys. Diagram 13.2 shows pictorially the basic model and data requirement for revenue forecasting.

Diagram 13.2: Forecasting tax revenues



Source: International forecasting practices, ATO

These forecasting methods are briefly described below:

- **Qualitative method** – In the past, it was common for tax departments to rely on a few experts to provide their judgments on revenue forecasting. These experts gain considerable working experience over a period of time and, based on some historical method and experience, forecast tax collections. But such methods are not transparent and are virtually impossible to replicate. In this approach, normally, work is not documented properly and there is no systematic way to pass on the method to new staff. This means that no training programme for new staff can be organised. By the very nature of this approach, there can also be inconsistency in the method and judgment over a period of time and the approach may not work if underlying economic conditions are changing too rapidly. It is for these reasons that most tax administrations look for the standard tax forecasting methods described below.

- **Quantitative methods** – Quantitative methods rely on numerical data, with explicit assumptions and procedures for generating tax revenue forecasts. There are many techniques for revenue forecasting that are being increasingly used by tax administrations. These methods are the following:

- "*Unconditional*" *time series analysis* – This includes trends and growth factor analysis. The method is called "unconditional" because the forecasts are based on past revenue data only. Econometric techniques are used for forecasts.

- "*Conditional*"/*causal models* – This method recognises the causal relationship between the tax base (the cause) and tax receipts (the effect). This relationship uses information from "the cause" (usually, tax bases) to explain the growth or shortfall in tax revenues. Due to the direct relationship, it becomes necessary to have forecasts of data on "the cause" included in the model. The method uses a large variety of methodologies, such as effective tax rate/tax ratio analysis and elasticity analysis, for forecasting. These models rely on least-square regression techniques.

- *Microeconomic analysis* – This method is most appropriate for estimating revenues from commodity taxes, such as excise taxes, or from trade taxes, such as import duties. This method can also be used to estimate the impact of discretionary changes in tax revenues and tax bases. These models often require the estimation of demand and supply elasticity.

- *Structural* (*input-output*) *analysis* – This method is most useful for analysing broad-based indirect taxes, such as general sales tax and value-added tax. It requires a detailed breakdown of national accounts to estimate the tax base in various sectors in the economy.

- *Micro-simulation analysis* –The focus in micro-simulation analysis is on the actions or behaviour of individual sectors and how they are affected by policy change. The advantage of micro-simulation modelling lies in its capacity to estimate the distributional effect of a given policy proposal on particular sectors of the population and identify the winners and losers. This method requires detailed information from tax return databases, and sometimes through surveys for specific purposes. The micro-simulation models are more appropriate for income tax, but their application can also be expanded to general sales tax/value-added taxes and import taxes.

- *Other specific approaches* – Some types of taxes, such as tax expenditure, property taxes and taxes on natural resources, may require other specific models not mentioned above.

The models described above have been ordered in terms of the increasing complexity of their data requirements – micro-simulation models require very detailed data while time series analysis requires less detailed data. The first two quantitative approaches mentioned above – time series analysis and causal models – are more useful for revenue forecasting, while the remaining four are more appropriate for tax analysis and revenue estimation. All these are elaborated below.

(i) Times-series forecasting

Tax administrations employ time-series models for projecting tax revenues. Extrapolating an established linear trend in tax receipts is often considered a straightforward method for forecasting. This method requires a set of values that a variable takes at regular time intervals. Time series models essentially involve recognising the patterns associated with past values in a data series and projecting future values based on a past trend. The trend can be a pattern of increase or decrease in a straight line or a curve. The method employed for time-series forecasting will depend on recognising the trend – whether it is a straight line or a curve. The trend can be visually depicted on a graph. But these trends can also have cyclicality, which result due to business cycles that affect tax revenues. Seasonality, a pattern of increases and decreases, is another cyclic phenomenon, typically observed in monthly or quarterly observations. But data can also be random, not following any trend. Randomness occurs due to unexpected events that distort the long-term trend. If these patterns can be identified and projected into the future, a time-series forecast can be made.

Based on different patterns, whether it is a simple, cyclical or seasonal trend, different time-series models can be employed. These time-series models can be divided into two broad categories – open-model time series models and fixed-model time series models – based on their suitability. Open-model time series models analyse the time series to determine the trend pattern and then build a unique model to project the pattern into the future to forecast tax revenues. This is in contrast to the fixed-model time series model, which is a fixed equation model, based on the *a priori* assumption that certain patterns do or do not exist in the data. Most time series models, which are more than 60, fall broadly into these two categories. Among the two, fixed-model time series models are simple and inexpensive to use and require little data. The basic technique used in this model is averaging and includes moving through the levels of moving average, exponential smoothing, adaptive smoothing, and incorporating trend and seasonality.

Moving average models are probably the most commonly used time series models. In this model, the future value to be forecast is based on the average of previous periods. It is a moving average because the oldest data points are dropped off as new ones are added. Thus, the model uses some of the more recent data, rather than use all the previous data in the calculation. To explain it further, the moving average time-series model uses the following formula:

$$F_{t+1} = (S_t + S_{t-1} + S_{t-2} + \ldots + S_{t-N-1})/N,$$

In the above equation, F_{t+1} is forecast for the period t + 1, and S_{t-1} is tax collection for the period t – 1 and N is the number of periods in the moving average.

The difficulty with a moving average is in deciding how many periods to use in the forecast. The more periods used, the more it starts to look like an average, and fewer periods used can result in giving weightage to higher values.

The exponential smoothing model is another forecasting technique, based on the moving average forecasting method. This model builds itself on iteration, correcting the error observed in preceding forecasts. In the first degree of the smoothing model, it is assumed that there is an observable trend or seasonal pattern. The model is given below:

$$F_t = F_{t-1} + \alpha\,(A_{t-1} - F_{t-1})$$

where, F_t is the forecast at time t, and A_{t-1} is the actual value at time $t-1$. The parameter $\boldsymbol{\alpha}$ is the smoothing coefficient and has an estimated value of between zero and one.[327] It is referred to as the exponential smoothing model because the value tends to affect past values exponentially. Thus, the forecast for the next period is a function of the last period's forecast and includes the observable value of the last period. If the data has seasonality, simple smoothing, as explained above, will have to be changed to include the additional dimension of seasonality smoothing. Thus, the series will have two components – trend analysis and seasonality analysis. This is best done using Holt's model. The Box-Jenkins ARIMA (Autoregressive Integrated Moving Average) model is also used for the purpose of removing seasonality.[328] In the ARIMA model, the projection is based on a linear combination of previous period values, as shown below:

$$F_t = \alpha_1 A_{t-1} + \alpha_2 A_{t-2} + \cdots + \alpha_p A_{t-p}$$

In the above equation, F_t is the required projection value at time t, A_{t-p} is the actual value at time t, and α_p's are the parameters to be estimated. The other part in the ARIMA model is the moving average component, providing forecasts based on prior forecasting errors. This part of the model can be shown as:

$$F_t = \beta_1 \epsilon_t + \beta_2 \varepsilon_{t-1} + \cdots + \beta_p \epsilon_{t-p}$$

where, F_t is the predicted value at time t, ε_{t-p} is the forecast error at time t, and the β_p's are another set of parameters to be estimated.

Normally, the ARIMA model will require a minimum of 50 data points to have meaningful results. The process of setting up the model has three stages. In the first stage – the model identification stage – it needs to be decided whether the time series will use an autoregressive model or a moving average model, or both. This is usually done by visually inspecting the data plot on a graph. Statistical techniques for checking auto-correlation or partial auto-correlation are employed at this stage. In the second stage, model estimation and diagnostic checks are carried out. During this stage, the model is tested to find out whether the choice of method, auto-correlation or partial auto-correlation, was correct. This is determined by a set of statistical figures from the exercise. If the method employed is correct, the forecasting, being the third stage, is done.

---

[327] The value of $\boldsymbol{\alpha}$ can be calculated by running a least-square regression between F_t and A_{t-1}.

[328] The vector auto-regression (VAR) model is one of the most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. The VAR model has proven to be especially useful to describe the dynamic behaviour of economic and financial time series and for forecasting. This model also does not require much knowledge of the factors influencing a variable. The only prior knowledge required is a list of variables that can be hypothesised to affect each other temporally.

Whatever the method employed for time-series forecasts, it remains to be "best" fitting past values of tax receipts[329] and extrapolating the trend, thus observed, to find future values. Time-series forecasting does not involve any knowledge of the tax system or the relationship between tax revenues and other economic variables; it is dependent only on the tax revenues observed in the past. So, any change in the tax structure – tax base or tax rate – does not get fully reflected in the forecasts. Even if a dummy variable is used to capture the change, it does not give very analytical and insightful information about the change. This is why the time-series method for revenue forecasts is not widely used and is considered weak; it is usually used for quick calculations, or sometimes as a standby, when nothing better is available.

(ii) Conditional forecasts or partial-equilibrium model

In conditional forecasts, the tax revenues are conditional on the tax base. Since the statutory tax base is usually too complex to be useful for economic analysis and forecasting, it is not normally used; instead, an alternative or proxy tax base is used for forecasting purposes. For example, GDP may be a proxy for personal income, assuming that household income grows at the same rate as GDP; aggregate consumption can serve as a proxy for the tax base of a value added or sales tax, if VAT is structured as a consumption tax; production could serve as a proxy base for excise duty; imports may be a good proxy for imports subject to tax. Table 13.1summarises the tax base and its proxy tax base.

Table 13.1: Tax base and proxy tax base

Tax base	Proxy tax base
Taxes on income, profits, capital gains	GDP at factor cost
Sales Tax or Value-added tax	Private and government consumption
Excise duty or CENVAT	Industrial production or private and government consumption
Import Duties	Value/volume of imports

Methods used for forecasting are:

- Effective tax rate method
- Buoyancy/elasticity method
- Micro-simulation method

[329] Fitting a "best" line means to plot a line through all the data points in such a manner that the sum of the square of the deviations of the proposed line from the data points is minimum.

Effective tax rate method

The effective tax rate method is based on observed data on tax revenue and movements in the proxy tax base. It does not require detailed knowledge of the tax system and the provisions of tax laws, such as exemptions or deductions. The method involves finding the effective tax rate by taking the ratio of tax revenues to the tax base and forecasting the tax base (based on an appropriate proxy base). Thereafter, the tax revenue is projected by using the basic formula of effective tax rate multiplied by forecast tax base. The effective tax rate is normally different from the statutory tax rate, due to tax exemptions, deductions, and non-compliance. This method assumes an unchanged tax base structure, i.e., no change in the composition of the tax base between taxed and tax-exempt items, an unchanged tax system, i.e., no change in the rate structure, tax rate level, coverage and exemptions, and an unchanged compliance ratio.

Tax buoyancy method

Buoyancy or the elasticity method for tax revenue forecasting uses a marginal approach, as opposed to the effective tax rate method, which is an averaging concept. But that apart, the variables used in this method are the same as in the effective tax rate method – tax revenues and tax base. Being a marginal approach, the method uses the changes in tax revenue and changes in the tax base; buoyancy is defined as the ratio of the percentage change in tax revenue and the percentage change in tax base. These percentage changes in tax revenue or tax base are measured in *real* terms, after adjusting for inflation.

Tax buoyancy is calculated using the revenue and the tax base proxy actually observed, without any adjustment on account of discretionary changes in the tax revenue reported over a number of years. Tax buoyancy is defined as the percentage change in tax revenue ($\%\Delta T$) divided by the percentage change in the tax base ($\%\Delta Y$), as shown below:

$$B = \frac{\%\Delta T}{\%\Delta Y}$$

Typically, GDP is used as the tax base proxy, but other bases are possible, for instance, consumption as the base for sales tax, or imports as the base for tariffs.

The method uses the understanding that growth in tax revenues arises from growth in the tax base and/or changes in effective tax rate. Tax buoyancy, thus, measures the responsiveness of tax revenues to changes in the tax base due to economic activities. This responsiveness also includes any impact due to discretionary changes,[330] either in the tax rate or the tax base. The tax bases used are shown in Table 13.2.

[330] Discretionary changes are the tax statute changes, normally brought out in the Finance Bill every year, to change the tax base, by reducing tax exemptions or deductions or by bringing new statutes, or the tax rate.

Tax buoyancy can vary from year-to-year. Calculating buoyancy for each year and then taking an average has the disadvantage of being influenced by outliers (those away from the general trend) and, in any case, tax buoyancy may vary from year to year. Thus, it is useful to measure buoyancy over a longer period – perhaps, five or ten years at a time. Sometimes, buoyancy is also calculated by using the growth of tax revenue and that of the tax base between the end years. The problem with this calculation is that the result is sensitive to the end years chosen, but it does have the advantage of the data of the intervening years. A variation of this method can also be by taking average of the end years (for example, the average of the first three years of the series, compared with the last three years of the series). Although this method of calculation requires more data, it is less sensitive to the choice of years than the earlier calculation involving only end years. But, none of the methods fully utilise all the data points and the averaging concept in entirety. Tax buoyancy can be calculated[331] using least-square regression analysis. This method is the most elegant method. The regression is of the log of tax revenue on the log of the tax base (GDP). The coefficient of the log of the tax base gives tax buoyancy. Another vairation could be to regress the log of tax revenue on the year to get the average growth rate of tax revenue and to do the same for the tax base and then using the growth rates from these regression equations (coefficients of the independent variable i.e. the year). But this method of calculation is least successful in cases where the coefficients in the regressions are not statistically significant or where the growth rate of the base is very small.

Tax buoyancy is a measure of both the soundness of the tax base and the effectiveness of tax changes in terms of revenue collection. The tax buoyancy number provides a measure to indicate how tax revenue increases with the growth of the economy. If buoyancy is more than one, it would imply that the growth of tax revenue is more than the growth in the economy and this will be in line with the basic goal of any tax system, which is that tax revenues grow at least at the same rate as the economy. If it is less, it can be a cause for concern.

While tax buoyancy is a useful concept to measure the performance of both tax policy and tax administration over time, it is tax elasticity that is more relevant for forecasting purposes. Tax elasticity measures the “natural” response of tax revenue to changes in the tax base without any discretionary changes. The main use of tax elasticity is to identify which taxes are naturally elastic, that is, which taxes will yield more revenue as GDP rises, even with an unchanged tax rate.

Tax elasticity method

Tax revenue in any year is the sum of tax revenue in the preceding year, the increase or decrease in tax revenue due to a change in tax law (also called discretionary change) and general growth in tax revenue due to growth in tax base or inflation.

[331] To calculate the tax base (for example, GDP) and tax revenue, values for at least 10 years are taken together. The nominal values are turned into real values and thereafter, their natural log values are taken. A regression analysis is carried out on the log values of the real tax base and the log value of tax revenues. The coefficient of the log of the real tax base will be the buoyancy.

$$T_t = T_{t-1} + \Delta T$$

$$T_t = \underbrace{T_{t-1} + \text{Revenue Impact of Discretionary Measure}}_{\text{Adjusted tax revenue } (A\,T_{t-1})} + \Delta T$$

To estimate tax elasticity, the revenue series is required to be adjusted to remove the impact of discretionary changes. Once that impact is removed, tax elasticity is calculated in the same manner as tax buoyancy. Even the definition of tax elasticity is similar to tax buoyancy, with the crucial difference that the revenue elasticity calculation is adjusted for change in tax laws and administrative capacity (i.e. the discretionary changes). Tax elasticity is, thus, defined as a ratio of the percentage change in adjusted tax revenue and the percentage change in the tax base:

$$E = \frac{\%\Delta AT}{\%\Delta Y},$$

where $\%\Delta AT$ is the percentage change in the adjusted tax revenue.

The impact of discretionary changes in tax laws and administrative capacity is removed through adjustment in tax revenue. The adjusted tax revenue series is, therefore, a hypothetical construct and it shows what the tax revenue would have been if there had been no change in tax rules or administrative capacity. In other words, this indicates what last year's (or, last five years') revenue would be if current tax laws were applied to that year. Tax elasticity, thus, gives an indication of the impact of economic growth on tax revenues, which is usually a smaller figure when compared to tax buoyancy; tax buoyancy indicates the rise in tax revenues not only due to economic growth but due to discretionary changes in tax rules or due to administrative measures. Like in the case of estimating tax buoyancy, estimation of tax elasticity would be more accurate if it is done over a long period (say, ten years or more).

Four methods, as given below, are normally used to adjust tax revenue series:

- Constant rate structure method
- Tax proportional adjustment method
- Divisia index method
- Dummy variable method

In the constant rate structure method, the current year's rate structure is applied to last year's tax base and the adjusted revenue series is constructed on the basis of the effective tax rate for a given reference year and estimates of the tax base for subsequent years. This approach requires detailed tax base series for each tax bracket with multiple rates (e.g., sales tax and excise) or different marginal tax rates (e.g., personal income tax). This approach, therefore, is suitable if the number of items is small, the range of tax rates is narrow, and the compilation of data is easy (e.g., single rate excise tax on a few items). In other words, this method cannot be applied to broad tax categories such as excise or customs, but to individual products within these categories. The method is extremely cumbersome if it is to be applied to the full range of tax instruments that exists, since its data requirements are necessarily very heavy. Consequently, the constant rate structure method is rarely used for analytical purposes, and is normally relevant only when substantial changes are being considered in the tax structure.

The proportional adjustment method does not demand large and detailed data requirement. It does not require disaggregated data on tax rates and tax bases, which is necessary for the constant rate structure method and can make do with actual tax collection data. It, however, requires the use of budget estimates of tax yield arising out of discretionary changes and the availability of such data may restrict the applicability of this method. If such data is available, this method can yield better estimates of tax elasticity than the other methods explained above. Calculation of tax elasticity using the proportional adjustment method has been given in Appendix XIII.1.

In the proportional adjustment method, the tax revenue series is adjusted for the revenue impact of discretionary changes to indicate what the tax revenue in earlier years would have been if here were no changes in tax rules or administrative measures. Detailed information on tax bases is not essential in this method, but it requires basic information about revenue collections and the estimated revenue impact of discretionary changes (*ex-ante*). Estimating the revenue impact of discretionary changes could be complex. The micro-simulation model, also called the discretionary change model, is normally used to estimate the impact of discretionary changes.

The Divisia index method is less demanding in terms of data, as it requires actual tax collections and tax base measures at fairly aggregated levels. But it has certain inherent weaknesses. Computation, in this method requires the underlying tax function to be continuously differentiable and homogeneous, preferably linear homogeneous.[332] Although these may not seem to be particularly demanding conditions, it requires detailed mathematical analysis for the validity of the tax function to be used for the method.

The dummy variable method relies on econometric techniques to control for the revenue impact of discretionary changes when estimating tax elasticity. This approach may not work if the government changes tax laws too frequently or major tax reforms takes place, since this can lead to a substantial reduction in the degrees of freedom and hence, the efficiency of the estimates. Even if the number of such discretionary changes is relatively small, serious problems can arise in the specification of the estimation equations unless there is information on the nature of the tax changes and the extent to which their effects are independent of one another.

Projections based on buoyancy or elasticity

Any projection based on buoyancy or elasticity assumes that the buoyancy or elasticity is constant for the projection period. This may often be a strong assumption because the ratio constantly changes. But,

[332] Hulten, C.R., (1973), "Divisia Index Numbers", *Econometrica,* vol. 41, November 1973.

econometric methods, using ordinary least-square regression analysis, can be used to arrive at buoyancy or elasticity. Buoyancy or elasticity can be used by multiplying the buoyancy or elasticity with the estimated percentage increase in the tax base[333] to project tax revenue for the next year. This gives the percentage change in tax revenue.

To arrive at the revenue projection for next year, the percentage change in tax revenue is used, as follows:

Tax revenue in year t+1 will be tax revenue in the year, t, multiplied by one plus the percentage change in tax revenue, calculated using buoyancy or elasticity. The can be shown as below:

$$TaxRev_{t+1} = \mathrm{TaxRev}_{\mathrm{t}} \times \left(1+\varepsilon \times \%\Delta Y_{t+1}\right)$$, where ε can be buoyancy or elasticity.[334]

It may be mentioned here that for calculation of buoyancy or elasticity, tax revenues or the tax base has to be calculated on real basis, indexed at the same price level. But taxes are normally calculated on nominal or current prices, so adjustment of tax revenue for year, t+1, will have to be made with the expected inflation in the year, *t*+1, as given below:

$$TaxRev_{t+1} = \mathrm{TaxRev}_{\mathrm{t}} \times \left(1+\varepsilon \times \%\Delta Y_{t+1}\right)\left(1+InflationRate_{t+1}\right)$$

Business cycles and tax revenue projection

Revenue estimates during a budget are normally based on specific assumptions relating to nominal GDP growth and other macroeconomic variables. Increasing market orientation of an economy often brings out the cyclical nature of macroeconomic variables, particularly that of GDP. While revenue stability is a key objective, tax collections tend to vary during crests and troughs. The other set of assumptions, which are implicit, tend to delineate the behavioural response of tax payees to changes in tax rates and regulations. Often, these are considered fixed in the absence of micro-level information. The propensity to pay taxes can vary depending on the rates and overall economic context, i.e., whether the units are facing credit or cash flow constraint, etc., leading to an increase or shortfall in revenue based on the stage of the business cycle the economy is in. The efficiency of tax collection or tax rate changes has limited impact. In fact, tax authorities during a period of economic contraction could be forced to chase a 'targeted' amount, which may be rendered irrelevant by a deceleration in the GDP growth rate, industrial output, etc. Under these circumstances, tax targets need to be visited periodically during the year and adjusted accordingly.

Tax revenue forecasts often depend on estimates of long-run elasticity or a more aggregate measure like buoyancy. These estimates are useful as long-term averages. However, elasticity may not be an appropriate parameter to use during specific periods of economic output expansion and

333 Tax bases are exogenous, provided by the Central Statistical Office or the Ministry of Finance.

334 To calculate elasticity, there is need to remove the effect of discretionary change, whether in tax laws or in administrative actions.

contraction. Tax revenue rises more sharply than the tax base during economic booms; during downturns, it collapses more sharply than the tax base does; consequently, using the long-run elasticity leads to over-estimating revenue during contractions and underestimating tax revenue during expansions.[335] It, therefore, is important to introduce the cyclical aspects of GDP into tax revenue projections. The response of tax revenue to business cycles can be conceptually presented in the following manner:

$$\text{Tax revenue}_t = \alpha_1 + \beta_1\text{taxbase}_t + \beta_2\text{taxrate}_t + \beta_3\text{outputgap}_t + \mu_t$$

where tax base depends on the nature of the tax base, whether it is the GDP or consumption, etc. Tax rate is the standard rate or it can be a weighted average of components. The output gap is defined as the difference between real GDP (actual) and potential real GDP, obtained through the HP filter. A positive gap indicates that actual GDP is above the potential. μ_t is the error term. All these are at the same time, t.

In its simplest form, one can use GDP as the tax base, and the tax/GDP ratio as the standard rate. It may be useful to carry out this exercise for key taxes and then aggregate the results for the total revenue forecast. This is because tax efficiency differs. Tax efficiency is often defined as follows:

$$\text{Tax Efficiency} = \frac{\text{Tax revenue}}{\text{Tax base}} / \text{Standard rate} \times 100$$

In the case of VAT, it is commonly known as C-efficiency.[336] Tax efficiency may differ during good times and bad as there could be shifts in the consumption pattern or income distribution. The change in the inherent structure of the tax base during good times and bad need to be taken into account

$$\text{Tax efficiency}_t = \alpha_1 + \beta_1\text{output gap}_t + \beta_2 X_t + \mu_t$$

X_t in the above equation can be a consumption shift, for example (it may be captured through a dummy variable). During bad times, people may shift from premium items to necessities, which have much lower rates. Consequently, tax collection would be low. This, it may be stated, will in no way reflect on the ability of tax collection officials. Similarly, there can be profit shifts and income shifts during prolonged recession/downturn leading to lower corporate tax or income tax collection. Further, during a downturn, the propensity to evade taxes may increase as credit constrained taxpayers may fail to pay taxes to help finance themselves.[337]

There is some kind of a consensus in public finance literature that government spending is typically pro-cyclical in developing countries and countercyclical or acyclical in advanced countries. As far

[335] Cemile Sanack, Ricardo Velloso, Jing Xing, "Tax Revenue Response to the Business Cycle", IMF Working Paper, March 2010

[336] Ebrill, Liam, Michael Keen, Jean-Paul Bodin, and Victoria Summer, "The Modern Vat", IMF 2001

[337] Brondolo, John, "Collecting Taxes During and Economic Crisis", IMF Position Paper, 2009

as tax revenue is concerned, the situation is more complicated. As we have seen above, policy makers do not have any say on the output gap, which is endogenously determined along with other macro variables. Similarly, the tax base itself can get influenced by non-policy factors like shifts in the consumption pattern or income distribution changes in an economic downturn. Only the tax rate is under the control of policy makers. It is possible that during an economic downturn, even if policy makers increase tax rates, the overall collection could fall due to the factors explained earlier. In other words, revenue collection will depend on both policy as well as non-policy variables.[338]

(iii) VAT aggregate model

Value-added tax or goods and services tax in India has been implemented as a consumption-type, multi-stage sales tax based on the destination principle. The potential VAT revenue will depend on whether it is origin- or destination-based, the method of assessment (credit-invoice or subtraction method), the scope of exemptions and zero-rating, the level of tax rates, and the degree of tax compliance. The key objective of VAT or GST modelling for forecasting tax revenues will involve steps to estimate tax expenditure in administering VAT and to estimate the compliance level. Three methods are normally employed to do that: the aggregate national account method, the sector-wise national account methods or disaggregate method, and the input-output (I-O) method. The aggregate method or the national accounts method follows one of two alternative approaches: (a) production and (b) consumption. The production method starts with the GDP. To arrive at the value-added included in the VAT base, the GDP needs to be adjusted for imports, zero-rating, exemptions, and turnover threshold. The consumption method recognises the fact that VAT or GST is ultimately paid by consumers, and so, the aggregate consumption figure in national accounts becomes the starting point, which is then adjusted for zero-rating and exempted goods and services to estimate the VAT base. Table 13.2 outlines the steps to estimate VAT base, using the production approach.

Table 13.2: Estimating VAT base using production approach

	Narration	Remarks	Likely data source
VAT base			
=	GDP	Sum of all value added of domestic production.	NIA
Adjustment A*: *Government expenditure			
–	Expenditures on wages and salaries	Non-taxable expenditure component in NIA.	NIA, GOV

[338] Carlos A Vegh, Guillermo Vuletin, "How is Tax Policy Conducted Over the Business Cycle?", NBER Working Paper, January 2012.

	Narration	Remarks	Likely data source
Adjustment B: Exempt sectors			
–	Value added of exempt domestic production (at factor costs)	e.g., financial services and farmers are typically exempt from VAT; imputed rents of owner-occupied dwellings are not taxed. Adjustment B does not include exemption threshold adjustment, which is done in Adjustment G.	NIA
–	Indirect taxes in exempt sectors	This is needed if the national accounts are in producers' prices.	NIA, IO
Adjustment C: Private capital formation			
–	Gross domestic capital formation	For consumption-type VAT, investments are not subject to tax.	NIA
+	New residential construction, including land value	Sales of new houses, including the land price, are typically subject to VAT. Gross capital formation includes new residential construction, but excludes land value.	NIA, TAX
+	Gross capital formation in exempt sectors	Input tax on investments in exempt sectors is not creditable.	NIA, IS
–	Change in inventories	VAT is not charged to unsold inventories.	NIA
Adjustment D: Border tax		For destination-based VAT.	
–	Exports	Exports are zero-rated.	NIA
+	Exempt exports	No input tax credit is given to exports of exempt commodities.	TAX
+	Imports		NIA
–	Exempt imports	Exempt imports are not taxed at the border.	TAX
Adjustment E: Intermediate sales (Cascading)			
+	Output of exempt sectors sold to taxed sectors	Including output from businesses below exemption threshold.	IO
Adjustment F: Final private consumption expenditure			
–	Expenditures on exempted goods and services	e.g., basic food items, health and education services, rents.	HHS
+	Taxed inputs in exempt goods and services	Due to cascading.	IO

	Narration	Remarks	Likely data source
+	Foreign expenditures in local markets	This addition only applies if foreigners are not allowed to claim tax refunds.	NIA
–	Expenditures abroad by residents	These consumption expenditures should not be in the VAT base.	NIA
Adjustment G*: *Exemption threshold		For reducing administrative and compliance costs.	
–	Value added of businesses below turnover threshold		GOV, IS
Adjustment H*: *Tax replacement			
–	Indirect taxes replaced by VAT	Indirect taxes replaced by VAT should not be in the base.	NIA, TAX
Adjustment I*: *Non-Compliance			
–	Estimated extent of leakage	Losses due to non-compliance.	GOV, TAX

NIA = National income accounts; I/O = Input-Output table, GOV = Government departments/agencies, TAX = Tax department, HHS = Household survey, IS = Industrial survey

To determine the VAT base using the consumption approach, the starting point is final private consumption expenditure and government expenditure on goods and services. Table 13.3 outlines the approach.

Table 13.3: Estimating VAT base using consumption approach

	Narration	Remarks	Likely data source
VAT Base			
=	Final private consumption expenditure		NIA
+	Government purchases of goods and services		NIA, GOV
Adjustment A*: *Final Private Consumption Expenditure			
–	Exempt final consumption expenditure		NIA, HHS
+	Taxed inputs in exempt goods and services	Due to cascading. Alternatively, this can be done through Adjustment D	IO
–	Zero-rated final consumption expenditure		NIA, HHS
Adjustment B*: *Government expenditure			

	Narration	Remarks	Likely data source
–	Exempt government expenditure		NIA, GOV
–	Zero-rated government expenditure		NIA, GOV
Adjustment C*: *Capital formation			
+	New residential construction, including land value	Sales of new houses, including the land price, are typically subject of VAT. Gross capital formation includes new residential construction, but excludes land value.	NIA, TAX
+	Gross capital formation in exempt sectors	Input tax on investments in exempt sectors is not creditable.	NIA, IS
***Adjustment D*: *Intermediate sales* (*Cascading*)**			
+	Input of exempt sectors purchased from taxed sectors	Including inputs purchased by businesses below the exemption threshold.	NIA, IO
Adjustment E*: *exemption threshold (For reducing administrative and compliance costs)			
–	Value added of businesses below turnover threshold	Alternative computation is to subtract the total sales of businesses below the turnover threshold and add back taxed inputs used in these sales.	GOV, IS
Adjustment F*: *Tax replacement			
–	Indirect taxes replaced by VAT	Indirect taxes replaced by VAT should not be in the base.	NIA, TAX
Adjustment G*: *Non-Compliance			
–	Estimated extent of leakage	Losses due to non-compliance.	GOV, TAX

NIA = National income accounts; I/O = Input-Output table, GOV = Government departments/agencies, TAX = Tax department, HHS = Household survey, IS = Industrial survey

VAT disaggregate method

The disaggregate method or sector-wise national account method is useful to estimate revenues from VAT or GST with multiple tax rates. Multiple tax rates are a common feature in many developing countries, including in India. Like the aggregate method, the disaggregate method can also use either a production approach or consumption approach. This method greatly relies on I-O tables. Since the I-O tables are static in nature, they do not allow for behavioural responses to policy changes. Consequently, the estimation of a VAT base using this method does not take into account behavioural responses to changes in the tax structure. These behavioural responses can be incorporated in the estimation of the VAT base only if appropriate own- and cross-price elasticity of demand for goods and services can be obtained.

(iv) Excise duties

Excise duties have historically been a fee for the privilege of carrying out some kind of activity –usually production of a commodity. Currently, excise taxes are typically charged on selected consumption items (but exempting exports) and structured in co-ordination with a broad-based tax on the consumption of goods and services. Excise duties are typically imposed in two ways – as a specific or unit tax, or as an *ad valorem* tax. Unit taxes are easier to administer as the base is determined by simply counting the quantity of commodities being taxed.[339] However, in the presence of inflation, the tax base gets eroded as the general increase in prices actually decreases the tax revenue in real terms. In addition, unit taxes tend to be regressive as they result in high effective tax rates on the cheaper items in a product category commonly purchased by lower income people. As compared to that, inflation effects are captured automatically by imposing *ad valorem* rates since the tax is levied as a percentage of commodity prices. This basically means that tax revenues increase at the same pace as general prices – maintaining constant revenue collection in real terms, holding everything else constant. Since it is based on product price, however, *ad valorem* excise may have to cope with valuation problems.

The model for excise, in a single commodity market, uses the basic economic theory that imposition of an excise tax reduces the quantity demanded and supplied, and changes the prices paid by consumers and received by producers. The change in quantity after the tax, T, is imposed can be expressed as

$$\Delta Q = T\frac{Q_0}{P_0}\left(\frac{\varepsilon\eta}{\varepsilon-\eta}\right)$$

where P_0 and Q_0 are the price and quantity demanded/supplied before tax, η is the own-price elasticity of demand, and ε is the price elasticity of supply. Both elasticities are measured at the market equilibrium without the tax (P_0, Q_0), assuming the supply and demand curves can be approximated by straight line relationships.

The change in demand price is

$$\Delta P^d = T\left(\frac{\varepsilon}{\varepsilon-\eta}\right)$$

The change in supply price is

[339] In India, close to 60 per cent of central government indirect tax revenues are collected from petroleum products, cigarettes and sugar, using unit or specific tax.

$$\Delta P^s = T\left(\frac{\eta}{\varepsilon - \eta}\right)$$

The formula for estimating the tax revenue, *R*, for a specific tax *T* is

$$R = TQ_0 + T^2 \frac{Q_0}{P_0}\left(\frac{\varepsilon\eta}{\varepsilon - \eta}\right)$$

From the formula given above, higher (absolute) values of either ε or η imply lower tax revenue; that is, if demand and/or supply are more elastic, then tax revenue will be lower, holding everything else constant. An elastic demand means that consumers will shift quickly to other goods if the price of a product rises. As tax rises, total tax revenue will first rise, reach a maximum and then fall. The idea is that as the tax rate becomes high, a large number of consumers will eventually stop consuming the taxed good, and tax revenue will decline.

The change in quantity demanded/supplied after an *ad valorem* tax *t* is imposed can be expressed as

$$\Delta Q = tQ_0\left(\frac{\varepsilon\eta}{\varepsilon - \eta(1+t)}\right)$$

The change in demand price is

$$\Delta P^d = tP_0\left(\frac{\varepsilon}{\varepsilon - \eta(1+t)}\right)$$

The change in supply price is

$$\Delta P^s = tP_0\left(\frac{\eta}{\varepsilon - \eta(1+t)}\right)$$

The tax revenue formula for an *ad valorem* tax *t is*:

$$R = tP_0Q_0\left[1 + \frac{t\eta}{\varepsilon - \eta(1+t)}\right]\left[1 + \frac{t\varepsilon\eta}{\varepsilon - \eta(1+t)}\right]$$

The equations as stated above are for a single market. But in the real world, the market has other goods and services, some of which are close substitutes for the taxed commodity; for example, tea and coffee, small and medium-sized cars, formal-sector and informal-sector provisions and taxable goods and smuggled goods. In such situations, the model becomes more complicated. A useful measure for analysing the behaviour of multiple markets is the cross-price elasticity of demand, which is defined as:

$$\eta_{ij} = \frac{\Delta\% Q_i}{\Delta\% P_j},$$

where η_{ij} is the cross-price elasticity of good *i* with respect to the price of good *j*; in other words, the cross-price elasticity is the percentage change in quantity of good *i* demanded due to a change in the price of good *j*. A positive cross-price elasticity will mean that goods *i* and *j* are substitutes – an increase in the price of *i* causes an increase in the quantity of good *j* demanded and if, the cross-price elasticity is negative, both goods are complements.

It can be shown for a change in the price of good *i* that $k_j + \sum_{i=1}^{n} \eta_{ij} k_i$ will be equal to zero, where k_i is the share of total spending on good *i*.

The total tax revenue from all commodities subject to *ad valorem* excise tax can be expressed as follows:

$$TR_1 = \sum_{i=1}^{n} R_i = \sum_{i=1}^{n} t_i P_i^s Q_i$$

When the tax rate changes – holding everything else constant – the total tax revenue becomes

$$TR_2 = \sum_{i=1}^{n} R_i \left(1 + \%\Delta t_i\right)\left(1 + \%\Delta P_i^s\right)\left(1 + \%\Delta Q_i\right)$$

Assuming the supply of all commodities is infinitely elastic, when the tax levied on good *X* changes, the equation above becomes

$$TR_2 = \sum_{i=1}^{n} R_i \left(1 + \%\Delta t_i\right)\left(1 + \eta_{iX} \%\Delta P_X^d\right)$$

$$= \sum_{i=1}^{n} R_i \left(1 + \%\Delta t_i\right)\left(1 + \eta_{iX} \frac{\left(1 + t_{X2}\right) - \left(1 + t_{X1}\right)}{\left(1 + t_{X1}\right)}\right)$$

where, η_{iX} is the demand elasticity of good *i* with respect to the price of good *X*, and $\%\Delta t_i = 0$ for $i \neq X$.

To estimate the change in the tax base due to the change in the tax rate, it is necessary to estimate the price elasticity of demand for the commodity. The quantity of goods consumed would be the dependent variable for the regression analysis to estimate demand elasticity, using the understanding[340] that the

[340] Ordinary least-square equation, $Log(Q_t) = a_1 \cdot Log(realP_t^{own}) + a_2 \cdot Log(realP_t^{subs}) + a_3 \cdot Log(realP_t^{comp}) + a_4 \cdot Log(RGDP_t) + C$, can be used to calculate regression coefficients giving own-price elasticity of demand *(a_1),* income elasticity of demand *(a_4)*, and cross-price elasticity of substitutes and complements respectively, from a_2 and

quantity of commodity is dependent on its own price, the price of substitute and complementary goods, the quantities consumed in previous periods and income:

$$Q_t = f\left(p_{own}, Y, p_{subs}, p_{comp}, Q_{t-1}, \ldots\right)$$

(v) *Trade tax model*

Trade tax revenue can be set up using the basic understanding that demand for imports (in quantity terms) is equal to the domestic demand for importable goods, which is in excess of domestic supply. This can be expressed in symbols as below:

$$Q_0^M = Q_0^d - Q_0^s$$

where Q_0^M = quantity imported, Q_0^d = domestic demand for importable goods, and Q_0^s = domestic supply of importable goods.

Using this understanding, combined with the demand elasticity of imports, $\eta^M = \Delta Q^M / Q_0^M \div \Delta P / P_0$ and using the definition of price elasticity of demand, $\eta = \Delta Q^d / Q_0^d \div \Delta P / P_0$, the trade forecasting model can be set up using the following two equations:

$$\%\Delta M = \left(\eta_{GDP}^M \times \%\Delta GDP\right) + \left(\eta_P^M \times \%\Delta RP_{GDP}^M\right)$$

$$TR_{n+1} = TR_n \times \left(1 + \eta_M^{TR} \times \%\Delta M\right) \times \left(1 + \%\Delta PriceIdx_{n+1}\right)$$

where $\%\Delta M$ is the percentage change in (the growth rate of) import values, or quantities, η_{GDP}^M is the elasticity of imports with respect to GDP

$$= \Delta M / M_0 \div \Delta GDP / GDP_0$$

η_P^M is the elasticity of imports with respect to relative prices of imports to GDP deflator $= \Delta M / M_0 \div \Delta RP / RP_0$.

---

a3. This equation is static in time, but quantity demanded can also have lag in time, as shown in this equation: $Log(Q_t) = Log(Q_{t-1}) + k \cdot \left[Log(Q_t^*) - Log(Q_{t-1})\right]$

Relative price of imports to the GDP deflator is calculated by dividing the import price index by the GDP deflator. If information on the import price index is not available, real exchange rate can be used in place of the relative price of imports to GDP deflator.

η_M^{TR} : Elasticity of tariff revenue with respect to import values.

RP_{GDP}^{M} : The relative price of imports to GDP deflator is calculated by Import Price Index $\div$ GDP Deflator.

TR_n : Tariff revenue in year *n*.

$\%\Delta PriceIdx_n$: Percentage change in price index in year *n*. Price index can be the import price index, inflation rate, or deflator.

The key to applying the model is to estimate the three elasticities, η_{GDP}^{M} , η_P^M , and η_P^M . The different elasticities can be estimated using regression analysis.[341] Based on the estimated elasticity, the growth rate of imports and tariff revenues can be estimated. The import growth rate or the percentage change in imports, $\%\Delta M$, is used to project import values. In practice, countries that have implemented the medium-term macroeconomic framework will have information about import projections. In such cases, it is often advisable to use macro parameters suggested by the macroeconomic framework to maintain consistency.

Normally, 10-year data of trade tax collections are used for setting up this model. If the import price index is not available, the rate of change in the nominal exchange rate can be used as a proxy to find the relative prices, which is estimated by dividing the import price index by the GDP deflator.

(vi) Input-output model

Input-output (I-O) data[342] is the starting point for setting up a VAT/GST micro-simulation model. This model is often used to estimate the disaggregated impact of a policy change on the VAT base and tax collection and compliance at each commodity level. For I-O-based VAT modelling, the starting point is the final domestic consumption expenditure on both goods and services. This will include both domestically produced as well as imported goods and services. This is reflected in the final demand matrix of the I-O table. This decomposition into import or domestic production will depend upon the information available on imports. Imported goods will also have to be classified into capital and transport equipment, industrial inputs, primary goods for industry or consumption, fuel, consumption goods, etc. If import data includes information about the type of goods and their final use classifications,

[341] log (M) = α + β log (RGDP) + c log (RP), log (TR) = a_0 + a_1 log (M), w*here*: M = Real imports, RGDP = Real GDP, RP = Relative prices and TR = Real tariff revenues. Please note that log here denotes natural logarithm.

[342] Input-output data gives supply of goods, and its use or consumption in an economy in a tabular form. The table helps in understanding the inter-relationship between different commodity and service industries in terms of supply and use, and final demand.

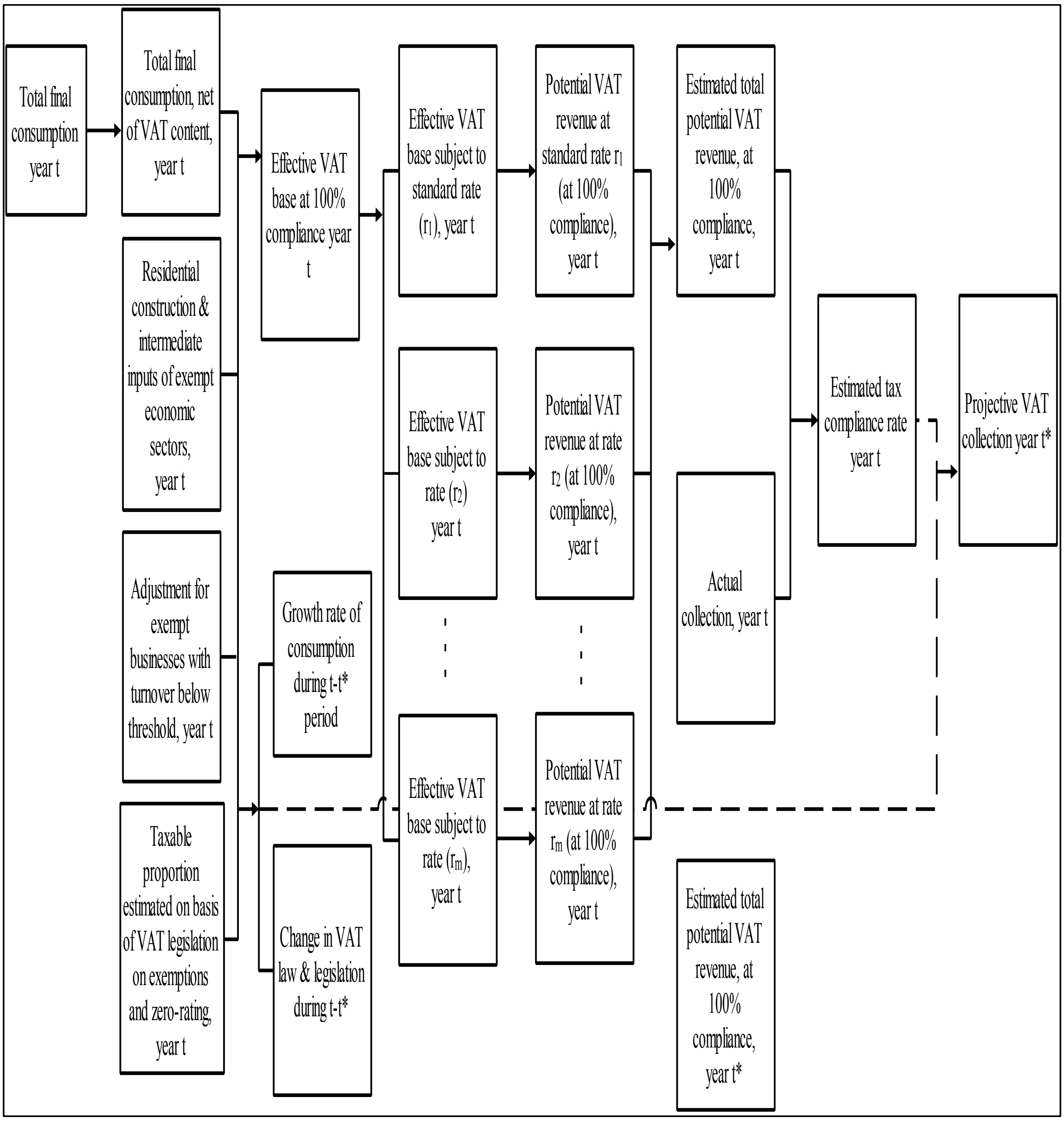
and the type of importer, then more or less strong inferences can be drawn about whether VAT paid on the import will be deductible or not, and where it is not deductible, whether it is going into final consumption or an exempt business. For example, most industrial machinery and capital equipment, and raw and intermediate inputs would have a high probability of being deductible.

To calculate the tax base, indirect taxes paid will have to be netted out since final consumption is inclusive of VAT. The other adjustment that will need to be made will be for exempt and zero-rated supplies. Typically, for any good or service, only a proportion of the supplies will be exempt or zero-rated. This taxable proportion can either be estimated on the basis of existing legal provisions stipulated in tax laws for levy of VAT or GST or through surveys.

Once this adjustment has been made, the taxable VAT base will have to be further adjusted for intermediate goods of exempt business activities and for small taxpayers who are normally exempt due to their turnover being below threshold levels. The effective VAT base will have to be net of sales of below-threshold small businesses, but inclusive of their inputs. The value added or difference between the outputs and inputs of below-threshold small businesses are normally unobservable and hence, will have to be estimated.

Based on these adjustments, the expected VAT revenue can be calculated by multiplying the tax base with the taxable proportion and the tax rates. But, there can also be more than one VAT rate (this will normally be the case). In that case, the estimation will need to be done separately for those commodities, which are subject to one rate, and those that are subject to other rates. Tax compliance will be another multiplicative factor to arrive at the tax figures. Non-compliance could happen for a variety of reasons, such as failure to register, non-filing of tax returns, under-reporting of sales or over-reporting of input tax credits, import smuggling, and such other mistakes, whether advertently or inadvertently. Potential VAT revenue is compared with actual tax collections to estimate the compliance rate or leakage risk factor. A few other adjustments to actual collections may be necessary, for example, to account for improvement or deterioration in the collection of arrears and accumulation of unpaid refunds. This will give the average compliance rate. For commodity or service-wise non-compliance rate, the ratio of actual tax collections from that particular commodity or service to the potential tax revenue from that commodity or service will have to be taken. It may be stated that services tend to have more non-compliance than goods and so, compliance rates for services are less than that for goods. Diagram 13.3 outlines the steps for setting up an I-O model.

Diagram 13.3: I-O model - framework for VAT forecasting



Source: Estimating the VAT Base: Method and Application, Tuan Le Minh, Tax Notes International, Volume 46, Number 2, April, 2007

(vii) Micro-simulation models

Micro-simulation is a general term for modelling the behaviour and interaction between micro units, such as persons, households, firms, etc. The model involves a category of computerised analytical tools that perform detailed analysis of a taxpayer's tax liability, through a set of rules operating on a representative sample of micro units. Despite the use of computer modelling, micro-simulation is distinguishable from other types of computer modelling in that it looks at the interaction between individual "units". Each unit is treated as an autonomous entity, and the interaction of the units is assumed to be on the basis of stochastic (randomised) parameters. These parameters normally represent individual preferences and tendencies. The model analyses the impact of policy changes on individual taxpayers, rather than on the mean, which happens when using regression techniques. Hence, micro-simulation models are used to evaluate the effects of proposed interventions before they are implemented in the real world.

Personal income tax

The personal income tax micro-simulation model is a computer-based algorithm to calculate tax liabilities for individual units, households or firms, in nationally representative microdata samples. The model builds an analytical framework at the micro level, simulates existing tax rules and calculates tax liability for each individual unit. The calculation takes into account the interaction between different elements of tax law. The resulting taxes and net income or profit measures for each unit are then provided a "weight" on the basis of the sample to provide results at the population level on an aggregate basis. The micro-simulation model, thus, uses a grossing up technique to arrive at the results at the macro level by using micro-level data on each unit. Since the model is built upwards from single taxpayer's data to an aggregate data, it is often used to study the tax liability at the level of individual units and the effects of tax policy changes on an individual unit as well as on overall tax revenues. It is for these reasons micro-simulation models are increasingly being used as analytical tools to estimate tax revenues and to evaluate the impact of tax changes. The other advantage of this model lies in its capacity to estimate the distributional effect of a given policy proposal on particular sectors of the population, showing the effects of tax policy changes on specific groups, breaking down the impact on the basis of income class, age group or tax jurisdiction.They are capable of providing distributional impact analysis by identifying potential winners and losers from given policy change proposals.

The steps followed in a micro-simulation model from the development of typical taxpayer's tax calculator model or personal income tax micro-simulation model to forecast revenue are the following:

i. Typical taxpayer tax calculator modelling

ii. Sample design and database construction

iii. Data ageing, updating, and validation

iv. Aggregate tax calculator modelling and impact distribution analysis

v. Result validation

A typical taxpayer model calculates the tax liability of a typical individual or family-based taxpayer and simulates the impact of proposed policy changes on the tax liability of the individual taxpayer. The underlying idea of a typical tax calculator model is to understand the application of tax laws on individuals under the existing tax regime and to compare it with the tax liability of the same individual under a proposed tax policy change. The impact of the tax policy change proposal on the individual is the difference between the individual's current tax liability and the tax liability if the proposed changes were implemented. The model usually follows the same logic used to fill out tax returns to calculate the income tax liability. A typical taxpayer may include an elderly person, a couple with children, a single person or a working woman.

Generally, the structure of a tax calculator model consists of three main components:

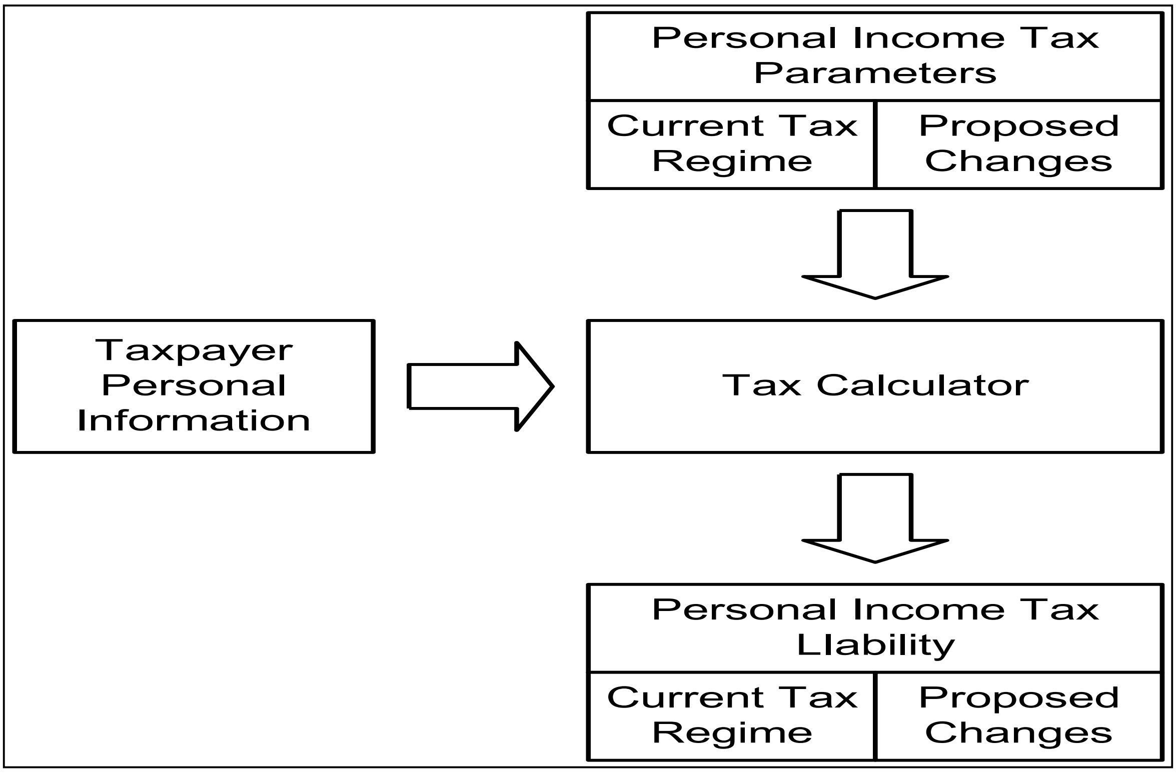
(a) Personal income tax parameters (tax code regulations)

(b) Taxpayer personal information (historical data) and

(c) Tax calculator module

Diagram 13.4 schematically gives the steps required to setup this micro-simulation model for personal income tax.

Diagram 13.4: Typical taxpayer's tax calculator model



The personal income tax parameters component contains detailed information regarding statutory tax parameters as described in the current tax codes, and their corresponding parameters in the proposed change proposals. These parameters include information regarding income brackets and their corresponding tax rates, exemptions, allowable deductions and their constraints, and tax credits. The level of information provided under these three components determines the degree of analysis that can be undertaken.

The taxpayer's personal information component captures personal data, which is required to assess income tax liability properly. This personal information may include taxpayer identification (PAN), age, marital status, number of children in school, wages and salaries, investment income, mortgage interest payments, donations, and other information, which may otherwise be required in the individual tax return. Since most of this personal information will be historical data, some adjustments may be required to update the personal information into the observation year. In addition, this component should include data according to categories, which are important for the impact distribution analysis. Such data may include age group, gender, area or region of residence, industry, foreign taxpayer, whether or not the taxpayer is in politically sensitive groups, etc., and allow the analyst to evaluate the impact of proposed tax policy changes on a certain taxpayer category.[343]

The tax calculator module will calculate personal income tax liabilities using the same mathematical logic as used in tax returns. The tax calculator module takes individual incomes from all sources and applies tax laws to determine allowable deductions, allowances, and credits based on personal information, yielding an assessment of total tax payable.

Sample design and database construction

Data requirements for micro-simulation modelling depend on the policy questions required to be investigated; how much time is available to provide those answers, and at what cost. These criteria determine the development of the typical taxpayer model. Because of these constraints, the sample size is often small when compared to the actual population of potential personal income taxpayers. The sample size may range from half a per cent of the total filing population to five per cent of the population, depending on the size of the total potential taxpayers and the resources available to collect and analyse the data.

While taxpayer personal information (historical data) may come from annual tax returns, sometimes household survey data are also used to create the database. Household surveys are often used as important sources of data, especially when data from tax returns are incomplete or inadequate.

Additionally, the relevance and accuracy of the data and the representativeness of the obtained sample need to be evaluated to maximise the reliability of the model. This may require data validation by crosschecking the data in the sample with data from external sources. This needs to be carried out at different stages of the model development process.

The reliability and validity of the data can be improved by using a structured (stratified) sample. Typically, the stratification will follow certain criteria, such as source of income (e.g., employees, self-employed, investment, pensioners, etc.), place of residence (e.g., urban, rural, provinces, foreign

[343] The impact distribution analysis is relevant to evaluate the impact of tax policy change proposals on groups of individuals at the aggregate level.

residence, etc.), and income levels (e.g., low, medium, high). The stratification criteria may also differentiate tax filers with special circumstances, such as politically sensitive groups, groups with volatile income, and taxpayers claiming refunds or large deductions.[344]

Data ageing, updating, and validation

Input data in the model will normally be from the past – usually, the latest available historical data lags considerably behind the time period for which the analysis has to be done, while estimates are needed for several years into the future. Hence, data "ageing" and updating of the latest available data to future periods are required before running the simulations. Data ageing can be done using simple, static growth factors or using more complex, longer-term dynamic methods, which take into consideration behavioural responses due to changes in tax policies. The static method is usually sufficient during the early stage of development to evaluate short-term effects of tax policy changes. This method uses information from macroeconomic and demographic forecasts, such as real GDP, inflation rate, and population and labour force growth rates to determine the growth factors for data ageing.

Aggregate tax calculator modelling and impact distribution analysis

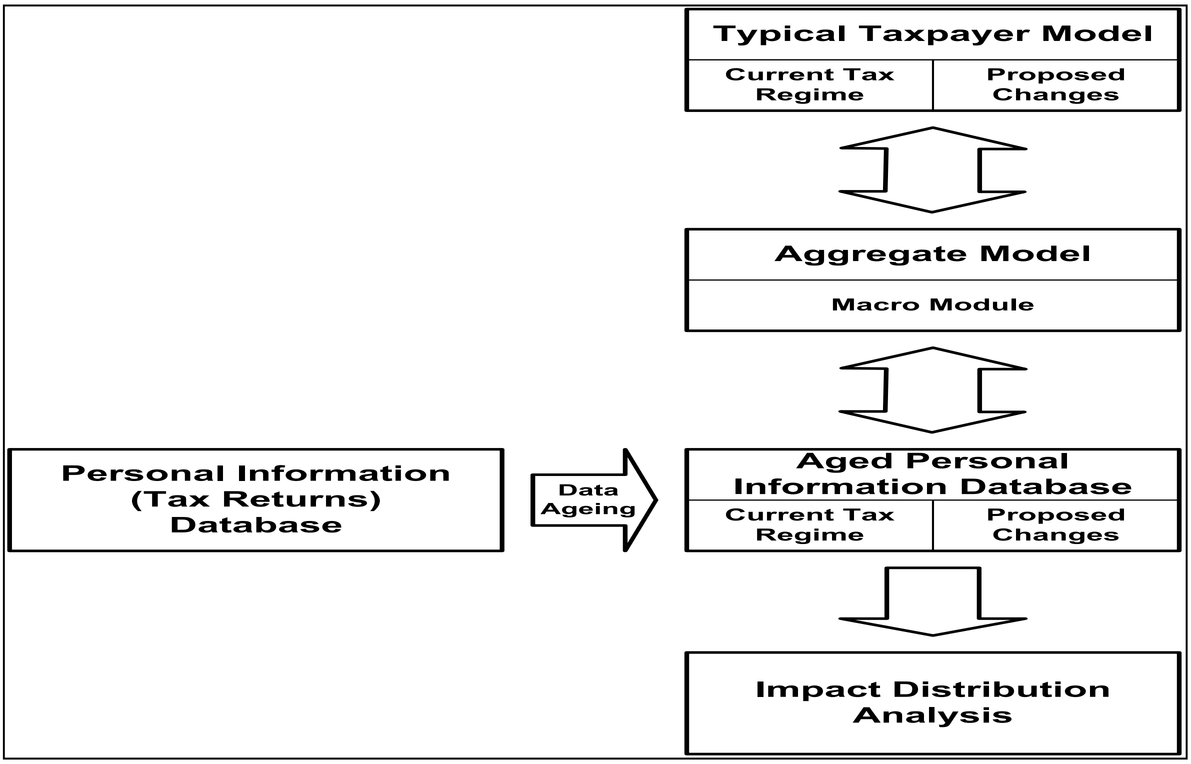
The aggregate tax calculator model uses personal income tax parameters and taxpayer personal information as its inputs. But, unlike the typical taxpayer model, the aggregate model is designed to handle a massive amount of the taxpayer personal information database and is capable of simulating the distributional impact of tax liability on various taxpayer categories as well as the fiscal impact on tax revenues for the nation as a whole. The distributional impact table provides a summary outcome of tax policy options by clearly identifying winners and losers in each case. Under the aggregate model, each record from the personal information database will be automatically fed into the typical taxpayer tax calculator model, and the output – essentially an impact table – will be generated based on predefined taxpayer categories. Diagram 13.5 schematically depicts the aggregate tax model and impact analysis.

[344] Stratified sampling is done by sub-dividing the total population into groups or strata, and each unit is assigned to a unique stratum according to its specific nature. The purpose of stratification is to increase the efficiency of sampling by dividing a heterogeneous universe in such a way that (a) there is some degree of homogeneity among samples within a stratum, and (b) there is some degree of difference between one stratum and another.

The strata may be created more than one level deep; that is, one stratum can be further subdivided into several strata. After the grouping is done, the sample within a stratum can be created using either systematic or random sampling. The numbers of items taken from each group are then taken in proportion to their relative weights. The weights are used to scale the sample to represent the population. Since this method has an equal probability of selection of items, it is known as proportionate stratified sampling.

This proportional stratification is simple and satisfactory if there is no difference in dispersion from stratum to stratum. Alternatively, sampling fractions can be set at different rates for each stratum. Use of disproportionate stratified sampling leads to an unequal probability of selection. Disproportionate stratified sampling requires additional work as the responses have to be weighted for the analysis. However, the payoff in terms of increased precision may justify the additional work. The use of proportionate stratification ensures the adequate representation of relatively small, but especially relevant groups, in the sample. In contrast, disproportionate stratification permits the analysis of particular members of a given stratum.

Diagram13.5: Aggregate model structure



To automate the simulation process, a computer program (macro) is often developed to read the personal income data from the sample database and calculate the tax liability for each individual sample in the database. Generally, the macro module automates the simulation process in three basic steps: (i) read the personal record from the database (ii) calculate the tax liabilities under the current and proposed tax structure using the tax calculator model and (iii) update the tax liabilities in the personal information database and calculate the weighted impact. Once the simulation process that evaluates the impact of discretionary changes on each individual taxpayer is completed, impact tables for distributional analysis can be easily generated.

Result validation

Any model, including the micro-simulation model, can contain errors of estimation caused either by biases in the sampling process, poor quality of input data or the straightforward failure of some of the components of the underlying model. Once the models are designed and the analysis is carried out, a validation system is needed to provide information on the likely margin of error. Every model, therefore, should include some form of data validation that is capable of producing minimum estimations of uncertainty in the outcome. Result validation increases the reliability of the model's output.

Three validation techniques are often used:

a) External validity studies – outputs are compared with external sources to ensure objectivity

b) Sensitivity analysis – models are evaluated under different alternatives of input assumptions to study the effects of changes in specific components of the model. These analyses enable assessing the relative effect of changes in each component on the overall outcome.

c) Computer statistical techniques to measure the variance of model estimates.

One major limitation is the possible effects of policy changes on the behaviour of affected units. These responses introduce elements of uncertainty that may bias model outcomes.

Database construction is often the most challenging task. It need not be emphasised that the quality of micro-simulation results will depend critically on the quality of data used. The sophistication of the model and the results it yields will depend upon data quality, sample size and its representativeness, level of information sought, and effects of policy changes on specific groups or on the larger population. Since policy simulations require sufficiently representative samples, the development of micro-simulation models for personal income taxation often involves a complex data collection process through modifications of tax returns and/or extensive household surveys.

Corporate tax

The micro-simulation model for corporate tax is complicated by the heterogeneity (in terms of types, size, age, business activity, corporate financial policy and structures) of firms, the volatility of annual profits or losses of companies and the skewed distribution of tax payments. Another problem is that macroeconomic indicators and their forecasts depend on business cycles; tax revenues from companies may vary due to historical profit or loss, which necessitates the use of past data to determine taxes due. Hence, micro-simulation methods may not forecast tax revenues accurately. But many countries, particularly some advanced countries, have used an iterative procedure. In this procedure, the macroeconomic forecast is made first and the tax revenues are estimated using simple tax revenue functions. The micro-simulation model is then used to adjust the revenue forecasts, using residuals in the macro tax revenue function, and the process is repeated several times until convergence occurs. In this method, a number of methodologies for revenue forecasting, including those conditional on macroeconomic variables, such as GDP, and those that are made unconditionally, are used along with the structural micro-simulation model.

The steps to setup a micro-simulation model for corporate tax will involve the following:

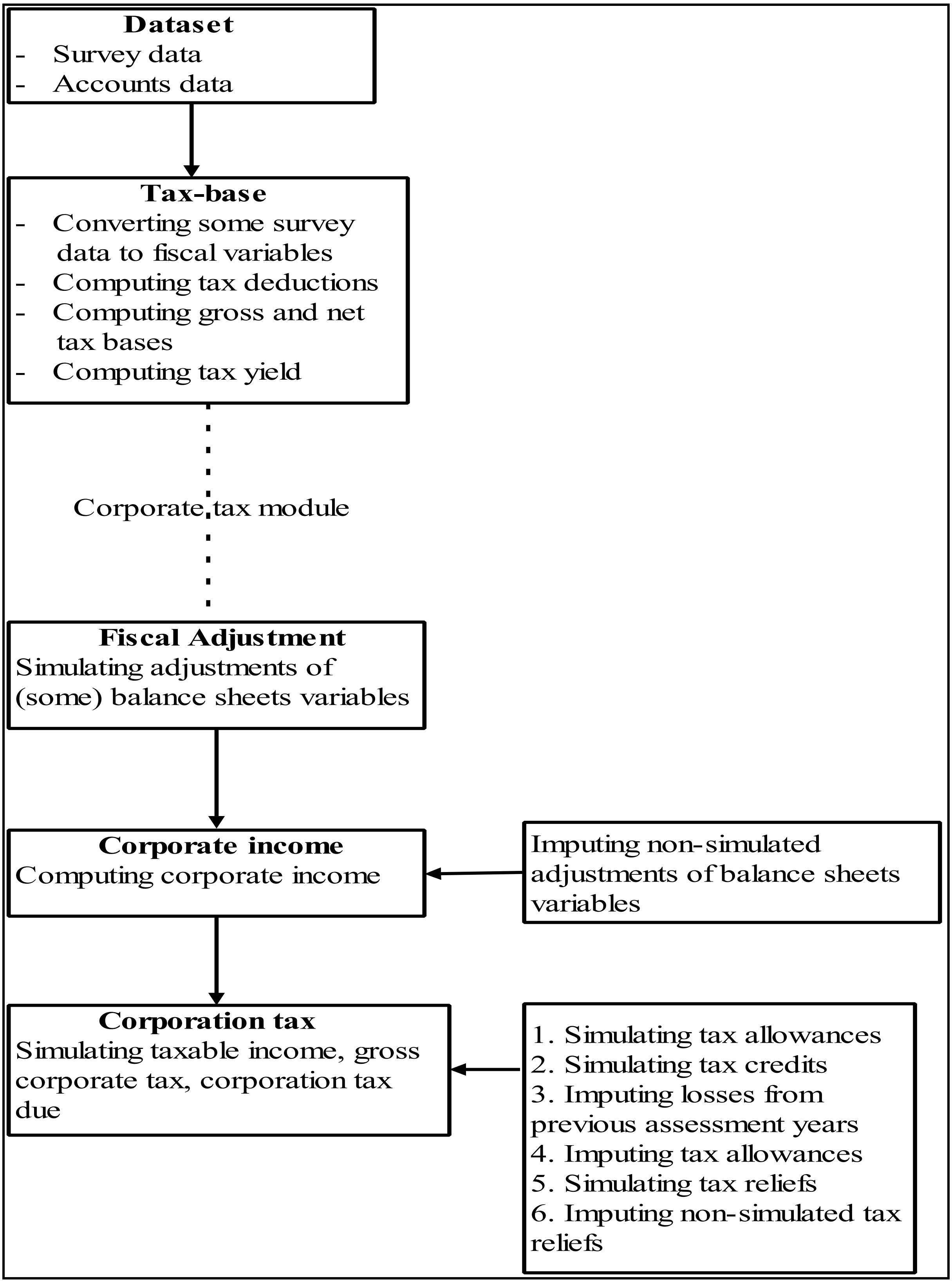
- Drawing up a sample of corporate tax returns – stratified sampling is usually used to draw a sample of tax returns. First, different strata are established on the basis of some critical characteristics, such as size of assets, income, industry, or region. These samples must include corporates that have different filing dates on account of delayed filing or have had tax return filing extensions.

- After the sample has been drawn, the data is required to be cleaned to ensure consistency and reliability. Missing data can filled either by using data from corporate financial statements or by imputing data based on past data.

- The corporate income tax base is derived from gross income from different sources minus itemised deductions including costs of goods sold, depreciation allowance, interest payments, overheads expenses, and any net operating loss from prior years.
- Using this tax base, the tax payable can be estimated as the product of the tax base and corporate income tax rate. Tax liability will have to be adjusted for any tax credits that are allowed. From this base tax calculator, the impact of a proposed change in corporate income tax law on a representative corporate firm and on government tax revenue can be simulated. Corporate expenditures like depreciation and interest payments will be deductible from the corporate income tax base. A time series of the stock of the unused portion of deductions will be required to be carried forward.
- This calculator model will present the calculation of individual corporate tax liability under the current and proposed law, as well as changes (impacts) in corporate tax liability due to discretionary changes. This representative taxpayer model will have to be aggregated and connected to macroeconomic data like GDP, GDP deflator and investment.

Diagram 13.6 outlines the steps for setting up the model.

Diagram 13.6: Steps to set up corporate tax micro-simulation forecasting model



Source: The effects of the Italian tax reform on corporations: a micro-simulation approach, Valentino Parisi, et al, 2003

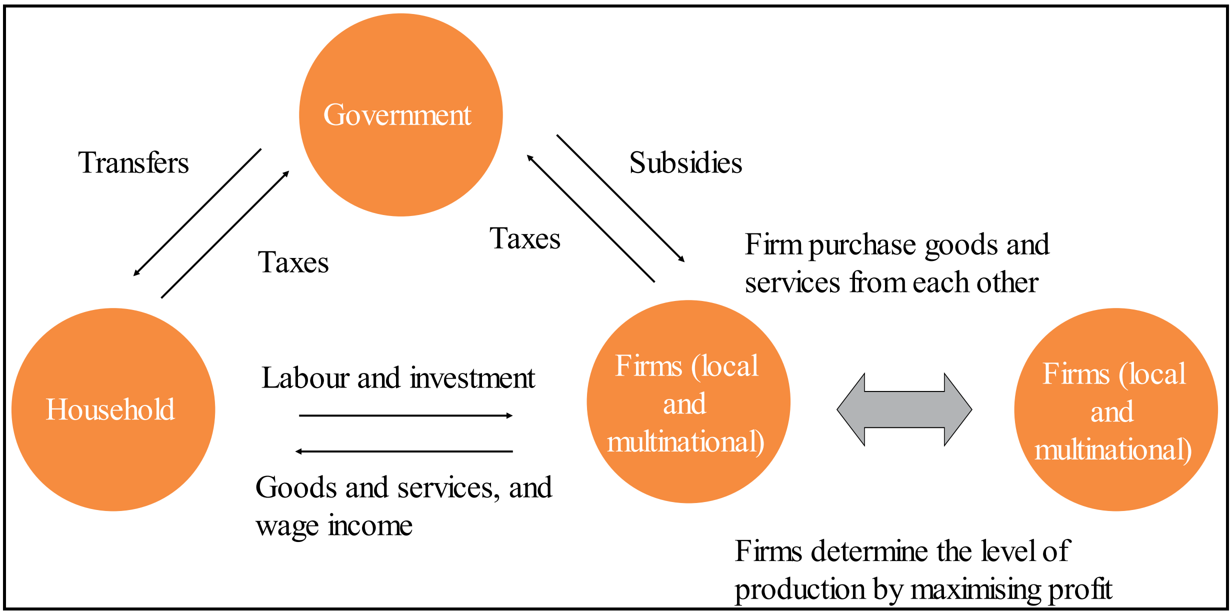
XIII.2.b Methods of revenue analysis

Computational General Equilibrium (CGE) model

Computable General Equilibrium (CGE) models have been used since the early 1970s to analyse the effects of changes in different economic variables on tax, trade and investments and to evaluate public policies in general. The CGE model is an analytical representation of transactions in a given economy that enables connecting each economic element of the model to some observed empirical data and simulates the core economic interactions in the economy by using data on the economic structure along with a set of equations to estimate the effects of fiscal policies on the economy. The idea is to have an instrument that is capable of describing numerically how the economy behaves and reacts to different external changes or shocks while being consistent with standard economic theory. CGE models are used widely by international institutions such as the World Bank, IMF (International Monetary Fund), OECD (Organisation for Economic Co-operation and Development) and the European Commission, as well as advanced economies. Their primary use is to assess the 'impact' of different government or institutional policies (e.g., changes in tax policy, government spending) or to investigate the effects of change in different economic indicators (e.g., change in real exchange rates, or the level of aggregate demand) on tax collections or the impact of changes on different segments of taxpayers, by including the behaviour of consumers and producers, and other market factors.

The first step for setting up the CGE model is to clearly define the problem to be analysed and accordingly choose the model's type and features. For example, if the type of problem that is desired to be analysed requires disaggregation into different economic sectors, such as industry types or taxpayer's types, the model will have to incorporate such disaggregation and an equation defining that sector will have to be put in place for each sector, which are then aggregated and connected to relevant macroeconomic indicators. The model combines economic data and a set of equations to capture interactions between the three entities in an economy – households, business and government – as taxes are transfers from households as well as from industry to government. Each entity will need to be defined and interlinked through appropriate relationships either in the labour market or household consumption or capital market flows, intermediate product demand, taxes or government transfers. Diagram 13.7 provides a basic illustration of these interactions.

Diagram 13.7: Relationships captured in a CGE model



Source: A multi-regional CGE model of the UK: a report by PwC for the HMRC, PricewaterhouseCoopers (PwC), March 2004

The choice of specification in each equation will limit output and thus, will constitute the objective function for the CGE model in a disaggregated manner. These objective functions, however, are constrained by the availability of data. The CGE model also uses econometric estimation, which although accurate, can pose difficulties in estimating some parameters of the econometric equation. To circumvent that, a calibration method is used, which is less precise but requires less data, and fewer observations and calculations.

In practice, data used for CGE models are taken from tax departments for taxpayer data and from national accounts for aggregate data. Data available in other government departments are also used in the CGE model.

Once the functional forms of all the economic entities in the model have been defined and calibrated, the CGE model is to be set for the benchmark equilibrium or the starting point equilibrium. The purpose of this is to replicate the observed economy in such a way that the model reproduces the equilibrium state of the economy. The outcome of the model is to be tested on the basis of observed levels of production of goods and services, their prices – inclusive or exclusive of indirect taxes, observed tax revenues – direct or indirect taxes – and other such observed variables. These observations may require that the model be calibrated and recalibrated in an iterative manner so that the parameters in the CGE model capture the real-life situation. The calibration or recalibration will lead to the desired model so that the proposed changes, either in the tax rate or imposing or withdrawing a tax deduction or exemption or a change in the price of one good or many goods, can be simulated. This may take more than one financial year of observation and calibration. The model, once set up, can be used for predicting comprehensively the effects of changes, including for instance, changes in prices, output levels, government revenues and income distribution

CGE models are often used to simulate future effects of policy changes, and not necessarily for revenue forecasting. Policies are evaluated by comparing the base-line economy with the impact of a policy. The model, therefore, provides tools to analyse the effects of policy on various sectors or entities of the economy in a comprehensive, yet disaggregated, manner.

XIII.2.c Methods of revenue monitoring

To ensure a balanced budget, it is important to effectively monitor tax collections on a regular basis, along with expenditure. Monitoring of tax collections is important as forecasts of revenues for a year need to be converted into receipts (cash); forecasts give revenue potential from economic activity but not how much is actually collected. Actual tax receipts are a sum of the tax forecasts, including the expected tax collection from the new revenue raising measures, and the share of tax demands, arrears and current, to be collected over year. The share of new refunds expected to be paid and outstanding refunds of previous year expected to be paid over the year will be subtracted from all the collections to arrive at the actual tax collections.

Depending on the level of sophistication of the monitoring system, it is possible to track major sources of revenue collection on a periodic basis – weekly or monthly or quarterly. To perform this task, it is necessary to establish an effective information system and a well-developed database to measure accurately actual versus expected revenue collection. The database provides the main input for analysis of tax functions, including behavioural responses to new tax measures, revenue forecasting, and tax expenditure analysis. Hence, systematic data collection and setting up a data base are pre-requisites for the establishment of an efficient revenue collection and monitoring system.

Many advanced countries, instead of monitoring tax collections directly, monitor the tax gap, which is defined as the difference between potential and actual tax revenues. This section outlines both the methods.

(i) Tax receipt model

The tax receipt model uses tax receipts for projecting tax collections. It is a simple, yet functional, tool to monitor and project tax revenues for short-term forecasting. While this model is targeted at short-term forecasting and monitoring monthly tax receipts over a financial year, it contains features similar to that of a simple macro-model for forecasting one-year or medium-term revenues as long as there are no major changes expected in the tax or economic structure over the forecast period. The model is used in many countries to closely monitor tax revenue collections. The results from this model are relatively accurate and hence, attractive to use. The data requirement for the model is also simpler, primarily only monthly tax collection data, which are usually available within the tax administration.[345] For monitoring purposes, the model provides the collections in each month to date, the projected revenues for each month over the remainder of the year, and the estimated revenue surplus or deficit by type of taxes.

The model uses past seasonal tax collection patterns to predict collections in the following years, incorporating increased collections due to economic growth, actual collections to date that deviate from the expected growth rate, expected changes in the seasonal pattern, and changes in the effective tax rate between the same months in successive years. The model, given below, uses actual monthly receipts data and projected GDP growth – or other tax base proxies (e.g., private consumption or imports) growth – to forecast collection.

$$T_y = \sum_{a=1}^{m} T_{a,y} + \sum_{\substack{a=m+1 \\ i=m+1}}^{12} T_{a,y-1} \cdot (1+\delta) \cdot \left(\frac{\tau_{i,y}}{\tau_{a,y-1}} \right)^{1+\beta}$$

The terms used in the equation above are as follows:

$T_{a,y}$: Actual monthly tax receipts in fiscal year ***y***, where tax collection data is available

$T_{a,y-1}$: Actual monthly tax receipts in fiscal year ***y***–1

$B_{a,y-1}$: Actual monthly tax base in fiscal year ***y***–1

m : Number of months up to which actual tax receipt data in fiscal year ***y*** is available

$\tau_{i,y}$: Projected effective average tax rate in month ***i*** of fiscal year *y*

$\tau_{a,y}$: Actual effective average tax rate in month ***i*** of fiscal year *y*–1.

As shown, the model recognises that in any given month, annual tax receipts for the fiscal year will be a sum of two parts – actual revenues collected up to the month for which receipts data are available and forecast receipts for each of the remaining months of the fiscal year. To project the second part of monthly receipts, the model takes into account the actual growth of year-to-date tax collections as compared to collections for the same period in the previous fiscal year, and the projected growth of tax base proxies (e.g., GDP, private consumption, imports, etc.). If actual tax receipts data for month ***i*** is available, then $T_{i,y}$ is equal to actual tax receipts for that month. Otherwise, $T_{i,y}$ is projected using last year's actual receipts for the same month multiplied by a growth factor and adjusted by a user-defined elasticity, to capture the behavioural effect if there were policy changes during the fiscal year that modified the effective average tax rate. If the seasonal pattern of receipts over a year is expected to change from one year to the next – for example, the timing of a required tax payment instalment is

345 Although typically used to track and project monthly receipts, it can also be used for weekly or quarterly receipts.

changed – then, adjustment factors need to be introduced to last year's tax revenues to form the revised basis of projecting this year's pattern of revenues.

The growth factor in the equation,δ can be measured by a weighted average of two growth factors:

(i) the actual growth of year-to-date receipts in the current fiscal year over that of the same months in the previous fiscal year; and

(ii) the expected growth rate of tax base proxies (e.g., GDP) in the current fiscal year.

The weight for the first growth factor is the fraction of the number of months in a fiscal year during which the taxes were actually collected. The weight for the second growth factor is the remaining fraction. If the weight for the first growth factor is α, then the weight for the second growth factor is (1–α). Therefore,

$$\delta = \alpha \cdot \left[\frac{\sum_{a=1}^{m} T_{a,y} - \sum_{a=1}^{m} T_{a,y-1}}{\sum_{a=1}^{m} T_{a,y-1}} \right] + (1-\alpha) \cdot g$$

The term $(\tau_{i,y} / \tau_{a,y-1})^{1+\beta}$ in the model represents the ratio between the current year's projected effective average tax rate over the previous year's actual effective tax rate for the same month adjusted by the elasticity of the base to changes in the tax rate. The effective tax rate for a particular month is defined as the actual tax receipts divided by the effective tax base for that month. If there were no tax policy changes affecting the tax rate or tax base over the fiscal year, this ratio would be equal to 1.

This model, although specific to tax collections on monthly basis, can also be suitably changed[346] to monitor quarterly tax receipts at the end of each quarter of the financial year and annual tax collection. This may also be required as the proxy tax base for direct taxes is GDP, for which estimates are available on a quarterly basis only.[347]

(ii) Tax gap analysis

The tax gap is defined as the difference between potential and actual tax revenues.[348] The potential tax revenue is the amount of tax that the revenue authority receives if everyone complies with the tax law; but that is never the case. In this definition, the tax gap is a result of non-compliance with

[346] See Projection of Quarterly Corporate and Income Tax Collection, (2004), Prof. A. L. Nagar, et al, NIPFP (National Institute of Public Finance and Policy) Working Paper No. 24.

[347]This may also be because advance taxes in direct taxes are to be paid by certain dates towards the end of each quarter. A large proportion of direct taxes is collected by way of advance taxes on quarterly basis.

[348] The US IRS defines it as "*the difference between the tax that taxpayers should pay and what they actually pay on a timely basis*", whereas UK's HMRC defines tax gap as "*the difference between tax collected and what should be collected.*"

tax laws. Some tax administrations also call this the compliance gap, and proceed to divide the calculation into three components: non-filing of tax returns, under-reporting of taxable base and under-payment of taxes.[349] The International Monetary Fund (IMF) refers to the impact of compliance issues on revenue as "the compliance gap"[350] and the revenue loss attributable to provisions in tax laws that allow exemption, special credit, preferential rate of tax, or a deferral of tax liability as the "policy gap."[351] The definition and sources of the tax gap, thus, are varied in different tax jurisdictions, and differ for type of tax. The tax gap in some tax jurisdictions also includes, for example, uncollected taxes (i.e. bad debts), unintentional error, the underground economy and illegal activities. Because the measurement of the tax gap involves measuring potential tax revenue, it gives an estimate of the 'hidden' (also variously called 'underground', 'shadow', 'informal', 'black', 'grey') economy.

Quantifying the tax gap provides a picture of the total revenue due and from whom it should be collected (or in relation to what transactions). Mapping this information is a powerful tool for any tax administration, since it can be used as performance indicators, both at the organisational level and at each jurisdiction or employee level.[352] A tax gap estimate is a conceptually more relevant and valuable indicator to determine the effectiveness of the tax administration. Undertaking tax gap estimates also informs the government about the integrity of the tax system, the risks to revenue buoyancy, performance of their tax collection agency and processes, the evolving risks to revenue (and potential failures by their tax collection agencies), problems with tax legislation, problems with the national statistics and the impact of the non-observed economy on revenue.[353] Revenue collections, on the other hand, are not considered a good measure of the effectiveness of the tax administration, although it can be easily measured and compared to a target, as it depends largely on economic growth and changes in tax legislation than on the general performance of the tax administration. This is why the OECD and IMF support tax gap studies as a performance measure for the effectiveness of tax administration.

Other advantages for the tax administration from regular tax gap studies is the identification of the types and level of non-compliance that contribute to the tax gap, improved efficiency of resource

[349]The US IRS allocates total tax gap across each of these types of non-compliance, and then disaggregates the tax gap further by type of taxpayer.

[350]Understanding the sources of and reasons for non-compliance is important to develop strategies to encourage and enforce compliance and deter non-compliance, the core business of any tax authority. This intelligence can be gathered from many different activities undertaken by the tax authority, particularly audits. External sources of information, such as national statistics and literature on taxpayer behaviour and risk management, will also contribute. An increasingly popular method to analyse and use this information has been through the generation of tax gap estimations.

[351] http://www.imf.org/external/pubs/ft/scr/2013/cr13314.pdf, accessed in January 2015

[352]Since the 1990s, many countries, particularly the OECD countries, have used tax gap estimates to monitor tax collections from their tax administrations.

[353]These benefits to the governments also act as an indicator to the public that taxes due are paid, with assurances that will have a direct impact on the equity and economic efficiency of taxes. The loss of such assurance can have a deleterious impact on public confidence in the integrity of the tax and the revenue authority.

allocation within a revenue authority to combat non-compliance and its use as a measure of the effectiveness of the revenue authority. Identifying sources of non-compliance, no doubt, is a complex and difficult task but it is a key aspect of managing tax compliance. In order to identify the sources of the tax gap, the tax authority needs to develop a sound understanding of the tax(es) administered and associated types of compliance requirements, taxpayers and their compliance behaviours, and the environment in which they operate. This requires access to various sources of information, which is typically gathered from audit and other compliance activities. This process of identifying sources of non-compliance (also called, risk management) helps in prioritising identified risks. This cross analysis of the possible sources of the tax gap and its causes based on intelligence, data and tax gap estimates with the revenue authority has proved very powerful in practice for tax administrations.

Direct taxes

There are a number of models and methodologies to estimate the tax gap for direct taxes; there can be separate methods even for a particular type of tax. The two approaches most often adopted to estimate the tax gap for direct taxes are the top-down approach and bottom-up approach. In the top-down approach, estimates of revenue are made indirectly by constructing an estimate of the total tax liability from sources independent of data collected for tax purposes, which is then compared with actual tax receipts. In the bottom-up approach, on the other hand, estimates of revenue are made directly from the department's data using operational data. Often, the top-down approach is considered better as it is independent of both the tax department as well as taxpayers, and is more comprehensive because the estimates are out of other indicators. But this approach often suffers from insufficient data for constructing a reasonable model to independently estimate tax liabilities. In contrast, the bottom-up approach is less comprehensive because the departmental data may not cover all activities, and some activities may not be part of the data.

The top-down approach comprises three methods – the monetary method, physical input method and MIMIC (Multiple Indicators and Multiple Cause) method – to calculate the tax gap for direct taxes. Many tax administrations use one of the three methods, but sometimes it is useful to calculate the tax gap using all three methods, and then compare the results. As explained below, each method has its own limitation, either in its scope or due to data availability. The availability of trained personnel as well as other physical resources can also impose limitations.

The bottom-up approach uses microeconomic data to estimate the size of the relevant tax bases. Micro-simulation models, as already discussed above, are used to calculate potential tax liabilities. This approach provides detailed estimates of the tax gap by type of tax.

Monetary method

The monetary method, also called currency demand method, is based on the concept that dishonest taxpayers largely transact in cash in order to avoid leaving detectable traces. The method uses the concept that the supply of money, measured by the stock of money multiplied by the velocity of its circulation, should be equal to the demand for money measured by the total value of transactions in the economy using money. Based on this understanding, the method provides an estimate of GDP[354] by calculating the 'excess' demand for currency and then compares that with the reported GDP to estimate the size of the tax gap. To build the model, different parameters such as the rate of taxation, per capita income and the interest rate on savings deposits are used. The regression equation for estimating the stock of money in the economy can be as below:[355]

$$\ln\left(\frac{c}{M_2}\right) = a_0 + a_1 \ln T + a_2 \ln\left(\frac{ws}{NI}\right) + a_3 \ln R + a_4 \ln Y + e$$

Where, $\frac{c}{M_2}$ is the ratio of currency to M_2,[356] T is the rate of taxation, $\frac{ws}{NI}$ is the wage share in national income, R is the rate of interest on time deposits and Y is the real per capita income; e is the residual. This equation assumes that the share of wages in national income will increase the demand for currency as wages are often paid in cash; an increase in interest rate will reduce the relative demand for currency as most people will put their money in deposits; and higher taxes pushes people to undertake transactions in cash. The method also assumes that there is a stable relationship between the value of transactions and the value added from these transactions, and the velocity of money is either known or constant and hence, estimable.

Physical input method

The physical input method premises that any physical input indicator, such as electric power consumption or transport, can appropriately indicate overall economic activity. The estimate of GDP thus arrived is then compared with the official GDP to calculate the tax gap. Since the physical input is the basis of estimating the GDP, its choice is important for the exercise. Many times, the choice is criticised on the ground that many unreported economy activities may not require the physical input at all or if used, may not use a significant amount. Electricity consumption has often been used as the proxy for the overall economy.[357] Some tax administrations also use labour force participation rates instead of electricity consumption.

MIMIC model

---

[354] Sometimes, instead of stock of money, cash-deposit ratio is also used. The reason for its use is that cash-deposit ratio will vary with changes in tax rate and other government regulations.

[355] Tanzi, V, 1983. "The Underground Economy in the United States: Annual Estimates, 1930-1980." Staff Papers, 30:2, IMF

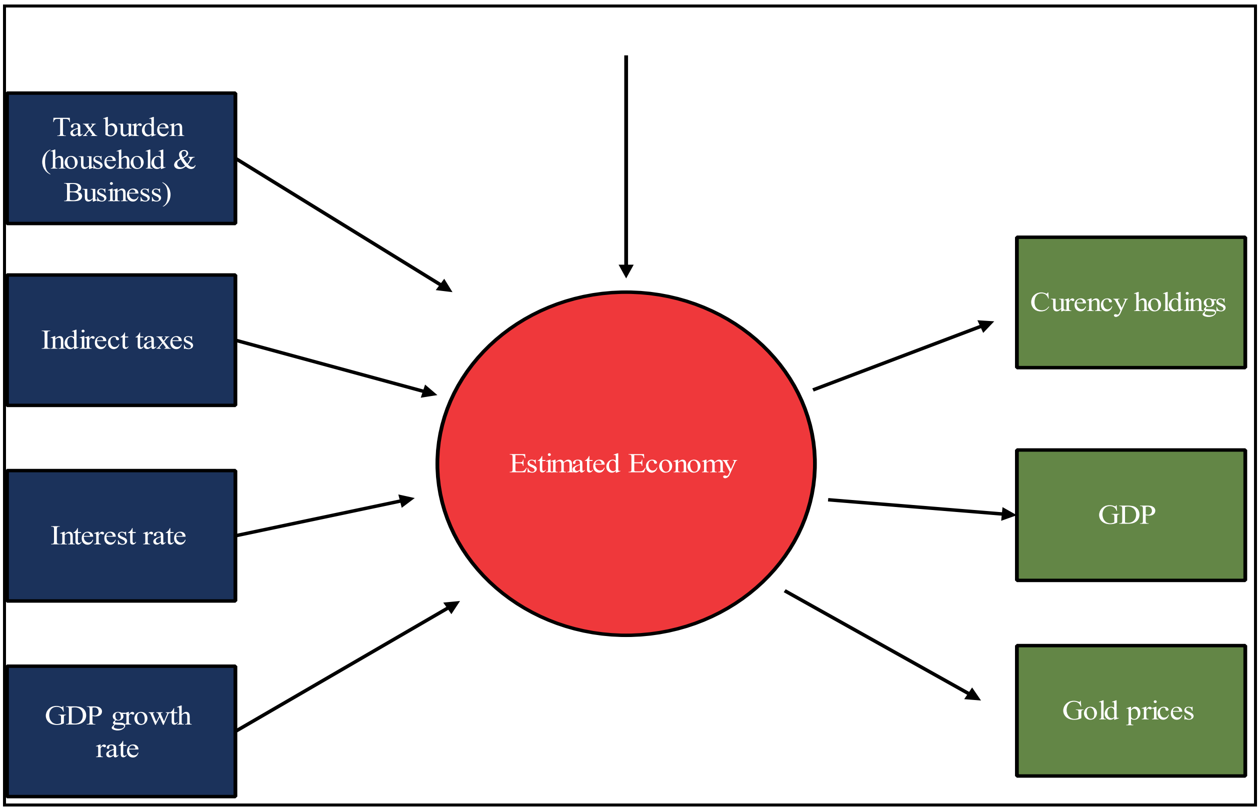
[356] M_2 is a measure of money supply that includes cash and checking deposits (together called M_1) as well as savings deposits, money market mutual funds and other time deposits, which are less liquid and not as suitable as exchange mediums but can be quickly converted into cash or checking deposits.

[357] Empirically, it has been shown throughout the world that consumption of electricity and GDP have an elasticity of nearly one.

The MIMIC method, also called structural equation model, is a more complex method. The idea behind the model is to represent the output (or income) of the underground economy as a latent variable or index, which has causes and effects that are observable but which cannot itself be directly measured. Thus, there are two kinds of observed variables in the model, "causal" variables and "indicator" variables. These variables are connected by a single unobserved index. Values of the index over time are inferred from data on causes and indicators by estimating the statistical model and predicting the index. The fitted index is then interpreted as a time-series estimate of the magnitude of the total economy, including the unreported economy.

Indicator variables are the observed variables, considered to have been influenced by a latent variable, and this latent variable is the only common variable that is reflected in all the indicator variables. Apart from this assumption, the MIMIC model also assumes that any correlation between these observed variables must be due to the latent variable. This correlation is used to recover the latent variable. Cause variables are also observed variables that affect the latent variables. There is no clear theoretical model for choosing the causes and indicators from the larger set of economic variables. Some variables, which can be logically included in the MIMIC model, are explained below. Diagram 13.8 shows the relationship pictorially.

Diagram 13.8: MMIC model - causal variables and indicator variables



Indicator Variables

The choice of indicators should be made in such a manner that the variables reflect the activity taking place in the unaccounted part of the economy. While any single measurement will contain an error, measuring the same thing multiple times can reduce the error, and hence, more than one indicator is often used in the model.

- ***Currency to M_3 ratio:*** As discussed in the monetary method, the presence of unaccounted incomes is likely to increase the demand for currency. Hence, the currency to M_3 ratio can be an indicator value.[358]

- ***Real estate investment:*** Real estate along with financial services like insurance and business services can be defined as another indicator variable. These variables are considered to have positive and negative linkages to unaccounted incomes, an important variable for calculating potential GDP. Some studies suggest that an increase in unaccounted incomes hinders the growth of financial instruments, and investment in real estate rises. Based on that understanding, it can be assumed that unreported income will be reflected in this indicator.

- ***Differences in gold prices:*** Gold is often an investment option for unreported income. If there is a supply constraint on gold, due to gold control or due to an increase in import duty on gold, there will be a price differential between the domestic and international market prices of gold. This also provides lucrative opportunities for gold smuggling, leading to the generation and utilisation of unreported income. The price differential, thus, can be an indicator variable for calculating the tax gap.

Causes variables

- ***Tax burden:*** A common hypothesis is that an increase in tax burden is a strong incentive to work in the unaccounted economy. Many studies, therefore, use the tax burden, a ratio of tax (both direct and indirect taxes) to GDP, as a cause variable. The tax rates used are usually the statutory tax rates.

- ***Growth rate in real GDP:*** The GDP growth rate reflects the extent of activity taking place in the economy, and so can be taken as a cause variable; a higher growth rate reflects a better state of the reported economy.

- ***Interest rate:*** This is a frequently used as a causal variable, since increased cost of borrowing is expected to motivate firms or individuals to use alternative sources of funds in place of borrowing funds from banks or financial institutions. This increases the incentive to remain unreported.

[358] M_3 money includes assets that are less liquid than other components of money supply, and are more closely related to the finances of larger financial institutions and corporations than to those of businesses and individuals. These types of assets are referred to as "near, near money." M_3 classification is considered the broadest measure of an economy's money supply and is used by economists to estimate the entire money supply within an economy and by the government to direct policy and control inflation over the medium and long-term.

To mathematically represent, the MIMIC model combines information to estimate an index for unreported income, using a simple regression analysis, in the form of:

$$Y = mx + c,$$

where, y represents a set of dependent variables and x, a set of independent variables; m and c are constants that are estimated. Based on that understanding, the MIMIC model can be represented by a set of two equations connected by an unobserved latent variable, η, as shown below:

$$Y_t = \lambda\eta_t + \varepsilon_t$$

$$\eta_t = \phi X_t + \nu_t$$

In these two equations, Y is a set of indicator variables and X, a set of causal variables. They are connected to each other through a latent variable, η; λ, ε, ϕ and υ are number variables that are estimated through regression analysis. The MIMIC model tracks the latent variable, η, over a period of time, and gives an estimate of unreported income. This estimate used with the reported income, and the tax thereon, gives the tax gap. Tax administrations try and minimise the tax gap to monitor tax collections. These figures of the tax gap are often broken into different geographical regions or into different trades or industry to monitor whether the tax administration has moved towards better tax compliance. The effort is not on how much tax is actually collected but on how quickly the gap is being closed.

Tax collection (for forecasting actual tax) for the prospective year will also have to be forecast. That can be done on the basis of a macroeconomic-based model or tax revenue receipt model. Besides, it is also important to keep in mind that potential tax is not a static number; it is dynamic due to changes in the economy and the tax base, discretionary tax changes, change in foreign exchange rate, changes in the taxpayer base, etc. Thus, potential tax will require to be forecast based on such factor changes, after having achieved a baseline figure on the basis of the three methods mentioned above. This will also mean that the tax gap will change. It, therefore, is important to incorporate these factor changes to arrive at a more realistic tax gap. After arriving at the headline tax gap, an analytical tool can be developed to break that headline tax gap figure on the basis of economic sectors, tax administration jurisdictions, and other such divisions that may be required by the tax administration.

Many tax administrations also have alternative formulations of the MIMIC model: DYMIMIC and EMIMIC. DYMIMIC model uses the data in the first difference, whereas the EMIMIC model uses the concept of co-integration analysis, particularly if the data series are related. Both these alternative models are improvements over the basic MIMIC model.

Indirect taxes

To calculate the theoretical liability to VAT, national consumption statistics can be used to estimate the overall potential collection for VAT, as shown earlier in Table 13.3. The tax gap can be estimated by subtracting actual collections from potential VAT liability.

Sectoral tax gap analysis

Any tax gap analysis is fraught with uncertainty and, therefore, requires a combination of several methods. There are two different types of methods used for calculating the tax gap. The first consists of calculations of the total tax gap from a macroeconomic perspective (top-down). The other consists of attempts to estimate the tax gap using micro data or results from major economic areas using specific models for them (bottom-up). Thereafter, the two results can be compared to arrive at an assessment of the tax gap. Typically, economic sectors that are more prone to have such a gap are taken for analysis as a first step. Many times, such micro models have to identify the forward and backward linkages of the sector, and the identification of such linkages needs to be done very carefully. For example, the construction sector depends on land transport, which in turn depends on petroleum products. Understanding the correlation will be the second key step in the exercise. The input-output model helps in this process. But to understand the comprehensive impact of the sectors of the economy on each other, it may be required to go beyond direct linkages, and explore total linkages. That complicates the model.

(iii) **Models for tax debt analysis**

Tax debt (tax arrears) collection is an important part of tax administration. Tax administrations have so far used traditional methods of collecting taxes, by applying a standard process on all debts more or less uniformly. In this standard process, either the tax debt is collected or after due process, is written off. But many advanced countries have started applying statistical methods to identify debtors rather than identify debts, and understanding their behaviour for collection of tax debts.

Cluster analysis

Cluster analysis is often used to segment tax debtors. In this method, a number of clusters, based on characteristics, such as tax debt size, taxpayer status, underpaid tax, total offset losses across all tax forms, taxes withheld at source, interest amount that has already been taxed, property income, self-employed business turnover, allowable expenses, and other deductions are identified. An average characteristic of each group is compared with other groups. Done iteratively, the clusters diverge and the groups become different. In practice, this means grouping taxpayers with similar characteristics in segments. Each cluster is then assigned a risk factor, to understand the level of attention that may be required. The assignment of the risk factor is on the basis of input from field formations as well as on the basis of data or information from the data warehouse.[359] The ultimate goal is to develop segments of these input characteristics that can be identified as high risk, based on particular patterns that emerge. In short, the analysis involves using a better understanding of taxpayer characteristics to tailor taxpayers' treatments on the basis of individual circumstances and behaviour.

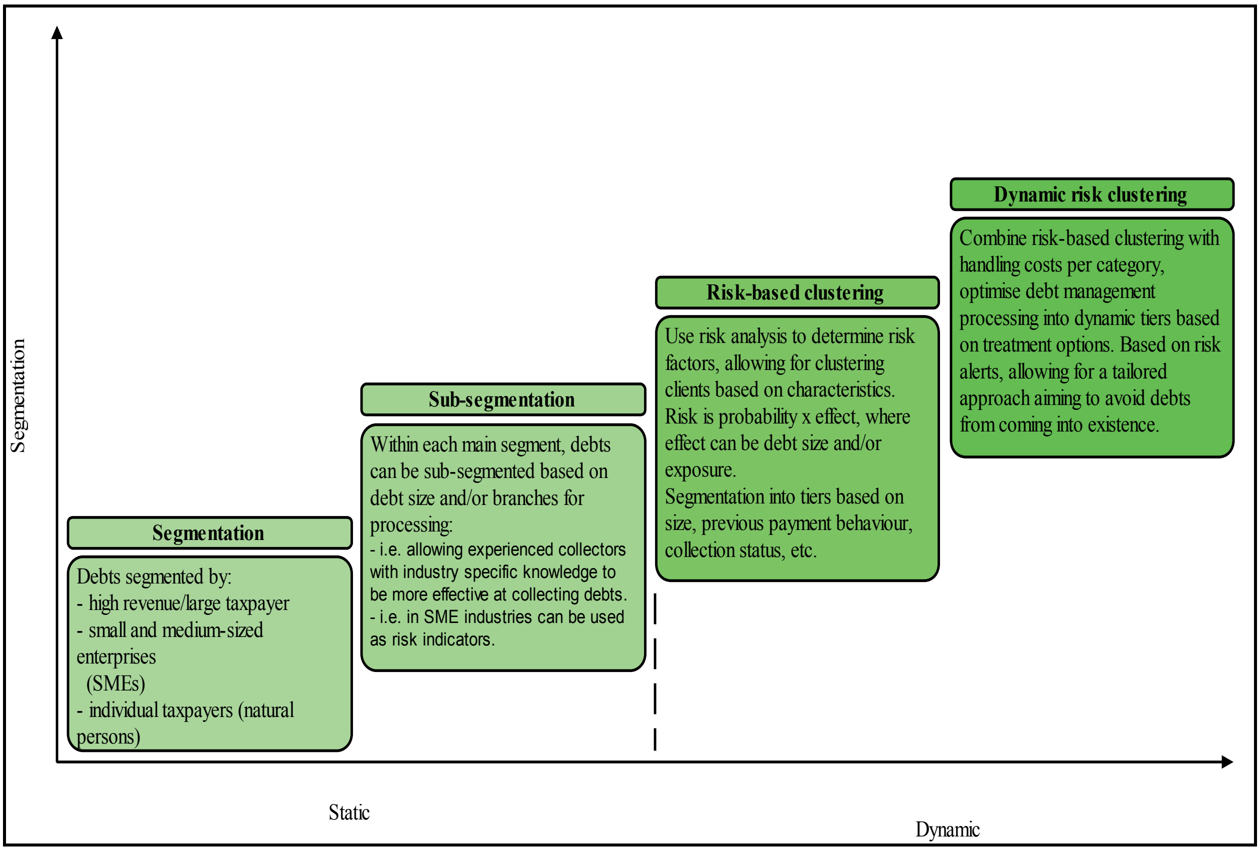
Risk-based clustering, as stated above, clusters together debtors with similar profiles and enables tax administrations to apply tailored tax collection and recovery methods based on the individual circumstances of taxpayers while also focusing on priority cases, such as those with higher yield potential or those in particular risk categories. While risk-profiling can be considered static, risk assessment and risk assignment can be dynamic. In dynamic risk clustering, the aim is to further improve the matching of tax collection and recoveries to each risk cluster and to prevent tax debt. Each payment obligation is classified using predictive analysis and some cases are subjected to action before a liability becomes due in order to reduce the rate of debt occurrence. The specific collection measure chosen for each risk cluster is the one that matches best the profile of the debtors in that cluster. Through this method, emphasis can be laid on identifying the best tax collection treatment, and prevent tax arrears by eliminating ineffective steps. This also makes the cost of collection fall and tax collections improve.[360] This approach enables the debt collection function to combine a tailored approach, providing flexibility of response at the organisational level, with expertise in the application of the tax treatment chosen at the level of the individual staff member dealing with a particular debtor. Tracking compliance performance is an important aspect of risk-based clustering. For a cross section of cases in any particular year, the tracking can be for information on beginning-of-the year stock of cases, in-year settlements and end-of-year stock.

Diagram 13.9 depicts how tax administrations have moved from simple taxpayer segmentation to risk-based segmentation.

[359] Risk models can be further improved by using third party data. Such third party data can be obtained or purchased (for example, from Credit Information Bureau (India) Limited (CIBIL) for early detection of financial difficulties or insolvency. Some tax administrations collect financial data on a regular basis for better tax debt management.

[360] Working Smarter in Tax Debt Management, OECD, 2014

Diagram 13.9: Risk-based segmentation for tax debt collection



Source: Working Smarter in Tax Debt Management, OECD, 2014

Discrete Event Simulation and System Dynamics models

Apart from tax debt management using cluster analysis, tax administrations have also moved on to tax debt modelling to forecast the overall debt balance on the basis of underlying economic factors. Different models are employed for debt analysis, such as discrete event simulation (DES) and system dynamics (SD). Both these models are based on simulation techniques; in the model based on discrete event simulation, the simulation is in discrete time and in system dynamics in continuous time. In the DES model, a flow of events is iteratively worked out on the basis of defined probability distributions, and so each debt is modelled individually and its movement tacked. In the SD models, the work is through continuous time, providing simulated solutions of high order differential equations. Thus, debts are modelled as stocks flowing through the system, with flow rates and routes decided by data inputs. In both models, debt items and debt values are kept separate, but in DES, linking discrete information pieces together is difficult in contrast to SD where information flows are continuous.

XIII.3 Current practices in India

At present, revenue forecasting is carried out independently, on an annual basis, by the TPL unit in the CBDT and the TRU in the CBEC for the taxes they administer. They largely use macroeconomic indicators and apply them to their respective models/methods, discussed below in detail. Besides, both TPL and TRU also get tax collection estimates by the Chief Commissioners for their regions/zones, who, in turn, ascertain the likely tax collection estimates from their respective Commissioners. The two figures – one, on the basis of their own model and the other on the basis of estimates communicated by the field formations – are then compared and some reasonable estimate is arrived at for the figures to be communicated to the budget division in the Ministry of Finance. The likely tax collections so arrived at are also adjusted for the projected tax gain or loss on account of legislative changes or tax administration changes. It was generally felt during the TARC's interaction with the CBDT and CBEC that revenue targets are being unrealistically fixed by the Ministry of Finance. This has happened particularly after the FRBM-related fiscal monitoring. The CBDT and CBEC felt that the revenue projections arrived at by the TPL and TRU are fairly accurate; tax departments' actual collections were within a reasonable error percentage –1 to 3 per cent of the projections. These projections are enhanced for FRBM-related deficit calculations. In a good economic environment, the gap may be narrow, but in a bad situation, the gap may be quite wide.

Macroeconomic figures, such as GDP, inflation, exchange rate, etc., for the above exercise are provided by the economic division in the Ministry of Finance; however, the economic division itself works on the likely tax revenue that the CBDT and CBEC will collect during the coming financial year. There is no information about the model used for the estimation. The figure in the budget for tax estimates is put in place after looking carefully into the two projections, one by the CBDT and CBEC and other by the economic division in the Ministry of Finance. This exercise of tax forecasting for the next financial year is done between November and February of the previous fiscal year. At that time, the tax departments also estimate the revised tax collections for the current fiscal year. The revised estimates are arrived at taking into account the "asking rate" of the last financial year, a rudimentary moving average method. This "asking rate", being the actual tax collection of the last fiscal year, incorporates the seasonality of tax collections and provides an assessment of what will be collected in future months. This is normally done in the month of January every year. It is generally seen that revenue collections are higher in the last quarter of the year as compared to that in the previous quarters.

Tax forecasting

The TPL in CBDT uses average tax buoyancy for the last four financial years to arrive at the tax forecast for the 5th financial year.[361] The tax buoyancy is the point estimate for each of the four financial years, using the growth rate of GDP (used as tax base proxy) and tax revenue, taken on

[361] This is on the basis of oral discussion with the TPL, CBDT on December 22, 2014, followed by written communication dated January 8, 2015.

a nominal basis; gross tax figures are normally used to calculate percentage tax growth. The buoyancy is calculated separately for personal income tax and corporate income tax; besides, separate projections are made for the two taxes. Refund figures, based on data from the Central Processing Centre (CPC), are plugged in separately to arrive at net tax calculations. The direct taxes figures are arrived at by taking sum of the two figures.

The CBDT has also developed a discretionary change model for some sectors of the economy on the basis of actual tax return data for at least 10 years for corporate tax and three to four years for individual taxpayers. This has helped them calculate deductions as well as exemptions for a "what if" analysis. At present, the tax department has 75 business codes that are used for taxpayer segmentation but many taxpayers do not quote this. No sampling exercise is done; whole taxpayer return data is used for any analysis.

The TRU in the CBEC makes revenue estimates based on inputs from many other departments, including the GDP growth forecast by the economic division in the Ministry of Finance. Revenue outflows on account of refunds, drawback and free trade agreements (FTAs), etc., are also taken into consideration before arriving at tax revenue estimates. Overall, tax revenue projections are largely based on the previous fiscal year's commodity-wise revenue collections and on the basis of inputs received from the CBEC's field formations. Before informing the CBEC about revenue projections, Chief Commissioners in their respective jurisdictions consult major taxpayers and, based on such consultation, they give estimates of revenue collection for the next financial year.

As part of tax revenue forecasting, the TRU also undertakes sectoral analysis taking into account past trends, growth prospects, the impact of policy level changes, sector analysis available in credible publications and information provided by the respective ministries (for example, growth projection in the petroleum sector is provided by the Ministry of Petroleum and Natural Gas). The import/export volume growth is estimated on the basis of historical trends. Based on these inputs, tax forecasts are consolidated and validated on the basis of figures reported by the Central Statistical Organisation, and due consultations with the budget and economic affairs division of the Finance Ministry taking into account the overall economic and fiscal scenario. The likely outgo of refunds and duty drawback are considered separately to arrive at net tax collections.

Overall, neither the TPL nor TRU employs detailed revenue forecasting models, unlike other developed and developing countries. The models are rudimentary and, at times, devoid of a theoretical foundation. Communications from both the CBDT and CBEC are placed in Appendix XIII.2.

Revenue foregone statement

Since 2005-06, a statement and analysis of revenue foregone for both direct and indirect taxes has been included in the annual budget as Annex 15 of the receipts budget. The revenue foregone statement for direct and indirect taxes is based on the potential revenue gain if exemptions, deductions, weighted deductions and similar measures are removed. The assumptions and methodology adopted to estimate the revenue foregone on account of different tax incentives are indicated in the statement itself in Annex 15.

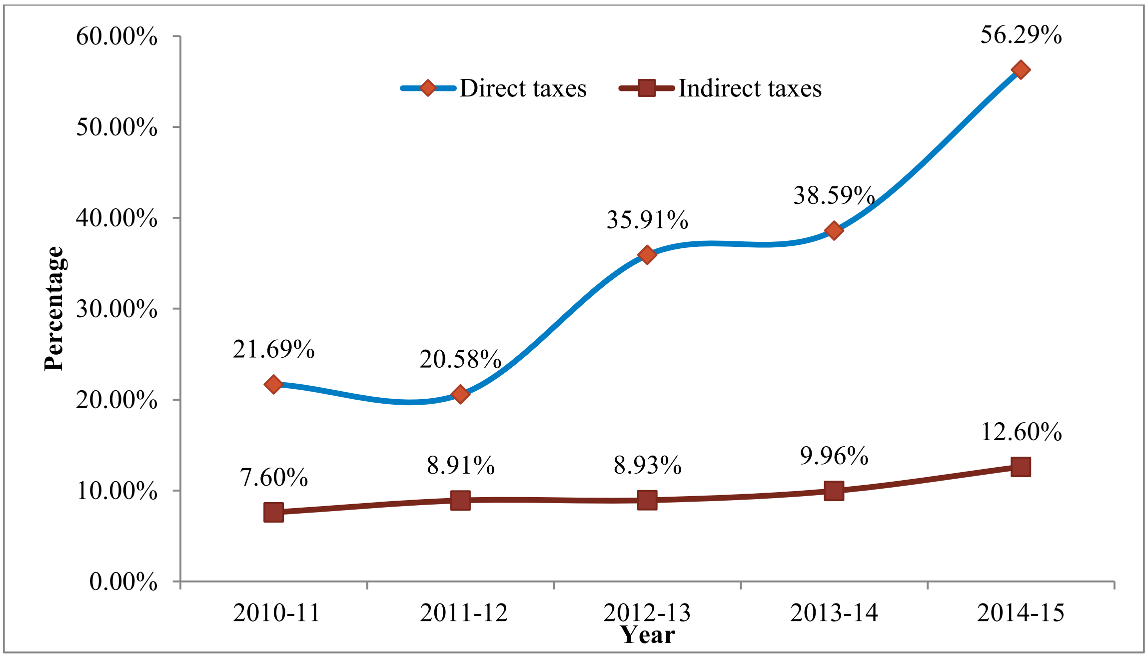
Staff position

At present, there is only one Director in each Board, who carries out tax revenue forecasting-related work. This assignment is in addition to other work assigned to him/her in the TPL or TRU, as the case may be. The Director works under the Member (Legislation) of the respective Board. The tax revenue monitoring for the year is done by a separate Director in both the Boards. In the CBEC, the Director responsible for tax revenue forecasting-related work is assisted by an Indian Statistical Service officer. No such help is available, at present, for the Director in the CBDT. In general, neither the Director in the CBDT nor the CBEC is required to have a formal education in economics or statistics for his/her appointment for the job. General tax administration experience is considered sufficient.

Tax debt management

Tax arrears constitute a sizeable proportion of total tax collections, more in the CBDT than in the CBEC. Graph 13.1 shows the growth of tax arrears as a percentage of the budget estimates of tax collection.

Graph 13.1: Arrear demand as percentage of the BE tax collections



Source: Annex 11 of the Receipts Budget, Ministry of Finance

Both the CBDT and CBEC employ well-delineated processes to collect tax arrears. The emphasis of these processes is on the amount to be collected rather than on the long-term attitude of the taxpayer towards the payment of taxes. The tax arrears recovery laws in both direct as well as indirect taxes are based on the Civil Procedure Code, 1908. Once the demand for tax, duty, penalty, or any other tax dues gets confirmed by the tax authorities, the taxpayer liable for payment of tax dues is required to deposit the amount within a prescribed time. When such government dues remain unpaid beyond a reasonable period of time, tax authorities are required to initiate recovery proceedings in accordance with the provisions of respective acts. Section 142 of Customs Act, 1962, (applicable also to central excise) provides that if any duty is demanded or if drawback paid or any amount erroneously refunded is recoverable from a person, it can be recovered either by deducting the sum from any amount payable by any customs officer to such person or by detaining and selling such goods belonging to the said person, which are under the control of the customs authorities. Chapter XVII of the Income Tax Act, 1961, deals with collection and recovery of taxes. Sections 222 to 232 of the Act and Schedules II and III of the Act, and the Income Tax (Certificate Proceedings) Rules, 1962, together constitute a self-contained code prescribing the various modes to recover arrears of taxes under the Act.

If full recovery of dues fails even after the application of these provisions, the Customs and Central excise department can issue a government dues certificate to the District Collector of the district where any property of the liable person is located or where he carries on his business. The concerned District Collector can then attach and auction the belongings of the concerned person to recover the amount as arrears of land revenue, a procedure known as certificate action. The property of the defaulter can also be seized and sold by the department. In the case of direct taxes, recovery proceedings can be initiated by the Tax Recovery Officer against an assessee when he is in default or deemed to be in default under the provisions of Section 220(4) of the Act by issuing a recovery certificate. The recovery processes in both direct as well as indirect taxes include attachment and sale of the defaulter’s movable and immovable property, and in the case of recalcitrant tax defaulters, arrest and detention in a civil prison. In both the cases, despite the legal methods available to tax officers to collect tax arrears, tax collection has not been efficient, leading to sizeable taxes remaining uncollected.

The organisational structure for recovery of tax demand is almost the same for both direct and indirect taxes. The difference is that in the case of direct taxes, the tax recovery officer is in every range of the commissionerate for recovery from tax defaulters after the recovery certificate is issued and for the other cases, the assessing officer is responsible for recovery of tax demands; In the case of indirect taxes, each commissionerate has a tax recovery cell, headed by an Assistant Commissioner level officer. At the apex level, in the case of direct taxes, there is a Directorate for recovery, an attached office of the CBDT, which monitors the collection of tax arrears and is responsible for suggesting changes in the recovery procedure, including amending the manual for tax recovery from time to time; in the case of indirect taxes, there is a unit of the CBEC, called Tax Arrear Recovery (TAR), headed by an officer of the rank of Chief Commissioner, responsible for collecting tax arrears.

XIII.4 International practices

Revenue forecasting is an essential part of budgeting in all countries and, hence, they make efforts to obtain reliable figures for expected revenues, with the clear understanding that revenue forecasts are associated with uncertainties, such as macroeconomic risks, or uncertainties about tax laws and their enforcement. Forecasting is made even more uncertain and difficult by changes in tax laws and structural changes in the economy resulting from domestic developments or from changes in the international economic scenario (for example, volatility in commodity prices and oil prices). Although these challenges are faced by forecasters in all countries, there are significant differences in the practice of revenue forecasting and in methodologies used to remove these uncertainties and arrive at more robust forecasts. There are also important differences among countries in the institutional aspects of revenue forecasting. In several countries, the government is directly in charge; other countries assign the forecasting task to research institutes, and emphasise the independence of forecasting. Given the efforts that some countries devote to ensuring independence from possible government manipulation, it is worth exploring whether this independence has a noticeable impact on the quality of forecasts. Another question that comes up in this context is whether forecasting performance is affected by the different practices and methodologies involved.

XIII.4.a Methods adopted

International practice on tax revenue forecasting is mainly centred on the broader issue of budget and deficit forecasting. For example, countries in the European Union entered into a Stability and Growth Pact, a rule-based framework for the co-ordination of national fiscal policies that was established to safeguard sound public finances and to monitor the budget deficit in a collective manner. This framework spurred them to set up revenue forecast and expenditure forecast models to arrive at budget deficit projections. The forecasting methodology used by these countries has several similarities; the forecasts are prepared in a disaggregated fashion for a number of single taxes; often individual taxes are aggregated into groups especially if they share the same income source or taxpayer. This partly reflects the need to employ up-to-date information on current revenues.

A typical feature of revenue forecasting methods employed by many advanced economies is that taxes that are strongly influenced by macroeconomic developments, such as corporation taxes or wage and income taxes, are forecast using elasticity. In these cases, revenue estimation relies on information on macroeconomic variables such as GDP, or components of the national accounts, like income from entrepreneurial activity and capital or wage bill. But in some countries, such as the United Kingdom, the Netherlands, New Zealand and Japan, econometric models are also used where the relationship between revenues and economic indicators is directly estimated using regression analysis. In some other countries, for example in Canada, the Netherlands and the United States, micro-simulation methods are used to forecast income tax, and in the United Kingdom and the United States, to predict corporation tax revenue. The method focuses first on a single tax entity for which tax liability is calculated; this is then aggregated to find the total tax liability. There are, however, some differences in these forecasting methods, particularly due to the country's economy characteristics and the tax structure. There are also differences in the timing of the forecasts. HMRC forecasts onshore corporation tax by breaking it down into three sectors – industrial and commercial companies, life insurance companies, and financial sector companies excluding life insurance. The model projects each component of corporation tax calculation, both on the income side (e.g., profits, capital gains) and deductions (e.g., capital allowances, group relief, trading losses carried forward). These are projected using the appropriate economic determinant or econometric equations. The corporate tax forecast uses accruals data, which is converted back into receipts. A comparison of the corporate micro-simulation models adopted in three European countries is given in Table 13.4.

Table 13.4: Comparative description of three existing corporate tax models in EU countries

Characteristics	Italian Corporation Tax Model	UK Corporation Tax Model	European Tax Analyser (ZEW)
Scope of the model	Italian business tax regime	UK Corporation tax regime	Germany, France, UK, Netherlands (extendable to other countries)
Methodological approach	▪ Micro-simulation model ▪ Application of tax rules to business profits ▪ Modules (Fiscal adjustment, corporate income, and corporate tax) to simulate tax rules and compute tax liability	▪ Micro-simulation model ▪ Transition probabilities assigned to each company ▪ Exploits macroeconomic forecasts ▪ Separate forecasts for capital gains, financial sector, giant companies, and North Sea Oil companies ▪ Aggregation at macro level	▪ Model firm approach ▪ Forward looking approach ▪ Computes effective average tax rate for investments generating economic rents
Purpose	To estimate tax liabilities of firms and ex-post marginal tax rates.	To forecast corporate tax revenues for the UK economy.	International computation and comparison of firm's tax burden.
Data description	SCI Data of 8279 enterprises (with more than 100 employees) subject to corporation tax for the year 1998.	Sample of 12,000 companies paying CT in manufacturing, distribution, and other sectors, selected	A German manufacturing company of medium size taken as 'Model firm'.

Characteristics	Italian Corporation Tax Model	UK Corporation Tax Model	European Tax Analyser (ZEW)
		randomly from a larger sample of 150,000 companies. Separate model for financial sector.	
Sources of data	SCI annual survey data (collected by ISTAT) integrated with published accounts data (from Italian Chamber of Commerce) available at ISTAT.	Actual company tax returns for past three years.	Official German statistics (DEUTSCHE BUNDESBANK) used to construct models firms of a certain size and type.
Limitations (biases) of data	▪ Sample covers only major firms (2 per cent of the total firms liable to CT). ▪ Firms in sample represent 52 per cent of total workers, 52 per cent of total turnover, and 68 per cent of the profits. ▪ Data excludes agriculture and fishing, and financial sectors.	▪ Sample not representative of companies' accounting periods.	▪ No actual data on firms. ▪ EATR depends on concrete individual cases; therefore, it is not possible to make universally valid statements regarding differences of the EATR across countries.
Base line period	1998	Forecasts done every year	Tax systems as on 1 January 1999.

Source: IAS/IFRS in Belgium: Quantitative analysis of the impact on the tax burden of companies, (2007), Jacqueline Haverals, Journal of International Accounting, Auditing and Taxation, 16, pp 69-89

For oil and gas revenues, comprising petroleum revenue tax and offshore corporation tax, an important tax stream for the UK, the HMRC uses another micro-simulation model that uses production and expenditure data on each individual oil and gas field. This model reflects the large swings in oil and gas prices. The model is based on econometric equations.[362]

HMRC has also set up a Personal Tax Model (PTM), a micro-simulation model based on a survey of taxpayers' liabilities. The model generates forecasts of marginal tax rates, taking account of reliefs and allowances. It also estimates the annual cost of indexation of tax thresholds. The self-

[362]For details, see Tax Shastra, Parthasarathi Shome, Business Standard, 2012.

assessment forecast is compiled on a liabilities basis and then converted to receipts. There are a number of other smaller components of income tax, such as the tax deduction scheme for interest and company income tax. Capital gains tax is estimated using a model separated into financial and non-financial assets. Historical data on the length of ownership and the financial/non-financial split are used in the model. These data are obtained from tax records.

The UK HMRC has set up a CGE model, using national accounts data published by the Office for National Statistics (ONS), survey data on households and firms, and other administrative data, to capture the complex transactions in the economy to analyse how GDP would be affected by raising alcohol duties, tobacco duty, air passenger duty, vehicle excise duty, insurance premium tax and fuel duty and what distortions they create. In their analysis, the HMRC has found that alcohol and tobacco duties are least distortive because these are close to the point of final consumption even if these duties are collected at the factory gate. Conversely, a tax on business insurance and petroleum products tend to fall near the upstream supply chain. This analysis is carried out by simple exercises on the CGE model. The more distortion a particular tax causes, the greater is the fall in GDP, initially as well as in its future trajectory. HMRC has also set up separate models to estimate the revenue impact of various excise duties, using the price elasticity of demand.[363]

Germany and Belgium employ macroeconomic revenue forecasts using buoyancy and elasticity methods. They also use regression analysis to forecast excise taxes or trade taxes. Methods, like trend-extrapolation or more formal time-series analysis or vector-autoregressive methods are also used by some countries to predict expected revenue for less important taxes, using data on revenue from of the respective tax from previous years.

In many countries, for example, the United Kingdom and the Netherlands, revenue forecasts are embedded in the overall macroeconomic model and they produce a consistent set of macroeconomic and revenue forecasts. Other countries, for example Ireland, New Zealand, and the United States (Congressional Budget Office), use these forecasts for mutual verification of the results. In New Zealand, tax forecasts are prepared more-or-less concurrently with economic forecasts. As each part of the economic forecast is prepared, tax forecasters prepare forecasts of the relevant tax types. The New Zealand Treasury uses mainly spreadsheet-based tax forecasting models.[364] New Zealand produces these forecasts twice a year. The tax forecasts are prepared in

---

[363]The rate of tax on these products should be related to the measure of consumer responsiveness to price, or the price elasticity of demand: (Change in quantity demanded / base value quantity demanded) / (Change in price / base price)

[364]It appears that the Treasury use a multiplicative model for the monthly, quarterly or annual levels of taxation revenue considered. A simple example of that model can be, $Y_t = \alpha X_t^{\beta} e_t$, where Y_t denotes a particular tax revenue, X_t denotes a macroeconomic predictor such as GDP, and e_t denotes a multiplicative error that varies about a mean of unity. If the transformed macroeconomic predictor X_t^{β} can be thought of as a proxy for the relevant tax-base, then α can be interpreted as a mean tax rate. Other multiplicative variables can be included and parameters such as α and β may also be time-dependent. Monthly or quarterly variables may have seasonal variation and all are likely to be affected, to some degree, by longer-term economic cycles.

terms of both receipts, the tax actually paid to the Inland Revenue Department, and revenue, the tax that is actually due, regardless of whether or not it has been paid.

Ghana has set up a micro-simulation based trade tax calculator that uses a large amount of actual raw trade data collected at the entry ports to analyse the impact of proposed policy changes on projected revenues from trade taxes. This model produces reports that, for example, show the amount of taxes that would have been collected from exempt goods that are imported by certain sectors/groups in the economy. This model uses the international 4-6 digit HS code for product identification and other details such as country of origin, and exempted import transactions, preferential tariff rates, etc. The model has three main components: (a) trade database, (b) tax policy tables and (c) the model output tables. The trade database is the core of the model and contains detailed import transactions captured by the customs system, such as ASYCUDA (Automated System for Customs Data) or any other country-specific system. The transaction information that is stored in the trade database includes the transaction date (month and year), entry port, H-S code, country of origin, import currency, import quantity, import value in foreign and domestic currency, and all taxes applied to the commodity at the point of entry. The tax policy tables contain both the current and proposed import duty rates and the rates of other types of taxes and fees imposed at the point of entry for each commodity in the harmonised system. Depending on the types of commodities imported into the country, the table, based on HS codes, can be quite large (more than 6,000 records). In many cases, however, the tax structure for commodities with the same first 4-6 digit HS code will be the same. The current and proposed relief and exemption rules for different transactions are also stored in the tax policy tables. The model output tables retrieve data from the trade database and summarise the projected trade tax revenues under the current and proposed tax structures. The summary reports generated by the model can be broken down by various categories, such as tax rates, broad economic categories, harmonised section, customs procedure codes, and country of origin.

Tax gap analysis

Tax gap is increasingly being used not only for the forecasting of potential tax revenue, but also for compliance management. The analysis has helped many tax administrations improve their efficiency in resource allocation within a revenue authority to combat non-compliance and hence, has been increasingly used as a measure of the effectiveness of a revenue authority. This is why the IMF and the World Bank use tax gap estimates as a performance measure for governments in their fiscal stabilisation programmes. But that apart, a number of advanced economies, such as France, Sweden, the US, and the UK, have unilaterally undertaken tax gap estimates for different tax types. Some of these countries estimate tax gap on an annual basis or even greater regularity. In these cases, the tax gap estimates are calculated to provide guidance on the magnitude of the gap rather than for precise year-to-year measurement. The US IRS uses tax gap estimates for compliance management. Many US states also undertake separate tax gap studies for sales tax (in Minnesota) and income tax (Minnesota and California). The UK HMRC has undertaken the estimation of the VAT gap[365] annually since 2001, on the basis of consumption data from national accounts and household data. These gap estimates are given wide publicity and are used as a key performance indicator of the staff for collecting VAT, as well as for performance agreements between employees and the government. These VAT gap studies also helped the UK's HMRC to identify areas that require legislative changes or amendments. For excise duties and VAT, the calculations go back even further. Calculations of the tax gap have been conducted since 2004. HMRC first published measures of the overall tax gap in 2009.

The Danish tax administration, SKAT, has conducted measurements using random audits since 2006. In both the United Kingdom and Denmark, tax administrations conduct recurrent measurements of the tax gap. While HMRC updates its measurements (including retrospective recalculations) every year, SKAT does this every alternate year. The measurements in the UK aim to place a quantitative value on all tax gaps, including the hidden economy. An annual update is produced and published to be used by a number of different stakeholders.

Since 2007, the Netherlands and Sweden have also set up annual programmes to map its tax gap. The analysis encompasses individuals and small and medium-sized businesses. In addition, national accounts data are used for top-down measurements of the VAT gap.

The general approach taken in the United States, Canada and New Zealand is a 'bottom-up' approach, which is based on the availability of taxpayer information. Individual tax liability, in a typical manner, is forecast and then aggregated to produce an overall revenue forecast.[366] HMRC and the US IRS have never used top-down methods to estimate direct tax gaps, the reason being that there are no reliable estimates of under-declared income. Most administrations focus on identifying and assessing risk factors and prioritising compliance resources to areas of highest risk. Top-down estimates are of no value for decisions on compliance priorities.

The Australian Tax Office (ATO) also makes revenue forecasts using a 'bottom-up' approach and uses econometric methods to estimate revenue.[367] The Revenue Analysis Unit (RAU) in the ATO is currently looking at a micro-simulation model for personal income tax. While data limitations had previously constrained the use of micro-simulation techniques, arrangements have now been put in place to allow the RAU to access detailed confidential personal income tax data from the ATO, with appropriate safeguards in place. Some other countries, including the United States, employ micro-simulation techniques in their personal income tax revenue forecasting.

---

[365] The VAT forecast is based on the concepts of theoretical liability and the VAT gap. The gap is the difference between actual VAT receipts and the total value of VAT that could be theoretically collected from the tax base were all tax payers fully compliant. The gap is made up of error, fraud, evasion, avoidance and debts owed by firms to HMRC. Subtracting the forecast VAT gap from VTTL (VAT theoretical total liability) gives the VAT forecast.

[366] Review of Canadian Federal Fiscal Forecasting : Processes and Systems ,Tim O'Neill, June 2005, available at https://www.fin.gc.ca/activty/pubs/Oneil/PDF/Oneil_e.pdf, accessed in December, 2014

[367] Appropriateness of forecasting methodology, available at http://treasury.gov.au/~/media/Treasury/Publications%20and%20Media/Publications/2013/forecasting_review/downloads/PDF/08_section2.ashx, accessed in December, 2014

New Zealand initially used the latent variable method to estimate the tax gap, but later decided not to estimate a tax gap at all.

The reason for not using the top down methods was that the data lacked reliability and did not indicate the tax gap. Scotland uses the 'top-down' approach to forecast total taxation revenue and then disaggregates the figures into different heads of revenue.[368] In Denmark, the personal tax gap was estimated as the difference between personal income in the national accounts and income declared in tax returns. In Latin America, work on the tax gap has largely focused on VAT rather than on direct taxes. Unlike Denmark and the UK, where the tax gap is estimated each year as a performance measure, the Latin American countries use tax gap estimates for broad comparisons between different taxes. Table 13.5 gives country-wise use of the 'top-down' method.

Table 13.5: Country-wise use of Top-down method

Country/ region	Method used and what is or was estimated	Application of results
Denmark	Discrepancy using the income measure of GDP, pre-tax personal income tax gap	Used to calculate a performance objective, but administration is seeking to measure objective bottom up
Latin America	Calculation of theoretical liability from income, corporate tax gap direct surveys of households, personal tax gap	Broad assessment of relative levels of tax gaps for different taxes
New Zealand	Latent variable method, total tax gap	None – the administration has considerable reservations about the methodology
Sweden	Discrepancy using the income measure of GDP, personal tax gap	Reconciliation with bottom up concealed income estimates. No plan for further top down work

Source: The predictability of the top-down approach to estimating the direct tax gap, Marcus Rubin, HMRC

The New Zealand Treasury relies heavily on mapping technology to produce revenue forecasts.[369] Mapping models appear to be a widely used technology for countries where revenue forecasts are prepared in conjunction with macroeconomic forecasts. This technology uses the econometric models to estimate tax revenue elasticity and, in some instances, micro-simulation models. The New Zealand Treasury uses mainly spread sheet-based tax forecasting models and procedures comprising four phases: (a) determine the nominal tax revenue for the previous available year,

[368] Scottish Tax Forecast , March 2012, available at http://budgetresponsibility.org.uk/wordpress/docs/Scottish-tax-forecast-March-2012.pdf

[369] An Analysis of Tax revenue Forecast Errors, New Zealand Treasury, W*orking paper 07/02,* available at http://www.treasury.govt.nz/publications/research-policy/wp/2007/07-02/twp07-02-pt1.pdf, accessed in January 2015

which is the base year, (b) adjust the nominal tax revenue for the base year by removing any known anomalies to establish the true underlying tax position for that year, (c) apply the forecast growth rates of relevant macroeconomic variables(s) to forecast tax for one to five years ahead, applying elasticity if required, and (d) adjust the tax forecasts for irregularities such as tax policy changes, expected shifts in payment dates or taxpayer behaviour, and include any judgmental forecasting adjustments that may be deemed appropriate.

Apart from the above, some developing countries, for example, Lithuania, have set up small tax forecasting models to generate receipts forecasts as part of their annual budget process and to monitor receipts over the course of the year relative to amounts projected in the budget, using a tax receipt monitoring and forecasting model.[370] The model uses actual monthly receipts data and forecasts receipts for the current and subsequent years based on expected changes in real GDP, the price level, tax rates, and any behavioural effects that may be associated with changes in tax rates. The projected receipts, being a sum of actual and projected, are adjusted for the expected change in the tax base holding the tax rate constant, change in the tax rate, and change in the tax base associated with the change in the tax rate. The model, simple yet useful, has been in use in some other developing or transition economies such as Bhutan, Thailand, the Philippines.

XIII.4.b Organisational/institutional arrangements

For most tax administrations, there are separate units or offices in the Ministry of Finance or the Treasury Department (often called Fiscal Policy Unit (FPU) or in some countries, Fiscal Policy and Analysis Unit),[371] which provide revenue and expenditure estimates, estimates of inter-governmental finances, evaluations of proposed legislative changes, and analyses of the government's budget. These units independently perform revenue forecasting and tax analysis. Besides, tax administrations also have their revenue forecasting units and these units base their revenue projections on data available to them. These two agencies hold annual or bi-annual meetings between them to compare their forecasts and achievements so as to effect a collaborative arrangement.

In some countries, independent agencies also aid the fiscal policy unit in the Ministry of Finance in arriving at tax projections. Independent reviews by the parliament or the legislative body of tax estimates arrived at by the government are also seen in some advanced economies. The US and the UK are examples. In the US, the Congressional Budget Office (CBO) staff carefully tracks the economy all year long, but the formal forecasting process for its January report begins by examining the forecasts generated by private forecasting companies, such as Macroeconomic Advisers. These forecasting companies use macroeconomic models that assume the structure of

[370] A tax receipts monitoring and forecasting model for Lithuania, William B. Trautman, HIID, 2000

[371] FPU, in some cases, is also with agencies outside the government. For example, *Instituto de Estudios Fiscales* (Institute for Fiscal Studies) in Spain performs the FPU function for the Ministry of Finance; Centre for Fiscal Policy in Russia is a non-governmental agency carrying out fiscal policy analysis on a contractual basis.

the economy to be constant.[372] Table 13.6 summarises the involvement of non-government agencies in tax forecasting work.

Table 13.6: Involvement of non-government agencies in forecasts

	Involvement of non-government agencies in forecasts		**Ex-post assessment of forecasting performance**		**Availability of fiscal performance information**
	Macro	Revenue	Self	External	Detail and Regularity
Australia	Medium	Low	Regular	Occasional	Medium
Canada	High	Medium	Regular	Occasional	High
Germany	Medium	High	Occasional	Occasional	Low
Netherlands	Medium	Medium	Regular	No	Low
Sweden	Low	Low	Occasional	No	Low
Switzerland	Low	Low	Occasional	Occasional	...
United Kingdom	Low	Low	Regular, legal	Regular	High
France	Medium	Low	Regular	Regular	High
Italy	Low	Low	...	No	Low
New Zealand	Medium	Medium	Regular	Occasional	High
United States	...	...	Regular	...	High

Source: OECD

In the US, the CBO and the US Treasury provide two forecasts per year. The former usually reports to the Congress in January and August while the latter reports in January or early February and in July. CBO's January forecast is normally the most important because the macroeconomic assumptions that it generates are used to derive the spending and revenue baseline that will be used by Congress throughout the year. The same assumptions are often important in estimating the effect of tax policy changes on revenues or programme changes on outlays. The Office of Tax Analysis (OTA) in the US Treasury Department develops, analyses and implements tax policies

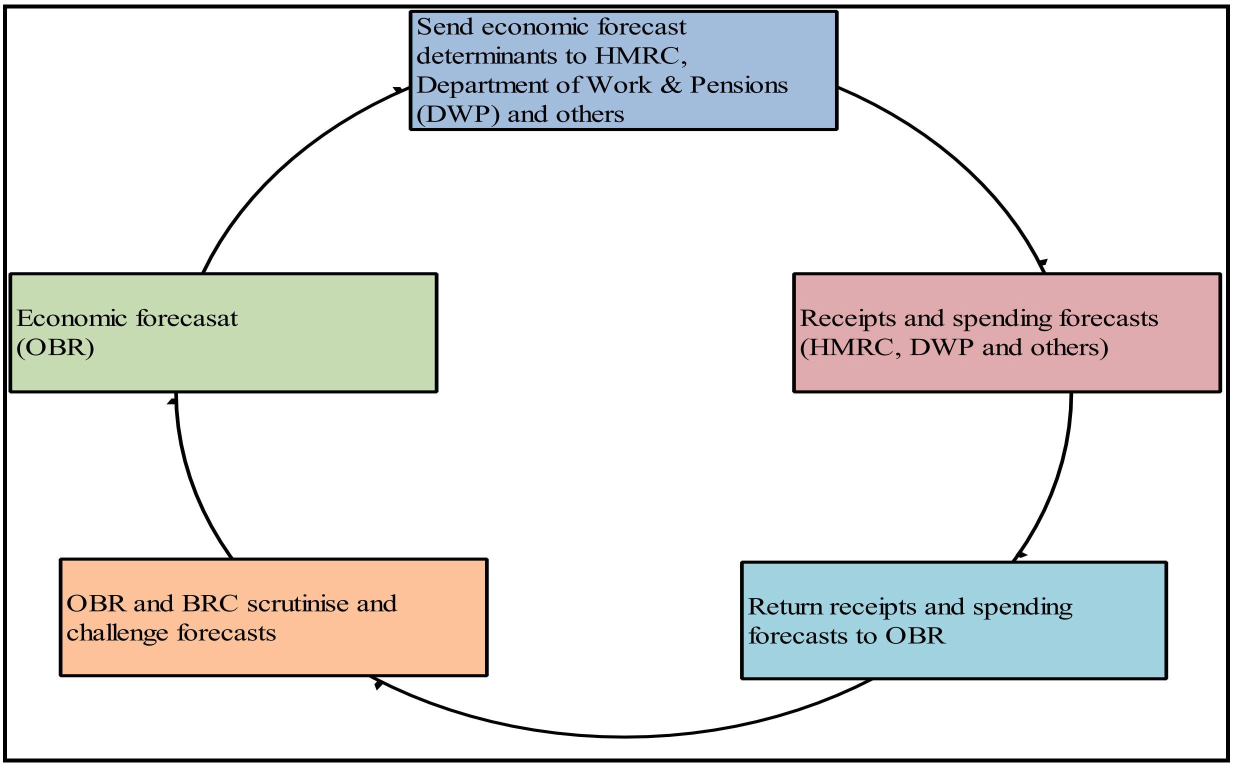
[372] This assumption has been challenged on the ground that that the structure of the economy is constantly evolving and attempts to estimate the parameters of equations that assume a constant structure can lead to meaningless results.

and programmes. The OTA uses modern analytical tools, including micro-simulation models, and maintains large statistical databases to analyse the economic, distributional, and revenue effects of alternative tax proposals. The OTA also has regular in-depth discussion with the US IRS and carries out studies on tax compliance and taxpayer compliance costs, and developing and revising tax forms. This, in turn, improves tax administration.

After the budget is presented to the Congress, but before it is discussed, the CBO provides estimates based upon its own model and informs the public at large, including the Congress and the senate, whether the tax revenue forecast is achievable or not.

The UK's budget making process provides a variant of the US process. In the UK, the HMRC forecasts tax revenue estimates on the taxes administered by them. These estimates are then verified independently by the Office for Budget Responsibility (OBR). The OBR is a newly constituted body set up in 2010 to review budgetary estimates. The OBR has been provided access to economic data by the government. Diagram 13.10 depicts the fiscal forecast process in the UK.

Diagram 13.10: Fiscal forecast process in the UK



Source: HMRC

The Australian Treasury forecasts revenue. The Revenue Analysis Unit (RAU) in the Australian Treasury works closely with the ATO to improve its technical understanding of taxable bases and the tax payments system, which helps them in arriving at tax forecasts. The forecasts are then subjected to a formal peer review at an annual revenue conference held jointly with the ATO and the Australian Customs and Border Protection.[373] In New Zealand, the Treasury Department forecasts the revenue for the government. The Inland Revenue Department (IRD) also independently forecasts revenue. But, the Treasury's forecasts are taken to be the Crown's official forecasts.[374] However, the Treasury department and the IRD discuss them.

The Canadian Department of Finance adopts a different method for tax revenue forecasts. The revenue projections as well as fiscal projections are made by private sector forecasters. These private sector forecasts are translated into a fiscal forecast, and they provide the difference between status-quo government revenues and expenses and the fiscal room available for new initiatives.[375]

XIII.4.c Staff resources

The staff for revenue forecasting in most tax administrations are trained technicians, with university level education in economics, statistics and other related disciplines. Staff are further trained to acquire specialised skills and knowledge. The senior staff normally use extensive analytical techniques for data analysis and examine the basis of existing models to see their relevance to extant fiscal policy issues, and to ensure that appropriate outputs for presentation of results are arrived at. These technical personnel are expected to perform basic tasks, such as running the models under the guidance of a senior member, gathering raw data from other agencies, and performing the more routine analytics.[376] The role of the fiscal policy unit leader, a Director, is to advise the Minister of Finance, other senior government officials and parliamentarians on a broad array of fiscal concerns, including macroeconomic issues, tax policy, revenue estimation and forecasting and other related issues. The Director also oversees the development of analytical models, manages day-to-day operations, and establishes capacity building programmes. The Revenue Analysis Unit in the Australian Treasury has around 10 staff members to perform these tasks. The Treasury has recently also hired a specialist to oversee the quality assurance of the output and examine the rigour of tax forecasting modelling techniques to ensure that the models are at the cutting edge of forecasting practice, within the overall modelling strategy adopted by the Treasury. In New Zealand, three staff members are directly involved in detailed tax forecasting, with the responsibility for preparing the forecasts.[377]

---

[373] Review of Treasury Macroeconomic and Revenue Forecasting, December 2012, available at http://www.treasury.gov.au/~/media/Treasury/Publications%20and%20Media/Publications/2013/forecasting_review/downloads/PDF/forecasting-review.ashx, accessed in December 2014.

[374] http://www.treasury.govt.nz/government/revenue, accessed in December, 2014.

[375] Briefing Note: The Government of Canada's Economic and Fiscal Forecasting Process, available at http://www.parl.gc.ca/PBO-DPB/documents/Forecasting%20Process%20-%20EN.pdf, accessed in December 2014.

[376] United States Agency International Development (USAID) Paper on Designing and Establishing Fiscal Policy Analysis Units: A Practical Guide, 2009.

[377] Examination of the New Zealand Treasury's tax forecasting methods and processes, October 2005 available at http://www.treasury.govt.nz/publications/informationreleases/forecastingperformance/accuracy/schoefisch-tax-oct05.pdf.

In many countries, if expertise is not available within the Ministry of Finance or FPU, a network of academic institutions and professional associations often support the Finance Ministry or FPU on relevant research and collaborate on the forecasting. The Research, Policy and Planning Department (RPPD) of the Tanzania Revenue Authority in Tanzania and the Centre for Fiscal Policy in Russia follow such strategy.

The Tax Analysis Division within the US CBO has a staff of 17 with wide experience in revenue forecasting, modelling and analysis of tax policy issues.

The UK OBR, another independent body like CBO, has a permanent staff of 17 civil servants. The OBR is led by the three members of the Budget Responsibility Committee (BRC), who have the executive responsibility for carrying out the core functions of the OBR, including responsibility for the judgments reached in its forecasts.[378] The OBR's Advisory Panel of economic and fiscal experts meets regularly to advise the OBR on its work programme and analytical methods. The fiscal and economic forecast process is co-ordinated by the OBR and overseen by the 'Forecast liaison group', attended by HMRC, Department of Work and Pensions, and the Treasury. The role of HMRC is to provide in-depth knowledge on the data, and to help in analysing taxpayer behaviour.[379]

XIII.4.d Finance Ministry interaction with tax administration for tax forecasting

While the interactions of fiscal policy units with tax administrations have been discussed above, it is important to highlight that regular and structured interaction with the tax administration helps make better forecasts and validate the results by systematic comparison and evaluation of differences in the forecasts between the fiscal policy unit and the tax administration. Thus, a section of the fiscal policy unit is often dedicated to evaluating the effects of current tax laws and proposed changes to the tax system. For this purpose, it interacts from time to time with the tax administration to meet its data requirements. This process also enhances the operation of the tax administration in several ways, mutually benefitting the two. In the UK, for example, the OBR's Budget Responsibility Committee meets officials responsible for running key tax receipts models to challenge and scrutinise the projections produced. These meetings frequently lead to changes in assumptions and forecasting judgments. The HMRC and DWP are also closely associated at the staff level in this process of interaction.

In Canada, the Department of Finance obtains monthly reports from the Canada Revenue Agency (CRA) on tax collections of all types. Anomalies and outliers in the ongoing data flow are discussed in detail with the CRA. Statistics Canada is also engaged as a partner to obtain current data, to understand anomalies and to anticipate possible revisions.

---

[378] Available at http://budgetresponsibility.org.uk/about-the-obr/who-we-are/ accessed in December 2014.

[379] Office for Budget Responsibility, Briefing Paper No. 6, March 2014, available at http://cdn.budgetresponsibility.org.uk/27814-BriefingPaperNo_6.pdf, accessed in January 2015

The New Zealand Treasury Department also monitors and interprets the latest actual tax flows, and exchanges information with the IRD. The IRD provides a detailed external review of the Treasury's tax forecasts by preparing its own forecasts based on the same assumptions regarding the macroeconomic outlook. The forecast methods applied by the IRD are not shared with the Treasury. Therefore, the Treasury has the ultimate responsibility for producing official tax forecasts.[380]

XIII.4.d Reviewing forecasts

Reviewing tax forecasts on a regular basis is an important part of fiscal monitoring. Most of the fiscal policy units in finance ministries and tax administrations, either jointly or independently, review their revenue projections or compare them with actual tax receipts. These reviews help both in removing the error in the model and validating their assumptions for forecast, and hence, moving to more robust forecasting models. In some countries, this is done twice in a fiscal year (for example, in Australia) and in some, on a regular basis (for example, in the UK, Canada).

In Canada, the Finance Ministry examines the accuracy of the federal government's tax forecasting to assess the basis for persistent forecast errors over previous years and to determine what changes can improve the accuracy of forecasts and improve the conduct of public policy.[381] The review normally contains a quantitative as well as qualitative analysis of forecasts.

In the UK, the forecast reviews are pursuant to Section 8 of the Budget Responsibility and National Audit Act, 2011. These reviews result in a forecast evaluation report, which assesses the performance of the forecasts, ensures transparency and accountability and helps users understand how they are made and revised. Identifying and explaining forecast errors also helps improve the understanding of the OBR as well as the HMRC for future improvements in forecasting techniques.[382]

In many countries, such as the New Zealand and the US, external reviews are also commissioned to examine the tax forecasting processes and methods, to identify potential areas of improvement and to suggest approaches to be taken to achieve the improvements identified. These reviews, which include background material leading to the forecasts and error analysis, are often supplemented by a wide range of interviews with members of the treasury's tax forecasting team as well as the head of the tax forecasting team in the tax administration. The reviews normally do not examine the underlying macroeconomic indicators, which in any case are exogenous to tax

---

[380] Examination of the New Zealand Treasury's tax forecasting methods and processes, October 2005 available at http://www.treasury.govt.nz/publications/informationreleases/forecastingperformance/accuracy/schoefisch-tax-oct05.pdf, accessed in December 2014.

[381] Review of Canadian Federal Fiscal Forecasting, Processes and Systems, June 2005, available at http://www.fin.gc.ca/pub/psf-psp/pdf/oneil-eng.pdf accessed in December 2014.

[382] Forecast Evaluation Report, 2014, available at http://budgetresponsibility.org.uk/wordpress/docs/Forecast_evaluation_report_2014_dn4H.pdf, accessed in December 2014.

forecasting. The Australian Treasury first carries out an internal review, before a formal peer review by other government agencies, such as the Joint Economic Forecasting Group, which has representatives from the Reserve Bank of Australia (RBA), the Australian Government's central agencies and the Australian Bureau of Statistics (ABS). In the US, the CBO subjects their forecasts to peer review by an external panel of experts.

XIII.4.e Use of revenue forecasts in management of performance

Net tax collections, after payment of refunds, are often considered the end-product of the work of the tax administrations; it is also considered a key performance indicator for tax officers in field formations. Sometimes, these officers collect more or less than the forecast revenue, leading to deviation from the revenue estimates. Accurate revenue estimates, down to each revenue collector, is a key element in the design of key performance indicator. Unrealistic tax collection plans, despite the fact that there may be forecast errors due to changes in the economic scenario, are also inconsistent with the basic principles of transparency and diminish the accountability of tax officers.

Many tax administrations, particularly in developing countries, use the principles of principal-agent model as a device to monitor tax collections.[383] Put simply, the higher the target for revenue collections, the greater is the effort of the tax administration to increase collections, as long as there is a penalty for failing to meet the forecast. As per these principles, the principal can encourage the fixation of high revenue targets, often not so accurate and fixed through covert methods, thus earning higher public rating for public service delivery and credibility for government expenditure plans. Sometimes, it also becomes rational to intentionally bias forecasts upwards as it allows the government to set higher levels of public spending. But upward biases are not the only possible outcome. The opposite bias of understated forecasts is also possible, if the public is mainly concerned with the credibility of fiscal plans and efficiency gains from higher effort by the agent are small.

Many advanced tax administrations, on the other hand, use tax gap estimates, as stated earlier, as a key performance indicator for staff involved in collecting taxes. The tax gap is also used to frame performance agreements between the employees and the government.

Many tax administrations have also used tax effort as a performance indicator. Tax effort is defined as an index of the ratio between the share of the actual tax collection in gross domestic product

[383] The standard solution to the principal-agent problem is the design of an incentive compatible contract for the agent, which links compensation to an observable variable that varies with the principal's objective function and thus counterbalances the agent's conflicting goals. The model, therefore, would suggest that the compensation of the revenue administration should be linked to the revenue collection performance (e.g., a fixed share of collected revenue is distributed as a bonus). In reality, however, such contracts have not been found practical, as they are costly and often face serious political opposition. The reasons why they have been found to be impracticable are first, that compensation schemes are likely to be expensive in discouraging individual rent taking as targeting would be a problem. Second, they would be inefficient, since the role of other factors affecting revenue, such as economic growth, is quite large. Finally, they would be hard to justify politically, as their prime function is to reward non-corrupt behaviour.

and taxable capacity. The taxable capacity of a country represents the average or normal share of income that is collected as tax in the country. It is calculated as the predicted tax-to-GDP ratio that can be estimated using regression analyses, taking into account a country's specific macroeconomic, demographic, and institutional features.[384] Taxable capacity is found using regression analysis of all tax handles, a ratio of tax from each economic sector (for example, the mining sector, construction sector) and the GDP of that sector, in the economy. Cross-country comparison with countries that have a similar economic structure is often made to understand the performance of the tax administration.[385] The use of such ratios is reasonable if one attempts to establish trends or to compare tax revenue performance across countries with similar economic structures and the same levels of income. However, when used to compare effectiveness in revenue mobilisation across countries in different income groups, the tax-to-GDP ratio could provide a "completely distorted" picture due to different economic structures, institutional arrangements, and demographic trends.[386]

XIII.4.f Tax debt management

One of the key elements of forecasting tax receipts is to ensure that taxpayers pay their taxes. Many tax administrations have been developing tax debt management practices that include strategies and approaches to improve the tax collection and recovery processes, so that they are more effective and cost less. Very promising and proven new practices have emerged, resulting in some spectacular improvements in performance in tax collection and recovery. The ultimate goal of debt management is to reduce the net present value of outstanding debt and to make sure that debtors pay by the due date or as soon as possible.

Many tax administrations use risk models and combine that with business intelligence to develop an optimal enforcement wing to improve their actual tax collections. Analytical tools, such as predictive modelling, have also been increasingly used to make use of big data. Using advanced analytics as a tool, tax administrations segment the taxpayer population and collect information about their behaviour to understand their tax payment performance. Information on taxpayer behaviour can be in terms of whether the tax due was paid on time; if not, when it was paid and following what action by the tax administration. Using data for past years, tax administrations analyse how tax debtors reacted to different treatments in the past and this helps them in predicting their actions and strategies for tax collections. ATO, for example, has segmented taxpayer

---

[384] A typical tax capacity equation will look like $T/Y = a + b \times (Y/N) + c \times (X/Y) + d \times (R/Y) + e \times A/Y$, where a, b, c, d and e are parameters; and N = Population, X = Exports, for example, (excluding mining and petroleum), R = Mining and Petroleum Exports, A = Agriculture output. In theory, it will be expected that b will have a positive value as tax collections tend to rise as a proportion of the national income as the latter grows, and c and d to have positive values with perhaps d larger than c because there are resource rents to be collected from mineral exports. Finally, as agriculture is difficult to tax, we would expect e to have a regressive impact on tax collection, and so will have a negative sign.

[385] Understanding Countries' Tax Effort, Ricardo Fenochietto and Carola Pessino, IMF working paper WP/13/244, November 2013

[386] Tax Capacity and Tax Effort: Extended Cross-Country Analysis from 1994 to 2009, Policy Research Working Paper no. 6252, World Bank, 2012

behaviour to target debt treatments[387] and differentiates its engagement with the taxpayers. In recent years, this approach has contributed to increases in the amount of debt collected and improvements in the efficiency and effectiveness of debt collection.

Many advanced tax administrations, for example, the Tax and Customs Administration of the Netherlands, have also integrated tax records enabling them to track a taxpayer's liabilities, payments and balances using the basic data in one system. Such a record system monitors the development of each debt, and the debtor, over time. The collection process is perceived as a series or chain of collection actions. Data in the system are structured to form a data layer that contains every debt and debtor in order of time, making it possible to follow them throughout the collection lifecycle. The chain of collection can identify where the tax debtor is in the chain at a certain time.

SKAT, the Danish tax and customs agency, has also created a model to predict debtor behaviour. The model comprises three sub-models: one for individuals, one for sole traders, and one for businesses. The sub-model for individuals has 2,700 variables, including the size of the debtor's family and the make and model of the car they drive. Considerable effort has been put to discover which of these variables are most significant.

The Inland Revenue Department of New Zealand has operationalised a model, called the PARE model, which includes the following four elements:

- **Prevent**: Create an environment where customers are aware of their obligations and have the skills and education to comply.
- **Assist**: Ensure customers with liabilities have the information, tools and incentives to file and pay on time. When they miss a payment, the department works with them to resolve their situation quickly and easily.
- **Recover**: Work with customers in debt who want to resolve their situation, and ensure future compliance.
- **Enforce**: Manage intensely the habitual non-compliers and deter them from future non-compliance.

This approach has also used additional early interventions and channels, such as "Just Pay" letters, outbound texting (SMSs) and enforcement rounds, which include garnishee or deduction notices.

[387] For example, the report of the Australian National Audit Office, The Management of Tax Debt Collection, states that, "*ATO research shows that there are three professional groups (lawyers, accountants and medical practitioners) with an average level of tax debt that is nearly five times the national average (barristers nearly ten times the national average). Whilst the average level of debt per occupation is only 2.6 per cent, over 20 per cent of taxpayers in these professions are tax debtors*."

XIII.4.g Time to pay arrangements

Many tax administrations make time-to-pay arrangements for tax debtors, ranging in duration from a few months to as long as ten years. Time-to-pay arrangements are considered to be an efficient and effective tax debt management tool. The ATO, for example, enters into a payment arrangement with the taxpayers so that they can pay their tax debt in a structured manner. This is an online service, currently available only to individuals with income tax debts up to AUD 50,000, and grants payment arrangement for 12 months. This arrangement is arrived at without needing to speak to the taxpayers and is made using either online services or automated telephone services. The automated telephone service, which uses interactive voice response (IVR) technology, is available to both individuals and businesses for both income tax debt and activity statement debts up to AUD 25,000. These self-help tools are designed to make it easy for taxpayers to self-manage their tax debts.

The US IRS has also established several ways for tax debtors to apply for a payment agreement online, through correspondence, by telephone, by e-mail and in person at the local revenue office. Payment in instalments is possible by payment cheque, through payroll deductions, electronic funds transfer and direct debit. Numerous programmes are available to taxpayers based on the type of taxpayer (individual, business, etc.) and the tax balance owed. This programme allows taxpayers to set up an instalment agreement through direct debit of their bank account. It helps taxpayers to resolve their tax liabilities without unnecessarily burdening them and without the involvement of additional staffing.

The Canada Revenue Agency also offers a range of payment arrangements that are based on the taxpayer's ability to pay. Most payment arrangements are made with the CRA debt management call centre but a taxpayer can also load his payment arrangement under some parameters online through the option of My Payment on the CRA website. A payment arrangement in Canada has to be made on the totality of the debt.[388]

XIII.5 Way forward

Forecasts are a basic input in the decision making processes of government; hence, accurate, reliable revenue and expenditure forecasts are essential to good budgeting. The revenue forecast sets the stage for building the budget every fiscal year. The closer the estimate is to actual revenue, the better a government is able to plan for action regarding the conduct of its regular or "general" activities as well as for any contingency. A dependable revenue forecast also provides an in-depth understanding of the government's tax revenue sources, which will include descriptions of the tax base and rate, and historic and projected revenue trends, collection methodology, accounting information, etc. The current tax receipt budget woes[389] are a testament to such difficulty as well

[388] OECD Paper on Working Smarter in Tax Debt Management, 2014

[389] Last year both direct and indirect taxes were overestimated in the budget estimates, and were subsequently revised. Similar situation is likely to be there during the current fiscal year.

as the importance of accurately projecting tax revenues. Qualitative methods of revenue forecasting adopted by the CBDT and CBEC employ human judgment, and that often demands greater objectivity on the part of different stakeholders (the budget division and the economic division in the Department of Economic Affairs) to arrive at an agreement on a particular forecast. Over-estimated tax receipt forecasts often leaves the CBDT and CBEC dissatisfied with the overall budget making process, and the government with an inaccurate tax revenue forecast. Credible revenue estimates, based on realistic assumptions, are thus required on a regular basis to achieve a tax environment characterised by better accountability to taxpayers. This is unlikely to be achieved by the existing method of tax revenue augmentation based on forcing taxpayers to pay more taxes during the last few months of a fiscal year or on the stoppage of refunds or duty credits to them.[390]

XIII.5.a Approaching revenue forecasting

There are several models that can be used for tax revenue forecasting with separate models for each tax type, as shown in Section XIII.2. Although budget forecasting theories and practices may vary from time to time and from government to government, it is important that the process is balanced, transparent, and trusted. It is also important that the process involves expertise and experience so that its dynamic character is properly harnessed to understand far-reaching policy implications. The experience of different countries has shown that the institution of an independent forecasting office, the adoption of consensus forecasting, and the establishment of a transparent process has helped generate accurate forecasts and, more importantly, build trust in and bring about accountability to the budget process. But this is not to say that these processes will result in 100 per cent accurate tax forecasts; the use of technically sound methods and approaches only helps make forecasts more accurate.

There are three key elements to the revenue forecasting process – transparency, formality and organisational simplicity. Transparency requires that the macroeconomic and other assumptions made while estimating the forecasts are properly explained and made public. Such detailed information in the public domain will have the salutary role of improving data quality as well as accountability in the tax forecasting process. This should result in increased accuracy and possibly, a reduction in ad hoc or discretionary adjustments during the fiscal year. It will also help improve the credibility of the forecasts and lend greater legitimacy to tax proposals, which so far has been sadly missing.

Formality, another key aspect of the forecasting process, refers to whether the forecast is formally defined, initiated, regularly reviewed, documented or not. Budget preparation practices in the TPL and TRU are to a large extent unstructured; the existence of formal rules on issues, such as forecasting responsibilities, time table and documentation will set in a well-structured process leading to more timely forecasts and reducing the scope for covert interference.

[390] Chapter VI of the TARC report also discusses this.

In the budget process, there are a number of agencies involved, each competing with the other with their internal forecasts. This affects quality and often requires more time to reconcile before a single forecast is arrived at. The process is often time-consuming, and may have an impact on the quality of the ultimate forecast. Such an open-door process, involving multiple agencies with competing forecasts, is a duplication of effort and involves a co-ordination cost, besides undermining the timeliness of such forecasts. Therefore, there is need for organisational simplicity, so that there is coherence in the result with less time spent on sparring over competing models and more on improving the forecast results, based on detailing the model itself. A core aspect of this will be how cohesive the forecasting process is. This will be discussed in detail later in this section.

A survey of comparable revenue forecasts reveals that there is no commonly accepted standard practice or model for revenue forecasts. Most tax administrations draw upon a combination of modelling techniques, consumer and business surveys and expert opinions to arrive at the forecasts. Differences, however, exist in the detailed modelling techniques, particularly the type of the models employed – whether long-term or short-term. It is common for forecasting agencies to maintain two forecasting systems – one focusing on the short-term horizon (up to, perhaps, six months), and the other that focuses upon longer time horizons (greater than six months). The first uses indicator or spread sheet models and relies heavily upon judgment supplied by in-depth analysis of major sectors of the economy undertaken by sector specialists; the second involves modelling the economy as a 'system' with the core forecasting technology typically being either a relatively small economy-wide structural macro-econometric model or a single structural econometric equation, with national account identities preserved within a spread sheet system, and consistency between the elements of the forecasts achieved by iteration.

Short-run revenue forecasts can begin with an economic forecast; hence, it is important to carefully track the economy, and start the formal forecasting process at least nine months in advance. As part of this exercise, the two Boards can also set up a tax revenue receipt model. This model should have the growth rate of the economy as one of the variables. Table 13.7 depicts typical revenue forecasting in a budget cycle. The table shows the time table and when the budget preparation sequence should start. In this table, month -12 is the first month of fiscal year 0.

Table 13.7: Revenue forecasting time table/cycle

Time	Activity
–9 Months	– Medium term macro forecast, revenue forecast, and debt policy. – Determine revenue gaps: review and revise basic policies (i.e., debt burden, and new tax policies – change in structure, administration, incentives, distribution, size of government expenditures, etc.).
–6 Months	– Revised macro and revenue forecast, including tax measures requirements to meet the tax debt policy.
–3 Months	– Refine forecast with emphasis on years 0 and 1 revenues, with final target for discretionary revenues for year 1.

Time	Activity
	− Revised year 0 forecast changes the basis for forecast for subsequent years 1, 2, 3.
+3 Months	− End-of-budget year review. − Review of previous fiscal year's revenue performance and analysis of *ex post* revenue impact of discretionary changes

This forecasting time-table will be underpinned by the recognition of the constantly changing structure of the economy and the consequent need to revisit the underlying parameters of forecasting equations. While time-series forecasting models, including vector auto-regression model or other types of time series analysis, can be the starting point, they do not provide, as stated in Section XIII.2, the same kind of logical check as macroeconomic forecasts do because they are based on historical data. Consequently, macro models, based on buoyancy and elasticity calculations, can play an important role in the forecasting process. For macro models, macroeconomic forecasts are vital and provide a fiscal baseline for medium-term fiscal strategy. But it should be kept in mind that economies are buffeted by shocks, for example natural disasters, such as droughts or floods, which by their very nature are not foreseeable. Besides, there are lags in data collection and survey errors. All these factors can play an important in the preparation of forecasts. Hence, discussion on risks and uncertainties is a fundamental part of any set of forecasts and a critical input into macroeconomic forecasting deliberations. The forecasts may also require continuous evaluation, internally as well as by a peer group,[391] so that forecasting results and performance improve, and the model becomes more robust. Forecasting models using a 'bottom-up' approach where individual heads of revenue are forecast and then aggregated to produce a total revenue forecast can be the next level of detailing. Data limitations can constrain the use of this approach.

Micro-simulation models operating at the individual taxpayer level can be the higher level of sophistication in forecasting. These models, as explained in Section XIII.2, simulate large representative populations of individual entities, such as a person, family, or firm, to draw conclusions that apply to higher levels of aggregation such as an entire taxpayer base. These models require detailed data and can be used to forecast personal income tax revenue as well as corporate tax revenue. The CBDT has all the taxpayer details in electronic form; it is time they set up these micro-simulation models without much delay. The CBEC, similarly, can attempt a trade simulation model and an excise model.

Since the tax gap provides a picture of the total revenue due and from whom it should be collected, mapping that even at the field formation level can be helpful in setting a tax gap reduction target as a performance indicator both at the organisational and jurisdictional level and at the assessing

[391]Benefits of a multi-dimensional approach to forecasting have been documented by the Australian Office for Budget Responsibility.

officer level. Both the CBDT and CBEC must move towards estimating the tax gap immediately to track tax collections and to determine the effectiveness of the tax administration. This will help the two Boards identify the areas, the types and the level of non-compliance that contribute to the tax gap. It will also help them develop better strategies to combat non-compliance – which is actually their objective.[392] Direct revenue collections, on the other hand, emphasises "target" chasing without any understanding of the underlying reasons for non-compliance. This approach has led the two Boards on a low productivity path, and has resulted in tax regime that is oppressive and burdensome for both taxpayers as well as tax officers.

To summarise, the TPL and TRU can use the following tools and models for revenue forecasting and tax policy analysis:

- **Revenue forecasting and monitoring models** – These will include quantitative methods used to project and monitor short- and long-term receipts over a single or multiple budget periods. The revenue forecasting models can be "unconditional" time-series models, which are primarily based on historical revenue data, or "conditional/causal" models, which use other historical and projected data, using tax bases, to predict revenues. The GDP-based forecasting models and key measures of tax elasticity and buoyancy are among the most common models for revenue forecasting. The short-term receipts forecasting model can be used for monitoring tax receipts on a monthly basis (for excise, customs and direct taxes), and a modified model for monitoring quarterly tax collections (for direct taxes).

- **Microeconomic models** – These models will require the estimation of demand and supply elasticity, which are typically done using econometric models. These models are most appropriate for estimating the impact of discretionary changes in narrow-based commodity taxes, such as excise, on revenue.

- **Models based on structural/national accounts data** – These models will be useful to analyse broad-based indirect taxes, such as the VAT/GST. Input-output tables can be used to analyse the impact of policy changes on VAT/GST.

- **Micro-simulation models** – The main advantage of a micro-simulation model lies in its ability to estimate the distributional effect of a given tax policy proposal on particular sectors of the population, and identify the gainers or losers. This model will require detailed information from tax returns, and perhaps also through surveys for specific purposes. The models are more appropriate for personal income tax and corporate tax, but their applications can also be expanded to import taxes. Micro-simulation models are also useful in setting up discretionary change models.

- **Tax capacity, tax effort, tax compliance, and gap analysis** – This involves the quantification of the overall level of tax evasion and avoidance in the tax system. This type of analysis allows

[392] Vision 2020 statements of the CBDT and CBEC

an assessment of the impact of compliance and exemptions on tax yield by type of tax and by economic sector. Tax gap analysis should be used to monitor tax collections.

- **Tax expenditure analysis** – Both TPL and TRU carry out tax expenditure analysis, quantifying the tax revenue losses attributable to various "non-standard" exclusions, exemptions, deductions, credits, deferrals, and preferential rates in tax laws. But these figures so far have time lags as actuals are available only after a gap of one or two years. These figures for the current year need to be forecast. This will require the development of models to forecast tax expenditure to get current and future estimates of tax expenditures.

- **Computable general equilibrium model** - These models are an abstract representation of an economy than can be used to conduct policy evaluations using actual data. These models have economy-wide application, and take into account the behaviour of all economic agents as primary responders, or because of feedback or secondary effects, using widely-accepted principles of optimisation and choice theory to estimate the directional impact of proposed tax measures on important macroeconomic variables and the economy when alternative tax reform options are being evaluated. This should also be set up by the TPL and TRU.

To develop and use the analytical models outlined above, the TPL and TRU should be sufficiently equipped with computer hardware, database systems, and other office software, including the most recent econometric and statistics software packages. Such equipment and software systems are necessary to conduct the most demanding data analysis and micro-simulation modelling for tax policy analysis and revenue forecasting purposes. The TPL and TRU should also have access to current and seminal publications relating to tax and fiscal research and to a database of journal articles to continuously upgrade the skills and knowledge of the officers working there.

Tax debt collection

Arrear demands today constitute a sizeable proportion of tax revenue estimates in both the Boards – in the CBDT, it is roughly more than 56 per cent of budgeted estimates of tax collections for FY 2014-15 and, in the CBEC, about 13 per cent.[393] Both the CBDT and CBEC employ traditional methods for collecting taxes. This approach allows non-compliant taxpayers scope to delude the system, resulting in actions by the tax departments that are ineffective, or unnecessary. A changed approach based on identifying debtors rather than the present method of identifying debts, and understanding their behaviour can give better results. Cluster analysis, based on characteristics, such as tax debt size, taxpayer status, underpaid tax, etc., can be employed as one of the methods. The characteristics of the taxpayers, in a cluster, will require risk assignment on the basis of inputs from field formations as well as on the basis of data or information from the data warehouse to tailor taxpayer treatments on the basis of individual circumstances and behaviour.

[393] Annex 11 of the Revenue Receipts Budget, 2014-15

It will be useful to carry out this analysis using the combined data of the CBDT and CBEC; this will be in line with the TARC recommendation in Chapter IX of "one data, many users". The combined data, based on common standards and taxonomy, will facilitate data exchange between the two Boards, enabling better data analysis to track the taxpayer's tax liabilities, payments and balances using basic data in one system. The combined data can be structured to form a data layer that contains every tax debt and tax debtor, making it possible to follow them throughout the collection lifecycle in a united manner, as ultimately there is only one taxpayer. This chain of collection can identify where the tax debtor is in the chain at a certain time. Such a system to monitor the development of each tax debt, whether in the CBDT or the CBEC, will improve the overall tax collection process.

XIII.5.b Reviewing and reporting forecasts and monitoring taxes

Monitoring tax revenues on a regular basis is important to assess the quality of the tax forecast and to evaluate the economic and fiscal health of the country. At present, the TPL and TRU prepare revised estimates during December or January of the fiscal year. Tax receipts are monitored on a regular basis by another division in the CBDT, the Director (Budget), and in the CBEC, by the TRU. These reviews of tax forecast in the TPL and TRU are done from the point of view of fulfilling the legal, budgetary requirement of providing revised estimates rather than to revisit tax forecasts. But that is a limited exercise. It is important that tax forecasts are revisited on a regular basis as key economic variables, such as gross domestic product, inflation, short- and long-term interest rates, employment growth, etc. change constantly, and may affect forecasts. There is also the possibility that structural relationship(s) may change due to changes in economic indicators.

Keeping this in view, it is important that tax collections are monitored and reviewed to enhance the accuracy of forecasts. This may involve a series of consultations with macroeconomists in the economic division of the Department of Economic Affairs to understand changing economic indicators, as they impact tax collections. These consultations will also help the CBDT and CBEC to evaluate whether the tax forecast will need to be revisited, and may help in preparing an assessment on whether budget estimates are likely to be met or will require reduction. This, at present, is done within closed doors, and more often than not, results in applying pressure on top taxpayers to pay more taxes or in the stoppage of refunds or input credits to augment tax revenue if it is felt that there is likely to be a short fall in tax revenues. This process needs to be made more transparent; to do so, a small unit,[394] comprising TPL and TRU officials, and officers from the economic division of the Department of Economic Affairs and the Reserve Bank of India can be set up within the TARC-recommended Tax Policy and Analysis unit[395] with the mandate to evaluate and prepare a report on the tax implications of macroeconomic changes. The report can

---

[394] An internal review mechanism, called the Joint Economic Forecasting Group (JEFG), ahead of a formal peer review, has been set up in Australia. The forecasts are discussed within this group with representatives from the Reserve Bank of Australia (RBA), Australian Government central agencies and the Australian Bureau of Statistics (ABS).

[395] See Chapter III of the TARC report.

then be presented to Parliament as part of the FRBM Act.[396] These reviews can be done twice in a fiscal year, one after 6 months and another after 10 months of the start of the fiscal year.

The report can have both qualitative inputs as well as quantitative inputs. The quantitative part will be a key focus area of analysis during the forecast review, with forecasting methodologies, data inputs, and tracking against forecasts being the foremost areas of concern. The qualitative part will comprise interaction with experts in forecasting and budget preparation, mainly from the private sector and academia. These consultations will reflect on the nature and causes of the gap in the results from forecasting of budget balances, including factual inaccuracies as well as interpretations and inferences that can be (and, in a number of cases, are) challenged by the research carried out by them. Such a review and honest assessment can lend credibility to the tax forecasting process and provide accurate results. Inaccurate forecasts often give rise to public and taxpayer concern on whether the tax departments base tax policy decisions on inaccurate and/or incomplete information. The fact that improvements in forecast accuracy can ameliorate these concerns will be an additional benefit of the review.

XIII.5.c Data requirements

Access to relevant and quality data in sufficient detail is probably the most common problem faced by most fiscal policy analysis units around the world. However, discussions with both TPL and TRU revealed that they have little problem in accessing macroeconomic time series data, national accounts statistics, tax and non-tax revenue series, and detailed tax returns and customs data. However, the quality of the data and the taxpayer master file used by them needs to be checked on quality parameters to ensure the reliability of the analysis and the results.[397]

Data limitations are a constraint on revenue forecasting. Many times, there are long lags in receiving data, particularly tax return data. It is thus imperative that tax data for the selected time period are procured from the concerned department or agency and examined for gaps. The period for which data is procured can depend on the availability and quality of data, the type of revenue to be forecast, and the degree of accuracy sought. The time horizon affects the techniques and forecasting approach. For example, short- and medium-term forecasts demand higher levels of accuracy because they will be used for detailed budgeting decisions. Longer-term forecasts are used for more general planning, not detailed appropriations. Monthly data are best for mid-term (one to two years) and long-range (three to five years) forecasting. But more importantly, good historical data are essential to good forecasting because past revenue patterns provide clues to future behaviour. The first step, thus, is to compile revenue data for as many years back as is practical. This will often require scrubbing the data to remove the impact of historical events that reduce their predictive value. Therefore, it is important to develop a strategy to systematically

[396] The Finance Minister had stated in Parliament at the time of presenting the Economic Survey for 2013-14 that, "India needs a sharp fiscal correction, a new Fiscal Responsibility and Budget Management (FRBM) Act with teeth, better accounting practices, and improved budgetary management." He also stated that, "the Act needs to take into account business cycles."

[397] Appendix XIII.3 gives details of information called for and available to the CBDT and CBEC.

clean up the taxpayer master file and tax returns database. This activity may be beyond the scope of work of the TPL and TRU. However, since they have dealt extensively with cleaning up taxpayer data, inputs from them will be invaluable from the user's perspective.

Data, once available, can be examined for the underlying pattern, rates of change, or trends, by making comparisons of data from different sources, linking data sets, comparing corresponding items, finding relationship and patterns, and constructing descriptive or hypothetical representational and/or functional relationships between different variables of which the data is composed.[398] These underlying characteristics often provide insights regarding the model that can be employed, and the appropriateness of the forecasting method. Preparation and presentation of the data in different formats can be a part of this step. Technological advances – powerful computers and algorithms – can help reveal new insights that may otherwise remain hidden. Since every situation will require a different treatment, success can also be in developing algorithms, which can make sense of the amorphous data and information.[399]

Tax return data for employing the micro-simulation model have very precise details on income and deduction sources, as well as income distribution. However, the latest return data available are usually several years old due to the lag in processing tax returns for a given year. There is also the problem that the composition of income and deductions and the distribution of income may have changed dramatically from year to year. As a result, greater reliance will have to be placed on recent years' estimates of tax liability from the most recent collections data. A receipts forecast can be produced by estimating the cash flow of the forecast liability across fiscal years, using the tax revenue receipt model. The micro-simulation model gives good estimates of tax liability. But in times of revenue surges, collections liability estimated by the micro-simulation model may be lower than actual tax collections. In such situations, forecasting judgment may sometimes be required.

After projections have been made, the estimates need to be evaluated for their reliability and validity. To evaluate the validity of the estimates, the assumptions associated with the revenue source need to be re-examined. If the assumptions associated with existing economic, administrative and political environment are sound, the projections can be assumed to be valid. Reliability can be assessed by conducting a sensitivity analysis. This involves varying key parameters used to create the estimates. If large changes in estimates result, the projection is assumed to have a low degree of reliability.

---

[398] Many computer-based methodologies have data matching and analytics capability. These software need to be integrated into the working of the tax administration, so that they can be used for improving compliance and increasing revenues.

[399] This has been discussed in detail in Section IX.6.d of the TARC report.

XIII.5.d Partnering non-government bodies and research institutions

Maintaining a regular dialogue with the academia and business communities is valuable for identifying emerging trends in the economy. Information gained from these engagements can be used to inform the final macroeconomic forecasts.[400] For these interactions, the academia and business communities will have to be provided data, albeit encrypted data, so that they have the requisite data for a more meaningful engagement. For that, the CBDT and CBEC need to identify academic institutions of national repute and business or professional associations having all-India membership that can collaborate for the dialogue. This collaboration will enrich the overall quality of their research, and will improve their submissions to the Finance Ministry at the time of budget preparation. Lack of quality data hampers their tax proposals, and so the macro data, with encryption, need to be shared with them. The dialogue will also give the academia and business communities a sense of participation, which is important to improve tax governance. Regular dialogue on tax analysis and revenue forecasting will also help identify errors and rectify them before it is part of the budget. It may also be pointed out that this process will help the tax departments to augment and fill data gaps that, at times, becomes a major handicap.

XIII.5.e Integrating revenue forecasting with policy

It is important to integrate tax forecast results, particularly those relating to tax analysis such as distributional impact out of the micro-simulation model or tariff impact on commodities from the trade tax calculator. Integrated results can be used as a tool to inform decision makers, including ministers, of the likely benefits and costs, identifying the key factors that should be taken into account in decision making, strengthening the quality of analysis and making the policy inclusive.[401] The process will also help improve the forecast. It is often seen that tax policies are not consistent with tax forecasts; for example, many key indicators may have changed but may not have been changed in the tax forecasting model. Another example is that tax-motivated behaviour, such as tax avoidance activity, may not get plugged in to record the responsiveness to tax rates and tax structure. Dynamic analyses capturing the interaction between taxes and the economy thus are required to be integrated through macroeconomic models of the economy and micro-simulation models of taxation. An important part of that integration will be calibrating both models to the same “baseline” forecast. Like many tax administrations, it will be useful to develop and run several major micro-simulation models and maintain large statistical databases to analyse the economic, distributional, and revenue effects of alternative tax proposals and tax systems to do so. This will also be helpful in conducting studies on tax compliance and taxpayer compliance burdens, including developing and revising tax forms.

XIII.5.f Organisational/institutional arrangements

The TARC in Chapter III of its report recommended that the existing TPL and TRU wings of the CBEC and CBDT should be subsumed in the Tax Policy and Analysis (TPA) wing. It was further recommended that the TPA should be expanded to include specialists, such as economists, tax law

[400] Some agencies also regularly publish their findings of these engagements.

[401] This has been discussed in detail in Chapter X of the TARC report.

experts, statisticians, operations researchers and social researchers, to form a multidisciplinary team. It had recommended that these specialists be supported by a large complement of analysts, with adequate skills to undertake advanced data analytics. It was envisioned in the recommendation that the TPA for its analytical work will carry out statistical modelling such as GDP-based forecasting (long-term as well as short-term) and micro-simulations along with discretionary change models so that decisions are based on sound analyses, and are data and evidence driven.[402]

Continuing on the above recommendation, the forecasting units under the CBDT or CBEC can do the following specified jobs:

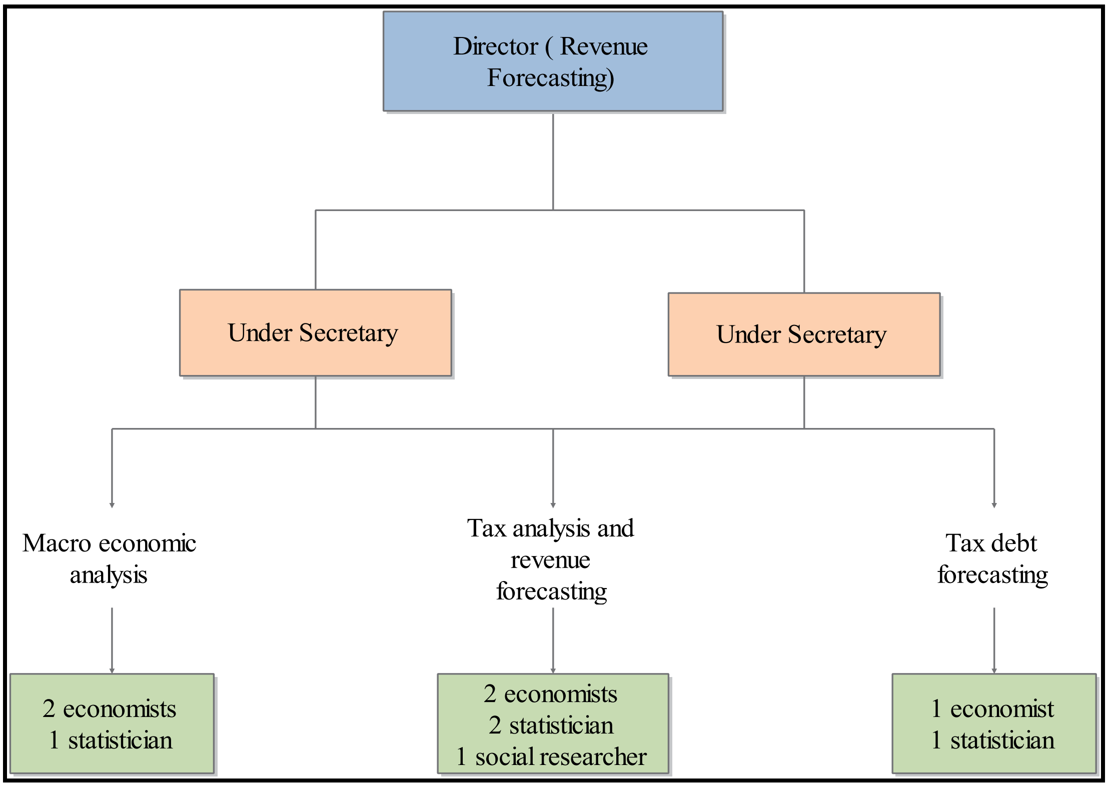
- running analytic models on forecasting
- analysing options based on data (proposed by the finance ministry or other agencies)
- generating baseline budget totals for budget preparation
- understanding the appropriateness, strengths, and weaknesses of current data sources, and recommending changes as needed
- monitoring current tax receipts and advising on the implications for future receipts
- analysing the current tax system and making recommendations for the future development of tax laws
- preparing position papers on tax issues for the Ministry of Finance
- anticipating and planning for future needs
- developing an agenda for ex-post analysis of existing tax proposals

Another important task of this unit will be to publish results from the analytical models it develops. These publications should contain objective and impartial analysis and should be helpful in guiding broad policy debates, so that the knowledge can be shared with peers and the policy community for discussion and feedback. As the forecasting unit accumulates experience and expands its tools and databases, and as individual staff become more knowledgeable concerning tax, income, and other data in their areas of expertise, the data so published will serve two broad purposes. First, it will ensure that important data are made available to a broad range of users, and second, it will enable staff and officers to further develop their skills and analytic capabilities through the discipline of writing and publication.

[402]See Section III.4.d of the TARC report.

The structure of the forecasting unit, separately for the CBDT and CBEC, will be as given in Diagram 13.11.

Diagram 13.11: Structure of the forecasting units in CBDT and CBEC



The economists and statisticians will work on macroeconomic statistics, including income aggregates and other macroeconomic indicators. He/she will examine data from the Reserve Bank of India, Central Statistical Organisation (CSO) and other government agencies and update them and provide them to the persons engaged in forecasting on a regular basis. These economists and statisticians will also write periodic reports on the extent to which the data differ from forecast data, and provide possible explanations for the divergence. They will have to be up to date with details of proposed tax and tariff changes, and the extent to which they might affect competitiveness and the economy. Besides, they will also be required to update and run analytic models, including a CGE model.

The tax analysis and revenue forecasting (TARF) unit will identify and evaluate the effects of current tax laws and proposed changes to the tax system and provide revenue forecasts. This section will include technicians, comprising economists, statisticians, social researchers, who will work on major taxes such as personal income tax, corporate or business income tax and indirect taxes, by updating and running tax models, creating periodic tabulations of tax return and other tax data. The unit will have to eventually move towards developing tax gap analysis at the all-India as well as the field level to help evolve measures to reduce the gap. All these will have to be made available online to the field officers. The TARF technicians will also be engaged in writing periodic reports for senior officials on each major tax and, therefore, will have to be familiar with any changes that affect these taxes. This unit will also have to develop tax forecasts for each jurisdiction

The tax debt management analysis group will be expected to have close links with the TARF unit and will be required to participate in the planning stage of the budgetary process since the collection of tax debt is part of tax receipts. This group can make a unique contribution to the overall budget process, considering that the use of analytics can facilitate correct and timely intervention, given that debtor behaviour is dynamic. A commitment to continuous improvement will ensure that the tax department is responsive to those changes and collection functions are correctly aligned to desired outcomes. There will be need to monitor tax debt collection online, for which both the CBDT and CBEC will have to set up online analytics to monitor tax debt collections by field officers and by the tax debt management analysis group on a more macro basis.

XIII.5.g Staff resources

The TARC in its first report had stated that the TRU and TPL wings, working in silos, were short on adequate research, analysis of data and multidisciplinary inputs and have become virtual legal repositories devoid of careful analysis, market surveys or use of macro-fiscal models.[403] The TARC also pointed out that the strength of staff and officers in the TPL and TRU were inadequate. Besides, the TARC had recommended that the unified Tax Policy and Analysis (TPA) unit should be expanded to include specialists such as economists, tax law experts, statisticians, operations researchers and social researchers to form a multidisciplinary team.[404] It had recommended that these specialists be supported by a large complement of analysts, with adequate skills to undertake advanced data analytics.

Continuing on the same lines, officers and staff of the TPL and TRU need to be selected on the basis of specified qualifications, which is not the case at present. Emphasis in selection, at present, is given to work experience in the tax administration, and past experience of TPL and TRU is considered an additional advantage. But it is important that officers and staff of the TPL and TRU include trained personnel with specialised skills and knowledge in the fields of revenue forecasting, analysis and monitoring. While Indian Revenue Service (IRS) officers at the levels of Joint Secretary and Director in the TPL and TRU may require wide knowledge of tax policy and macroeconomic issues, deeper understanding of revenue estimation and forecasting will be an advantage as they need to interact with the Minister of Finance, other senior government officials, and parliamentarians on a broad array of tax and fiscal concerns at the time of budget making and at the time of monitoring tax receipts.

For officers below the level of Joint Secretary and Director, the knowledge and skill to gather data from different agencies, perform routine analytics for forecasting revenue, and prepare results that

---

[403] See page 117 of the TARC report.

[404] See page 120 of the TARC report.

could be used for drafting tax memoranda will be useful. These officers and staff can be from the disciplines of statistics, economics, and social science. The duties of these officers will be to conduct macroeconomic analysis to forecast, monitor and analyse revenue and tax debt on a regular basis, and evaluate the economic and revenue impact of introducing new and/or maintaining existing tax policies. They will need to have some knowledge of taxation policies, direct and indirect tax laws, knowledge of technical report writing and skills in handling computer databases and statistical packages, and econometrics software, such as SAS, EViews, STATA, etc.

While selecting officers and staff may provide the necessary framework, this may not be sufficient for revenue forecasting work. It is recommended, therefore, that on-the-job training be considered the most important route to developing and deepening the necessary skills required to perform such functions. A university-level education in economics, public policy, statistics and other related disciplines may be necessary, but actual strengthening of capacity can be attained through regular internal seminars, workshops and specialised short courses. These capacity building exercises, particularly internal seminars and workshops, will have to be organised on a monthly or quarterly basis so that officers and staff get adequate opportunity to present his/her work on a particular issue to the rest of the unit. These presentations will provide an opportunity to present and explain specific tax forecasting models and tools of analysis and to receive feedback. This will not only benefit the presenter, but also the audience as a whole. It may also be beneficial to invite members from academia or research institutions, working on the same or similar topic, to these seminars or workshops so that the presentations have the benefit of being peer-reviewed. To do so, the TPL and TRU will also have to put in place an online database of such experts.

Customised and specialised short courses on revenue forecasting can also be organised for staff and officers of the TPL and TRU to update their knowledge on the subject. These courses should invariably be graded to ensure that the learning is imbibed and can be immediately utilised in work. Long-term training can also be considered on the subject for long-run viability of the analytic unit of the TPL and TRU.

XIII.6 Summing up

Tax forecasting, by no means an easy process, can be approached in a step-by-step manner. Structured steps to the forecasting process can help develop robust forecasting models. It would also help explain to others what the forecast does, and that can have a role in increasing the acceptability of the forecast results. Since a structured process will encourage adherence to using best practices more diligently than an unstructured method, the chances of getting more accurate results will be greater.

The first step towards initiating the process will be to define the fundamental issues, an objective function, of why the forecast should be made and what a forecast affects. Such questions may also provide an insight into which forecasting method will be the most appropriate and how the forecast can be analysed and provide a common understanding of the goals of the forecasting process. Gathering data, including statistical data, and information to support the forecasting process will be another step. Effort should invariably be made to gather at least ten years data for macro-based forecasting, and at least three years data for tax receipt forecasting. It should be kept in mind that data requirement will depend on the forecasting technique and the desired level of accuracy in the result. Data quality will also be of utmost significance. If there are data gaps, that will present its own problems.

The next step will be to conduct a preliminary/exploratory analysis, looking for seasonality, business cycle or outliers. Correlation analysis for determining important relationships between variables can aid in forecasting as it provides a strong basis for using quantitative forecasting. This can be followed by a decision on the quantitative and/or qualitative forecasting method to be used.

Forecasts are often expressed as "point forecasts" because they express future revenues as a single number, although this number is actually made up of averages – averages of historical experiences and perhaps averages of different forecasting techniques. Presenting one number can obscure variability, since the possible future outcomes are many.[405] Hence, it will be useful to give the results as "prediction intervals", illustrating the uncertainty in the forecast by showing a range around the baseline forecast in which the actual value is likely to fall.

XIII.7 Recommendations

The TARC recommends the following.

i) Approaches to revenue forecasting

a) It is important that the budget forecasting processes are balanced, transparent, and trusted. (Section XIII.5.a)

b) It is also important that the revenue forecasting process involves expertise and experience, so that its dynamic character is properly harnessed, to understand far-reaching policy implications. (Section XIII.5.a)

c) The revenue forecasting process should observe three key elements – transparency, formality and organisational simplicity.(Section XIII.5.a)

d) All macroeconomic assumptions or other assumptions in the forecasts should be properly explained and made public to ensure transparency. Such detailed information in the public domain will have the salutary effect of improving data quality and accountability in the tax forecasting process. This should result in increased accuracy, and possibly a reduction in ad hoc or discretionary adjustments during the fiscal year. Transparency will also enhance the credibility of the forecasts, which so far has been sadly missing. (Section XIII.5.a)

e) Budget preparation practices in the TPL and TRU are to a large extent unstructured and the existence of formal rules on issues, such as forecasting responsibilities, time table and

[405]Appendix XIII.4 presents a short discussion on errors.

documentation, will establish a well-structured process leading to more timely forecasts and will reduce the scope for covert interference. (Section XIII.5.a)

f) There is need for organisational simplicity, so that there is coherence in the result with less time spent on sparring over each competing model and more on improving forecast results, based on detailing the model itself. (Section XIII.5.a)

***ii)* Revenue forecasting methods**

g) Tax revenue forecasting has many models with a separate model for each tax type. There is no commonly accepted, standard practice or model for revenue forecasts. Tax administrations generally draw upon a combination of models, consumer and business surveys and expert opinions to arrive at tax forecasts and analysis. Both the Boards should also adopt a bouquet of methods and not rely on only one method. (Section XIII.5.a)

h) The two Boards should maintain the two forecasting systems – the first one focusing on the short-term forecasting horizon (say, up to six months), and the second focusing on longer time horizons (greater than six months). (Section XIII.5.a)

i) For short-run revenue forecasts, the two Boards can set up a tax revenue receipt model. (Section XIII.5.a)

j) The two Boards can use different tax forecasting and tax policy analysis models, depending on data availability and the rigour desired in analysis. Commonly used models are conditional/causal models, using historical and projected data for tax bases to predict tax revenues. To begin with, both Boards should set up conditional models. (Section XIII.5.a)

k) As the Boards gather more experience, they should move towards micro-simulation models. This will allow them to examine the distributional effect of a given tax policy proposal on particular sectors of the population, and identify the gainers or losers. Micro-simulation models are also useful in setting up discretionary change models.(Section XIII.5.a)

l) Both Boards must carry out tax expenditure analysis, quantifying the tax revenue losses attributable to various "non-standard" exclusions, exemptions, deductions, credits, deferrals, and preferential rates in tax laws. These figures so far have a lag of two years, but with data being captured on real time basis, it should be possible to move towards capturing current year data. Models should also forecast future estimates of tax expenditures. (Section XIII.5.a)

m) Both Boards must set up a CGE model to conduct policy evaluations using actual data. The CGE model will estimate the directional impact of proposed tax measures on important macroeconomic variables and the economy whenever alternative tax reform options are being evaluated. (Section XIII.5.a)

n) Both Boards must move towards estimating the tax gap to track tax collections and to determine the effectiveness of the tax administration. This will also help the two Boards to identify the areas, types and level of non-compliance that contribute to the tax gap. This will also help the Boards develop better strategies to combat non-compliance. (Section XIII.5.a)

o) It is useful to introduce the cyclical nature of tax revenue collection in a wider macro econometric model using output gap along with income/consumption shift variables/dummies for good and bad times. (Section XIII.2.a)

***iii)* Tax debt collection**

p) Both the CBDT and CBEC employ traditional methods to collect taxes. This approach allows non-compliant taxpayers scope to duck the system, resulting in actions by the tax departments that are ineffective, or unnecessary. A changed approach, based on identifying debtors rather than debts and an understanding of their behaviour, can give better results. (Section XIII.5.a)

q) Cluster analysis, based on characteristics, such as tax debt size, taxpayer status, underpaid tax, etc., can be employed as one of the methods. The characteristics of taxpayers in a cluster will require risk assignment on the basis of inputs from field formations as well as on the basis of data or information from the data warehouse to tailor taxpayer treatments on the basis of individual circumstances and behaviour. (Section XIII.5.a)

r) It will be useful to carry out the tax debt analysis using combined data with the CBDT and CBEC. This will be in line with the TARC recommendation in Chapter IX of "one data, many users". The combined data based on common standards and taxonomy will facilitate data exchange between the two Boards, enabling better data analysis to track the taxpayer's tax liabilities, payments and balances using the basic data in one system. (Section XIII.5.a)

s) The combined data can be structured to form a data layer that contains every tax debt and tax debtor, making it possible to follow them throughout the collection lifecycle in a united manner, as ultimately there is only one taxpayer. This chain of collection can identify where the tax debtor is in the chain at a certain time. Such a system to monitor the development of each tax debt, whether in the CBDT or the CBEC, will improve the overall tax collection process. (Section XIII.5.a)

***iv)* Reviewing and reporting of forecasts and monitoring of taxes**

t) It is important that tax forecasts are revisited on a regular basis as key economic variables, such as the gross domestic product, inflation, short- and long-term interest rates, employment growth, etc., are changing constantly, and that may affect forecast results. Due to change in economic indicators, there are chances that the structural relationship(s) may also change. (Section XIII.5.b)

u) It is important that tax collections are monitored and reviewed to enhance the accuracy of the forecast. This may involve a series of consultations with the macroeconomists in the economic division of the Department of Economic Affairs to understand changing economic indicators as they affect tax collections. These consultations will also help the CBDT and CBEC to assess whether the tax forecast will require to be revisited, and may help in preparing an assessment on whether the budget estimates are likely to be met or will require reduction. (Section XIII.5.b)

v) The forecast process needs to be made more transparent. It is recommended, therefore, that a small unit, comprising TPL and TRU officials and officers from the economic division of the Department of Economic Affairs and the Reserve Bank of India, be set up within the TARC-recommended Tax Policy and Analysis unit with a mandate to evaluate and prepare a report on the tax implications of macroeconomic changes. The report can be presented to the Parliament as part of the FRBM Act. These reviews can be done twice in a fiscal year, one after 6 months of the start of the fiscal year and another after 10 months. (Section XIII.5.b)

***v)* Data requirements**

w) Data availability can impose constraints on the revenue forecasting process. The two Boards must develop a clear strategy to systematically clean the taxpayer master file and tax returns database. This activity may be beyond the scope of work of the existing TPL and TRU. However, inputs from them will be invaluable from the user's perspective since they have dealt extensively with cleaning up the taxpayer data. (Section XIII.5.c)

x) Once the data is available, it can be examined for underlying pattern, rates of change, or trends, by making comparisons of data from different sources, linking data sets, comparing corresponding items, finding relationship and patterns, and constructing descriptive or hypothetical representational and/or functional relationships between the different variables of which the data is composed. (Section XIII.5.c)

y) After projections have been made, the estimates need to be evaluated for their reliability and validity. To evaluate the validity of the estimates, the assumptions associated with the revenue source need to be re-examined. If the assumptions associated with the existing economic, administrative, and political environment are sound, the projections can be assumed to be valid. (Section XIII.5.c)

z) Reliability can be assessed by conducting a sensitivity analysis. This involves varying key parameters used to create the estimates. If large changes in the estimates result, the projection is assumed to have a low degree of reliability. (Section XIII.5.c)

aa) The TPL and TRU must be sufficiently equipped with computer hardware, database systems, and other office software including the most recent econometric and statistics software packages. These equipment and software systems are necessary to conduct the most demanding data analysis. (Section XIII.5.a) bb) The TPL and TRU should also have access to current, seminal publications on tax and fiscal research and a database of journal articles for the continuous upgrading of the skills and knowledge of the officers working there. (Section XIII.5.a)

***vi)* Partnering non-government bodies and research institutions**

cc) Maintaining a regular dialogue with academia and business communities is valuable for identifying emerging trends in the economy. For these interactions, the academia and business communities must be provided data, albeit encrypted data, so that they have the requisite data for a more meaningful engagement. The CBDT and CBEC, for the purpose, should identify academic institutions of national repute and business or professional associations having all-India membership that can collaborate for the dialogue. (Section XIII.5.d)

***vii)* Integrating revenue forecasting with policy**

dd) It is important to integrate tax forecasting results, particularly relating to tax analysis such as distributional impact out of the micro-simulation model or tariff impact on commodities from the trade tax calculator, to estimate how much each tax policy will cost taxpayers and the tax that it will raise. This can be used as a tool to inform decision makers, including ministers, of the likely benefits and costs, identifying key factors that should affect the decision, strengthening the quality of analysis and making the policy inclusive. (Section XIII.5.e)

***viii)* Organisational/institutional arrangements**

ee) In line with the recommendation in Chapter III of the TARC report, it is reiterated that the existing TPL and TRU wings of the CBEC and CBDT should be subsumed in the TPA wing. The TPA wing should be expanded to include specialists, such as economists, tax law experts, statisticians, operations researchers and social researchers to form a multidisciplinary team. (Section XIII.5.f)

ff) An important task of this unit will be to publish results from the analytical models it develops. These publications should contain objective and impartial analyses and should be helpful in guiding broad policy debates, so that the knowledge can be shared with peers and the policy community for discussion and feedback. (Section XIII.5.f)

gg) The forecasting units in the TPL and TRU should be separate and should have three divisions for macro analysis, tax analysis, and revenue and tax debt forecasting. Each division should have economists, statisticians and social researchers (as per the requirement), along with tax administrators. (Section XIII.5.f)

ix) Staff resources

hh) The officers and staff of the TPL and TRU need to be selected on the basis of specified qualifications. The officers and staff of the TPL and TRU should include trained personnel with specialised skills and knowledge in the fields of revenue forecasting, analysis and monitoring. (Section XIII.5.g)

ii) IRS officers at the levels of Joint Secretary and Director in the TPL and TRU should have wide knowledge of tax policy and macroeconomic issues and deep understanding of revenue estimation and forecasting. (Section XIII.5.g)

jj) For officers below the level of Joint Secretary and Director, knowledge and skill to gather data from different agencies, perform routine analytics for forecasting revenue, and prepare results that could be used for drafting tax memoranda will be useful. These officers and staff can be from the disciplines of statistics, economics, and social science. Duties of these personnel will be to conduct macroeconomic analysis for revenue and tax debt forecasting, monitoring, and analysing tax receipts on a regular basis, and to evaluate the economic and revenue impact of introducing new and/or maintaining existing tax policies on the tax base and tax revenues. (Section XIII.5.g)

kk) Officers below the level of Joint Secretary and Director need to have some knowledge of taxation policies, and direct and indirect tax laws, knowledge of technical report writing, and skills in handling computer databases and statistical packages, and econometrics software, such as SAS, EViews, STATA, etc. (Section XIII.5.g)

ll) On-the-job training should be considered the most important route to developing and deepening the necessary skills required to perform such functions. A university-level education in economics, public policy, statistics and other related disciplines should be considered necessary. Actual strengthening of their capacity can be attained through regular internal seminars, workshops and specialised short courses. (Section XIII.5.g)

mm) These capacity building exercises, particularly internal seminars and workshops will have to be organised on a monthly or quarterly basis so that officers and staff get adequate opportunity to present his/her work on a particular issue to the rest of the unit. (Section XIII.5.g)

nn) Customised and specialised short courses on revenue forecasting can also be organised for staff and officers of the TPL and TRU so that they gain up-to-date knowledge on the subject, and are aware of new developments. These courses should invariably be graded to ensure that the learning is imbibed and can be immediately utilised in the work. Long-term training can also be considered on the subject for the long-run viability of the analytic unit of the TPL and TRU. (Section XIII.5.g)

Chapter XIV

Predictive Analysis

Chapter XIV

Predictive Analysis

List of contents

Chapter XIV

Predictive Analysis

XIV.1 Introduction

In its previous reports, particularly in Chapters VII and IX, the TARC has dealt with the critical importance of Information and Communication Technology and the use of analytics in a modern and effective tax administration. This chapter delves deeper into the use of analytics. Even though the term of reference dealt with in this chapter refers to the use of predictive analysis for detection and prevention of tax/economic crime, it is important to remember that progressive tax administrations treat information as their key asset and, recognising the power of data to assist informed decision making, use analytics in a very strategic way in all key aspects of their management, of which detection and prevention of tax or economic crime is but one dimension. As noted in Section VII.4 of the TARC's report, analytics add immense value in many areas of tax administration by:

- Enabling compliance measurement and accurate identification of non-compliance
- Providing support for effective risk management
- Enabling sectoral/industry wise analysis of trends and tax collections
- Enabling robust revenue forecasting and identifying tax gap
- Improving recovery of arrears or tax debt collection
- Enabling impact analysis prior to and after legislative and policy changes
- Providing predictive support for intelligence and fraud detection
- Enabling tax payer profiling and segmentation for better taxpayer services
- Tracking taxpayer behaviour
- Tracking performance and identifying areas for improving business processes

The TARC would wish the discussion and recommendations that follow to be read in this wider context as well.

In today's strained fiscal environment, tax administrations everywhere face the challenge of achieving cost reduction while at the same time reducing the tax gap by effectively plugging revenue leakages, whether such leakages are through fraud or error. Every tax department has limited resources. It is not possible to examine all tax payers' affairs minutely; audits necessarily have to be selective in order to make optimum use of available facilities. The need for a targeted and cost effective approach to compliance management underlines the need to make robust and reliable risk analysis an integral part of the efficient functioning of a tax administration.

The old adage of prevention being better than cure reflects an eternal truth. Traditionally, analytics was used to identify and correct non-compliance after a transaction had been completed – for example, using analytics to identify cases for investigation or audits. This involves spending already scarce resources to correct or punish non-compliance after it had occurred, e.g., fraudulent or erroneous payments that should never have been issued. If these payments were instigated by criminals, the transactions often involved a stolen identity making recovery of the amounts difficult through traditional collection practices. On the other hand, predictive analytics, embedded in core business operations and integrated with relevant processes, allows risks to be assessed more accurately and in real time, during the processing cycle itself, to mitigate the risk of fraud or non-compliance. By identifying and preventing non-compliance before the transaction is complete, the tax departments can achieve direct savings by freeing personnel from audit to focus on more complex cases and freeing up tax collectors to pursue outstanding arrears. And the deterrent effect of a robust and credible risk management system on potential fraudsters itself has a big positive influence on the compliance environment.

Historically, fraud detection systems have relied primarily on business rules. This has proven useful in identifying recurring instances of fraud through past experience. But there are three main issues with utilising only this methodology. First, business rules create a lot of noise. Legitimate customers constantly do things that look suspicious. False positives take time to triage and result in operational inefficiency. Second, business rules become common knowledge to the fraudsters. Either by trial and error, or worse, by infiltrating the organization, people can gain knowledge of or make reliable guesses about business rules and outwit the system. And third, business rules are not forward looking. They are reactive and are not there to catch tomorrow's fraud. This is where predictive analysis scores over traditional fraud detection systems, as it enables an organization to anticipate events and manage risks proactively.

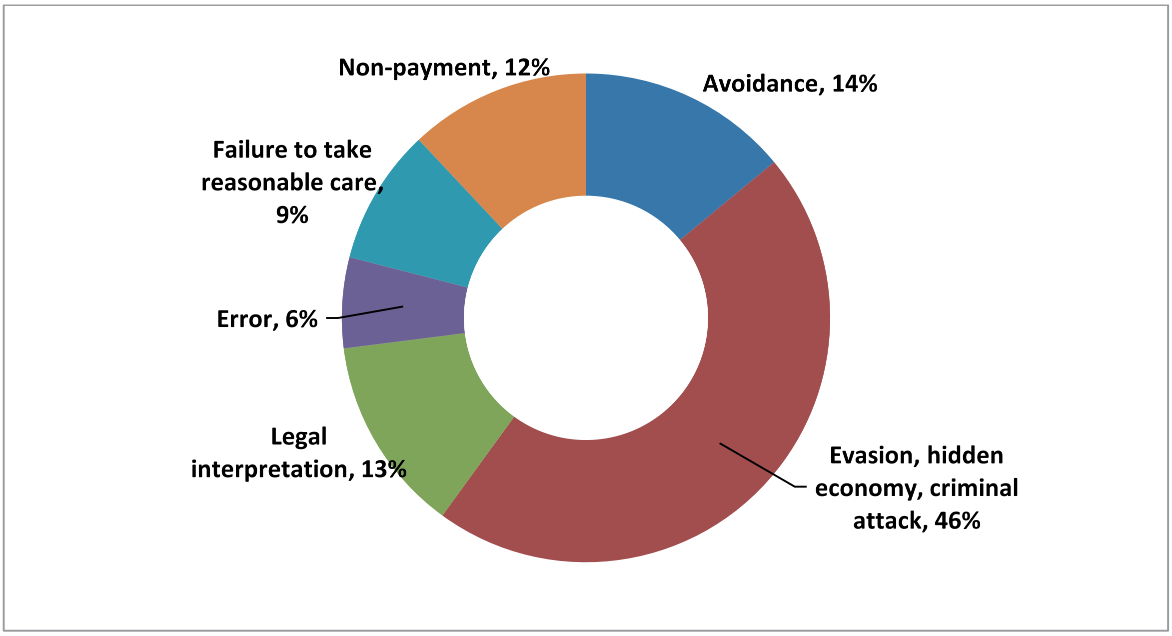
According to the Tax Justice Network, total tax evasion in excess of USD3.1 trillion (or 5 per cent of world's GDP) occurs as a result of the operation of the shadow economies found in every state and fraud and non-compliance threaten government's already burdened bottom line.[406]

In the United Kingdom, fraud accounted for 46 per cent of the HMRC's estimate of the tax gap of £32 billion for the year 2010-11, as depicted in Diagram 14.1 below:[407]

[406] Tax Justice Network, The Cost of Tax Abuse, November 2012 cited in Accenture ibid

[407] Closing in, HMRC's approach to tax evasion, December 2012

Diagram 14.1: HMRC's estimate of the tax gap for the year 2010-11



In the United States, the Internal Revenue Service (IRS) estimated its tax gap in one year at $385 billion.[408]

Around €1 trillion euro is estimated to be lost to tax evasion and tax avoidance in the EU every year.[409] The VAT gap for the EU was estimated at approximately €193 billion (1.5 per cent of GDP) in 2011.[410]

Many studies indicate that there is huge potential to unlock value through the exploitation of big data. As the TARC had mentioned in Section VII.2 of its report, a study by McKinsey Global Institute estimated that big data analytics has the potential to reduce Europe's tax gap by 20 per cent.[411] Another McKinsey benchmarking study of 13 tax administrations estimated the potential to generate additional revenues through enhanced effectiveness at US$86 billion and a saving of costs of around US$6 billion through improvement in efficiency in four core functions of the tax administration.[412] A study by the ACT-IAC Institute for Innovation in the US found that the US IRS could, through appropriate ICT based intervention using big data analytics to combat fraud and eliminate waste, save USD50 billion annually by recouping uncollected taxes and avoiding

---

[408] http://www.irs.gov/pub/newsroom/tax_gap_map_2006.pdf accessed January 2015

[409] http://europa.eu/rapid/press-release_IP-12-1325_en.htm accessed January 2015

[410] http://europa.eu/rapid/press-release_IP-13-844_en.htm accessed January 2015

[411] Big Data: The next frontier for innovation, competition and productivity, McKinsey Global Institute, May 2011

[412] The Road to Improved Compliance, A McKinsey benchmarking study of tax administrations – 2008-09, McKinsey and Company, Sept. 2009

improper payments.[413] By using advanced analytics, the Belgian tax administration is reported to have reduced VAT carousel fraud by 98 per cent.[414]

It is widely recognised that the key lever that can enable tax administrations to successfully negotiate this daunting challenge is information and consequently, all modern, forward looking administrations have made the harnessing of ICT a central pillar of their strategy. They are addressing the challenge on two fronts by using analytics: detecting, preventing and stopping transactions before they are processed and using the information gleaned from analytics to significantly reduce operating costs and drive business results.

Availability of data today provides a great opportunity to use the power of analytics. The digital world is constantly expanding as ICT penetrates deeper in our lives and becomes an integral part of the personal, social and business world. Individuals and organisations leave their digital footprint in all manner of places as they go about their business. Hence, along with the data from taxpayers' returns, data from diverse sources is accessible and available to tax departments. By capitalising on this availability of taxpayer data and advances in technology, tax departments can better understand the characteristics and motivations of different types of taxpayers and quickly tailor their response accordingly.

Setting up an early warning system for fraud, therefore, should be a critical part of the action plan against tax frauds. Because of the complexity of frauds, an approach dealing with contemporary and emerging challenges requires the tax administrations to develop the capability to identify networks of individuals and businesses involved in fraud. Detecting fraudulent networks requires analysis of a huge number of taxpayers, returns filed and intra-community transactions. Frauds these days, being high velocity frauds, require a quick response. Tax departments have limited analytical ability – the inability to perform advanced analysis of transactions in order to identify suspicious entities, and relationships and behaviours make authorities susceptible to organised crime. Manual processes – manually gathering and preparing data for fraud detection is time consuming, tedious and highly error prone. A reliable system for gathering and managing data from diverse systems and a sustainable capacity for advanced analytics, therefore, are the bedrocks of a high performing tax administration.

XIV.2 Current situation

As the TARC noted in its earlier reports, the CBDT and CBEC are currently functioning in separate silos with only sporadic sharing of information between them. Both have taken steps to set up data warehouses – CBEC's data warehouse is already in operation while the data warehouse of the CBDT is in the process of being set up. While both have taken major strides in automating key business processes, there are important areas of their operations that are yet to be automated. Leave

[413] Unleashing the Power of Information Technology Innovation to Reduce the Budget Deficit, Institute for Innovation, American Council for Innovation – Industry Advisory Council, 2012

[414] http://www.sas.com/en_us/customers/tax-fraud-belgium.html accessed in January 2015

alone comprehensive integration of data across direct and indirect taxes, the sharing of data internally across different taxes and functions within the two Boards remains well below par. Hence, a significant proportion of data that could be used for predictive analysis lies scattered either in different data bases or in paper records, where processes continue in the manual environment. Consequently, the use of predictive analysis in the detection and prevention of tax offences is limited and it would appear that there is considerable scope and an urgent need to advance in that direction.

CBDT

Enforcement

Under the CBDT, the primary responsibility for enforcement rests with the Director Generals (Investigations) located in different cities in the country. Cases with potential for search and seizure action are identified at present on the basis of leads from diverse sources, most commonly the following:

i. Lead from an informant: Some specific information in respect of the modus operandi, including evidence such as bank statements, copy of agreements and information about booking of bogus purchases or unaccounted sales may be provided by an informant. These are analysed and discreetly investigated to determine whether the information provided is actionable.

ii. Lead from an earlier search case: Sometimes, evidence gathered during the course of a search action points to tax evasion by another entity. The evidence is analysed to arrive at a conclusion about method of tax evasion by the concerned entity.

iii. Lead from an STR: Suspicious Transaction Reports (STRs) that are received from banks through the (Financial Intelligence Unit (FIU) contain details of suspicious transactions through banking channels, which provide leads to tax evasion cases. These mostly relate to huge cash deposits, circuitous or layered transactions and transactions without economic rationale. By analysing the transactions and going into the layering of bank accounts, the actual beneficiary is identified. Cases of such beneficiaries may then be processed for search.

iv. Lead from a TEP: Tax Evasion Petitions (TEPs) are complaints of tax evasion received by the Department form the general public. These sometimes contain specific information about tax evasion by an individual or a group including the modus operandi of tax evasion, details of bank accounts used for depositing unaccounted money or details of investments out of unaccounted income. The details contained in TEPs are analysed and discreet enquiries conducted to determine whether the case warrants search.

v. Media news and other publications: Leads for developing a search case are also taken from news on cash transactions, corruption and frauds appearing in media. One such example is the International Consortium of Investigative Journalists (ICIJ). Similarly, information about foreign bank accounts received from various sources also provides leads in processing a case for search.

vi. Market information: Field enquiries and reconnaissance activities may give local information about the business practices of a group with respect to the generation of unaccounted wealth. This information is correlated with actual business results to decide on search action in the case of a particular group.

vii. Information from other agencies: Sometimes, information emanating from action taken by other law enforcement agencies such as the CBI and Enforcement Directorate, and the Sales Tax, Excise and Customs Departments provide leads in developing a case for search. Similarly, information received under the Spontaneous Exchange of Information from a foreign jurisdiction is also helpful in processing a case for search.

These leads are developed by gathering information and evidence through both external and internal means.

External Means

External means of intelligence gathering includes information available in public domain such as on the internet, information through field visits and physical reconnaissance, factory/office visits and information about modus operandi in the trade.

Internal Means

Internal means include databases available/accessible to the Department through the Integrated Taxpayers Data Management System (ITDMS), which is a tool for generating a 360 degree profile of a taxpayer, and corporate database available from the website of Registrar of Companies (ROC) under the Ministry of Corporate Affairs (MCA).[415]

The ITDMS collates information relating to a particular entity from different internal and external sources and builds a 360 degree view of its affairs, including details of all expenditure/investment, returned incomes, taxes paid, particulars of group companies, directorship details, family relationships, linkages with other entities etc. Besides, processing cases for search, the ITDMS also helps in the realisation of uncollected tax demands.

At present, various fields available in the ITDMS software and the process of updating of these fields are:

- PAN, AIR, CIB, AST, TAN and OLTAS – These fields are being updated with data received from the Directorate of Systems on a request basis.
- Director Info – The field is updated from the data provided by the Registrar of Companies. The data has been updated until October 19, 2012, and ROC, Delhi and Haryana, has been approached to provide the updated information. However, the data has not been obtained from the concerned authority.

[415] The ITDMS was awarded the Prime Minister's Award for Excellence in Public Administration in 2010 for its impact, innovation, the re-engineering and its relevance for other enforcement agencies.

- Mobile information –The field is updated on the basis of data received from various telecom service providers on a monthly basis.
- AADHAR, Passport, ration and Voter ID – these fields have yet to be loaded on to the ITDMS system.

However, there is little predictive analysis being undertaken on the basis of data available to identify potential frauds. The Data Warehousing and Business Intelligence module being developed in the department is expected to provide for predictive analysis.

Compliance management

Besides enforcement, the Income Tax department also uses ICT for compliance management. The selection of cases for scrutiny is done through the Computer Assisted Audit Selection (CASS) system, in which parameter-based scoring is done and cases are picked up for scrutiny based on certain predetermined criteria. CASS is reported to use certain parameters to pick up returns that reflect high refunds, certain exemption claims, introduction of fresh capital funding into the balance sheet etc. for deeper scrutiny u/s. 143(3) of the Income Tax Act. Third-Party data from several agencies such as banks, sub-registrars, mutual fund managers and so on are also integrated into the analysis. Selection of parameters and assigning a fixed score to a particular parameter in the programme are based almost exclusively on the experience of departmental officers and there is little, if any, use of modelling or business intelligence techniques based on data analytics.

As already noted in section XII.2.e of the TARC's report and according to anecdotal feedback received from field formations, the efficacy of the CASS is yet to be consistently established. A persistent complaint from the Assessing Officers has been that the criteria, which are primarily transactional, often results in the selection of cases that do not yield meaningful results in terms of detection of evasion or concealment. Correspondingly, from the taxpayers' side also, there are complaints that the returns are picked up for scrutiny from year to year apparently based on the same criteria although scrutiny assessments in earlier years find the concerned transactions to be acceptable. A possible reason for this is that the CASS (together with the ITDMS) is more of a profiler rather than a predictive model. CASS appears to collect certain data primarily based on transactional criteria, and to apply a few ratios in order to determine a particular profile which it perceives as risky. However, historical data available with the Department in the system on the particular taxpayer (this data in the case of most taxpayers today is available for more than 10-15 consecutive years) is apparently not utilised in the analysis to obtain a clearer understanding of the compliance behaviour of the taxpayers and a better assessment of risks. Purely transaction level criteria inevitably suffer from too many "false positive" hits and undermine the effectiveness of the system. This is where predictive analysis, by use of appropriate models, could be of invaluable help by identifying the right type of cases. Further, CASS has been known to make erroneous selection for scrutiny by stipulating wrong interpretation of selection criteria. For example, 'investment' in mutual funds uses 'gross' rather than 'net' investment so that, an investor with an investment of say Rs. 1 crore may be interpreted to have invested Rs. 4 crore since the outgo of an investment is not netted out. This yields a non-match between income and investment, or a false positive. Such errors are not attended to by the department, leave alone corrected, putting honest taxpayers (retirees) to the utmost unwarranted difficulty.

The central processing centre for tax deducted at source (CPC-TDS) appears to be more advanced in the use of analysis of TDS payments, in conjunction with the OLTAS and related third party data. More than 15 lakh deductors have deducted and deposited tax to the tune of 40 per cent of the total direct tax collections and have reported more than 50 crore financial transactions during 2013-14 through around 60 lakh quarterly TDS statements. This data captures financial transactions pertaining to 94 per cent of the taxpayers. TDS data covers 17 types of payments (salary, bank interest, professional charges, contract payments, rent, commission, payments to non-residents etc.) amounting to more than Rs.40 lakh crore (35 per cent of GDP in value terms) during 2013-14. Similarly, OLTAS has details of tax payments made by deductors and taxpayers. There are around 4.5 crore challans received during the year. The challan information includes major head (corporate/non-corporate), minor head representing type of payment (advance tax/self-assessment/regular/TDS) and the section code in the case of TDS. Further, the AIR data contain the details of 6 different types of high value investments based on the annual information return filed by prescribed agencies under section 285BA of IT Act. CIB data also has information on various codes including commodity stock exchange transactions, cash deposits etc.

Using these data, the CPC-TDS has segmented the population of deductors by geography, size (the number of financial transactions in TDS statements and amount of tax deducted), forms used reflecting nature of payment (Forms 24Q, 26Q, 27Q, 27EQ) and the status of the deductor (e.g. corporate/non corporate). The insights based on the analysis have given the department a better understanding to the taxpayer population and are being used to tweak processes/ procedures to improve compliance as well as customer services. For instance, analysis revealed that different deductors have different types of filing behaviour – some deductors file all four forms (Forms 24Q, 26Q, 27Q and 27EQ) for each quarter, while others file only one form for one or more quarters. This analysis has facilitated streamlining the issue of notice to non-filers. The CPC-TDS portal has enabled a new functionality for deductors to report instances where the deductor is not required to file a TDS statement as there were no transactions liable for TDS for a particular quarter, thus avoiding infructuous and irksome notices.

Analysis has also enabled more accurate risk profiling of deductors as well as the nature of defaults and has enabled better performance management. Based on an analysis of defaults, the CPC (TDS) has recently rolled out a functionality to enable the deductors to rectify minor errors in TDS statements, if any, within a window of 7 working days of filing of their statement before processing the statement for TDS defaults. This step will minimise defaults arising out of typographical mistakes and avoid unproductive efforts in error rectification. The responsiveness of the deductors to resolve the defaults is being monitored and is factored as a variable in risk profiling. Similarly, analysis showed that bank branches were responsible for 70 per cent of PAN errors and 25 per cent of total TDS defaults. CPC-TDS, therefore, introduced the concept of "Corporate Connect" to the respective corporate headquarters to manage TDS defaults at the level of their branches, thus improving performance.

Through analytics, CPC-TDS is able to cull out the data of the deductors on various aspects of TDS compliance like deductors paying late deposit of tax deducted at source, quoting invalid/no PANs, late filing of TDS statement, stop filers, TDS defaults, paying tax but not filing TDS statement etc. CPC-TDS has already sent more than 2 crore advisory emails to deductors to sensitise them on key general and specific issues related to TDS compliance. This communication has found encouraging response from various taxation related websites, who have given wide publicity to it thus creating a multiplier effect, facilitating a movement towards greater compliance in a non-intrusive, non-adversarial manner.

CPC-TDS has also undertaken analysis of call centre and grievance portal data to identify the areas where there appears to be lack of clarity in understanding and based on the analysis, educational and awareness material on the relevant areas is being put out and continuously improved through means such as e-tutorials, FAQs etc. Similarly, process improvements are being undertaken based on an analysis of feedback received through e-mails.

The non-filers monitoring system (NMS), which was implemented as a pilot project under the Data Warehouse and Business Intelligence (DW & BI) project, is another example of the use of data analysis to identify non-filers with potential tax liabilities and prioritise action based on risk assessment. The salient features of this initiative are the following.

i. Data analysis was conducted to identify PAN holders who had not filed Income tax returns despite conducting high value transaction as reported in AIR, CIB data and TDS/TCS returns.

ii. A bulk data matching exercise was carried out with the Financial Intelligence Unit (FIU) to include non-filers who had conducted high value cash transactions.

iii. Rule-based algorithms were applied to classify and prioritise cases for graded monitoring.

iv. A Compliance Management Cell (CMC) was set up to send letters and capture responses from non-filers.

v. Bulk letters were sent to PAN holders communicating the information summary and seeking to know the submission details of Income tax return

vi. An online monitoring system was implemented to ensure that information related to non-filers is effectively used by the field formation

vii. Standard Operating Procedures (SOPs) were issued to ensure that field formations maintained consistency in their approach.

Two NMS processing cycles (in January 2013 and January 14) identified 1.22 and 2.21 million non-filers respectively, with potential tax liabilities. The project also deployed a compliance module on the e-filing portal and information related to non-filers was made available to PAN holders. SMS and email were sent to the target segments asking them to access the e-filing portal and to provide their response electronically.

As a result of this initiative, a large number of taxpayers have submitted their Income tax returns and commensurate self-assessment tax and advance tax has been collected. The pilot project also provided valuable learning and identified the challenges, including issues bearing on data quality and delays in capture of manual returns, in effective integration and utilisation of data.

CBEC

Enforcement

Currently, the Directorate General of Revenue Intelligence (DGRI) and Directorate of General of Central Excise Intelligence (DGCEI) and the various customs and central excise and service tax field formations are involved in the detection and investigation of offences, under relevant laws. While DRI's mandate covers the Customs Act and other allied Acts relating to import and export of goods, DGCEI has the mandate to investigate and prosecute offences relating to central excise and service tax. In addition, the Directorate General of Audit plays a critical role in compliance management as it is the nodal organisation for the management of the CBEC's audit programme. The extent of use of ICT and data analysis has been relatively stronger in customs than in the central excise and service tax. Customs have the advantage that with the Indian Customs EDI System (ICES) being operational in most of the signification ports/airports in the country, they have access to transaction level import and export data.

As mentioned in the case of DGs (Investigation) in CBDT, human intelligence and complaints about evasion from different sources continue to be an important basis of enforcement action, and the processes followed for verification of source information are more or less similar. In view of the greater richness of customs data, the use of analysis appears to be relatively greater in the case of DRI than the DG (CEI).

In the context of DRI's mandate, predictive analysis is fundamentally about exploiting historical and current data for (i) unearthing outright smuggling and detection of commercial frauds (ii) risk assessment and management of import and export entities/transactions and (iii) decision-making support for giving recommendations on legislative/policy changes that would curb non-compliance. Although the use of predictive analysis within DRI is in a nascent form, consistent efforts are on for greater use of such analysis and to increase its level of maturity. Some examples of use of analysis by DRI are given below.

Outright Smuggling – Based on historical data of cases booked, officers developed intimate knowledge on the profile of suspect shipments in the case of smuggling of Red sanders out of India. This knowledge, when applied to export data, has resulted in identifying many shipments which were intercepted at either the ports of export or even at a foreign port of transit / import.

Commercial Frauds – Analysing variations in the quantum and commodity mix reflected in the import data under FTAs, the profiles of importers who availed the benefits of export incentive schemes along with the nature of their exports and imports transactions and unusual trends noticed in the country of origin of imports, to name a few, have all helped officers detect commercial frauds.

Based on its analysis, the DRI has also been able to make recommendations for policy changes. Some of these are given below.

i. After studying the changing pattern in the quantum of imports together with the profile of importers and the extent of gold smuggling detected, DRI had recommended policy level changes with regard to encouraging legitimate imports of gold and curbing its smuggling. The recommendations of DRI were accepted and implemented.

ii. DRI had also provided valuable inputs on curbing trade-based money laundering and generation of black money after a thorough analysis of import and export data of select countries.

The above illustrations are a few instances of predictive analysis or data-informed decisions that have yielded substantial outcomes. From the experience of the DRI, it can be said that the successful outcomes out of predictive analysis squarely depend upon (a) extent and quality of data and (b) domain experts and dedicated resources available for data analysis. DRI is constrained on both accounts.

The data for analysis is currently limited to import and export data that is collected as part of the customs clearance process and the offence data that is maintained by DRI. This provides at best an incomplete view of the importer/exporter profile. Predictive analysis would be more successful when regulatory agencies have a full 360 degree view of the compliance profile of the entity involved in imports and exports. This would require that external sources of data are integrated with the data currently available to DRI. Some of the external sources of data that would be of immense use for DRI in discharging its mandate are as follows:

i. PAN, assessment, audit and offence data maintained by the CBDT;

ii. Registration, annual audited financial statements and offence data maintained by Ministry of Corporate Affairs in MCA 21 project and also data maintained by Serious Frauds Office;

iii. Returns and offence data maintained by state VAT administrations;

iv. Foreign exchange data maintained by the Reserve Bank of India;

v. Passenger details maintained by the Bureau of Immigration;

vi. Import and export details from Special Economic Zones;

vii. Suspicious transactions in foreign exchange maintained by Financial Intelligence Unit

viii. Offence data maintained by the Directorate of Enforcement;

ix. Licensing data maintained by DGFT; and

x. Risk indicators maintained by other border regulatory bodies such as food safety and security, animal and plant quarantine, drug controllers, Ministry of Environment etc.

Compliance management

To achieve an optimal balance between the concerns of facilitation and enforcement, the CBEC has created the risk management system (RMS), which is managed by the Risk Management Division (RMD). In Chapter VIII of its report, the TARC has highlighted the need to revamp the customs core process if customs in India are to achieve global benchmarks. The TARC had noted the large gulf that exists between the levels of interventions in India and those in advanced as well as in emerging economies. If customs in India are to advance, the role of the RMD will be critical, as already highlighted in Section VIII.4.d of the TARC report.

The current approach to predictive analysis is based on identification and profiling of risky entities involved in import/export transactions and suspicious transactions that could potentially be non-compliant. Combinations of entities and transactions are analysed to make hypotheses or predictions of risk. Domain experts predict/identify possible areas (transactions/transacting entities/interplay of both) of non-compliance (risk) and treat it by employing the RMS tools like rules, targets and interventions, after a due impact analysis. The Valuation Risk Assessment Module (V-RAM) managed by the Directorate General of Valuation (DGoV) is integrated with the RMS to spot outliers on the fly, before clearance of goods by customs. Similarly, RMS is also equipped to handle Risky Descriptions (RDH) and Compulsory Compliance Requirements (CCRs) under the customs law and other government departments' regulations. A parallel predictive analytical set up to select the declarations for post-clearance audit (PCA) is also in place. Besides, inputs from various external sources also get into the RMS to make the predictive analysis a little more diverse and robust. RMD maintains close co-ordination with field formations, DG (Audit) and DRI to get inputs that could be actionable for interdictions in the RMS. The RMD team also analyses clearance trends and reviews interdiction tools periodically for improving and refining its risk analysis. Offence case details collected from the DRI and other field formations, market trends and general intelligence gleaned from diverse external sources are other inputs used in predictive analysis. Currently, on the whole, the techniques rely predominantly on the knowledge of domain experts and no state-of-the-art statistical/analytical tools seem to be employed for predictive analysis. Further, the data used are largely internal data and there is very little access to external sources of data, which has a direct bearing on the effectiveness of predictive analysis and its outcomes.

The DRI has a dedicated team of officers working in the Commercial Intelligence Cell for detection of frauds and a National Risk Management team for risk assessment and evaluation. The officers have domain expertise and are assisted by vendor teams in analysing data. The officers use sophisticated tools available in Enterprise Data Warehouse, tools within Risk Management

Application and various other in-house applications developed by the DRI. The Customs Overseas Information Network (COIN) officers, who are posted outside the country, are also entrusted the task of analysing the pattern of imports and exports originating from countries under their jurisdiction and developing actionable intelligence. Data sharing mechanisms, although not automated, are also in place vis-à-vis certain countries.

However, the total complement of officers involved in data analysis per se is very small and efforts are underway to scale up the size of resources. It is reported that there is a proposal to set up a National Targeting Centre (which was also recommended by the TARC vide Section VIII.4.d of its report) that is under active consideration and which would house an analytical wing with the mandate to build risk profiles, track emerging commercial and contraband risks and developing intelligence inputs. The proposed National Targeting Centre would be responsible for:

i. Increasing precision in interdictions

ii. Collecting and collating information and developing intelligence and risk profiles for tracking non-compliant entities and transactions

iii. Undertaking risk assessment, risk evaluation and risk mitigation techniques

iv. Updating the risk parameters

v. Knowledge management with regard to statistical and analytical methods, tools and technology

vi. Assessing hardware/software requirements

vii. Identifying different data sources, both external and internal including historical data, offence data, declarations, periodical reports and returns.

viii. Identifying, interacting and co-ordinating with agencies/sources to procure/access and manage the data

ix. Integrating various applications for obtaining 360 degree view of entities.

x. Data mining and predictive analysis of internally and externally sourced data

Recognising the significance of predictive analysis as a vital tool in fulfilling its mandate, the DRI has been giving it due importance and encouragement within the organization. However, as mentioned at the beginning, predictive analysis is still in a nascent form in DRI. This is basically due to the fact that while officers have sufficient domain expertise, the knowledge of statistical techniques/predictive models and the operational capacity to handle sophisticated analytical tools have to be improved. The DRI is focusing on requisite skill development and capacity building in this regard. Simultaneously, the feasibility of hiring data analysts who would support domain experts is also being explored.

XIV.3 Advanced practices in the use of analytics

As was noted earlier, tax administrations in advanced and many emerging economies are increasingly focusing on the use of data analytics to improve tax administration performance and to detect and suppress frauds proactively. Besides fraud detection, the power of analytics is fast becoming central to supporting and protecting taxpayers, tailoring service delivery and operating an efficient tax administration.

This section reviews some of the widely applied methodologies in this area.

Considering the highly technical nature of the subject, it would be useful to begin with an explanation of the field of analytics and data-mining, terms which are often used interchangeably.

The EU Compliance Risk Management Guide[416] provides the following description of data mining: "An analytic process designed to explore data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data."

According to the OECD Guidance Note on Compliance Risk Management,[417] "data mining is the process of exploration and analysis, by automatic means, of large quantities of data in order to discover meaningful patterns and rules."

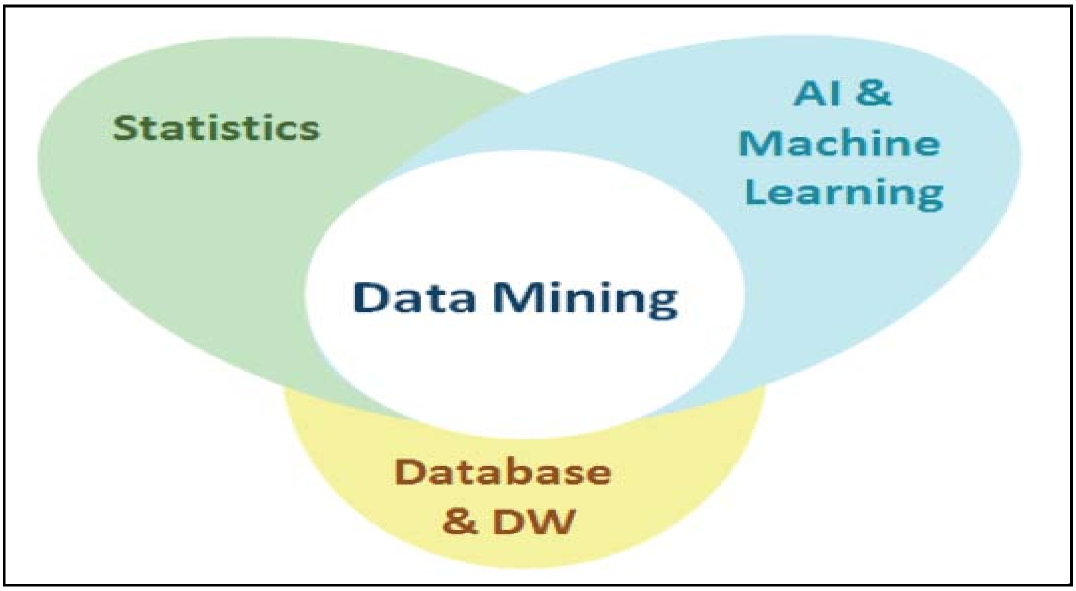
According to Sayad,[418] data mining is about explaining the past and predicting the future by means of data analysis. As illustrated in Diagram 14.2 below, data mining is a multi-disciplinary field that combines statistics, machine learning, artificial intelligence and database technology.

[416] Compliance Risk Management Guide for Tax Administrations, Fiscalis Risk Management Platform Group, 2010

[417] OECD Forum on Tax Administration Guidance Note: "Compliance Risk Management – Managing and Improving Tax Compliance", 2004, Committee on Fiscal Affairs

[418] http://www.saedsayad.com/data_mining.htm

Diagram 14.2: Data mining a multi-disciplinary field



Unlike simple analysis, analytics is the discipline of using statistics in combination with computer programming to gain insights contained in large quantities of data and undertaking actions based on those insights. Analytics is data driven in that it relies on information and knowledge contained in data holdings. Analytics uses historical and current data to predict future events and operates using data on a very large scale and is suited to understanding, using and communicating complex data.

In the mid-1990s and for several years thereafter, the term used for this kind of activity was "Business Intelligence" (BI). Even before the 1990s, statisticians were actively involved in building statistical models in order to analyse and forecast certain aspects of economic activities so that decision making could be aided. However, with the exponential growth data following the internet and the rapid emergence of the digital world, it became a challenge for conventional data base technologies and statistical applications to manage the models. It was then that the specialised practice of data mining emerged, which evolved into a more sophisticated incarnation as BI, crystallising into the highly-refined practice of predictive modelling. Today, analytics are broadly classified as descriptive and predictive and now business intelligence refers to a more limited practice of culling out reports from data and performing certain ad-hoc queries upon data stored in warehouses or data marts. Predictive modelling, on the other hand, identifies and mathematically represents underlying relationships in historical data to explain the data and make predictions about future events. In other words, BI typically tells you what happened, analysis tells you why it happened and analytics tells you what will happen.

Predictive analysis is a tool that transforms data into valuable knowledge. Data without analytics has limited utility; it is analytics that finds patterns in data to make it actionable. It is a form of data mining which seeks to achieve a dependable level of forecasting in relation to trends and the behaviour of stakeholders. It is a system that enables computer models to draw certain conclusions from large amounts of data by finding complex and hidden patterns that are not otherwise easily apparent. Predictive models typically analyse current and historical data to produce easily understood metrics such as scores. These scores rank-order individuals or transactions by the likely future performance, e.g., the likelihood of a transaction being fraudulent (risk detection). A number of analytic methodologies underlie solutions in this area including the following:

- Application of linear and nonlinear mathematical programming algorithms, in which one objective is optimised within a set of constraints,
- Advanced "neural" systems, which learn "complex"[419] patterns from large data sets to predict the probability that an individual will exhibit certain behaviour of business interest. Neural Networks (also known as Deep Learning) are biologically inspired machine learning models that are being used to achieve breakthroughs in speech recognition and visual object recognition.
- Statistical techniques for analysis and pattern detection within large data sets.
- Regression techniques, which are used for estimating the relationship among variables.

Analytics, thus, is a multi-dimensional discipline. There is extensive use of mathematics and statistics, the use of descriptive techniques and predictive models to gain valuable knowledge from data analysis. The insights from data are used to recommend action or to guide decision-making rooted in a business context.

Predictive modelling today is used widely within Information Technology systems. For instance, in spam filtering systems, predictive modelling is issued to identify the probability that a particular message is spam. Similarly, such predictions are widely being used for customer-relationship management, change management, disaster recovery, city planning, meteorology and various forms of security management. And, as noted earlier, it is finding increasing use in forward looking tax administrations.

The central building block of predictive analytics is the "predictor" which is the characteristic value related to the desired prediction. Whenever a prediction goal is set, there could be a large number of predictors that could have varying degrees of relationships with the goal. After identifying such relevant predictors, the next step would be to combine two or more of them with a formula. Such a combination forms a predictive model and hence the term predictive modelling.

---

[419] As explained in *Taxation – Principles and Application - A Compendium* by Parthasarathi Shome, Lexis Nexis, 2014(P 442) "The theory of Chaos and Complexity is best explained through an example of throwing a pebble in a pond. The ripples have no pattern; therefore, are chaotic. At the edge of chaos, where the ripples are dying down, however, complex patterns, based on the Mandelbrot sets $z \rightarrow z^2 + c$ (after the scientist who programmed it), may be discerned (Gleick 1988). On the basis of this, hypotheses of behaviour may be formulated. Complexity theory can thus be, and is being, used to explain biological cell growth, international trade patterns or galactic formations. See Johanson (2009) and Beinhocker (2007). The Indian Statistical Institute, Kolkata, and Santa Fe Institute, New Mexico, have pioneered research in this area though it remains nascent."

It becomes obvious that it is the quality and the correct combination of predictors that will yield a rigorous results when multiple aspects of the customers' status and behaviour are covered. Therefore, a really effective predictive model must be more complex and richer than simple linear models involving summations or two-variable regressions alone.

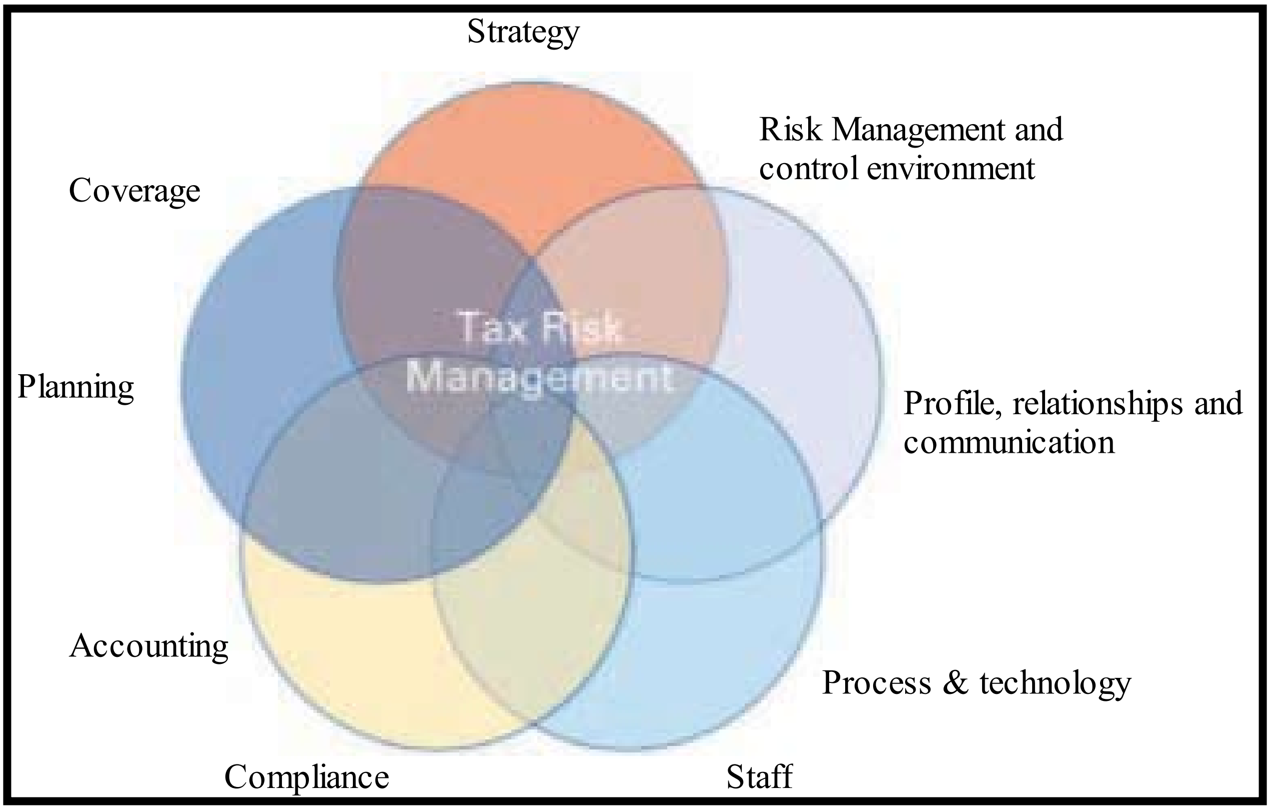
The real challenge, therefore, would be to identify the best predictors in a given context in order to populate the most effective predictive model. The predictive model suitable to a particular organisation can be built on the basis of the extent and quality of data, viz., customer data, available with it, in terms of its extent, depth, indicative versatility and quality. This process is illustrated in Diagram 14.3 below.

Diagram 14.3: Predictive modelling

While predictive analytics software applications could build models automatically, the effective alignment and integration of models with the various facets of the organisation and its peculiar characteristics require close experiential intervention. For example, the predictors, being the critical variable factors that are likely to influence future behaviour or indicate the quality and purpose of past conduct, need to be carefully chosen, based on field experience and the exposure of the selectors, and relevant additional data as and when available has to be integrated into the predictive system at the most appropriate time and level of the process.

Predictive analysis has to be a very important component of the larger risk management strategy of a tax administration. Both the identification and classification of risks facing the organisation need to be carefully analysed in the context of the organizational goals and related to a simultaneous classification or segmentation of the tax payer base. It is only when these two relational areas are determined that an effective analysis will result. Therefore, segmentation of risks as well as of tax payers is a necessary first step in any risk management process where predictive analysis may be applied. The basic elements contributing to tax risk analytics (within the genus of which predictive analysis operates) have been illustrated in Diagram 14.4 below[420].

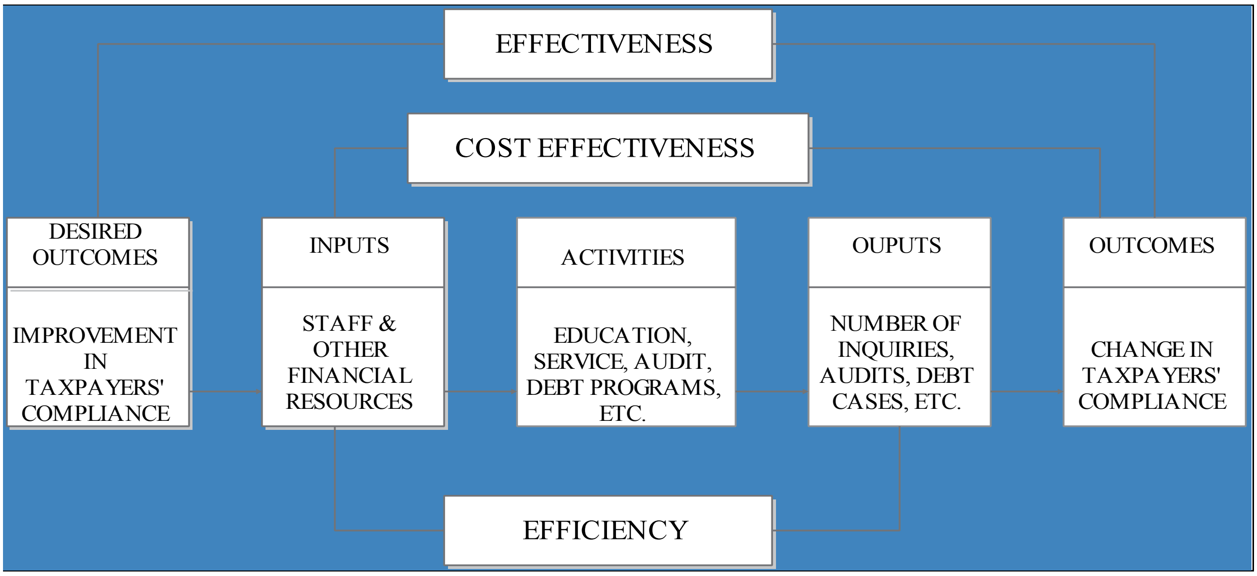
Diagram 14.4: Context for analytics



It is important to mention that the best practicing tax administrations now give analytics (of which predictive analysis is a very important component) a central place in their strategies, and use it as an underpinning of their programmes for continuous improvements in efficiency as well as effectiveness. The Australian Tax Office (ATO), for instance, has delineated the logic of their compliance management programme as shown in Diagram 14.5 below.

[420] Erle Bernd, "Tax Risk Management and Board Responsibility"(2007),http://www.itdweb.org/documents/Erle.pdf

Diagram 14.5: ATO's compliance management programme



In advancing their compliance management, the ATO makes extensive use of analytics across the core areas of their administration, including suppression of fraud, effective debt collection and continuously improving customer services by well researched segmentation.

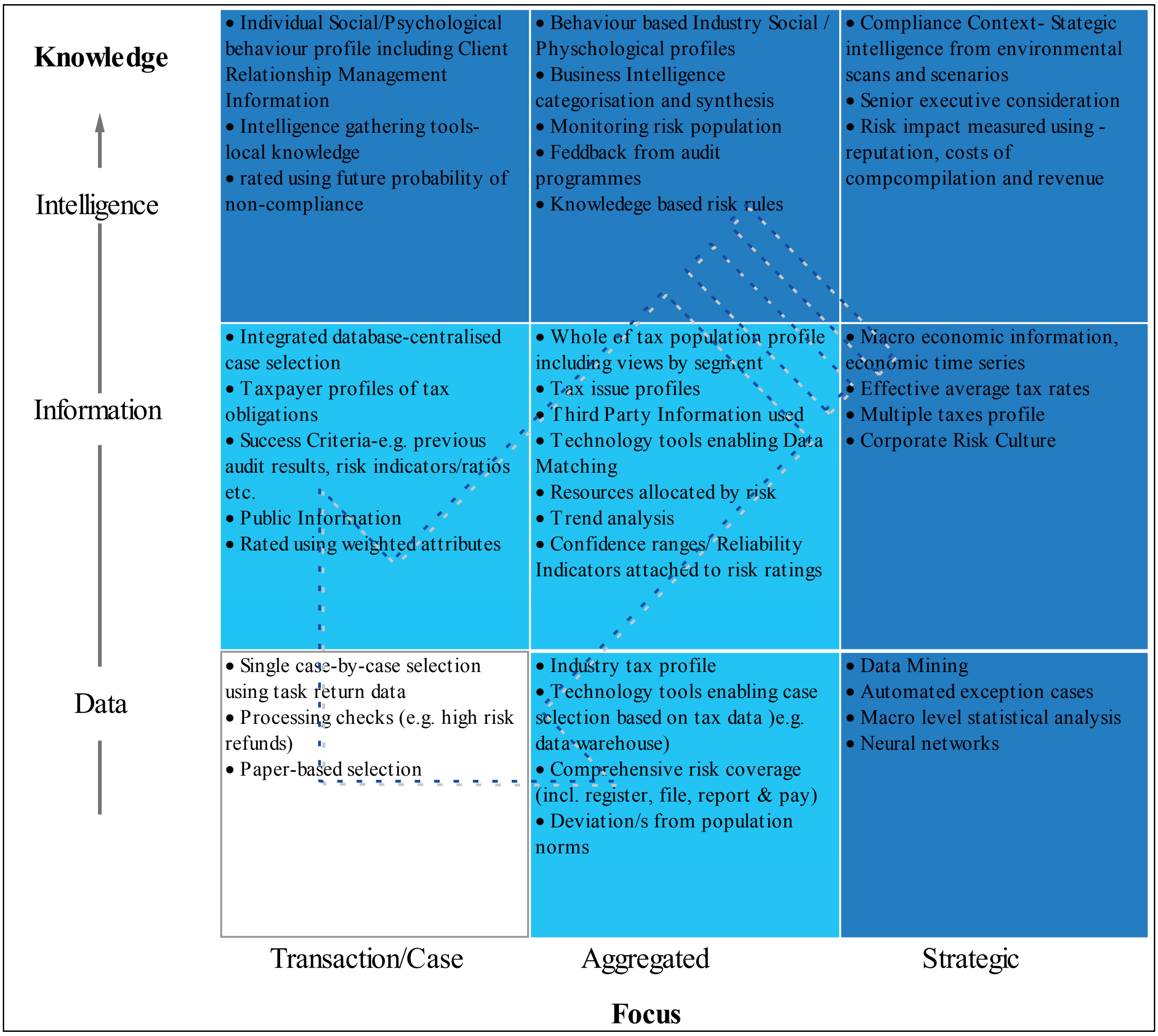
The OECD's popular Risk Identification Diagnostic Model (Diagram 14.6) indicates the broad aspects involved in running a risk analytic model. The OECD paper[421] further notes:

Most authorities use a range of data sources and data manipulation techniques accompanied by analytical tools and indicators to identify emergent risks and assess their significance. The use and manipulation of data is an important activity for both risk identification and risk assessment and sizing ... Most authorities use the analytical tools and techniques to evaluate the effectiveness of compliance treatment strategies.

Appropriate use of analytics thus drives all key actions of the tax administrations, whether strategic or operational. It may be noted that there are four progressive sources for use in customer profiling, viz., data, information, intelligence and finally knowledge. This is set off against specific transactional data and the strategic goals of the organisation. Such integrated profiling contains an effective mix of macro and micro variables and psychological assessment that could provide the basis for drawing reliable conclusions about the likely future conduct and behaviour of a particular tax payer or segment and tailor the administration's response accordingly.

[421] OECD Forum on Tax Administration Guidance Note: "Compliance Risk Management – Managing and Improving Tax Compliance", 2004, Committee on Fiscal Affairs

Diagram 14.6: Detailed risk identification model



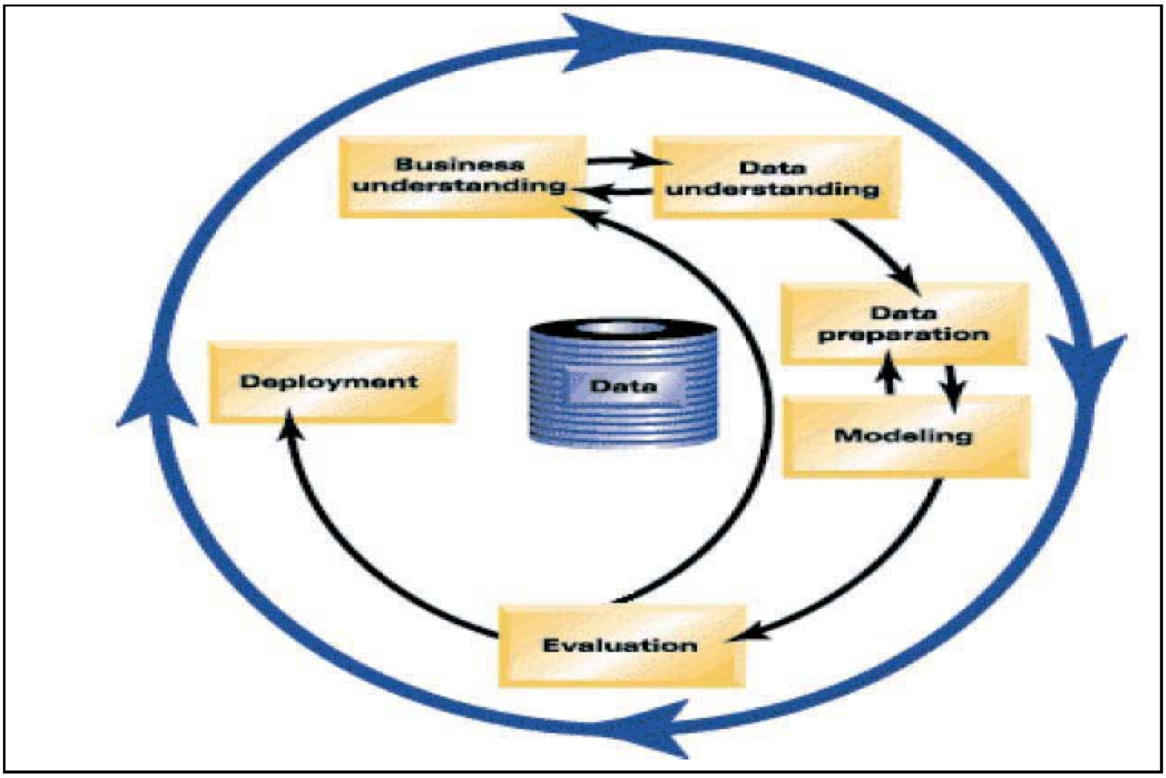
Data mining process and techniques

Data mining includes use of query languages such as SQL for data preparation and the use of procedural and statistical computer languages to apply mathematical and statistical functions to data. Data mining results include a model, such as a predictive model for assessing a risk or taxpayer segmentation. Another type of data mining result is a programme that automates a business process by processing data and making high-volume decisions.

Predictive analytics leverages specialised data warehousing systems that integrate internal and external data sources to enable a variety of applications, from trend analysis to non-compliance detection and revenue forecasting that can help tax departments answering questions such as how resources are to be allocated among tax types, which taxpayers are higher audit priorities, which codes are associated with higher rates of non-compliance, which taxpayers are likely to evade tax or claim fraudulent refunds etc.

Data mining is understood as a process, ideally integrated in a tax administration's day-to-day business. The challenge is to find insightful knowledge from large data sets and to make this knowledge actionable. A well-established process model is the Cross-Industry Standard Process for Data Mining (CRISP-DM) developed by a consortium of data mining consultants and technology specialists,[422] illustrated in Diagram 14.7, which provides an overview of the data mining life cycle.

Diagram 14.7: overview of the data mining life cycle



As shown this is a six phase model. The sequence of the phases is not strict; the projects can move back and forth between phases. Arrows indicate the most important and frequent dependencies between phases.

The Business understanding phase seeks to capture the business requirements and set the business context for the project. It leads to a business plan that documents in concrete terms the expected gains with data mining in the form of business success criteria. It ensures that all

[422] Improving tax administration with data mining, Daniele Micci-Barreca and Satheesh Ramachandran, SPSS Executive Report

participants understand the project goals from a business or organisational perspective. These business goals are then incorporated in a data mining problem definition and detailed project plan.

The data understanding phase involves a close look at the data available for mining, its quality and adequacy for the purpose, its attributes and identification of additional data that might be required. In other words, this phase is designed to assess the sources, quality, and characteristics of the data and the result is a detailed understanding of the key data elements that will be used to build models.

Data preparation is the phase where the data sets are placed in a format suitable for building models. The analyst uses the business objectives determined in the business understanding step to determine which data types and data mining algorithms to use. This phase also resolves data issues uncovered in the data understanding phase, such as missing data.

The Modelling phase involves building the data mining algorithms that extract knowledge from the data. There are a variety of data mining techniques; each is suitable for discovering a specific type of knowledge. A tax agency could use classification or regression models, for example, to discover the characteristics of more productive tax audits. Each technique requires specific types of data, which may require a return to the data preparation phase. The modelling phase produces a model or a set of models containing the discovered knowledge in an appropriate format.

Evaluation is about reflecting on the results obtained in the modelling phase against the business success criteria established at the beginning of the project. Data mining algorithms can uncover an unlimited number of patterns; many of these, however, may be meaningless. The idea is to ensure that the organisation can make use of the results. A key consideration here is the applicability of the conclusions or inferences drawn from the models and the data mining process to business goals. This phase helps determine which models are useful in terms of achieving the project's business objectives. In the context of audit selection, a predictive model for audit outcome could be assessed against a benchmark set of historical audits for which the outcome is known.

Deployment means the incorporation of data mining results into the organisation's decision making and business processes. It could mean, for example, the implementation of a data mining model producing scores that are used in its transaction processing systems. An important further consideration here is how to measure and monitor the validity and accuracy of each model. Alternatively, deployment can mean using the insights gained from data mining for business planning and decision-making. Depending on the significance of the results, this may require only minor modifications, or it may necessitate a major re-engineering of processes and decision-support systems. The deployment phase also involves creating a repeatable process for model enhancements or recalibrations in the light of experience and changing requirements. Analysts need a standard process for updating the models accordingly and deploying new results.

The appropriate presentation of results ensures that decision makers actually use the information. This can be as simple as creating a report or as complex as implementing a repeatable data mining process across the enterprise. It is important that project managers understand from the beginning what actions they will need to take in order to make use of the final models.

The six phases described above are integral to every data mining project. Although each phase is important, the sequence is not rigid; certain projects may require movement back and forth between phases.

The next phase or the next task in a phase depends on the outcome of each of the previous phases. The inner arrows in Diagram 14.7 above indicate the most important and frequent dependencies between phases. The outer circle symbolises the cyclical nature of data mining projects, namely that lessons learned during a data mining project and after deployment can trigger new, more focused business questions. Subsequent data mining projects, therefore, benefit from experience gained in previous ones.

To create the models, the analyst typically uses a collection of techniques and tools. Data mining techniques come from a variety of disciplines, including machine learning, statistical analysis, pattern recognition, signal processing, evolutionary computation, and pattern visualisation.

Supervised and Unsupervised learning

The whole object of data mining is to reveal new insights and discover unknown risks. These broadly fall into three categories.

i. Known Knowns – These are hypotheses that tax departments know to constitute suspicious behaviour, and therefore can be encapsulated as rules.

ii. Known Unknowns – These refer to activities which indicate their existence in the form of the adverse impact (loss in tax revenue) they present, but further investigation is needed in order to identify and examine the elements (process, factors, root causes) that lead to these losses. These generally are identified through anomaly detection and association of events.

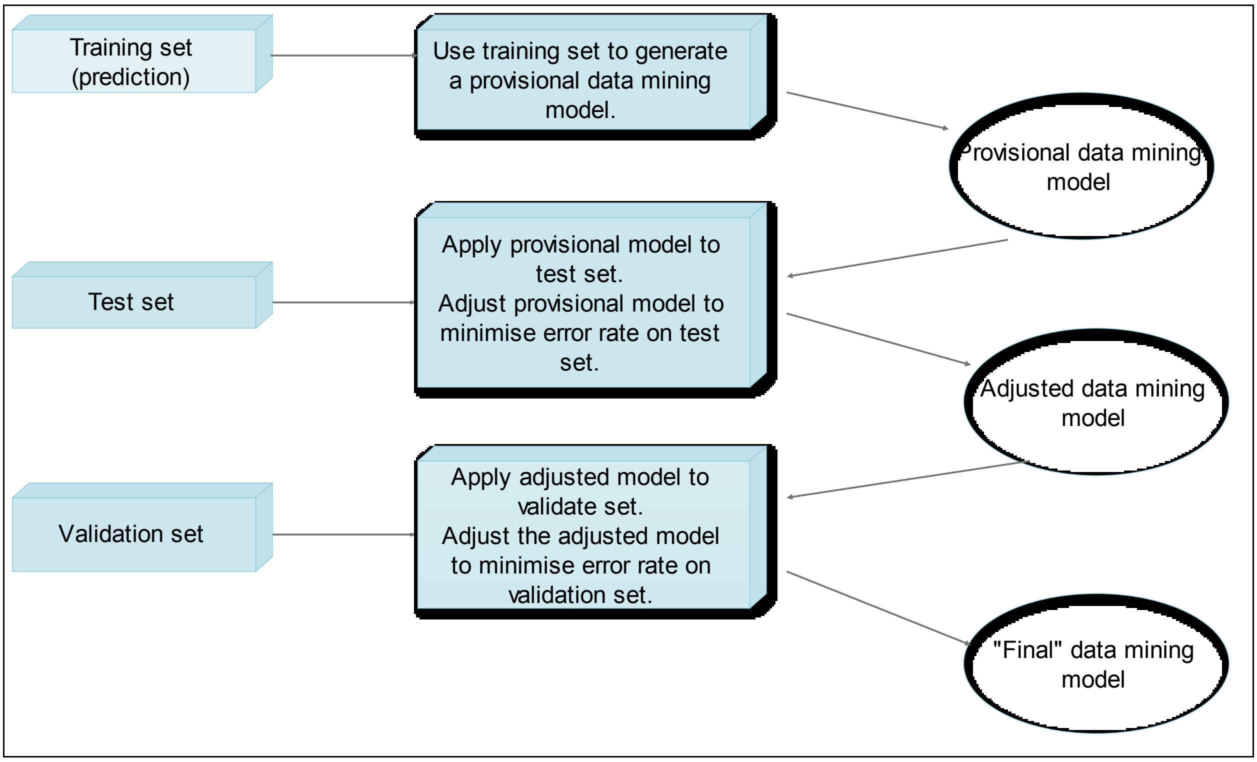
iii. Unknown Unknowns – These refer to any new events or actions that lead to possibly adverse results. The hypotheses formed on unknown unknowns are generally flexible and identified on the basis of multiple signals that emanate from a varied information pool (for example, information on financial transactions between two suspicious entities clubbed with knowledge on the demographic links between these entities identified from other information). It, therefore, is important to ensure that these linkages are drawn using more sophisticated analysis than known knowns or known unknowns.

Data mining methods can be broadly classified as supervised and unsupervised methods. The former is generally based on a prior hypothesis (the known knowns), while the latter require a problem statement rather than a hypothesis (known unknowns) or highlighting a pattern not considered before (unknown unknowns). In the supervised methods, the data mining algorithm first examines a sample data set where the values of the outcome variable of interest are known.

The algorithm learns from this training data the relationship between predictor variables and the outcome variable. Once the learning has taken place, the algorithm is applied to another sample data set where too the outcome is known, but not revealed to the algorithm (validation data), to see its performance in predicting the outcome. Having validated the performance, the model can then be used to classify or predict the outcome of interest in new cases where the outcome is unknown. In case of unsupervised methods, there is no outcome variable to predict or classify. Supervised learning is used in classification and prediction whereas association rules and clustering techniques are unsupervised methods.[423]

Most data mining methods are supervised methods, that is, (1) there is a particular pre-specified target variable, and (2) the algorithm is given many examples where the value of the target variable is provided, so that the algorithm may learn which values of the target variable are associated with which values of the predictor variables. Most supervised data mining methods apply the methodology given below for building and evaluating a model.

Diagram 14.8: Methodology for supervised modelling



Source: http://dm-dingwang.blogspot.in/2007/05/supervised-versus-unsupervised-methods.html

First, the algorithm is provided with a training set of data, which includes the pre-classified values of the target variable in addition to the predictor variables. For example, if we are interested in classifying income bracket, based on age, gender, and occupation, our classification algorithm

[423] Data Mining in Tax Administration - Using Analytics to Enhance Tax Compliance by Jani Martikainen , 2012 , available at http://epub.lib.aalto.fi/fi/ethesis/pdf/13054/hse_ethesis_13054.pdf , accessed in January ,2015.

would need a large pool of records, containing complete information about every field, including the target field, that is, income bracket. In other words, the records in the training set need to be pre-classified. A provisional data mining model is then constructed using the training samples provided in the training data set. However, the training set is necessarily incomplete; that is, it does not include the "new" or future data that the data modellers are really interested in classifying. Therefore, the algorithm needs to guard against "memorising" the training set and blindly applying all patterns found in the training set to future data.

The next step in supervised data mining methodology is to examine how the provisional data mining model performs on a test set of data. In the test set, the values of the target variable are hidden temporarily from the provisional model, which then performs classification according to the patterns and structure it learned from the training set. The efficacy of the classifications is then evaluated by comparing them against the true values of the target variable.

The provisional data mining model is then adjusted to minimise the error rate on the test set. The adjusted data mining model is then applied to a validation data set, where the values of the target variable are again hidden temporarily from the model. The adjusted model again adjusts itself to minimise the error rate on the validation set. Estimates of model performance for future, unseen data can then be computed by observing various evaluative measures applied to the validation set.

Unsupervised learning or self-organisation is one in which an (output) unit is trained to respond to clusters of pattern within the input. In this paradigm, the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli.

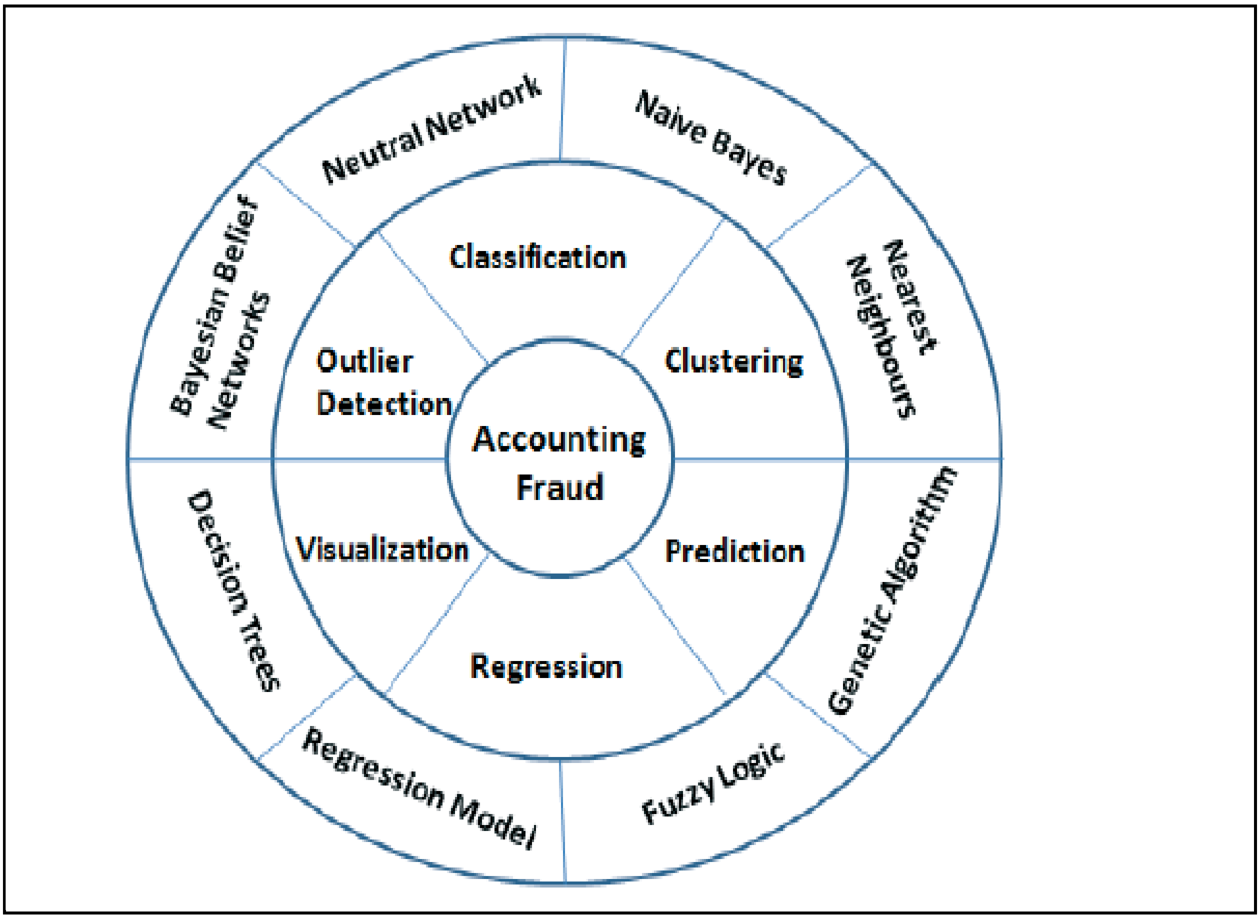
Unsupervised techniques can be a powerful means of understanding when it comes to risks that fall in the category of "unknown unknowns" and there is so much data available that it is difficult to understand the underlying structure of the population without using such methods as cluster analysis and segmentation. Especially if a target is not available, cluster analysis can help identify groups in the population that are alike within the group but different from members of other groups. Once segmented, cases can be assigned a group membership. This label can be used to determine treatment strategies and monitor the effectiveness of the efforts to change taxpayer behaviour over time.

Additional insight can be gained from overlaying outputs from unsupervised techniques with supervised techniques and vice versa. Some segments might emerge as inherently more risky, thus semi-supervised techniques are often useful where some business knowledge can be used, even where only minimal training data is available.

Classification of data mining applications

Generally, the data mining applications used for fraud detection as shown in Diagram 14.9 consist of two layers. The first layer comprises techniques of classification, clustering, prediction, outlier detection, regression, and visualisation. Each of the data mining applications is supported by a set of algorithmic approaches to extract the relevant relationships in the data. The second layer consists of data mining techniques used for determining the main algorithms used for fraud detection.

Diagram 14.9: Conceptual framework for application of data mining to fraud detection



Source: "Application of data mining techniques for financial accounting fraud detection scheme" by S.S. Chintalapati and G. Jyotsna, 2013.

Classification is the process of identifying a set of common features or patterns, and proposing models that describe and distinguish data classes or concepts. Common classification techniques include neural networks, the Naïve Bayes techniques, decision trees and support vector machines. Such classification tasks are used in the detection of credit card, healthcare and automobile insurance, and corporate fraud, among other types of fraud, and classification is one of the most common learning models in the application of data mining in fraud detection.

Clustering is used to partition objects into previously unknown conceptually meaningful groups (i.e. clusters), with the objects in a cluster being similar to one another but very dissimilar to the objects in other clusters. Clustering is also known as data segmentation or partitioning and is regarded as a variant of unsupervised classification. Cluster analysis decomposes or partitions a data set (single or multivariate) into dissimilar groups so that the data points in one group are similar to each other and are as different as possible from the data points in other groups. Generally, data objects in each cluster should have high intra-cluster similarity within the same cluster but should have low inter-cluster similarity to those in other clusters. The most common clustering techniques are the K-nearest neighbour, the Naïve Bayes technique and self-organising maps.

Prediction estimates numeric and ordered future values based on the patterns of a data set. The attribute for prediction which value being predicted is continuous-valued (ordered) rather than categorical (discrete-valued and unordered). This attribute is referred as the predicted attribute. Neural networks and logistic model prediction are the most commonly used prediction techniques.

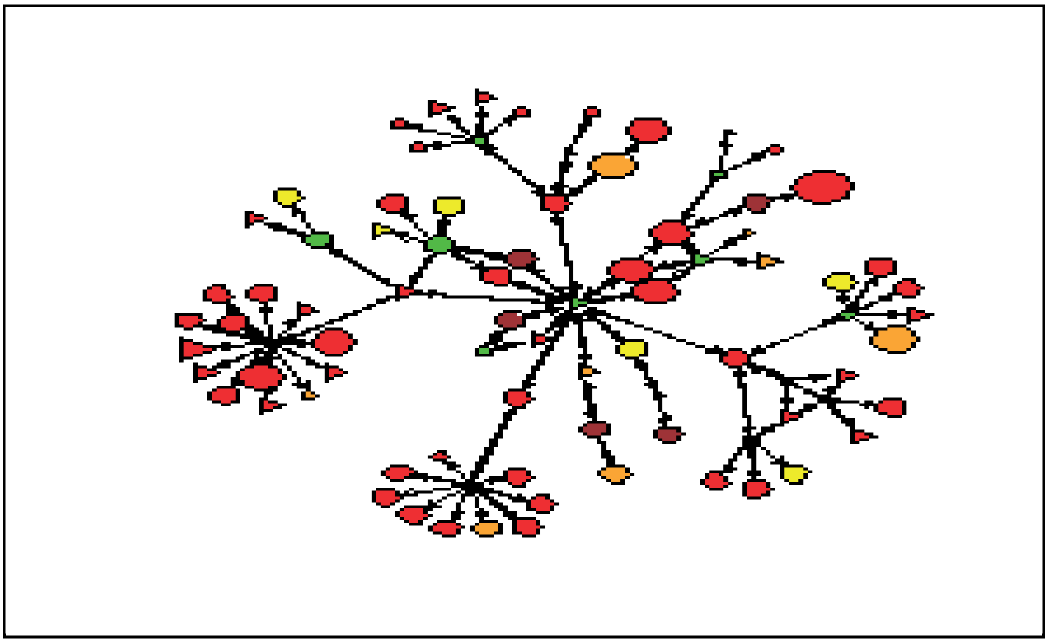
Outlier Detection is employed to measure the distance between data objects to detect those objects that are grossly different from or inconsistent with the remaining data set. Data that appear to have different characteristics from the rest of the population are called outliers. The problem of outlier/anomaly detection is one of the most fundamental issues in data mining. A commonly used technique in outlier detection is the discounting learning algorithm.

Regression is a statistical methodology used to reveal the relationship between one or more independent variables and a dependent variable (that is continuous-valued). The regression technique is typically undertaken using such mathematical methods as logistic regression and linear regression.

Visualisation refers to the easily understandable presentation of data and to methodology that converts complicated data characteristics into clear patterns to allow users to view the complex patterns or relationships uncovered in the data mining process. A suite of tools and applications that flexibly encode data using colour, position, size and other visual characteristics, present data in a manner that is suited to the pattern detection capabilities of the human visual system. Visualisation is best used to deliver complex patterns through the clear presentation of data or functions. Visualisation tools provide intuitive interfaces that allow users to query complex data sets and use their natural abilities to detect patterns and trends in visual information to increase the speed of interpretation of information and analysis. They bring the power of analytics to a wider group of users without investment in learning computer languages and mathematical algorithms.

The Diagram 14.10 below is an illustration of network visualisation.

Diagram 14.10: Visualisation of a network



Source - ATO

The above diagram shows the relationships between several companies, partnerships and trusts that are all inter-connected. The diagram also shows the structure, its relative size, the type of entity and lodgement (return filing) status of each of the entities in the group. Companies are in red, individuals in green, super funds (superannuation funds) in yellow, partnerships in brown and trusts in orange. Triangle means returns not filed, circle means filed. Size reflects market segments based on total business income. Such network visualization is a powerful way to track the activities of groups of networks whose connections with each other would not be easily visualized without the collation of information from diverse sources.

Data Mining Techniques

The most frequently used techniques are logistic regression, neural network, the Bayesian belief network, and decision trees. Some of the techniques are discussed in more detail in the following paragraphs.

Regression Models

Regression-based models are mostly used in fraud detection. Most are based on logistic regression, stepwise-logistic regression, multi-criteria decision making method and exponential generalised beta two. The logistic model is a generalised linear model that is used for binomial regression in which the predictor variables are either numerical or categorical. Logistic analysis and clustering analysis are jointly used to establish a detecting model of fraud based on four aspects – financial indices, company governance, financial risk and pressure and related trading. After cluster filtering significant variables, prediction models are established using standardisation, non-standardization Bayes and Logistic methods.

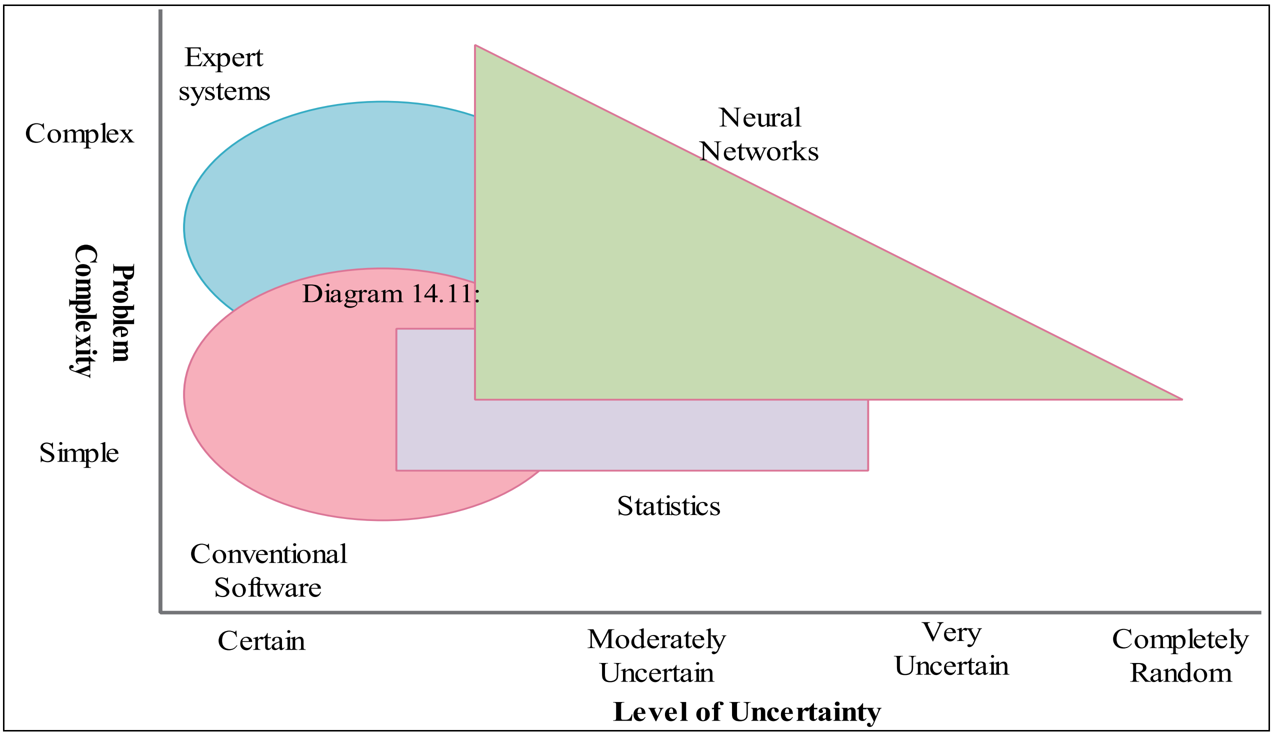
Neural Networks (NN)

The neural networks (NN) are non-linear statistical data modelling tools that are inspired by the functionality of the human brain using a set of interconnected nodes. They are used to model complex relationships between inputs and outputs or to find patterns in data. Using neural networks as a tool, data warehousing firms are harvesting information from datasets from data mining. The difference between these data warehouses and ordinary databases is that there is actual manipulation and cross-fertilisation of the data helping users makes more informed decisions.

Neural networks essentially comprise three pieces: the architecture or model, the learning algorithm, and the activation functions. An NN is programmed or trained to store, recognise and associatively retrieve patterns or data entries to solve combinatorial optimisation problems, to filter noise from measurement data, and to control ill-defined problems. In other words, it is used to estimate sample data when the form of the function is unknown. It is precisely the ability to recognise patterns and estimate functions that make artificial neural networks so useful in data mining. Neural networks, depending on the architecture, provide associations, classifications, clusters, prediction and forecasting to the data mining industry.

An artificial NN processes data just as a human brain does. NN is a type of nonlinear statistical analysis program that utilises a learning process to repeatedly analyse a collection of historical data sets to recognise patterns in that data and to automatically produce a model for that data. It processes data the way the brain processes data – in a multiple parallel processing mode. This multiple processing capability enables NNs to execute operations much faster than serial methods, which process one operation at a time. NNs solve problems by recognising patterns in data that may be too complex for humans or other types of automated methods to discern. As indicated in the diagram 14.10 below, an NN is used to solve a very broad range of problems that are almost completely random in nature and it is suitable for simple to complex, structured to unstructured problems.

Diagram 14.11: Neural network



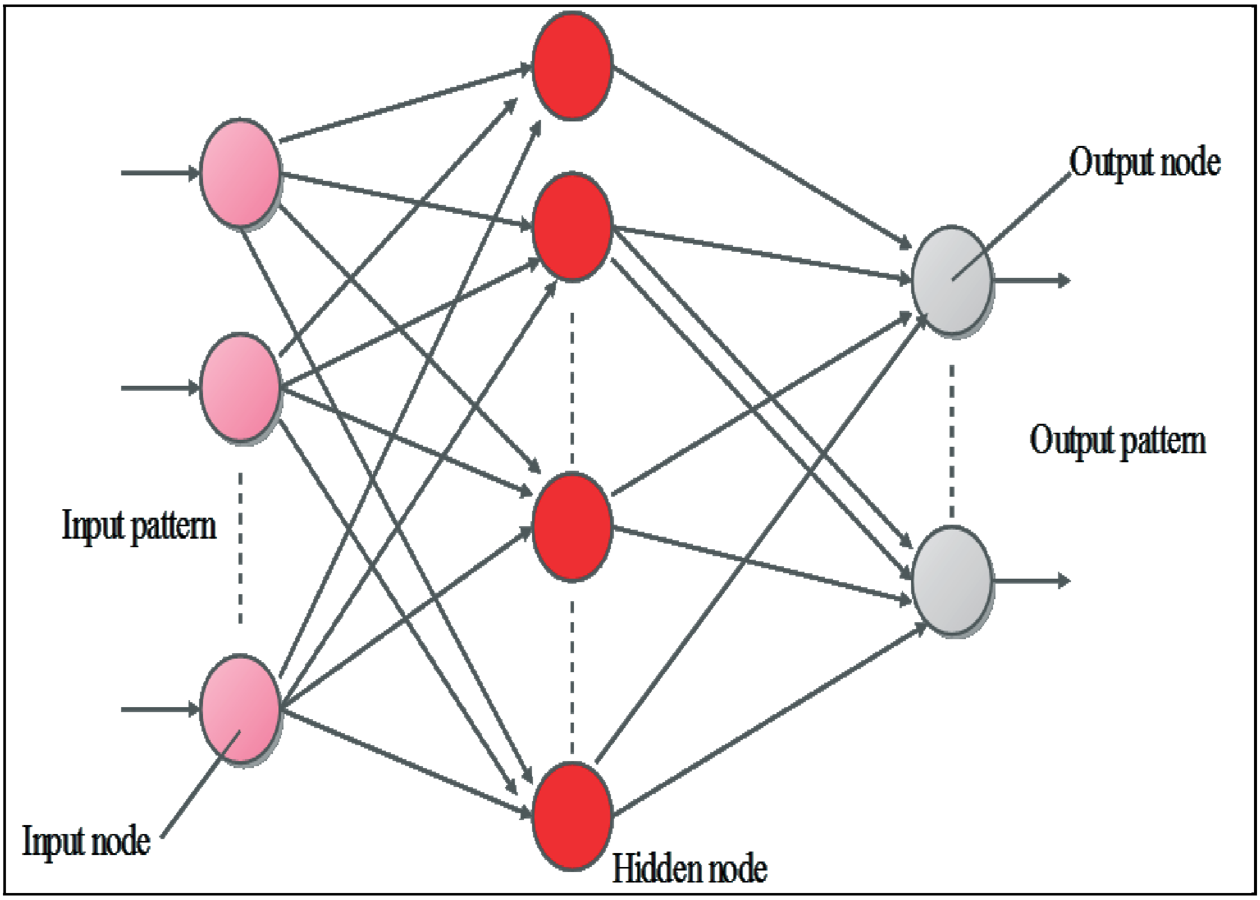
Source: "Using Neural Network Software as a Forensic Accounting Tool" by M. C. Cerullo and M.V Cerullo.

Generally, an NN creates a mathematical model from a historical database of examples of input and output values. After learning the relationship between the variables, the network has been trained, and a mathematical model is constructed that recognises patterns in the sample data, such as correlations between seemingly unrelated data. The resulting model, when used with new input data, provides projections of future outputs. For example, by collecting historical data of commercial loans made to organisations, a bank can determine which organisations have defaulted on repaying the loans. A model can be built that contains the relationship, if any, between a firm's selected financial ratios and the outcome of the loan. After the model is synthesised, it can be used to predict if a new commercial loan applicant is likely to default on repayment.

NN technology has been used in a few firms to develop models for various forms of fraud detection. Studies have shown that NN techniques are an effective method for developing a fraud classification model to assess the risk of errors and irregularities in financial statements. The variables used in developing a particular fraud detection model may differ from one company/industry to another or according to the targeted type of fraud. NNs, with their remarkable ability to derive meaning from imprecise data, are used to detect trends that are too complex to be noticed by humans or other computer programs. A trained NN can be thought of as an expert in the category of information it has been given to analyse.

The Diagram 14.12 below presents a simple NN structure, which comprises multiple inputs and a single output. The basic building block in an NN is the neuron, depicted by the nodes. Every NN has an input layer, a hidden layer and an output layer. Each artificial neuron in a layer receives its input from the output of the previous layer nodes or from network inputs. The connections between nodes are associated to adjustable weights that are adjusted as the network is trained. Each neuron in the network processes the input data, with resulting values steadily seeping through the network layer by layer, until a result is generated in the output layer.

Diagram 14.12: Structure of a simple neural network



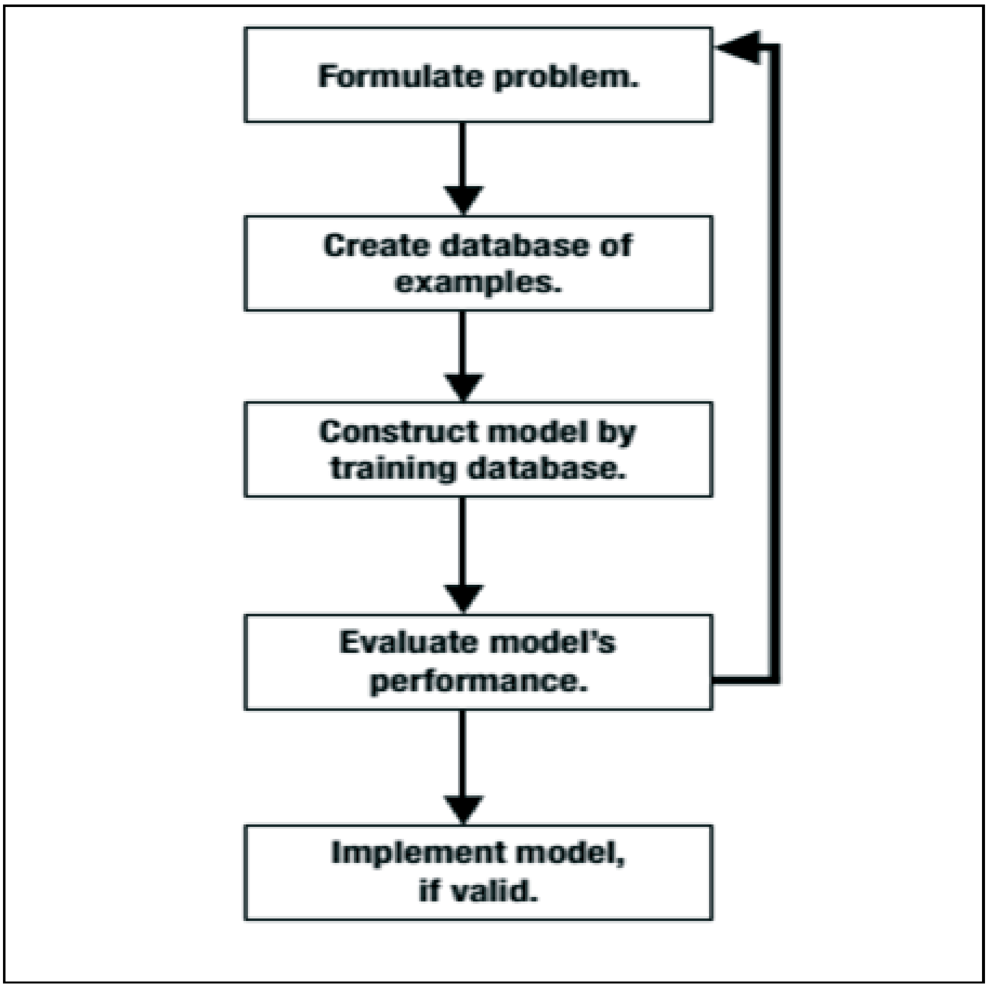
Source: "Neural network in data mining" by Yashpal Singh and Alok Singh Chauhan

Input data is presented to the network and propagated through the network until it reaches the output layer. This forward process produces a predicted output. The difference between the predicted value and actual value gives the error value of the network. Then, the NN uses supervised learning, which is back-propagation[424] or propagation of error, to train the network. This means that the artificial neurons are organised in layers, and send their signals "forward", and then the

[424]Back-propagation, or propagation of error, is a common method of teaching artificial neural networks how to perform a given task. It is a learning algorithm for adjusting weights.

errors are propagated backwards. After the end of this back propagation, the forward process starts again, and this cycle is continued until the error between predicted and actual outputs is minimised.

Diagram 14.13: Modelling fraud detection using neural network



Source: "Using Neural Network Software as a Forensic Accounting Tool" by M. C. Cerullo and M.V Cerullo.

Diagram 14.13 depicts a typical flow chart of NN model construction. The first step to develop a NN model to predict fraud is to formulate problems and then select a problem that is suitable for an artificial NN. There are categories of business problems which can be classified as follows.

- Classification problem – fraud detection, loan default and credit card applications
- Time series application – financial forecasting, stock market prediction and bankruptcy prediction
- Data mining application – looking for patterns in customer databases, targeting customers and analysing demographics

Typically, NN models are useful for detecting incidence of frauds from analysis of very large amounts of data. In tax context, some areas where neural networks would be relevant are returns processing, processing of import/export documents, tax arrears collections etc.

The second step is to create a database of historical numerical inputs into the NN software to generate the fraud prediction model. The database must contain independent input variables and corresponding dependent output variables. Creating the database is the most important step in the process of developing an NN.

The third step is to construct a model. A table of input and output values is provided to an NN package to construct a network depicting the relationship of the inputs to the output. The software package automatically splits the table into a training subset of observations and a test subset of observations. The NN software uses the training file to train the network into a mathematical model that recognises patterns between the independent variables and the dependent variable (i.e., fraud or no fraud). The test data determines how well the synthesised network model is able to generalise on a new, unseen data training file to train the network into a mathematical model.

The fourth step is to evaluate the performance of the network. The model is evaluated by comparing the output summary statistics for the test data subset with the training data subset. The average absolute error and maximum absolute error are two good indicators to determine if the model constructed is acceptable. If the evaluation of the output summary statistics indicates that an unacceptable model was developed, further iteration should be done, providing new variables, increased sample size etc. and its output again evaluated. If, despite iterations in this fashion, the results still appear to be unacceptable, an NN probably cannot be constructed for the given problem. If the model gives acceptable results and all input values are significant, the model can be used for its intended purpose of predicting fraudulent cases and conducting further investigations.

Generally, an NN computer software model is constructed and "trained" from a database of historical examples of input and output variables. During model training, an NN learns the patterns and correlations from a sample of input and output data representing actual fraud occurrences and non-fraud occurrences, respectively. The creation of the database is the most important step in the process of developing an NN. The input and output values are keyed into the NN package to construct a network that can then be used by a user to make a decision about the possible occurrence of fraud.

Bayesian belief network

The Bayesian belief network represents a set of random variables and their conditional independencies using a directed acyclic graph, in which nodes represent random variables, each arrow represents a probabilistic dependence and missing edges encode conditional independencies between the variables. If an arrow is drawn from node A to node B, then A is parent of B and B is a descendant of A. In a belief network, each variable is conditional independent of its non-descendant, given its parents. For each node X, there exists a conditional probability, which specifies the conditional probability of each value of X for each possible combination of the values of its parents. The network structure is defined in advance or inferred from the data. For classification purposes, one of the nodes is defined as the class node. The network calculates the probability of each alternative class.

Bayesian classification is based on the statistical theorem of Bayes. The Bayes theorem provides a calculation for posterior probability. According to the Bayes theorem, if H is a hypothesis – such as the object X belongs to the class C – then the probability that the hypothesis holds is $P(H_jX) = (P(X_jH) * P(H))/P(X)$. If an object X belongs to one of i alternative classes, in order to classify the object, a Bayesian classifier calculates the probabilities $P(C_{ij}X)$ for all possible classes Ci and assigns the object to the class with the maximum probability $P(C_{ij}X)$.

Naive Bayesian classifiers make the class condition independence assumption, which states that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption simplifies the calculation of $P(C_{ij}X)$. If this assumption holds true, the Naive Bayesian classifiers have the best accuracy rates compared to all other classifiers. However, in many cases, this assumption is not valid, since dependencies can exist between attributes.

The Bayesian belief network is used to develop models for credit card, automobile insurance, and corporate fraud detection. Research indicates that the Bayesian belief network model correctly classified 90.3 per cent of the validation sample for fraud detection. The Bayesian belief network outperformed the neural network and decision tree methods and achieved outstanding classification accuracy.

Decision Trees

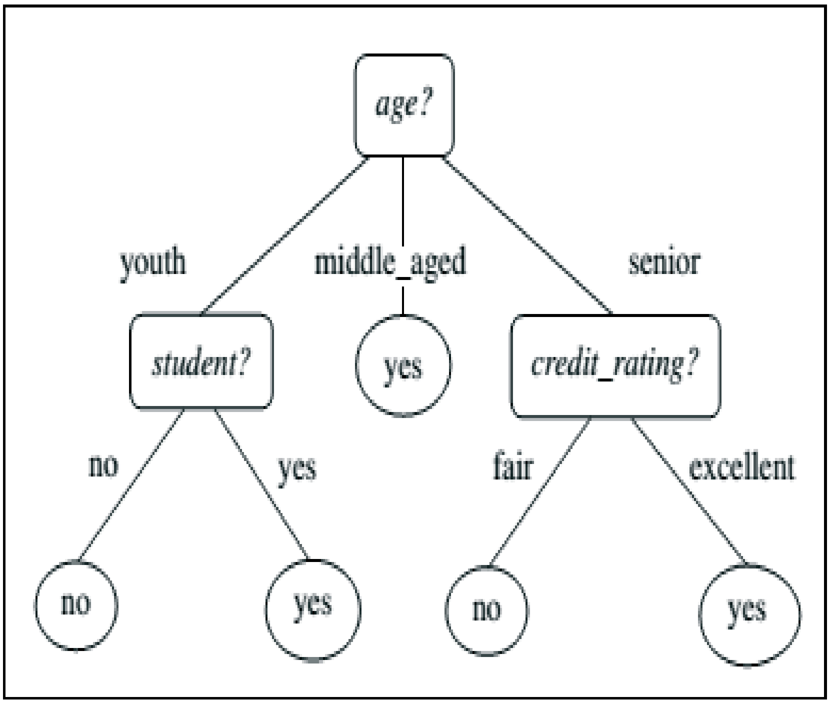
A decision tree is a tree structure, where each node represents a test on an attribute, and each branch represents an outcome of the test. In this way, the tree attempts to divide observations into mutually exclusive subgroups. The goodness of a split is based on the selection of the attribute that best separates the sample. The sample is successively divided into subsets, until either no further splitting can produce statistically significant differences or the subgroups are too small to undergo meaningful division. There are several proposed splitting algorithms such as automatic interaction detection (AID) and the classification and regression trees (CART).

Decision tree learning, used in data mining and machine learning, uses a decision tree as a predictive model, which maps observations about an item to conclusions about the item's target value. More descriptive names for such tree models are classification trees or regression trees. In these tree structures, leaves represent classifications and branches represent conjunctions of features that lead to those classifications.

A decision tree is constructed from a training set, which consists of objects. Each object is completely described by a set of attributes and a class label. Attributes can have real or Boolean values. The concept underlying a data set is the true mapping between the attributes and the class.

A noise-free training set is one in which all the objects are "generated" using the underlying concept. Construction of a tree from the training set is called tree induction, tree building and tree growing, which is a top down process.

Diagram 14.14: Decision Tree



Source: Data Mining: Concepts and Techniques by Jiawei Han and Micheline Kamber

In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data but not decisions; the resulting classification tree can be an input for decision making. Decision tree learning is a common method used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables. Each interior node corresponds to one of the input variables; there are edges to children nodes for each of the possible values of that input variable. Each leaf represents a value of the target variable given the values of the input variables, represented by the path from the root to the leaf.

A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has the same attribute value as the target variable, or when splitting no longer adds value to the predictions. In data mining, trees can be described also as the combination of mathematical and computational techniques to aid the description, categorisation and generalisation of a given set of data. Data comes in form of records described below.

$$(X, Y) = (x_1, x_2, x_3, \ldots\ldots, x_k, Y)$$

The vector x is composed of input variables x_1, x_2, x_3, etc., that are used to understand, classify or generalise the dependent variable Y.

The algorithms that are used for constructing decision trees usually work top-down by choosing a variable at each step that is the next best variable to use in splitting the set of items. "Best" is defined by how well the variable splits the set into subsets that have the same value as the target variable. Different algorithms use different formulae for measuring "best". These formulae are applied to each candidate subset, and the resulting values are combined (e.g., averaged) to provide a measure of the quality of the split.

When a decision tree is built, many of the branches will reflect anomalies in the training data due to noise or outliers. Tree pruning methods address this problem of over-fitting the data. Such methods typically use statistical measures to remove the least reliable branches. Pruned trees are likely to be smaller and less complex and, thus, easier to comprehend. They are usually faster and better at correctly classifying independent test data than un-pruned trees. There are two common approaches to tree pruning: pre-pruning and post-pruning. In the pre-pruning approach, a tree is "pruned" by halting its construction at an early stage. In post-pruning, sub-trees are removed from a "fully grown" tree.

Nearest Neighbour Method

The nearest neighbour method is a similarity-based classification approach. Based on a combination of the classes of the most similar k record(s), every record is classified. Sometimes this method is also known as the k-nearest neighbour technique. K-nearest neighbour method is used to detect fraudulent automobile insurance claims and to identify defaults by credit card clients.

The nearest-neighbour classifiers are based on learning by analogy, that is, by comparing given test observations with training observations that are similar to it. The training set data are described by 'n' attributes. Each observation represents a point in an n-dimensional space. In this way, all of the training observations are stored in an n-dimensional pattern space. When given an unknown observation, a k-nearest-neighbour classifier searches the pattern space for the 'k' training observations that are closest to the unknown observation. These k training observations are the k "nearest neighbours" of the unknown observation. Closeness is defined in terms of a distance metric, such as Euclidean distance.[425]

For k-nearest-neighbour classification, the unknown observation is assigned the most common class among its k nearest neighbours. When k = 1, the unknown observation is assigned the class of the training observations that is closest to it in pattern space. Nearest neighbour classifiers can also be used for prediction, that is, to return a real-valued prediction for a given unknown

[425] The Euclidean distance between two points or observations such as, X1 = (x11, x12, x1n) and X2 = (x21, x22, ..., x2n), is defined as: $dist\ (X_1, X_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})2}$.

observation. In this case, the classifier returns the average value of the real-valued labels associated with the k nearest neighbours of the unknown observation.

Nearest-neighbour classifiers use distance-based comparisons that intrinsically assign equal weight to each attribute. They, therefore, can suffer from poor accuracy when given noisy or irrelevant attributes. The method, however, has been modified to incorporate attribute weighting and the pruning of noisy data observations. The choice of a distance metric can be critical.

Fuzzy logic and Genetic algorithms

Fuzzy logic is a mathematical technique that classifies subjective reasoning and assigns data to a particular group, or cluster, based on the degree of possibility the data has of being in that group. The expert fuzzy classification techniques enable one to perform approximate reasoning that can improve performance in three ways. First, performance is improved through efficient numerical representation of vague terms, because the fuzzy technology can numerically show representation of a data item in a particular category. The second way performance is enhanced is through an increased range of operations in ill-defined environments, which is the way that fuzzy methodology shows partial membership of data elements in one or more categories that may not be clearly defined in traditional analysis. Finally, performance is increased because fuzzy technology has decreased sensitivity to "noisy" data, or outliers.

Machine Learning

Machine learning is the ability of a program to learn from experience – that is, to modify its execution on the basis of newly acquired information. There are two types of machine learning, supervised learning and unsupervised learning, which have already been discussed earlier.

Text Mining

Data mining mostly focuses on structured data, such as relational, transactional, and data warehouse data. However, in reality, a substantial portion of the available information is stored in text databases or document databases, which consist of large collections of documents from various sources, such as news articles, research papers, books, digital libraries, e-mail messages, and web pages. Text databases are rapidly growing due to the increasing amount of information available in electronic form, such as electronic publications, various kinds of electronic documents, e-mail, and the World Wide Web. Nowadays, most of the information in government, industry, business, and other institutions are stored electronically in the form of text databases.

Data stored in most text databases are semi structured data in that they are neither completely unstructured nor completely structured. Traditional information retrieval techniques are not adequate to make use of vast amounts of text data. Typically, only a small fraction of the many available documents will be relevant to a given individual user. It is difficult to formulate effective queries for analysing and extracting useful information from the data without knowing the content of the documents. Users need tools to compare different documents, rank the importance and relevance of the documents, or find patterns and trends across multiple documents. Thus, text mining has become an increasingly popular and essential theme in data mining.

Text mining specialises in identifying, locating and extracting meaning from unstructured information. This can include documents, computer hard drives, internet websites, phone call transcripts produced by speech recognition software and scanned hard copies of documents. The types of analysis can include the identification of a document type, for example spreadsheets, proprietary software files, e-mails etc. It also includes automated searching and finding of documents with particular content such as balance sheet, a letter, a bank statement, a contract or legal advice. Further, text mining can find concepts within unstructured data such as negative or positive sentiment in sentences, common reasons for action in survey results, and themes from a large body of comments, and categories from free form descriptions. Descriptions of transactions may hold clues to the fact that a business may be involved in fraud, e.g., particular categories of goods that a business is importing or exporting. Text mining is a technique that allows that information to be included in detection models.

There are many approaches to text mining, which can be categorised from different perspectives, based on the inputs taken in the text mining system and the data mining tasks to be performed. In general, the major approaches, based on the kinds of data they take as input, are: (1) the keyword-based approach, where the input is a set of keywords or terms in the documents, (2) the tagging approach, where the input is a set of tags, and (3) the information-extraction approach, which inputs semantic information, such as events, facts, or entities uncovered by information extraction.

A simple keyword-based approach may only discover relationships at a relatively shallow level, such as rediscovery of compound nouns (e.g., “database” and “systems”) or co-occurring patterns with less significance (e.g., “terrorist” and “explosion”). It may not bring much deep understanding to the text. The tagging approach may rely on tags obtained by manual tagging or by some automated categorisation algorithm. The information-extraction approach is more advanced and may lead to the discovery of deep knowledge, but it requires semantic analysis of text by natural language understanding and machine learning methods. This is a challenging knowledge discovery task. Various text mining tasks can be performed on the extracted keywords, tags, or semantic information. These include document clustering, classification, information extraction, association analysis, and trend analysis.

Anomaly Detection

This technique is used to identify businesses that are not behaving in the same way as other businesses within their peer group. Anomaly detection algorithms examine the data to identify unusual behaviour. Anomaly detection seeks to normalise events and set thresholds to identify outlier behaviour. Using distributional analysis, variables are analysed to identify claims and practice patterns that are extreme outliers relative to the rest of their respective distribution. In this manner, statistical outliers are used to identify new patterns of fraud that are otherwise not known.

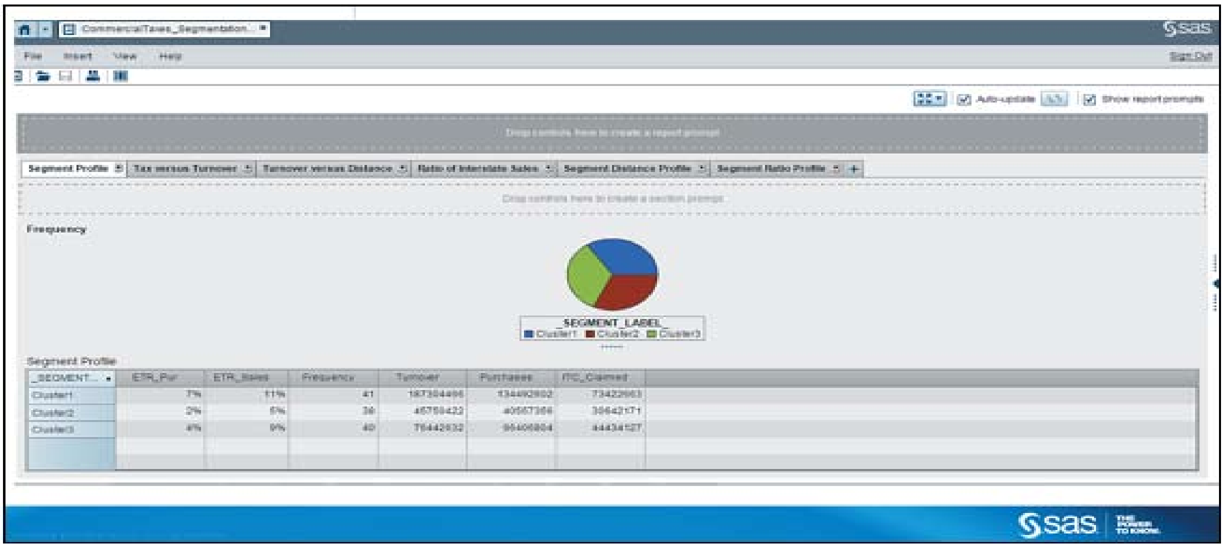
For example, with regard to income tax returns, various ratios can be constructed based on information provided in the returns and other documents, and a range or cut-off identified, beyond which the transaction would be suspicious. An example would be that of an income tax payer who states an income of Rs.3 to Rs.3.5 lakh, but also states that he/she pays Rs. 50,000 per month as rent, which is very high as per his income profile. Analysing this income tax payer, along with other members of the same peer group (salary, job profile etc.) would throw up this case as an outlier. However, anomaly detection also suffers from the following limitations when viewed in isolation.

- Too many false positives make it difficult for an investigator to distinguish between legitimate transactions and fraudulent transactions.

- Insiders, such as procurement and contracting specialists, are well aware of threshold detection. They use this insight to alter their behaviour to elude anomaly detection, for example, by setting up multiple bank accounts and depositing separately.

- The effectiveness of anomaly detection depends on the effectiveness and comparability of the peer group being matched. That is, the peer group for comparison should be as homogenous and similar to the entity as possible.

Diagram 14.15: Commercial tax dealers' segmentation

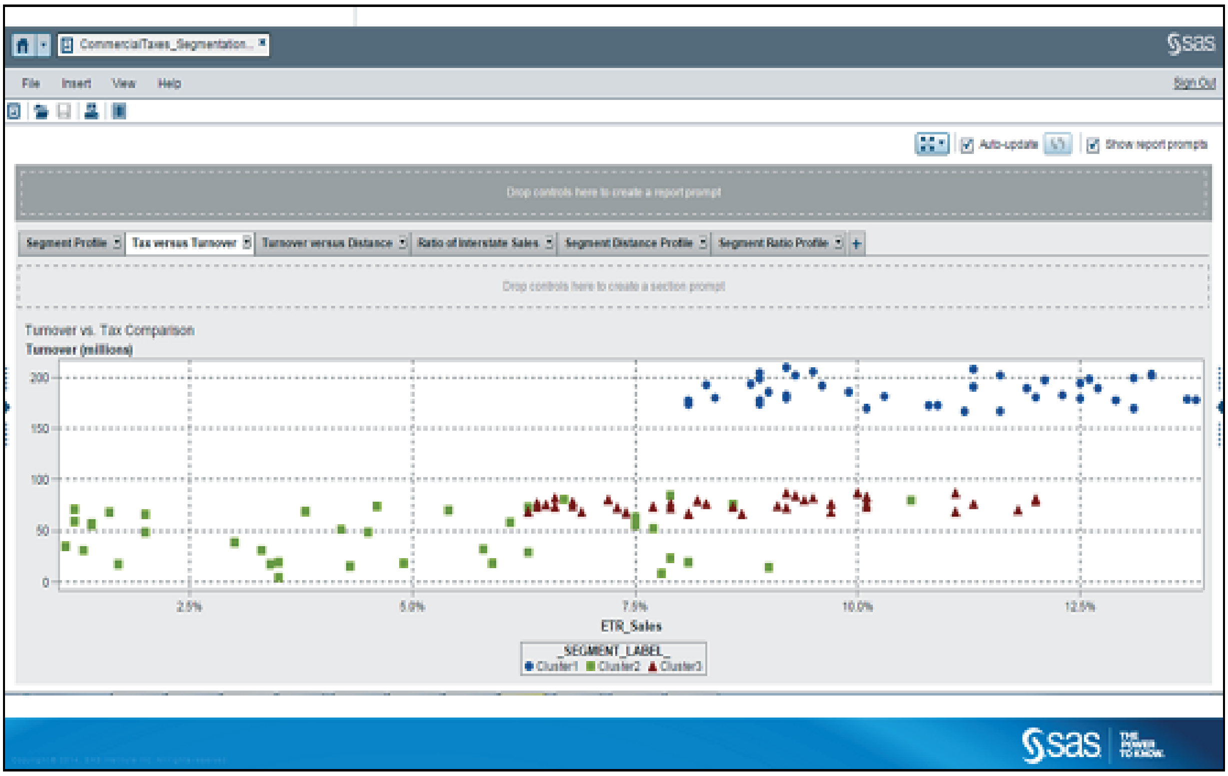


Source: SAS presentation to TARC

Therefore, anomaly detection needs to be combined with cluster analysis, wherein groups of homogeneous (similar) entities are identified out of a universe of heterogeneous (disparate and varied) entities. The purpose of cluster analysis is to place objects into groups, or clusters, suggested by the data, not defined a priori, such that objects in a given cluster tend to be similar to each other in some sense, and objects in different clusters tend to be dissimilar. Cluster analysis can also be used to summarise data rather than to find "natural" or "real" clusters – this, in a sense, will establish normal or the baseline behaviour. For example, group of commercial tax dealers can be grouped into three natural clusters, based on various characteristics, such as stated turnover, stated purchases, effective tax rate (VAT) on purchases, effective tax rate on sales, and input tax credit (benefit) claimed.

Cluster analysis is, thus, a technique where segments are naturally derived, without imposing pre-set rules, limits or constraints. For example, in the case illustrated in Diagram 14.16 below, no cut-off on turnover, purchases, etc. has been applied; these are naturally derived based on a distance measure (taking all variables into account) of each dealer/entity with one another.

Diagram 14.16: Profiling of dealers in different clusters against turnover and ETR-sales



Source: SAS presentation to TARC

Profiling the Clusters

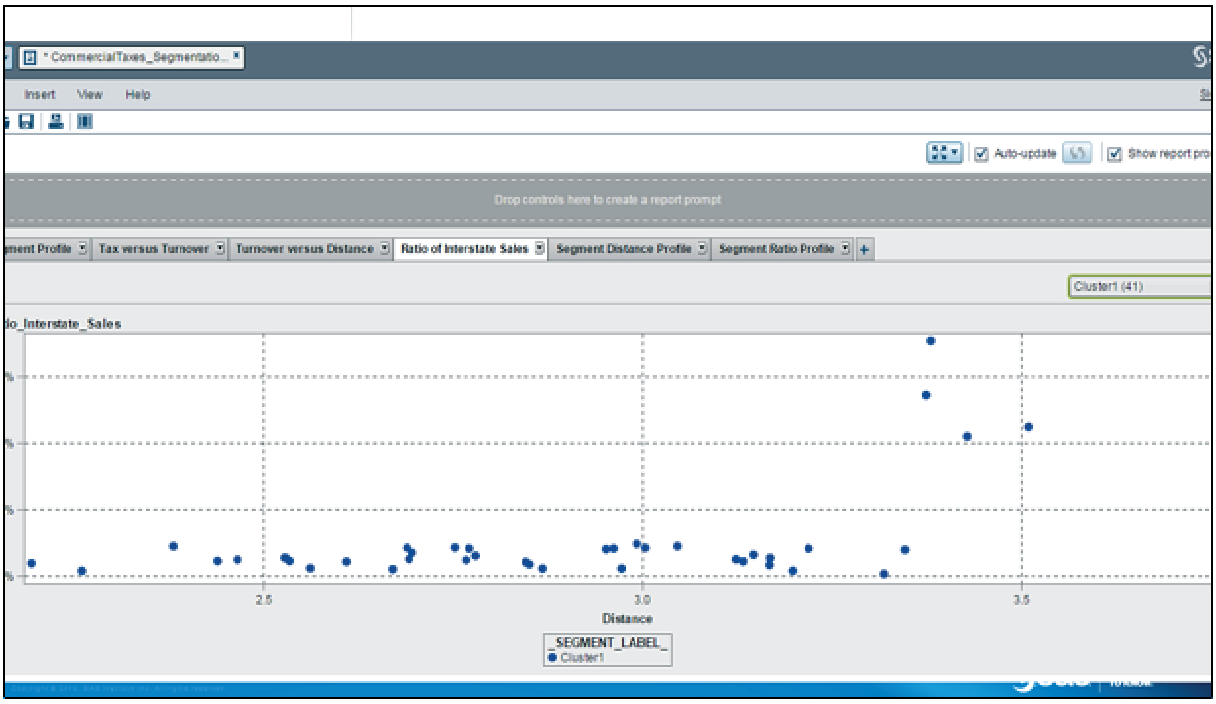
After deriving the clusters, the dealers in various clusters (denoted by a different symbol in the scatter plot), can be profiled across different variables in order to understand their innate characteristics. For example, when profiling dealers in different clusters according to turnover and effective sales tax rate in the diagram above, one can understand:

- Cluster 1 dealers are generally high turnover dealers (large dealers) who primarily sell finished products/high value-added products (high ETR).

- Cluster 2 dealers, also sell products in the 6 to 11 per cent range of tax rate, but are mostly medium scale dealers with lower turnover.

- Cluster 3 dealers are mostly small dealers, dealing mainly with input commodities (since these are in the low tax rate bracket).

Diagram 14.17: Cluster distance measurement from centroid



Source: SAS presentation to TARC

Cluster Distance Measure

Once one has a good understanding of the characteristics of an "average dealer in a cluster", it is possible to find out anomalies (i.e. dealers who do not conform to the normal behaviour expected in the cluster), by making use of the distance from the cluster centroid. As part of the cluster formation, each dealer is assigned a distance or a measure of deviation based on how different his characteristics are from the cluster he is assigned to. This is represented by the cluster centroid.

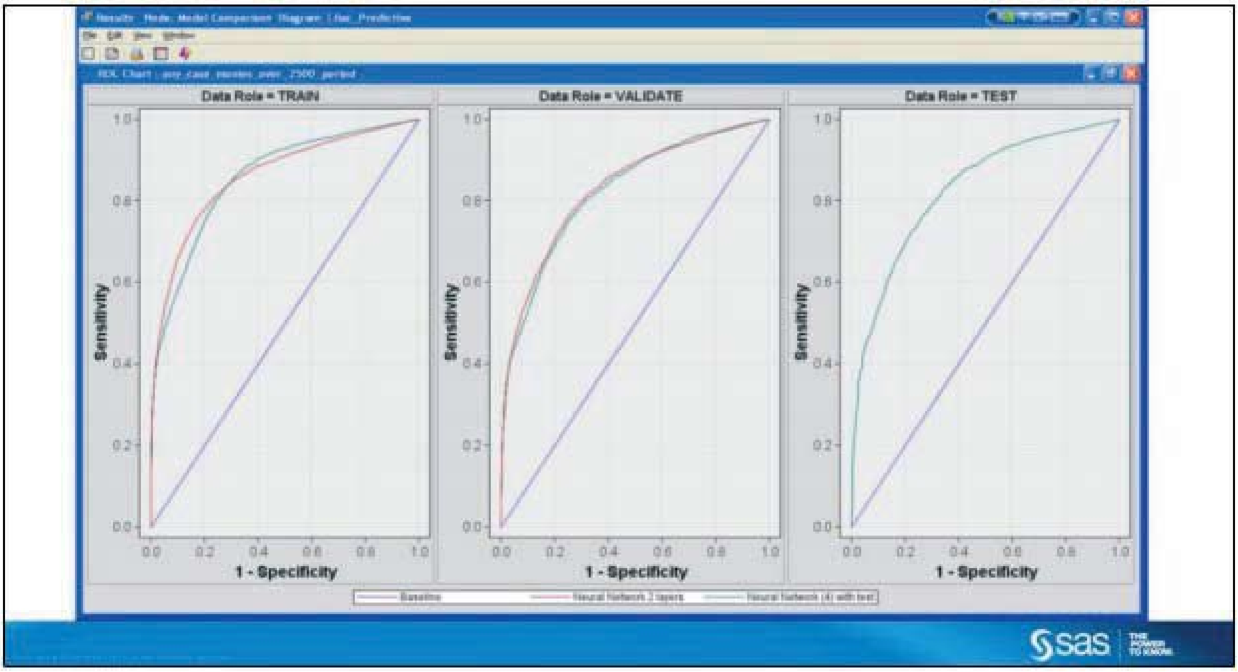
If we are to analyse dealers on the periphery of each cluster, such as for Cluster 1 in the diagram, we find that their characteristics differ from that of the cluster in certain measures. In this case, one notices that many of the dealers in Cluster 1 tend to have a low ratio of interstate to overall sales, but these tend to be noticeably higher for dealers who are at a greater distance from the cluster centroid. Therefore, it is possible to derive a rule that identifies any dealer in this cluster (the cut-off might be different for other clusters), with a ratio greater than 20 per cent, as someone who is flagged for investigation or further monitoring.

Consolidating upon the findings, it is possible to analyse the clusters, using a cut-off based on the distance measure across various ratios and metrics, and identify where aberrations might happen. Hence, when one applies a cut-off, and classifies dealers as outside the threshold and within the threshold, one notices that dealers outside the threshold tend to have higher ratios of interstate sales and purchases, as compared to effective tax rates (where there is not much of a difference). Therefore, one would tend to frame rules applying this threshold on dealers with high ratios of interstate trade.

The Diagram 14.18 below, known as a lift chart, or a ROC (Receiver Operating Characteristic) curve helps to understand the efficacy of the model generated – how good the model is in achieving its primary objective – that of separating the bad (fraudulent/suspicious/liquidated etc.) from the good. In a situation with no model being used to target taxpayers (in such a case, it is assumed that taxpayers shall be selected for audit on a random basis), the probability of capturing a good percentage of all fraudulent cases will also be random. This is represented by the blue line which diagonally passes through the chart.

However, were one to use a model to prioritise the taxpayers targeted, it would tend to yield better results than a purely random selection. Therefore, if we were to look at the red/green lines, which are shaped like a D in the diagram, we understand that the model, if used, would tend to capture a higher percentage of all fraudsters, with a lower targeted base for audit (the x-axis/specificity). This increases the yield of the audit process, since a lesser percentage of all taxpayers need to get targeted in order to capture a good percentage of the fraudsters.

Diagram 14.18: Lift Chart showing model effectiveness



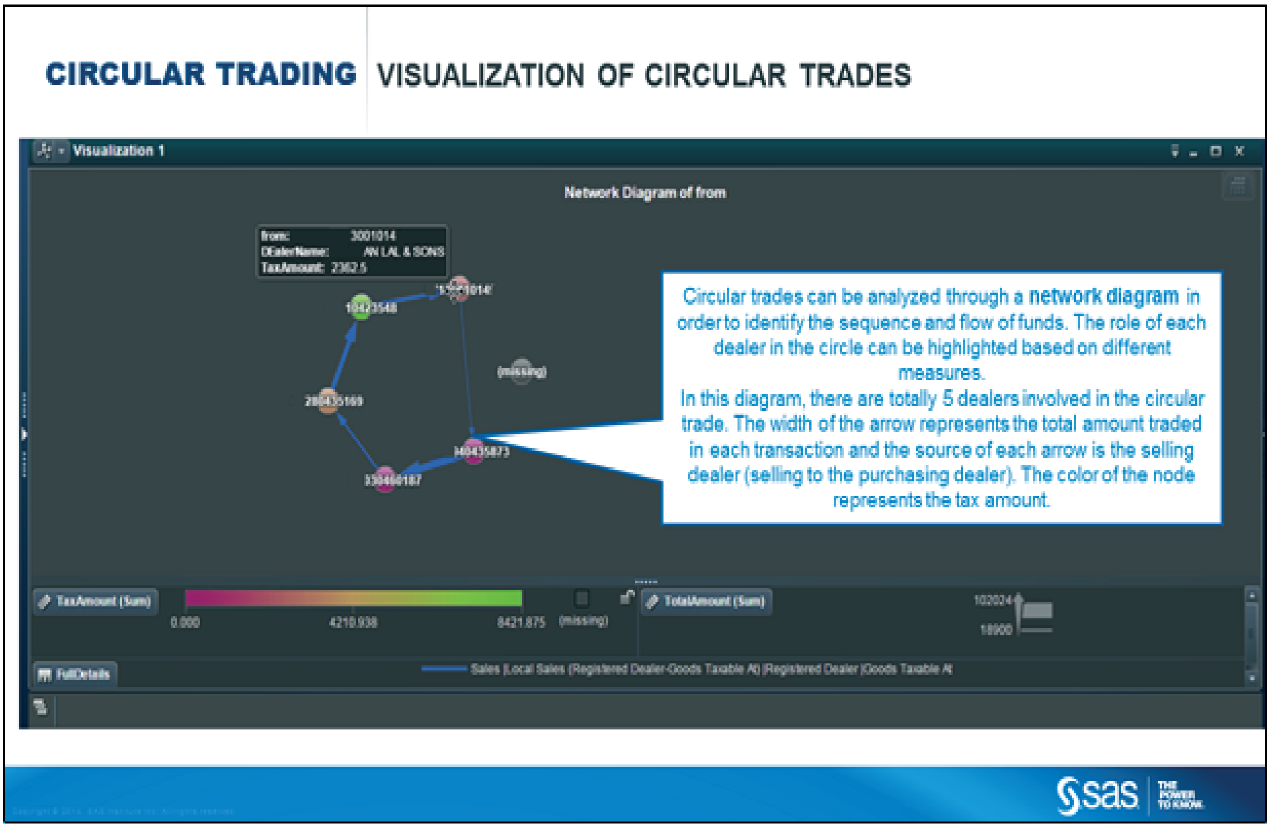
Source: SAS presentation to TARC

Social Network Analysis

The goal of social network analysis is to establish connections between people and businesses through associative linkage analysis. This technique is used to discover the networks of organised crime. As data is fed, the system automatically builds networks of relationships in the data – e.g. which transactions link different businesses together, which directors have directorships of other companies, etc. Where shared details are given such as addresses or telephone numbers, fuzzy matching methods link similar addresses together. Social network analysis, or linkage analysis, models the linkages established between disparate source data and groups of potentially fraudulent actors into a single malignant social network. Entities may include locations, service providers, beneficiaries, family members of fraudsters, addresses or telephone numbers, etc. In the past, constructing networks manually has been quite labour intensive and only highly skilled investigators, working over long time periods with manual tools, pushpins and string, have been able to identify enterprise-level activity. However, powerful tools now fully automate the analysis, with the system continuously updating the interrelated networks and rescoring for fraud, revealing so to speak, the elephant hidden among the mice. As such, identification of only one suspicious activity or entity in the ring can literally expose an entire network of highly collusive behaviour.

Social network analysis establishes frequencies among the various linkages of entities in the network and then runs those linkages back through anomaly detection and predictive modelling to identify collusive networks that are statistically unusual. This is a powerful tool to facilitate early identification of collaborative fraudulent networks, as it enables programme managers to act early to limit losses, prioritise their efforts, and aggregate losses to enable high-impact prosecutions. Social network analysis provides a holistic view of fraud. Typically, organisations only look at individuals and accounts to analyse activities. By combining social network analysis with rules, anomaly detection and predictive modelling, investigators can go beyond the typical view of fraud to see a bigger picture, including related perpetrators, and gain a clearer understanding of all activities and relationships at a network level. The hybrid approach also gives investigators a better understanding of emerging threats so that they can take action to prevent substantial losses before they happen.

Diagram 14.19: Network Visualisation

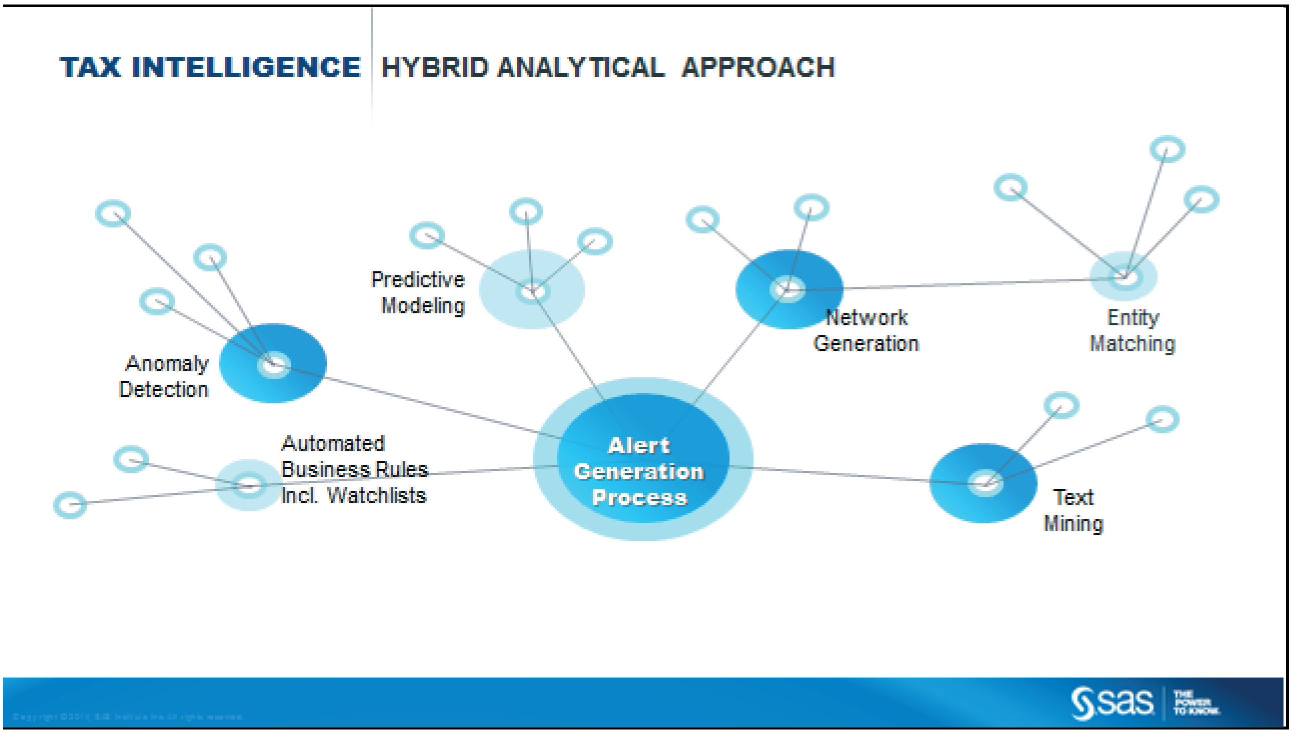


Source: SAS presentation to TARC

The above diagram illustrates one of the outputs of Social Network Analysis, viz. a circular trade or a cyclical pattern. This is relevant in the tax scenario because it is possible that individual dealers engage in collusive behaviour, buying and selling among themselves (or at least reporting transactions among themselves) in such a way that they all claim credit for inputs purchased. It is possible in such cyclical trades that the actual transaction did not take place and the transactions are getting reported only to enjoy the input tax credit.

It will be clear from the discussion above that the utility and effectiveness of different data mining techniques and tools depends upon the nature of the problem addressed in a given data mining project. Hence, organisations typically need to adopt a hybrid approach combining different techniques to get the best results, as depicted in Diagram 14.20 below.

Diagram 14.20: Hybrid analytical approach



Source: SAS presentation to TARC

XIV.4 Some practices in international tax administrations

Different tax administrations around the globe are now using analytics to predict and analyse areas of risk and accordingly detect emerging patterns in taxpayer behaviour specific to non-compliance, tax avoidance, tax evasion and tax fraud. Analytics is a data-driven approach that relies on information and knowledge contained within the data stores of tax administration and helps tax administrations allocate resources better and provide better services, improve operational efficiency, revenue collection and compliance.

As the range of data available in both the private and public sectors increases, tax administrations are taking opportunities to constantly improve their ability to detect hidden income and wealth using sophisticated technology. To achieve this, analytics plays an important role in enhancing a tax administration's capacity to identify potential fraudulent activities by using taxpayer data, both historical and current, in combination with established results from previous reviews and audits to predict probable outcomes of future cases. Tax administrations try and close in on such evasion and related economic crimes using and analysing data from tax returns, property purchases, loans, bank accounts and employment data to identify property assets, spot and track suspicious financial transactions and highlight connections that identify those trying to hide their income and wealth in order to evade tax. New age technology and analytics require tax administrations to dedicate resources and to bring in expertise from the private sector where needed to use analytics to sharpen the focus on tax evaders.

Frauds cost tax administrations billions in revenue each year. Accurate and rapid detection of potential problems help tax administrations increase tax collection and revenue. By collecting and analysing data from a wide variety of structured and unstructured sources, big data analytics has helped tax administrations across the world to automatically identify hidden relationships and activities that may point to fraud or errors.

Table 14.19 below depicts some of the data mining methods used by a few tax administrations around the world.

Table 14.1: Data mining techniques used by tax administrations to detect tax fraud

Technique Applied	USA	Canada	Australia	UK	Brazil	Peru	Chile
Neural Networks	✓	✓		✓		✓	✓
Decision Tree	✓	✓	✓			✓	✓
Logistic Regression	✓		✓	✓			
Self-organizing maps[426]			✓				✓
K-means			✓				✓
Support Vector Machines	✓		✓				
Social Networks Analysis combined with Visualisation Techniques	✓		✓	✓	✓		
Bayesian Networks			✓				
K-Nearest Neighbour			✓				
Association Rules						✓	

[426] The self-organising map is one of the models most widely used in artificial neural networks for analysis and visualisation of high dimensional data, based on unsupervised competitive learning.

Technique Applied	USA	Canada	Australia	UK	Brazil	Peru	Chile
Fuzzy Rules						✓	
Markov Chains					✓		
Time Series		✓					
Regression				✓			
Simulations	✓						

Source: Based on "Characterization and detection of taxpayers with false invoices using data mining techniques", Pamela Castellón González, Juan D. Velásquez, Expert systems with applications 40(2013)

Predictive analysis is about connecting the dots and finding a pattern in data. Often, different methods such as deviation analysis, clustering of transactions, time series analysis, regression analysis, decision trees and other statistical techniques are used in predictive analysis. The larger and more comprehensive the data, the greater is the success of predictive analysis.

The United States (US) National Targeting Centre (NTC) gets regular data feed from 15 agencies including direct links to their data base. Even the personnel of these agencies are stationed in the NTC. Using advanced analytics, they are able to home in on the movement of suspect passengers or suspect cargo across the borders.

The United States (US) Internal Revenue Service (IRS) has been using data mining techniques for various purposes like measuring the risk of taxpayer non-compliance, detection of tax evasion and criminal financial activities, electronic fraud detection, detection of housing tax abuse, detection of fraud by taxpayers who receive income from tax credits and money laundering. The IRS is using SAS Analytics[427] to help detect, prevent and resolve criminal and civil non-compliance with tax laws. The technology provided by SAS supports the IRS's new electronic Return Review Programme (RRP) system.[428] SAS scores tax returns through a hybrid approach of business rules, anomaly detection, predictive modelling and social network analysis. Users set up business rules that detect possible fraud and immediately alert investigators or auditors about suspect returns. The software searches data for anomalies that could indicate fraud or error. Predictive modelling uses historical behavioural information to identify suspicious behaviours that are similar to known

[427] SAS is a software suite developed by SAS Institute for advanced analytics, business intelligence, data management, and predictive analytics. In a gesture of exemplary co-operation, SAS shared with the TARC methodologies, approaches and international experience. The TARC acknowledges their contributions.

[428] The RRP is designed to help reduce the $345 billion tax gap – the difference between what taxpayers owe and what they pay voluntarily and on time. SAS is helping the IRS to reduce the number and amount of fraudulent tax refunds, discover emerging fraud schemes and increase tax collections to shrink the tax gap, identify tax evaders and fraud.

fraud patterns. Social network analysis uncovers hidden relationships or linkages that suggest collusion and organised fraud rings.

The Internal Revenue Service-Criminal Investigation (IRS-CI) uses three software programmes – Reveal, Web Currency & Banking Retrieval System (Web-CBRS)[429] and Electronic Fraud Detection System (EFDS)[430] – that perform sophisticated search and analytical tasks to search for specific characteristics that have been identified as potential indicators of criminal activity.[431] These programmes are used to perform data mining activities by searching databases of internal and external information.[432]

"Reveal" is a data query and visualisation tool that provides CI analysts and agents with the capability to query and analyse large and potentially disparate sets of data through a single access point, enhancing the user's ability to develop a unified overall picture of suspicious or criminal activity. Information is presented to the user visually, exposing associations between entities in the data that might otherwise remain undiscovered. Visualisation diagrams are built by using the "Visual inks" tool based on the data queried. The analyst is not required to manually construct the link analysis charts. The system is used in IRS-CI Lead Development Centers (LDC), Fraud Detection Centers (FDC) and field offices to identify and develop leads in the areas of counterterrorism, money laundering, offshore abusive trust schemes and other financial crime.

Another weapon used by the US IRS against fraud is the SAS text miner. The SAS text miner scours unstructured data, such as call centre data, to detect suspicious activity. Alerts and results are reported via a customisable dashboard. SAS' case management capabilities help investigators prioritise and assign cases.

The US uses the following data sources:

i. IRS: Third-party data store (TPDS), business master file (BMF), individual master file, information returns master file and questionable refund programme.

ii. Taxpayer: The source is the electronically/paper filed return.

iii. Employee: Source of employee information is the Online 5081.

[429] Web-CBRS is a web-based application that accesses a database containing Bank Secrecy Act (BSA) forms and information. IRS-CI access the database for research in tax cases, tracking money-laundering activities, investigative leads, intelligence for the tracking of currency flows, corroborating information, and probative evidence.

[430] In 2009, the IRS developed the RRP to replace the EFDS to automate many of the IRS's tasks that were previously performed by employees manually, available at http://www.taxpayeradvocate.irs.gov/userfiles/file/FullReport/Implementation-of-the-IRS%E2%80%99s-Return-Review-Program-Is-at-Extreme-Risk-Which-Could-Cause-Significant-Harm-and-Cost.pdf, accessed in December 2014.

[431] Federal Agency Data Mining Report, the Department of Treasury, 2010.

[432] SAS Analytics to help IRS identify tax evaders, reduce fraud

iv. Other Federal agencies: Federal Bureau of Prisons, Bank Secrecy Act data.

v. State and local agencies: All states and the District of Columbia prisons deliver prisoner listing information annually to CI in electronic format.

vi. Other third party sources: Commercial public business telephone directory

vii. listings/databases are purchased by CI to contact employers for employment and wage information, for example, Accurint.

The Australian Taxation Office's (ATO's) compliance programme is based on a risk model which uses statistical techniques and data mining in order to make comparisons to find associations and patterns by logistic regression, decision trees and support vector machines. Similarly, the existing model used by the New Zealand Inland Revenue Department (IRD) associates the degree of compliance with attention to auditing. This plan includes an analysis of the economic, international, population, ethnic diversity and family structures.

The ATO uses network and visualisation tools to rapidly and comprehensively understand complex structures used by some high wealth individuals. Every year, the ATO produces over 10,000 visualisations that support their compliance and intelligence activities. The ATO uses visualisation tools to map linkages and complex relationships and business structures, for example, for high wealth taxpayers with complex structures. Visualisations are prepared to support investigative activity in relation to taxpayers who are avoiding or evading their tax obligations through complex arrangements.

Text mining enables ATO case officers to rapidly identify relevant information from a mass of records to progress reviews more quickly. For example, computer hard drives contain a large amount of information, not all of which is relevant to the ATO's investigations. They use text mining to remove files that have no relevance to the case and sort through the remaining files to determine, among thousands of files, which files they should focus their attention on.[433]

ATO has also been undertaking extensive data matching. On receipt of information from its taxpayers, the ATO verifies the information and identifies discrepancies, anomalies and indicators of risk. The ATO does this by matching new information against previously verified (and therefore high integrity) ATO and external data.

The external data used by the ATO for data matching (and other activities) is acquired via four main avenues. These areas are as follows.

[433] In one example involving more than 800,000 files, text mining identified the files of interest reducing the number of documents that the case officers were required to review to 130,000. Obviously, this provided significant benefits in terms of the quality and efficiency of the ATO's investigation.

- Datasets provided by external parties to the ATO as a result of specific legislated requirements (for example, Annual Investment Income Reports (AIIR) provided by financial institutions). These legislated data sets cover major income streams such as salary and wages and interest and dividends, but do not cover other income streams such as capital gains, net rents, and income from self-employment.

- Datasets a government agency provides to the ATO under a memorandum of understanding (for example, state revenue offices).

- Commercially available datasets the ATO purchases (for example, electoral roll data).

- Datasets the ATO requisitions on a non-routine basis, for data matching projects (for example, data is requisitioned from sellers of luxury vehicles to address undeclared income and other tax risks).[434]

The effectiveness of data matching activities is influenced by the quality of internal ATO data, ATO's identity matching engines and external data. Having quality internal data is critical for the ATO to maximise the potential of its data matching activities. In order to ensure the quality of data, the ATO cleans its tax database from time-to-time, removes extraneous records and upgrades data quality, tightens up identity requirements, brings TFN registration arrangements up to a higher standard and undertakes effective data matching, especially with the use of extensive third party datasets.[435]

HMRC's approach to fraud and error has changed over the years. From the approach of "Pay Now and Then Check Later', which had led to excessive losses, a private sector approach has been adopted namely 'Check First Then Pay' involving segmentation analysis of its taxpayers to get a better understanding of those interacting with the system. From core analysis, HMRC has moved to embedding the principle of Check First, Then Pay into its business process and leveraging data-matching of both the Departmental data and the third party data.

Its strategy is based around 5 key elements viz. prevent, detect, correct, punish and deter supported by a real-time information system. HMRC data-matching capability is supported by an integrated and intelligence unit, which is a hub for data and intelligence on fraud. It uses best data matching techniques drawing on private sector best practice and has at its disposal high quality analysts using intelligence acquired to target high risk cases. A feedback loop is created which enables it to gather information on the success of the referrals. This information is used in a process of constant re-evaluation to continually review, refine and improve data-matching techniques and risk scoring models.

[434] The Auditor-General, Performance Audit on The Australian Taxation Office's Use of Data Matching and Analytics in Tax Administration, 2007-2008.

[435] The Auditor-General, Performance Audit on The Australian Taxation Office's Use of Data Matching and Analytics in Tax Administration, 2007-2008.

Where HMRC cannot prevent fraud, it identifies and stops it as soon as possible. Under this element of the strategy it drives efficiencies in its detection work through the creation of a single, integrated fraud investigation service which investigates fraud across HMRC and local authorities.

It also has a cross-departmental Identity Fraud Unit using industry experts to target the problem of identity fraud. This integrated approach enables HMRC to tackle fraud in its entirety, across the range of services, sharing information and analysis where appropriate. HMRC and the Department of Works and Pensions are working together to counter fraud in the benefit and tax credit system.

Alongside this strategy, HMRC has strengthened its analysis of the causes of fraud and error and put in place mechanisms for ensuring focus on prevention.

HMRC proactively educates taxpayers to ensure that they are aware of their obligations to report changes in circumstances, backed up with media campaigns and customer guides and fact-sheets on its website.

HMRC has developed the Connect software, which is its strategic risking tool that cross-matches over one billion internal and third-party data items and presents the information as a network of associations. Connect's visualisation tool has impressive features that combine the integrated compliance environment (ICE) and the analytical compliance environment (ACE). ICE provide graphical analytical tools and infrastructure that provides risk and intelligence analysts and provides tax investigators with technology to follow threads of information and highlight anomalies in stated tax liability. ACE risk profile tools provide technical analysts and statisticians the tools to interrogate and risk profile large volumes of data. They create profiles based on key risks identified within the data and are used regularly to produce a target population for investigation. These profiles are maintained and adapted to reflect changes in customer behaviour.[436]

HMRC's 'Connect' takes in data from over 28 different data sources. It then cross-matches this data over a billion internal and third-party items. These include items such as property purchases, tax returns, loans, bank accounts and employment data. By doing so, the system can uncover hidden relationships across organisations, customers and their associated data sources.[437]

The Canada Revenue Agency (CRA) uses neural networks and decision trees to distinguish the characteristics, based on the results of past audits, of taxpayers who evade or commit fraud to detect patterns of non-compliance or evasion.

---

[436] Business Intelligence Technology helps HMRC Increase Yield, Capgemini in collaboration with HMRC, available at http://www.in.capgemini.com/resource-file access/resource/pdf/ss_Business_Intelligence_Technology_helps_HMRC_Increase_Yield.pdf, accessed in December 2014.

[437] Capgemini Consulting Paper on Digital Leadership: An interview with Mike Hainey, Head of Data Analytics at HMRC.

The Irish tax and customs authority used a predictive modelling workbench such as SAS to develop data mining techniques to assist in the better targeting of taxpayers for possible non-compliance, tax evasion, or liquidation. This was carried out by harnessing the organisation's risk analysis programme, which was already running on 300 business rules and on approximately 800,000 taxpayer entities. Based on the data generated by these rules, the organisation created an analytical base table (ABT), which consists of a transposed set of multiple characteristics defining a taxpayer's performance before a stated event. The event to target/measures was defined by the tax authority as one of the following (as per appropriate model):

- Whether an alert on the entity resulted in an audit which yielded revenue
- The quantum of revenue yielded once audited
- Whether the entity under question liquidated his business (business failure) or not

Having established the target variable, the tax authority then proceeded to create a predictive model that highlighted the variables and characteristics that separated the targets (1) from the non-targets (0). This was done using an analytical methodology known as SEMMA (sample, explore, modify, model, & assess).

- Sample – Sampling the dataset to be analysed, and dividing it into a development and a validation dataset
- Explore – Carrying out preliminary explorations and statistical analysis
- Modify - Carry out transformations necessary for analysis
- Model – Fit an appropriate model as per the target variable to the relevant ranges and weightage assigned to the explanatory variables
- Assess – Test the model for consistency, robustness and efficacy

Peru's *Superintendencia Nacional de Administración Tributaria* (SUNAT)[438] was one of the first to apply neural networks, decision tree, association rules and fuzzy rules to detect tax evasion, adding to the selection system of the Maritime Customs of Callao, an artificial intelligence tool based on neural networks. During 2004, this model was improved through the application of fuzzy rules and association for pre-processing variables, and classification and regression trees to select the most relevant variables.

The Brazilian Federal Revenue has developed a risk analysis project, involving the use of artificial intelligence, jointly with some universities throughout the country. This project consists of a detection system of atypical points to help regulators identify suspicious transactions based on a

[438] Tax morale in Latin America: Public Choice, B. Torgler, 2005

graphic display of information on historical imports and exports and a system of export product information based on Markov chains,[439] to help importers in the registration and classification of their products, avoid duplication and to calculate the probability that a string is valid in a given domain.

The Chile Inland Revenue Service (IRS) developed its first trial in 2007 using self-organising maps and k-means to segment its VAT taxpayers. Later, following international trends, they built risk models for different stages of the lifecycle of the taxpayer in which neural networks, decision trees and logistic regression techniques were applied. The first trial was further developed to identify potential users of false invoices through artificial neural networks and decision trees, mainly using information from tax and income declarations in micro and small enterprises.

The Swedish Tax Agency (STA) uses analytics to help it achieve two important objectives – to ensure taxes are correctly reported and to ensure that tax debt is paid. Even though Sweden enjoys one of the most tax compliant cultures in the world, the STA regards the maintenance of high compliance as their key task. Using advance analytics, they have easily identified risky profiles and indicators, preventing tax fraud and ensure focus where the risk of non-payment is greatest.

According to STA extent of the tax discrepancy (defined as the tax that should be reported but isn't) is estimated to be roughly 10 percent of the theoretically correct tax and one of its goals is to half the tax discrepancy. And one of the Analysis Unit's tasks is to identify anything that could jeopardize meeting this goal. To reduce the percentage of tax errors, the agency has endeavored to make life easier for tax payers wherever possible, gradually simplifying its tax returns. It has also introduced various control measures and conducted information campaigns. Going further, it publishes information about some of our controls online, so that individuals and companies can run checks themselves before submitting their returns and avoid being flagged in their control system, thereby saving time for the taxpayers as well as the administration.

Legislative and Policy Framework for data analytics

In a self-assessment tax regime, tax legislation places responsibility upon taxpayers to declare all their assessable income and claim only deductions and/or offsets to which they are entitled in deriving their income.

In the US, the use of all tax data is governed by the Internal Revenue Code - 26 U.S.C. 6103. Subsection (a) sets out the general rule of confidentiality. Subsection (b) sets forth definitions of terms commonly used throughout Section 6103. Subsections (c) through (o) of Section 6103 contain exceptions to the general rule of confidentiality. In addition to disclosures permitted under provisions of Section 6103, other provisions of the code also authorise disclosure of tax

[439] A Markov process is a stochastic process (random process) in which the probability distribution of the current state is conditionally dependent of the path of past states, a characteristic called the Markov property. The Markov chain is a discrete time stochastic process with the Markov property.

information. The information contained in Web-CBRS is gathered under the guidelines dictated by the Bank Secrecy Act, 31 U.S.C. 5311.[440]

The ATO essentially derives its legislative authority from two areas of the law – the first, specific provisions in legislation requiring annual reporting of certain incomes (for example, wages, interest, dividends) and the Tax Commissioner's information gathering powers and the second, memoranda of understanding (MoUs) covering acquisition of data, where prepared, under the legal authority of the Commissioner's powers. Some external datasets are acquired on a normal commercial basis without regard to any legislative authority.[441]

Compliance with privacy requirements

In the US, the head of each department or agency of the Federal Government that is engaged in any activity to use or develop data analytics is required to submit a report to Congress on all such activities of the department or agency under the jurisdiction of that official.[442] Additionally, since citizens expect that the tax information provided to the IRS is private, an abuse of that reasonable expectation of privacy or the loss of public confidence would seriously impair the tax system. Hence, all tax information is protected as required in 26 U.S.C. 6103. The use of Bank Secrecy Act information is strictly controlled under the statute that directed its collection.[443]

Since the UK HMRC uses a variety of data sources that they are legally entitled to utilise, they take privacy issues very seriously. All data used is proportionate and appropriate in tackling the range of risks and issues that HMRC faces. They apply rigorous audit against people who use the 'Connect' system and have strong controls on movement of data from 'Connect' to other environments.

In Australia, the Office of the Privacy Commissioner is an independent office outside the ATO that has responsibilities under the Federal Privacy Act, 1988, to promote an Australian culture that respects privacy. The Privacy Commissioner notes that data matching poses a particular threat to personal privacy because it involves analysing information about large numbers of people without prior cause for suspicion.[444]

Organisational Structure

In the US, two bureaus of the Department of the Treasury are engaged in data mining activities – the IRS and the Financial Crimes and Enforcement Network (FinCEN). The IRS data mining

[440] Federal Agency Data Mining Report, The Department of the Treasury, 2010.

[441] The Auditor-General, Performance Audit on The Australian Taxation Office's Use of Data Matching and Analytics in Tax Administration, 2007-2008.

[442] Section 804 of Title VIII, Privacy and Civil Liberties, Public Law 110–53, 121 STAT. 363,

[443] Federal Agency Data Mining Report, The Department of the Treasury, 2010.

[444] The Auditor-General, Performance Audit on The Australian Taxation Office's Use of Data Matching and Analytics in Tax Administration, 2007-2008.

programmes focus on the identification of financial crimes including tax fraud, money laundering, terrorism, and offshore abusive trust schemes. The FinCEN's data mining activities focus on money laundering activities and other financial crimes.[445]

The UK HMRC has a dedicated analytics team that has blended three skill sets together – operational research, data specialists and frontline tax expertise. This combination has proved effective in delivering results and provided practical insights to evolve HMRC's big data solution.[446] Its Fraud Investigation Service (FIS) acts on referrals from a number of channels to investigate potential cases of fraud. The organisation is driven by the work of around 1800 fraud investigators, a dedicated national organised fraud unit and a national operational intelligence to provide evidential support for fraud investigators. FIS focuses its resources on investigating cases which are most likely to result in a criminal sanction.

The ATO undertakes two categories of data matching – large-scale, post-issue, semi-automated system processing and data matching projects. The two categories of data matching are administered by different areas of the ATO. The micro enterprises and individuals business line is responsible for data matching in the first category. However, the Data Matching Steering Committee (DMSC) co-ordinates and facilitates the initiation and conduct of data matching projects in the second category. There are similarly two broad categories of analytics projects. Advanced analytics projects, largely based on data mining and modelling, are administered by the analytic group in the Office of the Chief Knowledge Officer. Basic analytics projects, using more conventional tools, are distributed broadly across most ATO business lines.[447]

Ireland Revenue and Customs have a dedicated Research and Analytics Branch (RAB) that conducts program-wide and macro-level research at a corporate level. The branch conducts analyses to transform data into information using analytical software. The branch's work in includes large sample surveys, data mining (population profiling/ customer segmentation, pattern recognition, forecasting, predictive modelling) data quality exercises, experimental design for evidence based decision support, economic research and risk analysis. The branch uses both Revenue data and data from other sources. This work enables Revenue to make better use of its data and provides an improved understanding of the taxpayer population. The results are used to better target services to customers and to improve compliance.

Sweden has created a dedicated analytical unit for advancing analytical work. The unit engages in extensive to develop and refine predictive models that enhance the agency's capacity to identify and address potential non-compliance. It had engaged two master's degree students from

[445] Federal Agency Data Mining Report, The Department of the Treasury, 2010.

[446] Capgemini Consulting Paper on Digital Leadership: An interview with Mike Hainey, Head of Data Analytics at HMRC.

[447] The Auditor-General, Performance Audit on The Australian Taxation Office's Use of Data Matching and Analytics in Tax Administration, 2007-2008.

Copenhagen School of Business to further its research to improve its risk models using network analysis.

In the US, tax data is self-reported by the individual/entity submitting the information to the government. Web-CBRS data is gathered from information compiled by the reporter based on information provided by their customer or based on the reporter's personal experience. Investigators scrutinise the SARs filed by the subject companies and request grand jury subpoenas for the underlying documentation. The supporting records are examined and individuals of interest are identified. IRS-CI applications are required to have internal auditing capabilities. The internal audits track user access and queries performed with checks to validate against misuse. In addition, the data is a read-only extract that is validated for missing or duplicative data before entering the CI systems and remains unchanged in the CI systems.[448]

XIV.5 Way forward

XIV.5.a Increasing the data pool for analytics

The primary object of predictive analysis is to use the information resources available to the tax administration to move from a reactive approach to non-compliance, including frauds, to a proactive approach that deters and prevents fraud. Predictive analysis also enables the administration to anticipate and address emerging risks by guiding and enabling it to develop and deploy other appropriate interventions such as improved customer services, sharply targeted scrutiny and audits. Further, by enabling automation of many routine functions, it frees up resources for better quality work and contributes to greater efficiencies. Therefore, it is essential that both direct and indirect tax administrations place increasing reliance on it in their operations.

The note in Appendix XIII.3 describes the current sources of data for the CBDT and CBEC. With the advances made by them in automation, a much larger pool of data is progressively becoming available to them, although in the absence of comprehensive automation, issues of data quality and data completeness continue to remain important constraints on the availability of data for reliable analysis. The biggest obstacle is that even the data that is available is not readily accessible as it is held different organisations or different wings of the same organisation who have not established routines of sharing information. This, coupled with silo-style working of the tax administrations, is a major contributor to the tax gap. Greater sharing of information would lead to the gradual filling of the tax gap.

Sharing of information is critical as the value of advanced analytics can be realised fully only when they are applied to large data sets constructed from as diverse and wide sources as possible.

---

[448] Analytics in Tax Administration, 2007-2008.

[448] Federal Agency Data Mining Report, The Department of the Treasury, 2010.

In Chapter IX, the TARC has dealt with the sharing of information and covered all key issues relating to it and made a number of recommendations relating to the following.

i. Common framework for sharing

ii. Common standards and taxonomy

iii. Third party exchanges

iv. Data storage, data quality and data usage

v. Safeguards and security

vi. Audit and accountability

vii. Human resources

All those recommendations are critical for preparing the framework for sustained use of analytics.

While noting that both the Boards had moved to set up their data warehouses, and acquired advanced analytical and reporting tools, in Section VII.4 the TARC had noted:

> *However, these again are separate projects, and the current constraints and under realisation of potential are likely to continue. It is learnt that discussions on data sharing are on between the Boards. Hopefully, these will yield results. It is of utmost importance that these are speeded up and concluded. Besides the internal data in the two tax systems, integration with third party data from which the CBDT has derived substantial value already, is equally important. Even though the CBDT continues to wrestle with the challenge of non-PAN data, available data could be of immense use in the CBEC, particularly in the area of service tax. Going beyond this, successful organizations are now travelling beyond the traditional data warehousing technologies to exploit the opportunities of "big data" by tapping into the data that is constantly being churned out in the world. The CBDT and CBEC will have to explore the potential in this area as well. Even assuming that hassle free data sharing is established, the two administrations doing their analyses independently would again mean a fragmented approach towards the taxpayer.*
>
> *Considering this, it is clear that it will not be sufficient to merely create a mechanism for data sharing (which, by itself, has proved to be a matter of considerable difficulty between the two Boards). The Boards will have to create mechanisms for joint analyses and exploitation of the data that they acquire and hold. Nothing short of a common database and a joint mechanism for exploiting the data would provide the answer. Quite apart from anything else, this will also avoid considerable (and quite expensive) duplication of effort.*

In Chapter IX, The TARC also gave a roadmap towards the eventual integration of data warehouses independently set up or being set up at present. Even though both departments appear to be taking steps to increase the use of analytics, the use of data accessed through data exchange with each other is sporadic and episodic. As long as this situation remains unaltered, the value of analytics even at a preliminary or elementary level will remain underexploited.

The first order of business for both the Boards, therefore, should be to take urgent steps to integrate data across the two tax administrations. This should ordinarily require no emphasis as even the sporadic exchange of information has yielded immediate benefit to both administrations as illustrated in Appendices X.6 and XII.2 of the third TARC report. Hence, data exchange needs to be put on a firm footing so that it becomes a regular routine between the two organisations. Going beyond data sharing, it is important that the teams in the two Boards begin working jointly on the analysis of pooled data so that their efforts are synergised and the combined talent in the two organisations is brought to bear on the tax. This will pave the way for mutual understanding that will lay the foundation for a sustained collaboration and cooperation. And this should be only the first step leading to eventual integration of the databases as recommended by the TARC earlier.

As noted, the use of third party data in the CBDT is relatively more advanced than in the CBEC. CBEC has been enabled to secure third party data by Section 15A of the Central Excise Act, 1944 (which has been made applicable for service tax also). However, its efforts appear to be in a nascent state. It is also not known to what extent third party data already available with the CBDT is being shared with the CBEC and is being utilised. As already noted by the TARC, data sharing between them is at best sporadic and this would be yet another example of the silo-style approach between the two. It may lead to situations in which persons or organisations may have to submit the same data to both. This can be avoided if the "one data many users" principle is adopted as recommended in Chapter IX of the TARC's report.

Further, in Chapter IX the TARC had also highlighted the need to enact a specific legislation, in line with best international practice, for inter-agency data sharing that could lay out the broad "umbrella" framework within which specific arrangements could be worked out. The marked lack of consistency and regularity in sharing of information between the two Boards, and between the Boards and other agencies, underlines the need for the creation of a framework that not only enables but also mandates such sharing. Hence, it is important that urgent steps are taken to develop and enact such legislation.

In Section VIII.4.e, the TARC had highlighted the need for revamping the customs clearance process in line with the global best practices. This will entail a substantial enhancement in the effectiveness of the Risk Management System through use of advanced analytics on a richer data set that includes the relevant external data.

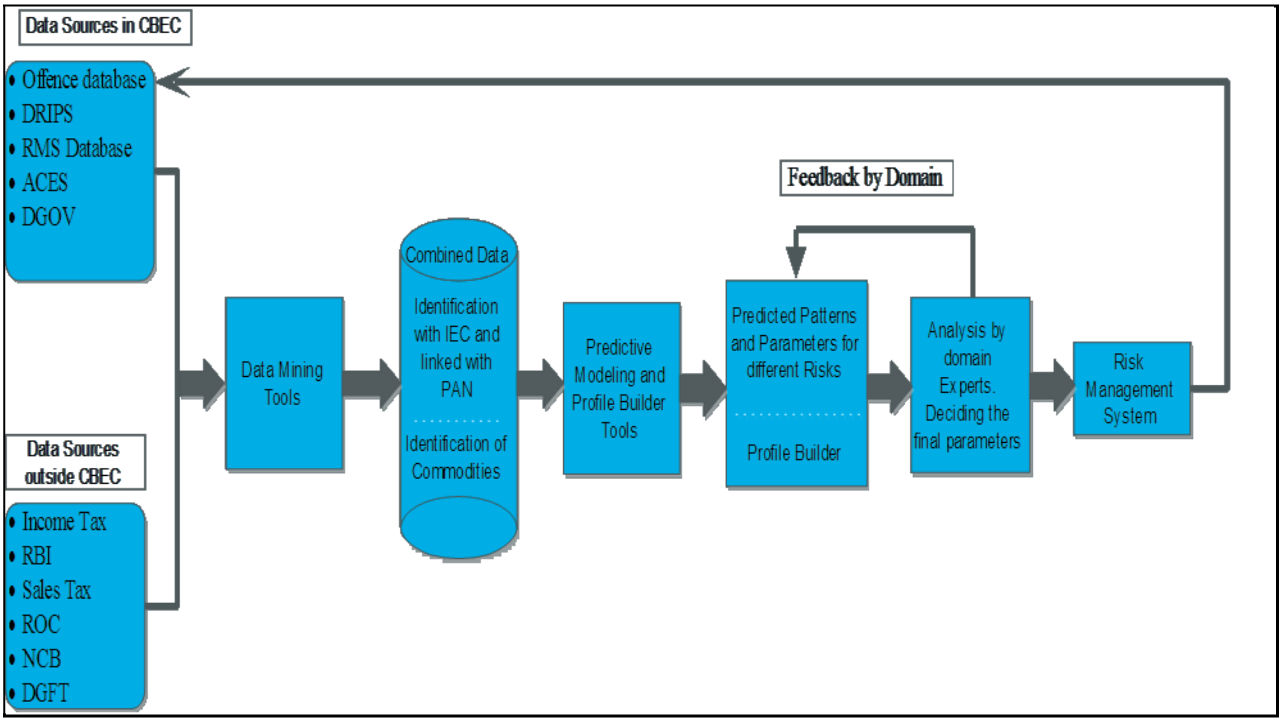
Table 14.2 below illustrates how integration of external data can enrich CBEC's RMS by enlarging the basket of risk indicators that can be used to profile entities and transactions for a more robust predictive analysis.

Table 14.2: Integration of external data

Sl. No	*Risk Indicators*	*External Sources*
1	*An Importer/Exporter is also a CHA/Freight forwarder*	*ITD-MS of CBDT, MCA-21*
2	*An Importer/Exporter is also a Supplier/Consignor*	*ITD-MS of CBDT, MCA-21*
4	*Family member of the Importer/Exporter is a CHA/Freight forwarder*	*ITD-MS of CBDT, MCA-21*
5	*Fictitious address of the Importer/Exporter*	*ITD-MS of CBDT, MCA-21*
6	*Frequent changes in the ownership of a company*	*ITD-MS of CBDT, MCA-21*
7	*Ports visited/Routing of the vessel*	*Shipping Line/Airlines*
8	*Abnormal duration of travel*	*Shipping Line/Airlines*
9	*Unusual Commodity from the Country*	*Shipping Line/Airlines*
10	*Weight difference between Customs declaration and data with shipping lines/airlines*	*Shipping Line/Airlines*
11	*Last minute booking of vessel/aircraft*	*Shipping Line/Airlines*
12	*High insurance cost*	*Insurance/Shipping line/Airline*
13	*Disproportionate transportation cost*	*Insurance/Shipping line/Airline*
14	*Offenders under income tax*	*CBDT*
15	*Person/Entities appear in STR of CBDT*	*CBDT*
16	*Cost as shown in Transfer pricing document*	*CBDT*
17	*Cost incurred overseas for the goods imported*	*CBDT*
18	*Services rendered on behalf of the supplier*	*CBDT*
19	*Abnormal remittances for the services received from supplier*	*FETERS data of RBI*
20	*Value of goods imported/exported shown differently in VAT returns*	*State Commercial Tax*
21	*Abnormal VAT benefits*	*State Commercial Tax*
22	*Discrepancies in the VAT registration form compared to IEC details*	*State Commercial Tax*
23	*Remittances made to the supplier of the imported goods through other purpose codes*	*FETERS data of RBI*
24	*Advance payments for imports*	*FETERS data of RBI*

Sl. No	Risk Indicators	External Sources
25	*Defaulters under BEF statements of authorised dealers*	*BEF statement of RBI*
26	*Importer/Exporter having accounts for remittances with multiple authorised dealers*	*FETERS data of RBI*
27	*Remittances to different countries*	*FETERS data of RBI*
28	*Remittances from different countries*	*FETERS data of RBI*
29	*Persons Entities involved in Narcotics smuggling*	*NCB*
30	*Offenders under SFIO*	*SFIO*
31	*Offenders under DGFT*	*DGFT*
32	*Persons involved in licensing fraud*	*DGFT*
33	*Person who was rejected by DGFT for exporting dual use goods*	*DGFT*

Diagram 14.21: Schematic depiction of the customs risk management process using both internal and external data



This is just one example how analytics can be leveraged by using external data sources in the customs context. Similar examples can be cited for other taxes as well.

In line with the recommendations made in Chapter IX of the TARC report, it is equally important to have an on-going exercise to continually scan the environment for identifying new and emerging sources of data to enlarge the information pool available to the organisation. This is to capture the potential generated by the ever-increasing digital data by different departments and organisations. Therefore, it is important for the two Boards to create a dedicated mechanism which will be responsible for continuous scanning of the environment and engaging with multiple agencies for securing data access. This will be an important responsibility of the DG (Systems) as CIO of the organisation. This will enhance the robustness and incisiveness of the analytics efforts.

As already noted, the digital coverage of the operations of the two Boards remains partial and large parts of their operations remain in the traditional paper environment. This is bound to prove a significant impediment to effective use of analytics. This is particularly relevant in the context of predictive analysis for detection or prevention of tax or economic crimes. For predictive analytics to work, analysts must have access to data related to non-compliance. Both organisations hold the data but do not share them regularly, and the same is true of most other government agencies. Besides, the offence databases suffer from significant weaknesses. The processes are not carried out in online systems. Hence, the primary mode of data capture is after the fact data entry, which inevitably suffers from delays, inaccuracies and incompleteness. As was noted earlier, a vital element in the construction of a predictive model is training of the model on a properly built and reliable data set and in the absence of a robust offence data base, the predictive ability of the model is bound to remain sub-optimal. Besides data sharing, there should also be proper data publication. As is the case with several countries, there also is a need to publish consolidated watch lists such as lists of tax defaulters, international fugitives and known offenders. Several of these lists are openly available and are routinely used in analytical databases along with additional attributes derived through analysis. The global practice to publish such lists supports the idea that they should be incorporated into any types of analysis.

These considerations will also apply when it comes to the use of analytics in other domains of critical operations. For the value of analytics to be realised, data quality is of the essence. Information should be accurate and complete and the harmful consequences of inaccuracies must be guarded against. TARC has dealt with these aspects in Chapter IX of its report.

Analytics rely extensively on data sources from the organisation's business systems, such as online transaction processing (OLTP) systems. The richer and more complete and reliable the data, the more effective is the analytic product, and *vice versa*. This underlines the importance of transforming the tax administrations to a "digital by default" status, a recommendation that the TARC had emphatically made in Chapter VII of its report. In fact, in Section VII.6, TARC had laid out the broad guidelines for the path in the journey to "digital by default". In the TARC's view, the importance of beginning that journey cannot be overemphasised if the tax administration in India is to reach the level of maturity in the use of data analytics that is comparable to its best international peers.

Looking beyond the immediate, there are certain key conditions which are essential for success in the deployment of analytics in an organisation.

XIV.5.b Mainstreaming and integrating analytics in the organization

(i) Leadership and culture

Analytics cannot exist as an isolated function. It must support core organisational functions such as intelligence, investigation, fraud prediction and prevention, forecasting and analytical research. In fact, these functions cannot be carried out effectively without proper analytical input. This position is supported amply by the examples cited in the previous section showing the extensive use of analytics in many tax administrations.

For analytics to be integrated into the organisation, it must, on the one hand, occupy a central place at the strategic core of the organisation while, on the other, it must inform its operational decisions and actions. Analytically focused organisations apply analytics with a clear strategic intent. This requires an enterprise-wide approach to managing information and analytics and cross-functional working to pull together the expertise, data, and systems that allow the optimisation of organisational resources.

The leadership of the two Boards will have to play a critical role in laying the foundation of analytics in the organisations' culture. While there has been frequent talk about the need for a data driven culture for a long time, there is little evidence of a change in attitudes and working practices at all levels of the two organisations. A common challenge faced by many organisations seeking to enhance the use of analytics is that they have to battle the common perception of analytics as a purely IT innovation, overlooking the need for managerial innovation to successfully use analytics. This is true of the two Boards as well. By and large, ICT functions are still seen as peripheral to the core functions of the organisation by the bulk of IRS officers as evidenced by the marked reluctance of officers to undertake specialisation in this area, leading to the DG (Systems) in both Boards experiencing difficulties in getting the required number of willing and capable officers. As for analytics, the appetite appears even weaker. In the popular lore of the organisations, these would seem to be areas that are best left to a few "geeks", who carry on their lonely toil.

This culture has to change if a data driven approach is to be imbibed in the organisations. And the onus of generating momentum for that change clearly lies on the top leadership of the two Boards. For a start, they must themselves learn to value information as the key asset of the organisation, embrace the idea that data are central to the organisation's business and demonstrate their appreciation of their value in everyday work by seeking areas where data analytics could deliver quantum leaps in performance. For this, they need to acquire a working knowledge of data analytics so they can understand what is rapidly becoming feasible. It needs to be ensured that the use of ICT and data analytics becomes a core ingredient of the strategic and operational plans in different functional domains. Experience shows that a key challenge for data-analytics leaders is getting their senior management teams to understand its potential, finding enough talent to build models, and creating the right framework to tie together the often disparate databases inside and outside the organisation. The importance of educating and sensitising the top leadership, therefore, cannot be overstated.

It is only when a top-level perspective is in place that the necessary behavioural changes can begin to percolate down through the ranks of the organisation.

(ii) Integrating with operations and getting buy in from frontline

As the TARC has already noted in Chapter IX, the organisations and agencies that collect data are not fully cognizant of the possible analytical uses of the data they hold. It is critical to get a buy-in from operational staff for productive use of analytics and this cannot happen unless they are sensitised to the power of analytics and see a connect between their jobs and the product of analysis. For most frontline users, ICT tools are often a "black box". They frequently lack confidence that analytics will improve their decision making or simply do not understand the analytics or the recommendations it suggests in the absence of adequate understanding and communication. Therefore, they understandably fall back on their historic rules of thumb when they do not trust the analytics, particularly if their analytics-based tools are not easy to use or are not embedded into established workflows and processes.

For example, the criteria used in the present CASS system are generally opaque to users and confined to the Directorate of Systems and a select Committee formed with a few officers from the field formation. As a result, at the field level, the Assessing Officers are today completing scrutiny assessments often without full and complete awareness of the very risks being sought to be covered, as the narrations on the system are often cryptic and confusing. Another instance of lack of communication can be seen where certain scrutiny cases are marked as "priority" on the system but the Assessing Officer or Range Head are not kept informed of the rationale behind such prioritisation and they complete these scrutiny assessments without this information. The reasons for prioritisation are neither conveyed clearly to the Chief Commissioner or Commissioner, who are nevertheless expected to review such cases from time to time.

Such barriers need to be overcome by improving the quality and accuracy of communication and improving the reliability of the system and user friendliness of the interface. Equally important is an ongoing programme of user training and education. Hands-on experience, with the required guidance, helps people grow accustomed to using data. That builds confidence and, over time, people's capacity and appetite for data-informed problem solving and decision support.

Similarly in selecting cases for audit or scrutiny, the analytics-based selection mechanism need to be designed to ensure that the probability of a productive outcome is high so that officers do not go after the cases predicted not to win. This will enhance audit or scrutiny productivity and infuse confidence amongst officers and taxpayers.

New technological solutions can provide help in meeting this challenge. The deployment of user friendly "self-service" tools can increase business users' confidence in analytics. They are rapidly increasing what is called democratisation of information – i.e., getting analytics out of the exclusive hands of technical specialists and into the hands of a broad base of frontline users, thus scaling up the use of analytics. Without needing to know a single line of coding, frontline users of new technology tools can link data from multiple sources (including external ones) and apply predictive analytics. As noted earlier, visualisation tools are putting business users in control of the analytics tools by making it easy to slice and dice data, define the data exploration needed to address business issues, and support decision making. This, of course, requires substantial investments in such tools.

Finally, it's becoming much easier to automate processes and decision making. Technology improvements are allowing a much broader capture of real-time data (for example, through sensors) while facilitating real-time, large-scale data processing and analysis. These advances are opening new pathways to automation and machine learning that are becoming mainstreamed across different sectors. By automating routine tasks, employees can be relieved of routine drudgery and released for more productive and challenging tasks. For example, as the customs risk management system matures, and is made more potent and effective through advanced analytical models and tools, customs can explore the option of "machine release" that will lead to release of types of cargo without any human intervention. Similarly, with improved CASS, supported by advanced analytics, the income tax department can vastly sharpen the selection of returns for scrutiny and avoid unproductive workload on their assessing officers, leaving them with a scrutiny basket that is more challenging and productive.

This requires that the top leadership assumes responsibility for sustained ground-level change, maintain the momentum of change and create hunger and enthusiasm for data-led and evidence-based decisions. A substantial part of the investment on analytics, by some accounts of best practices, nearly half needs to be devoted to user friendly frontline tools and measures such as training and coaching in order to make frontline employees eager to use the products of analysis.

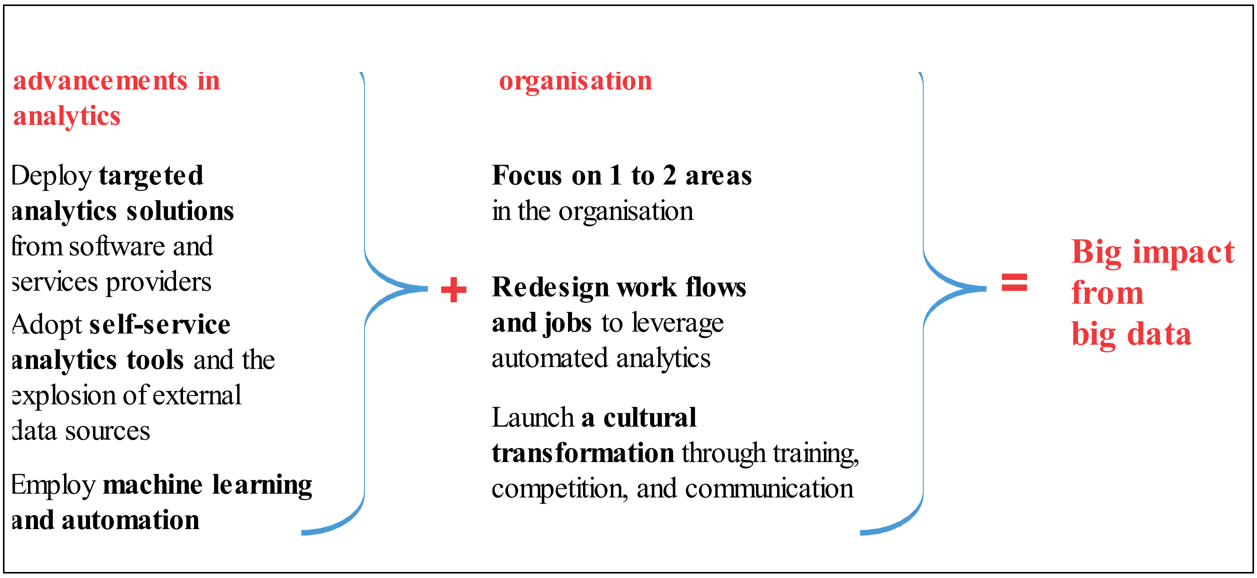
(iii) Developing strategic plans and delivering through well-defined and designed projects

Data analytics will under deliver on its potential without a clear strategy and well-articulated initiatives and benchmarks for success. Capturing the potential of data analytics requires a clear plan that establishes priorities and a well-defined path to business results. Therefore, it must be an integral part of the strategic planning process. As the organisation draws up plans, the planners must ask themselves what role analytics can play in making a big impact in achieving planned outcomes. It is essential to build this business-analytics efforts linkage while the plans are on the drawing board and analytical efforts need to be integrated with programme and project management.

Based on such strategic plans, an analytics plan needs to be developed in a project mode – setting out the resources needed, defining the business goals and the expected business outcomes. The team responsible for the analytics efforts should have a clear view of their targets and expectations.

An example could be the need to expand the taxpayer base. This will necessarily involve concerted action across multiple fronts including changes in legislation, enforcement and customer services. A clearly laid out analytical plan, beginning with a review of data requirements including plans for identification and acquisition of data can lay the ground work for the application of analytics that will discover potential taxpayers so far hidden from the administration and help develop appropriate and effective actions or interventions for the administration to take. By delivering concrete results, the analytical efforts will establish its credibility and help in embedding analytics in the organisation. Diagram 14.22 below from McKinsey and Company graphically depicts such an approach.

Diagram 14.22: A best practice approach



Source: McKinsey Analysis

Starting with a small achievable project with a clearly defined goal can often be the first step on the path to success. The results do not need to be spectacular, but if they show how data mining can add value and potentially reduce fraud and error, then the case will be made and analytics can start to become a core part of the agency's business processes.

Therefore, an important aspect is that planning must be done to capture low hanging fruit along the way. Therefore, it is important to begin with small scale pilots that establish the benefits and show visible results that build the confidence of the organisation in the value of analytics, and scale up the projects as they progress. The aspect of scalability – shifting from exploratory analytics to a full-scale rollout – should be an important consideration in planning for projects.

It is equally important to recognise that this is a painstaking journey and the organisation has to map a careful and well defined path towards progressive maturity in terms of complexity and sophistication – building capabilities along the way. It necessarily involves lot of experimentation and plans for development need to accommodate sufficient flexibility for the purpose.

Implementation also needs to be accompanied by an ongoing programme and project evaluation to ensure continuing effectiveness of the analytics efforts.

(iv) Role of KAIC

Being a highly specialised and multifaceted discipline, analytics requires the necessary structures to be created. The Knowledge Analysis and Intelligence Centre is the structure that has been recommended by the TARC. In Section VII.4 of its report, in relation to the setting up of KAIC the TARC had noted:

> *Best performing organisations, including many tax administrations, are creating dedicated units for analytics. Typically, the mechanism adopted follows the "Centre of Excellence" model and the unit is a high-valued shared resource in the organisation. Multiple skills need to be brought to bear on the task of advanced analytics. These include deep data analysts, domain experts with sharp analytical skills and ICT capabilities, statisticians and economists, behavioural scientists and even, on occasion, data scientists. These skills are not normally resident in a tax administration. They are also in short supply. Considering the high premiums such resources command in the market, there are challenges in attracting and retaining the right talent. A major motivation for people in such areas is the content, challenge and varied nature of the problems they are asked to work on.*
>
> *Hence, the "Centre of Excellence" approach is the best bet for setting up such a unit successfully. The ICT function, through the SPV we are recommending, will have to support it by providing the ICT platform, tools and technologies and perhaps, some expert resources as well. Lest there be confusion, it should be clarified that it is not our recommendation that the KAI centre should be the sole repository of analysis. It would be highly valued and to entrust normal analysis to it would be a serious case of under-utilising a scarce resource. Good quality analysis, relevant to each function, should continue to happen in each functional vertical and the ICT support for this should be provided by the ICT function. In fact, an analytical and data driven approach should drive operational decisions in each functional vertical.*
>
> *Only complex problems with strategic implications should normally get referred to the KAI centre. It is, in fact, possible to create a financial model for chargeability for KAI centre services that will ensure that it is optimally utilised. In such a scenario, each functional vertical will budget for and pay for KAI centre services.*

The KAIC needs to be a hotbed of learning and innovation as teams in it share ideas on how to construct robust data sets, build powerful models, and translate them into valuable business tools. It should also support the analytical efforts of the different functional verticals through collaboration and guidance to projects that they may on their own undertake. Its goal should be to be so successful at building data-analytics capabilities across the organisation that they can tackle increasingly ambitious priorities. As these verticals build their analytics capabilities, the KAIC will increasingly focus on longer-term projects involving sophisticated R&D, with an emphasis on analytics innovation and breakthrough insights.

With regard to staffing, the KAIC will need to be staffed with a combination of multidisciplinary expertise such as ICT specialists, data scientists, statisticians, behavioural or social scientists and domain experts with rich operational experience. The specialised talent in some cases will have to be recruited from outside, either directly or through engagements with private firms or vendors specialising in data analytics.

However, it needs emphasis that while such outside talent would be necessary, there is no substitute for a cadre of domain experts who also possess the required ICT knowledge and statistical and analytical capabilities. Analytics work requires skills that must be acquired by the line managers. It is therefore necessary that such a cadre of analysts is developed internally by identifying and training suitable departmental officers who would form an important part of the analytical set up, both in the KAIC and in the respective functional verticals. The need for such in house analysts is particularly important in areas of intelligence and investigation, wherein use of predictive analysis critical. These are sensitive areas and access to information has often to be restricted on "need to know basis" and officers would need appropriate security clearances before they are given access to data and information resources that could include data flowing not only from the departmental processing systems, but also inputs from their intelligence gathering activities, whether through physical or through other technology assisted means. Such information necessarily needs to be closely guarded as the consequences of leaks can be severe. On the other hand, the value of joining such highly sensitive information with information available from open sources for the purpose of analysis can have inestimable value.

As the TARC has already noted, this will need to be achieved through changes in HR policies to encourage and sustain the creation of such specialisations within the department.

(v) Training

The importance of training and development in spreading the use of analytics in the organisation cannot be overestimated. In Chapter IV of its report, the TARC has already dealt with this. Both the KAIC and the training institutes need to collaborate in fulfilling this need. Training should be imparted at multiple levels.

At one level, business users need to be sensitised to the value of analytics and equipped with knowledge about the use of tools that they would be expected to use. This should be delivered through appropriately designed curricula at all levels of the organisation. And the trainees must include the top levels of the organisation. As already noted, the required knowledge about the potential of analytics among the top leadership is an essential prerequisite for its successful use. Hence, it is necessary to invest in properly designed and mandatory in-service courses for senior and mid-level officers to equip them to lead transformation efforts. Frontline officers will need to be given more task oriented training, including hands on training, in the use of ICT tools.

At another level, training efforts will have to be directed at building a cadre of data and analytics savvy officers, who will meet the analytical needs both in the KAIC and the functional verticals. As already noted, a key ingredient for success in the use of analytics is the presence of such a cadre within the organisation. This strength is necessary not only to successfully deploy analytics but also to manage projects robustly and in the best interest of the organisation. It becomes critical when issues of transition management, such as change in the vendor or project partner, arise.

Such a cadre of officers will also need to play a key role, that of "translators", who provide the link between the ICT/analytics domain and business domain to both top and line management. They will also be a key resource for training efforts.

For building such a skill set, the department cannot rely purely on internal resources. Hence, there is a need to build partnerships with industry, professional trainers as well as academic institutions as well as sponsor chosen officers for specialised training courses. Similar partnerships will be necessary to enrich the research efforts of the KAIC.

Finally, at the induction level, it is essential that young IRS inductees, the future leadership of the two Boards, are sensitised early to the use of ICT and analytics as they begin their careers. Therefore, foundational training in this area should be part of their training curriculum.

To sum up, the successful use of analytics needs what Davenport and Jarvenpaa[449] call the DELTA model.

D – Accessible high quality **data**
E – An **enterprise** orientation
L – Analytical **leadership**
T – A long-term strategic **target**
A – A cadre of **analysts**

XIV.6 Recommendations

The TARC recommends that:

i) Increasing data pool for analytics

a) Both the Boards need to take urgent steps to integrate the data across the two tax administrations and begin working jointly on the analysis of the pooled data so that their efforts are synergised and the combined talent in both organisations is brought to bear on the task. This will pave the way for mutual understanding that will lay the foundation for sustained

[449] Strategic Use of Analytics in Government, Thomas H. Davenport and Sirkka L. Jarvenpaa, IBM Centre for Business of Government, 2008

collaboration and co-operation. And this should be only the first step leading to eventual integration of the databases as recommended by TARC earlier. (Section XIV.5.a)

b) With regard to third party data, the "silo" approach should be avoided and the "one data many users" principle must be adopted as recommended in Chapter IX of the TARC's report. (Section XIV.5.a)

c) To enable reliable predictive analysis, the offence data base needs to be made robust, reliable and shared. Besides data sharing, there should also be proper data publication. As is the case with several countries, there is also a need to publish consolidated watch lists such as lists of tax defaulters, international fugitives and known offenders. (Section XIV.5.a)

d) As recommended in VII.6 of the TARC's report, both the Boards need to commence the journey to the "digital by default" status in order to reach a level of maturity in the use of data analytics comparable to the best international peers. (Section XIV.5.a)

e) Data analytics should be made an integral part of the strategic planning process and the analytical efforts should to be integrated with the programme and project management. (Section XIV.5.b.iii)

f) Based on such strategic plans, analytics plans needs to be developed in a project mode – setting out the resources needed, defining the business goals and the expected business outcomes. (Section XIV.5.b.iii)

g) An important responsibility of the CIO will be to expand the data pool for analytics by continuous scan of the environment for identifying new and emerging sources of data and securing data access from them, in line with recommendations in Chapter IX of the TARC report. (Section XIV.5.a)

ii) Leadership and culture

h) The leadership of the two Boards needs to play a critical role in laying the foundation of analytics in organisations' culture through visible actions to move the ICT functions to the core of the organisations. They need to do this by

- Imbibing a data driven approach, acquiring a working knowledge of data analytics and learning the value of information as a key asset.
- Embracing the idea that data are central to the organisation's business and seeking areas where data analytics could deliver quantum leaps in performance.
- Ensuring that the use of ICT and data analytics become a core ingredient of the strategic and operational plans in the different functional domains. (Section XIV.5.b.i)

i) Frontline employees' buy in needs to be secured by

- Improving the quality and accuracy of communication and improving the reliability of the system and user friendliness of the interface.

- An ongoing programme of user training and education including hands-on guidance.

- Investments in user friendly "self-service" tools that can increase business users' confidence in analytics.

- Progressive automation of routine processes, for example, as its risk management system matures and is made more potent and effective through advanced analytical models and tools, customs can explore the option of "machine release" that will lead to release of certain types of cargo without human intervention. Similarly, with improved CASS supported by advanced analytics, the income tax department can vastly sharpen the selection of returns for scrutiny and avoid unproductive workload on their assessing officers, leaving them with a scrutiny basket that is more challenging and productive. (Section XIV.5.b.ii)

j) Analytics-based selection for audit or scrutiny need to be designed to ensure that the probability of a productive outcome is high so that officers do not go after the cases predicted not to win. This will enhance audit or scrutiny productivity and infuse confidence amongst officers and taxpayers. (Section XIV.5.b.ii)

iii) Strategic plans and delivery

k) Analytical capacity needs to be developed on the basis of clearly laid out plans with defined business outcomes. (Section XIV.5.b.iii)

l) Recognising that progress towards maturity is a painstaking journey, there are benefits in starting small, launching pilots and capturing low hanging fruit along the way to build confidence and then scaling up. (Section XIV.5.b.iii)

m) The planning, however, should factor in the requirements of scaling up. (Section XIV.5.b.iii)

n) The implementation should be accompanied by regular programme and project evaluation. (Section XIV.5.b.iii)

iv) Role of KAIC

o) The KAIC needs to be made a hotbed of learning and innovation with continuously increasing R & D capabilities to achieve breakthrough insights and should be staffed by an adequate strength of multidisciplinary expertise, including domain experts with strong ICT/analytics skills. (Section XIV.5.b.iv)

p) It should also support the analytical efforts with the different functional verticals through collaboration and guidance to projects that they may on their own undertake. (Section XIV.5.b.iv)

q) It should also support training and skill building in the organisation. (Section XIV.5.b.iv)

v) Training

r) A cadre of data analysts should be groomed within the organisation to service analytical needs in both the KAIC and the functional verticals. They can also act as "translators" who interpret technology for business leaders and operational staff. (Section XIV.5.b.v)

s) Focused and well-structured training courses should be designed and mandatorily implemented for all levels in the organisation to equip them with the necessary knowledge and skills. (Section XIV.5.b.v)

t) This should include the top and middle-level management who must be imparted broad knowledge of analytics and its potential. (Section XIV.5.b.v)

u) A well designed course on ICT and analytics should be made a part of IRS officers' induction training in the two academies. (Section XIV.5.b.v)

v) Partnerships need to be built with industry, academic and research institutions to build and sustain highly specialised skills and to promote research in KAIC. (Section XIV.5.b.v)

Chapter XV

Research for Tax Governance

Chapter XV

Research for Tax Governance

List of contents

Chapter XV

Research for Tax Governance

XV.1 Importance of research in tax governance

There exists a virtuous circle that begins with improvements in the tax administration that lead to better governance – facilitating revenue mobilisation, yielding better financing for public services, encouraging investments, and leading to more sustainable economic growth. These links, no doubt, are complex and context-specific, but they nonetheless indicate a strong synergy between tax reforms and governance. The TARC in its three reports and the last two chapters of this report has looked carefully and deeply into the tax administration processes, structure, and people function to recommend fundamental reforms in the way different programmes and functions of the CBDT and CBEC are delivered to taxpayers. The reports looked into the economic costs to society, including costs of compliance by taxpayers and distortions in economic decision making due to inconsistency in tax administration procedures, poor design of tax regulations, inappropriate skill sets and work process of tax organisations, and insufficient use of technology, and suggested measures to improve tax compliance and tax governance. The TARC's recommendations not only looked into the current situation and at existing international practices, but also into the foreseeable future.

Tax administration is not static; it has a dynamic character, requiring constant evaluation and assessment, so as to seamlessly modernise itself and look into its future needs. The dynamic character of the tax administration is necessarily linked to demand for greater responsiveness and accountability, better management, improved customer services, an organisational structure oriented towards more efficient and prompt delivery, and stronger oversight. These demands require continuous, ongoing research in tax governance so that there is sufficient and modern thinking available to improve the processes, structure, and people functions in the tax administration, leading to better tax governance. These research components need to include international comparisons by identifying good practices adopted by different tax administrations, understanding them and drawing lessons from them to raise the standards of tax governance.

The current Indian tax administration is attempting to enhance delivery, although much is desired in terms of initiating and expanding such attempts on the basis of research outcomes. Research in tax governance, with a multi-disciplinary approach involving disciplines such as economics, accounting, finance, law, management, behavioural science, ICT, and statistics, can review current practices and design new approaches to expand the tax base and increase tax revenue, while remaining cognizant of the government's aim to promote economic development and tax justice. Research is evidence-based and will have to be meticulously designed and conducted so that the output is relevant to and usable by the tax administration. This will require an organisational framework, finance, staff, development of linkages for collaboration – internal-governmental mechanism or with outside agencies – and an evaluation of the output to ensure their utility to the tax administration. These exercises for undertaking research cannot be a one-off exercise, but have to be embedded in the tax administration so that there is a two-way movement, top-down as well as bottom-up, within the various tax departments building an ecosystem for undertaking meaningful research. This will ensure that departmental decisions to change the structure and processes are well-informed, and backed by definitive findings. It will ensure transparency and credibility, and result in meaningful gains to both the administration and taxpayers. These policy-driven analytical exercises will bolster the desired impact on the economy by successfully building sound tax governance.

XV.2 Current practices in India

Analytical capability in the Indian tax administration is at a nascent stage. A bulk of the officers gets assigned to the administration function and have rarely been involved in detailed analysis based on research. A small group of officers in the TPL and TRU of the CBDT and CBEC, respectively, or in the Directorates carry out top-level analysis. Officers in the TPL and TRU, to a great extent, are involved in the legislative formulation of direct and indirect taxes. Most of the research by them is for framing tax policy, laws and procedures for inclusion in the Finance Act and/or for framing delegated legislation in the form of circulars and notifications. It is difficult to say much about the quality and quantity of research input as they are largely internal to the Ministry of Finance and not known to outsiders. Nonetheless, it is acknowledged that research input, if any, is sporadic and sketchy, and is at the initiative of individual officers involved in the process.

A beginning in research and analysis was made to estimate tax expenditures, i.e., the loss to the exchequer due to tax incentives. This exercise was carried out meticulously in India by the CBDT and CBEC and began to be reported from the 2006 Indian budget. But the methodology of the exercise remains rudimentary and the comparability of the indirect and direct tax estimates has some limitations. While the beginning was good, a number of issues in tax expenditures estimation were to be resolved and the approach needed to be sharpened and deepened in terms of further subdividing sectors for gathering data. But, the analysis has remained where it was at the beginning. In comparison, advanced countries, for example, the US, carries out detailed analysis on tax expenditures and reports them in voluminous detail.

Some analytical and research capability also resides in a few other offices; for example, the international taxation directorate focusing on transfer pricing and arm's length issues, or the recovery directorate on improving tax collection methods. But, these again are not systemic, but based on individual initiative. The main reason often identified for almost no analysis and research is the absence of specialisation. While in advanced tax administrations, transfer pricing or tax treaties would be typically considered as specialised fields in which an officer would be allowed to specialise throughout his/her entire career, in India, officers are transferred approximately every three years across a broad spectrum of cities at varying levels of convenience or hardship.[450] This is intended to maintain equity across employees but it renders specialisation difficult.

Apart from the above, the Income Tax Department, until a few years ago, used to publish "Investigation of Accounts" on a regular basis, but that seems to have been discontinued. These publications could not be called research input for tax governance, as their contents were largely intended to develop better industry understanding among the officers, and were useful for improving compliance audit. Nonetheless, they had details of how industry worked, including production processes, and used to succinctly bring out the non-compliance methods generally adopted by selected industry segments. This has now been replaced by the 'Let Us Share', a knowledge management initiative, which has the best-case experience of officers on select audit cases. Some independent and sporadic initiatives have also been taken by the Mumbai and Ahmedabad Chief Commissioners, who have compiled experiences on select audit cases in their respective zones.

A separate directorate for legal issues and research was set up by the CBDT in 2001. The research side of the directorate produced only one paper on external audit efficacy, following which it did not carry out any work worth mentioning.[451] The job on the legal side was so engaging and humungous that the research side just got subsumed in the overwhelming work on the legal side. In fact, before 2001 too, the publicity and publication directorate in the CBDT, called the Directorate of Research, Statistics, Publication and Public Relations, had a research component. It had a separate complement of officers and staff from the Indian Statistics Service. But even at that time, research work was not given any priority, and hence, research output was almost non-existent. The Income Tax Department's attempts at introducing research brings out two points: first, the CBDT did recognise the need for research, and second, it did not realise the need to keep the directorate for research separate, treating research instead as add-on work.

There is no separate directorate for research in the CBEC. The Centre of Excellence under NACEN has been created to carry out research, but at present, it is lying dormant for want of staff and officers and proper focus. The Directorate General of Service Tax in the CBEC has placed on its website profiles of different services. They are exhaustive on operational issues – whether there is any exemption available for a particular service, details on valuation of taxable services for charging service tax, etc., – but by no stretch of imagination does the website contain material that could be considered even close to research input for tax governance.

---

[450] This has been pointed out and discussed in detail in Chapter I and IV of the TARC report.

[451] The CBDT had, some time back, also considered setting up a Direct Tax Research Institute, as an autonomous society, to carry out research on direct tax policy and administration. It was proposed to be governed and managed professionally by select former and in-service officers under the overall supervision of the CBDT. The institute was to carry out research on topics given to it by the CBDT and its attached directorates. The institute was also to engage experts on different subjects on contractual basis, to enhance the quality of research. But the stage of its processing of the proposal is not known.

Overall, it can be safely stated that both the Boards have not really focused on research for tax governance.

Some input is also commissioned from research institutions, again in a scattered manner and not on a regular basis, on issues generally asked for by parliamentary committees.[452] Since these commissioned reports are mainly for the work of parliamentary committees, the conclusions arrived at may serve only a limited purpose and are intended to provide inputs on questions posed by the committees.

XV.3 International practices

Advanced tax administrations recognise the importance of research as an input to improve tax governance. They have well-laid out institutional arrangements to ensure that tax governance gets research and analysis of high quality on matters relating to tax policy and tax administration. In the UK, both the HM Treasury and the HM Revenue and Customs have their own research teams. The Knowledge, Analysis and Intelligence (KAI) team, comprising professionals from mainly four disciplines – statisticians, economists, operational researchers and social researchers – is a key contributor to research and analysis. The KAI identifies, builds upon and shares knowledge to deliver evidence-based research on tax policy, needed customer service improvements, and departmental efficiency. The multi-disciplinary team in the KAI works closely with operational and policy persons within the HMRC and the HM Treasury to prepare their research work.

Apart from the research being carried out internally, the HMRC also gets external research for the Department's strategic and operational business needs and delivery to HM Treasury. Such analytical research is helpful in improving the design of HMRC systems, products and processes, targeting better compliance, and developing a more robust and strategic HMRC and HMT knowledge base. External research also serves the needs of the HMT/HMRC policy partnership. To build in flexibility and cater to changing priorities, resources for research are allocated on a twice yearly cycle, with planned and current projects published on the HMRC Research webpage. Published findings from research and analysis projects are also available on the website. Some of the recent research work of HMRC, as available on the website, are the following:

[452] Three reports recently commissioned by the CBDT on black money generation and widening the tax base were on the basis of recommendations of the parliamentary committees. Commissioned research on compliance cost for the taxpayers is at their own initiative.

- Migration to PAYE real time information: customer experience research
- HMRC tax agent segmentation research
- Evasion Publicity Post-Campaign Tracking 2014: Individuals
- Evasion Publicity Post-Campaign Tracking 2014: Small and Medium Enterprises
- Compliance Perceptions Survey 2013
- Annual Tax Summaries: Testing of web content with customers

In the US, the Office of Tax Policy in the Treasury, the US President's Office of Budget and Management and the Congressional Budget Office together perform research and analysis on tax questions of all types, including policy, administration and operations. These agencies produce regular reports – some annual, and some once every few years. The Office of Tax Analysis within the Office of Tax Policy analyses the effects of existing tax laws and alternative tax programmes and prepares a variety of background papers, position papers, policy memoranda, and analytical reports on economic aspects of domestic and international tax policy.[453] These reports are generally in the public domain. The Office of Tax Policy brings out the Fiscal Year Green book and Fiscal Year Tax Expenditures annually to accompany the President's budget proposals. The Green book, among other things, provides an analysis of current law, reasons for change and an explanation for each tax-related proposal.

In Canada, the Tax Policy Branch (TPB) of Canada's Finance Ministry is responsible for the development of tax policy at the federal level, and their personnel conduct internal research and analysis, including economic analyses and reviews of case laws, to generate ideas for tax initiatives and to analyse ideas for tax changes from other sources. The Canada Revenue Agency (CRA) carries out separate research, having quantitative and qualitative content, with a focus on contemplated policy options. The TPB's staff strength is about 150, comprising economists, lawyers and accountants. The Department of Finance in Canada brings out *Tax Expenditures and Evaluations,* an annual publication, containing estimates of tax expenditures and analytical research articles on various tax policy issues.

In Australia, the Australian Tax Office along with the Treasury plays an important role in carrying out research and analysis for tax governance. The research is on tax policy options and their economic and social impact, tax revenue forecasts and the cost of taxation policies. The research group has specialists from law, economics, finance and statistics. Besides internal research and analysis, private sector consultants are also engaged from time to time. Recently, this group has taken to conducting semi-annual consultations, including early stage pre-policy consultations, with stakeholders.[454] As part of research, the ATO regularly conducts surveys to monitor perceptions

---

[453]http://www.treasury.gov/resource-center/tax-policy/Pages/Tax-Analysis-and-Research.aspx, accessed in January, 2015

[454]Tax Policy Formulation in Australia, Rob Heferen, Nicole Mitchell, and Ian Amalo, accessed in January, 2015, available at

of taxpayers generally, and of the business community and tax professionals about the way the ATO administers the tax systems, and to gauge satisfaction levels. Some of the key research works carried out by the ATO during 2014 are as follows:

- Attitudinal and Behavioural Research on the Prevention of Aged Debt
- ATO Perceptions and Service Satisfaction survey
- Professional to Professional (P2P) Service User's Survey
- Perceptions of Fairness in Tax Disputes
- Identity Crime Support Research

The tax governance process in New Zealand is a formalised multiphase 18-month process from the conception of fiscal and revenue strategy to legislation and post-implementation review. The process includes a role for the Treasury as well as the Inland Revenue Department with strong research and consultative components, including data analysis. The National Research and Evaluation Unit, within the Inland Revenue Department, carries out specialist research, evaluation and analysis of various aspects of tax administration programmes and services from time to time. There is also an effort to build a role for academic institutions into the process. In 2009, a Tax Working Group was set up in Victoria University, Wellington, in conjunction with the Inland Revenue and the Treasury to bring together expert tax practitioners, academics, business people, and officials to consider key problems with the current tax system and options for reform. Some of the recent research works done by the Inland Revenue Department are the following:

- Potential impact on the Integrity of the Tax System of Sharing Taxpayer Information with Other Government Departments to Identify, Stop, or Disrupt Serious Crime
- Information Sharing Between Government Agencies – Cultural Perspectives
- Impact on the Integrity of the Tax System of IR Sharing Information with Other Public Sector Organisations
- Immigrant Entrepreneurship and Tax Compliance
- Large enterprises tax compliance costs
- The Cost of Business Tax Compliance

XV.4 Way forward

Improving tax governance revolves around the competing expectations and goals of taxpayers and tax administrations; tax administrations desire certainty of tax compliance and certainty of attaining a revenue target while the taxpayer seeks a 'fair' tax burden, low compliance costs and

---

http://www.treasury.gov.au/~/media/Treasury/Publications%20and%20Media/Publications/2013/Economic%20Roundup%20Issue%202/Downloads/PDF/1-Tax-Policy-Formulation-in-Australia.ashx.

certainty in tax rulings. The tax department seeks to device procedures and laws and to design governance structures that promote tax compliance and garner revenue to government, and the taxpayer reacts with his own strategies to reduce the tax burden, lower compliance costs and bring finality to tax matters. Tax departments employ strategies on three fronts – defining tax base and imposing tax rates, setting up tax compliance procedures, and employing enforcement measures based on risk identification. Such strategies are pursued through legislation or through operational administrative measures. The taxpayer has two types of strategies at his disposal – lobbying (for tax shelters or lower tax rates, for example) and choosing the level of compliance. An ideal tax governance process will bring about convergence in attaining the expectations of the tax department as well as the taxpayer. But, it is also important for the tax administration to orient their strategies to guide the tax system towards that ideal equilibrium. Research has an important role to play in the process of attaining that equilibrium. Changes to the tax system need to be guided by knowledge of direct and indirect taxes, and the non-tax effects that changes would trigger. It is only with such knowledge that the tax department can fashion its strategies to progressively move towards the ideal equilibrium. In other words, the tax department's pursuit of its policy strategies has to be guided by research that provides answers to questions like the following. Will the strategy increase or decrease compliance? Will it widen or shrink the tax base? Will it increase or decrease compliance costs? Will it increase or decrease the effectiveness and credibility of its enforcement? Does the tax administration have the capacity to implement the strategy? The tax department's strategies in the tax governance process needs to be informed by qualitative and quantitative answers to the above questions. Meaningful research in tax administration should enable the evaluation of the relative efficacy of alternative strategies in achieving the objectives. The need for continuous research into taxpayer behaviour and into tax administration's effectiveness and how they affect the tax department's objectives should be part of research.

In fact, such research cannot be sporadic and event-oriented, but will have to be systematic and continuous with a much wider ambit to cover issues of tax administration and tax policy. A review of advanced tax administrations, in Section XV.3 above, shows that these tax administrations have built their foundation of tax governance on high quality of research inputs, with an institutional arrangement for carrying out research. These research inputs, whether framed in-house or through external collaboration, are evaluated and then utilised for improving tax governance. The Indian tax administration will also have to embark on a similar route by identifying critical areas of research to start with, setting up an institutional mechanism with a robust staffing pattern to carry out and evaluate the research and to use the results of such research gainfully, etc.

XV.4.a Areas of tax research

Research on important areas of tax administration structure, its compliance programmes and tax laws determines the quality of a tax administration's governance. Some of the key areas of research often undertaken are how to tax more, how to reduce or increase tax exemptions or deduction structure, etc. While these areas may be critical for a resource-starved nation like India, it is also important that research on issues of tax governance is carried out on a regular basis in both the tax administrations. Some of the research work can be carried out jointly by the CBDT and CBEC.

Examples of important areas of research, only indicative by enumeration and by no means exhaustive, can be as follows.

i) Compliance tracking

Compliance strategy is best pursued through random enquiry programmes. The first step towards developing the strategy is to segment taxpayers into types. The extent of tax compliance can then be measured for each of these types. Compliance tracking for each type can then be used for resource allocations – how much of the overall resources should be allocated to compliance activities. This is critical because the marginal compliance yield may fall off quickly as compliance activity thins out across too many areas. Thus, compliance activities pertaining to large business corporation tax scrutiny or large civil investigations would tend to have high marginal yields. Other activities may not have comparable marginal yields. Tracking compliance and yields and allocating resources on the basis of such tracking can be an area of regular research in tax administration. This will also pave the way for business process re-engineering if the monitoring results so demand. For instance, the following would typically have low yields: (a) compliance by large businesses in their role as employer withholding tax (b) charities (c) scrutiny on particular aspects of corporations or (d) self-assessment individuals. Spending a lot of administrative resources erroneously on the last aspect, even though it possesses moral hazard, may not be wise.

Carrying out regular research and analysis on the success of random audit or scrutiny selection can provide evidence on the impact of random audit or scrutiny on taxpayer behaviour. The research and analysis will provide a study on the impact of the preventive measures, tracked, may be, for three years after the audit. Such tracking will also provide a basis for resource allocation. Many tax administrations, in particular the UK's HMRC,[455] have used such a strategy and reported positive preventive effects on non-compliant medium-sized and small businesses. For this research exercise, the data comprised randomly selected samples of medium and small-sized (S) businesses and non-complex individuals for two fiscal years, 2000 and 2001. The data were fitted to a difference-in-differences model. The model helped in identifying a treatment group of audited taxpayers and a control group of taxpayers not audited in that year. The change in the outcome (declared tax liabilities) before and after the treatment (audit) for the two groups – treatment and control –was compared through regression methodology.

ii) Identifying the rich and wealthy

Identification of high net worth individuals (HNWIs) and developing a separate tax treatment for them are among the top priorities of many tax administrations. Tax treatments often revolve around

[455] Also see Parthasarathi Shome, Tax Shastra: Administrative Reforms in India, United Kingdom and Brazil, Business Standard Publishers, New Delhi, 2012,Chapter 4

minimisation of tax avoidance and evasion by this category.[456] But, tax administrations do not set up analytical units that could use intelligent yet just methods of identification, since the use of third-party information can typically address only, and up to, the upper middle classes leaving out the top end of the wealthy. Lack of will to pursue a suitable tax treatment based on detailed analysis is often the reason for poorly addressing the task and developing a meaningful strategy to identify the large loopholes being used by HNWIs to avoid or evade taxes. An unfocused compliance strategy and bad results often disturb the social justification for collecting taxes from different taxpayer segments.

Identification of HNWI can be based on at least two approaches, one calculated on the basis of ranking, and the other based on predictive analysis. The first approach may be a simple compilation of a set of names by drawing on the increasing number of lists of wealthy individuals in newspapers and magazines, as well as manual enquiries and electronic intelligence. A simple definition that meaningfully classifies the wealthy and rich may need to be developed for the purpose. The parameters often used to define the set of wealthy individuals typically include income, real estate property, business ownership savings, capital owned including stocks and shares, bonds, inheritance, and the nature and extent of complexity in the overall portfolio and the role a tax adviser plays in managing it. Weights, thereafter, have to be assigned to each component. This could be done through a Wealth Conversion Weight (WCW); for example, employment – salary, self-employment, partnership – income should be given a low WCW compared to foreign income, profits from property, trusts and estates or savings. The weights could have a huge impact on the way in which the list is drawn up and hence, have to be selected judiciously. The modelling exercise must also be done by professionals as it calls for high analytical skills.

A second approach, based on predictive analysis, will rely on a variety of tools such as risk analysis based on risk rules, audit service cluster analysis, time series profiles, business trends related to the case at hand, detection of discrepancies and anomalies, and other analytical techniques.[457] A combination of the two approaches is often recommended to effectively penetrate the complex and global wealth structures that lie beyond the scope of simple searches.

The use of cluster analysis for segmenting taxpayers can be part of the analytical approach. In this approach, data mining techniques can be utilised to segment, and further sub-segment taxpayers and attach risk to each such sub-segment and to understand their behaviour and attitudinal patterns (mistake, lack of reasonable care to pay taxes or deliberate inaccuracy) or the reasons underlying the behaviour (inadequate records, deliberate suppression of information of income or difficulty in understanding the tax rules and obligations).[458] To analyse changes in compliance behaviour, social research methodology can be used. The objective should be to track behavioural responses

---

[456]Also see Section XI.5.g in Chapter XI of the TARC report.

[457] This has been discussed in detail in Chapter XIV of this report.

[458] See Parthasarathi Shome, Tax Shastra: Administrative Reforms in India, United Kingdom and Brazil, Business Standard Publishers, New Delhi, 2012

or voluntary compliance. This will help tax administration undertake suitable measures to improve customer services, depending on change in compliance.

iii) Measuring voluntary compliance

Improvement in tax administration seeks to secure maximum tax revenue effectively and efficiently given the tax rates. In an ideal situation, people would pay taxes they owe, and the tax administration would amount to no more than providing facilities for tax payments. But such a situation does not exist. Effective tax administration requires an environment in which citizens are induced to comply with tax laws voluntarily. An efficient tax administration also needs to detect and penalise non-compliance, and facilitate voluntary compliance through the provision of quality taxpayer service. Thus, taxpayers' education and service, collection, collation, storage, retrieval and verification of information, along with collection of taxes and grievance redressal systems create synergies for an efficient and effective tax administration. Some of these functions encourage voluntary compliance, and some are required for enforcing compliance. Given their synergistic role in building an efficient tax administration, it becomes very difficult to segregate one from the other; it also results in difficulty in measuring voluntary compliance. Voluntary compliance is one of the key elements in the Vision 2020 of both the CBDT and CBEC. It is thus important to measure voluntary compliance on a regular basis to know where the tax administration is in terms of realising the goal of Vision 2020. Research will have to be carried out so as to develop analytical methods to measure voluntary compliance. This research will provide inputs to develop long-term strategies, and also short- and medium-term programmes to achieve organisational goals.[459]

iv) Identifying audit/scrutiny risk elements

Both the CBDT and CBEC have traditionally focused on taxpayer audit, without being too selective about the cases for audit. The methods employed are often traditional, lack a scientific basis, and largely depend on individual perception. Of late, however, a beginning has been made in developing a risk matrix for the identification of cases – CASS (Computer-assisted Audit Selection System) in the CBDT and RMS (Risk Management System) in the CBEC for customs. However, this has come under criticism by both taxpayers and tax administrators on the grounds of gaps in efficiency and equity.[460] Nevertheless, it is necessary to develop up a risk management system since effective audit strategies cannot be laid out unless there is reliable risk identification, with emphasis on identifying aggregate risks as opposed to selection of individual cases. As depicted in Diagram 12.1 of Chapter XII of the TARC report, the first step to developing a robust risk management system is to identify the risks, evaluate them and prioritise them on the basis of the objective of the tax administration. The evaluation of risk elements is part of this step, as evaluations can be continuously used as feedback to improve the effectiveness of the audit system

[459] This has been discussed in detail in Chapter XII of the TARC report.

[460] For example, CASS is essentially unable to catch non-PAN activities in the tax net.

and of risk management. To be successful, the entire process needs to be based on sound measurement and carried out in an objective and scientific manner.

Risk elements can be identified using either top-down techniques such as macro-economic analysis or by bottom-up processes such as case-based risk assessment systems. Typically, tax administrations consider both approaches to develop an audit matrix. Commercial databases, along with the department's database often provide information about key financial ratios and sectoral variations. Such data and comparisons, using data mining techniques, are often used to identify risk as they provide potential comparables within sectors by indicating deviations from industry standards.[461] These comparisons can help develop a specific business model for different taxpayer segments, and a compliance risk model. This statistical model will identify historical relationships between compliance risk (e.g., audit results) and a set of predictor variables. But, the model will need to be further developed into a simulation model, refining the compliance risk model, so that the model is not based only on historical data but based on relatively more current data, and takes into account structural differences between companies to work out a predictive model to address future compliance risks. The simulation model will rely less on actual audit results and to a greater extent on proxies for compliance risk, such as changes in estimated tax liability.

For customs in the CBEC, the effectiveness of RMS and the performance measures to verify the effectiveness of RMS in terms of selection and detection rates for various techniques of selectivity can be part of the compliance risk model. A detailed statistical approach can help improve the existing RMS model. The simulation model can provide information on the optimality of the levels of examination, scrutiny and audit vis-à-vis the constraints of resources and the levels of compliance attained with that level of examination. Research can also be on the type of non-intrusive inspection systems and the adoption of new inspection methods for different types of prohibited goods.

Setting up the compliance model and simulation model with an ongoing scan of the changing global environment and its impact on the tax environment is a data-intensive exercise and will need continuous research to stay ahead in developing strategic interventions, including changes in tax laws. This level of research-based analysis will help the tax administration to identify significant risks for different taxpayer segments.

v) Impact analysis

Taxes need to be imposed more consensually to encourage voluntary compliance. But that, unfortunately, is not the way public revenues are raised in India. Little attention is given to one of the most fundamental drivers of the relationship between a state and its citizens. It was

[461] Using data mining techniques, tax administrations analyse data from hundreds of thousands of taxpayers to identify common attributes and then create profiles that represent different types of activity. They can, for example, create profiles of high-yield returns, so that tax administrators can concentrate resources on outliers with similar attributes. The technique, thus, leverages the data to understand, analyse and predict non-compliant taxpayers. Section XII.4.b of Chapter XII of the TARC report explains this in detail.

recommended in Chapter X of the TARC report that an impact assessment process that recognises the need for public proposals, comments, complaints and communication, and acts on the feedback thus received, will have to be engrained in the administrative system so that public consultations are mandatory before any changes in laws or procedures are brought about. This will enable increased participation of the general public in the process, and will help set up a different paradigm of governance.

The process involves critical research on the topics of change, so as to capture the complexity and comprehensiveness of the topics. The research inputs will have to be on the economic impact, impact on individuals and households, impact on businesses, impact on equality and/or impact on the tax administration's delivery mechanism, including the involvement of other departments/ ministries for delivery. Research input should not only be to assess impact, but also to throw up the kind of questions that need to be asked to each segment to get honest responses. The process will require continuous research, and the questions cannot be based on the one-size-fits-all assumption.

vi) Taxpayer surveys

Many tax administrations periodically conduct surveys to measure the time and money that taxpayers spend on pre-filing and filing activities. The survey data is used as an input to developing cost models, which is a micro-simulation model, having econometrically estimated relationships between compliance burden and the tax characteristics available from the tax returns of the taxpayers.[462] Typically, the objectives of the model are to assess the impact of programmes on taxpayer burden, to assess the role of burden in tax administration, and to improve services to taxpayers. The model also supports tax policy decision making through "what-if" type analysis, which allows the administration to understand better the effect of changing rules or laws or processes. The survey questionnaire often requires considerable thinking and research, and can be divided into sections representing pre-filing or filing activity and the time or money spent in completing tax returns.

Analysing survey data can be done either by outside agencies or in-house. In case it is done in-house within the tax administration, analytical methods will have to be developed with care so that the results are properly evaluated. If it is done by outside agencies, care will have to be taken to ensure that adequate expertise exists for monitoring to make sure that conclusions reflect survey responses. All these will require considerable research as well as expertise.

Taxpayer surveys of importers and exporters to assess trade costs, including the clearance costs at gateway ports in India and the measures to reduce these costs are important questions to encourage them to make India a part of the global supply chain, will be an important trade facilitation measure to improve on the Ease of Doing Business/Trading Across Borders and Logistics Performance

[462] The need for taxpayer surveys has been highlighted in Chapter II, and cost-benefit models have been discussed in detail in Chapter X of the TARC report.

indices.[463] The survey should also identify which of the trade facilitation measures have proven to be more effective, and which measure is relatively easy, quick and cost-effective to implement. This will also help the country in building better trade competitiveness.

vii) Trade and customs

India has entered into a number of trade agreements of varying comprehensiveness that focus largely on duty reduction. These agreements have also led to the development of a complex web of rules of origin. There is need to review the impact of these trade agreements on growth in trade and industry competitiveness. Many times, the agreements also provide a route for trade diversion instead of trade creation, which is the objective of the trade agreements. The extent of diversion of trade needs to be examined to check whether there are advantages in pursuing alternative approaches (for example, unilateral liberalisation). Another area of research in the above context can be on how to reduce the administrative burden of examining the misuse of rules of origin.[464] Surveys and research can also identify the locations of origin of the goods, which attract frequent misuse of rules of origin. Identification of ports can be part of that research. Since this is a dynamic exercise rather than a static one, the research cannot be a one-off exercise.

Since trade facilitation is an integral part of the work of the customs administration, regular research should be undertaken on the costs of clearance at gateway ports; this will include both time and money. A reduction in these costs will result in developing a better climate for doing business with India. The study will also help identify ports that are efficient and those which are not and need to be helped to improve efficiency. It may also be pointed out that clearance should not be taken in terms of only customs clearance, but must include other regulatory clearances in terms of a 'whole-government' approach, and must include the transportation and logistics sector. The study can reveal the preferred routes and modes of transportation. This is particularly important as trade agreements today are focused on building global value chains; hence, research has to be holistic and not one-dimensional. This research will be useful for the Commerce Ministry as well as the transportation ministries, such as rail, shipping, road and air transport. Such studies can be done in collaboration with these ministries.

viii) Taxing the hard-to-tax sectors

In India, as in other countries, the informal sector is quite extensive, heterogeneous and hard to tax. Nonetheless, experience from different countries suggests that these small taxpayers can be encouraged to pay a "single tax" or presumptive tax. In fact, many reports also suggest that the introduction of a single tax has helped tax administrations reduce tax evasion. In India, a scheme to levy Rs.1,400 from these groups was discontinued, as it was found that some of the normal taxpayers moved to the single tax regime and the segment never graduated to the regular taxpayer

[463] Chapter VIII of the TARC report discusses customs measures needed for trade facilitation.

[464] Chapter VIII of the TARC report deals with capacity building in customs and discusses the issues connected with trade and customs in detail.

segment. These observations, to a large extent, can be met by improving the design of the single-tax system and keeping a watch on this hard-to-tax sector. This requires research and analysis to identify such taxpayers by gathering information about their income from third parties and designing attributes to closely target this taxpayer segment. Understanding taxpayer behaviour along with how taxes affect them can also be a critical area of research. Such research inputs will provide more in-depth understanding on how tax payments impact taxpayers – whether this allows them to move towards a formal economy – and help the tax administration arrive at a more appropriate implementation plan with possibly better outcomes.

ix) Tax effects on investment

Use of tax incentives – such as tax holidays, partial profit exemptions, free trade zones, etc. – through income tax laws or indirect tax laws is often considered a cost-effective way of overcoming investment impediments. Revenues foregone because of these tax incentives often tend to exceed revenue costs expected before the tax concession is put in place by a wide margin. Examining the effect of taxes on investment is clearly an important question. Traditionally, the question has been analysed with aggregate data focusing on time-series changes in tax rates or tax regimes. But, it is also important to consider whether and to what extent investment varies with taxes. Economists generally use what they call a "marginal effective tax rate", which is based on tax codes and measures the wedge between the pre-and post-tax return on the marginal investment project, taking into account the return on investment, depreciation, investment tax credits, inflation, and the discounting of cash taxes. Their concern is with the tax effect on the marginal investment. In contrast, accountants and finance researchers are often interested in the tax effect on the marginal dollar of income and use the "marginal tax rate", which is the present value of tax on an additional dollar of income after taking into account the carry back provisions of net operating losses, and is a firm-specific rate. Recently, there has been a move in economics to more cross-firm and cross-asset type studies to avoid endemic problems plaguing aggregate time-series analysis. These are created by the fact that a change in tax rates is endogenous, meaning that a change in the tax rate in response to some macroeconomic factors could also affect investment, and that it is difficult to control for contemporaneous non-tax shocks that affect investment.

Geographically-targeted tax incentives remain popular initiatives in Indian tax laws to tackle stagnant or deteriorating economic conditions in specific areas. An estimate of the impact of various tax incentives on a number of dimensions of local economic growth, therefore, will be important. Econometric analysis can be used to find out the growth outcomes of gross investment flows, separately accounted for by new, existing, and vanishing businesses in the target areas. The results can offer empirical evidence to support a number of specific policy recommendations and can reveal whether the tax incentives have a more dynamic outcome on growth than projected. Such studies can provide ex-ante analysis of tax laws, and if the results are positive from a policy perspective, the revenue foregone can be justified. Otherwise, it might be better to focus on actual impediments to investment and address these directly.

x) Tax incentives for R & D

Research and development (R & D) has not received the desired emphasis so far. Some incentive has been given, but without much focus. But with the focus of the government being on improving the manufacturing climate in the country, there is a strong case to have a relook at the present tax incentive for R&D. R&D efforts will need to be expanded to focus on multi-dimensional initiatives such as developing an innovation culture, engaging prospective participants, knowledge management and compliance assurance, and understanding the transitions and boundaries of new R&D tax incentives. The tax regime thus would have to be clear, comprehensive, and consistent between sectors and, over time, relevant to the needs of different sectors and aligned with policy intentions and tax compliance requirements. The role of the tax administration may also need to be redefined and moved towards facilitating meaningful engagement between industry, the two tax departments, taxpayer companies, and their tax advisors. Tax compliance, in this context, will need to be a two-way journey. Companies need to adopt practices to be 'compliance-ready' and tax administrators need to undertake quality compliance assurance to ensure the programme's ongoing integrity. All these questions and the guidance processes will need detailed research input to support compliance work, while at the same time encouraging taxpayers, new and existing, to undertake R&D.

xi) **Taxing natural resources**

In many countries, there is a separate taxation mechanism for natural resources, such as oil, gas, coal and minerals. Income tax laws, at present, treat resource industries more favourably than most other industries – through favourable treatment of such capital expenses as depletion, exploration and development, and the cost of acquiring resource properties. Properly designed income taxes can attempt to include capital income on a uniform basis. The case for special resource taxes is to tax resource rents over and above the levies implicit in general income taxes. This can have two justifications: (a) the efficiency-based argument that a tax on resource rents is non-distorting and complementary; and (b) the equity argument that the property rights to resources ought to accrue to the public at large rather than to private citizens since the rents represent the bounty nature has bestowed on the economy rather than a reward for economic effort. If the main purpose of a resource tax is to capture rents for the public sector, it can be argued that the base of resource taxes should be economic rents (or their present value equivalent).

xii) **Taxation of the insurance sector**

Taxation of insurance companies is at present governed by Section 44 of the Income Tax Act, 1961. Under this statute, the tax on the income of an insurance company is the aggregate of the amount of income tax calculated on the amount of profits and gains of the insurance business included in the total income at the rate of twelve and one-half per cent, and the amount of income tax with which the assessee would have been chargeable had the total income of the assessee been reduced by the amount of profits and gains of the life insurance business. Section 115B of the Income Tax Act, 1961 and Rule 2 of the First Schedule mentions that the profits and gains of the life insurance business is to be taken to be the annual average of the surplus arrived at by adjusting the surplus or deficit disclosed by the actuarial valuation made in accordance with the Insurance Act, 1938, in respect of the last inter-valuation period ending before the commencement of the assessment year so as to exclude from it any surplus or deficit made in any earlier inter-valuation period.

These tax legislations were framed for a particular environment and the legislation does not address the issues that have emerged in the changed situation. Significant changes were brought about under the Insurance Regulatory and Development Authority of India (IRDA) Act, 1999. New accounting formats were introduced. Two separate accounts were prescribed: policyholders' account (PHA) known as the revenue account (a technical account), and shareholders' account (SHA) known as profit and loss account (non-technical account). A new format for the valuation balance sheet (i.e. Form I) was introduced. Rule 2 is not in sync with these changes for life insurance busniess. These changes have made the term 'annual average' referred to in Rule 2 redundant since actuarial valuation is carried out on a yearly basis by insurance companies. Further, 'actuarial valuation' has not been defined in the I-T Act.

All these have led to non-uniform assessment practices by the Assessing Officers and to lack of clarity or uniformity in approach by taxpayers in computing taxable profits as per Rule 2 of the First Schedule. Most modern tax administrations, on the other hand, are introducing actuarial methods to determine the reserves required, and bringing tax laws in sync with insurance regulation. A similar effort should be made by the CBDT, which should not set the required reserve level on the basis of figures set many years ago. These reserves may not bear an exact relation to the reserves that an insurance company would need to meet its obligations. Its obligations may have dynamic characteristics, which need to be recognised.

xiii) **Performance management**

A robust performance management system (PMS), characterised by active and continuous engagement between the organisation and its employees, plays a significant role in building the skills needed by staff to perform their tasks efficiently and to build their competency. This includes active employee engagement, communicating the expectations of the organisation through the goals set for a function, creating a performance measurement system, specifying the rigour with which performance appraisals can be conducted, and setting up an incentive system that recognises an employee's contribution.[465]A PMS cannot just concentrate on measuring and managing individual effectiveness, but will also have to pay attention to the effectiveness of teams and to the alignment of individual objectives to team goals. If PMS is not actively and continuously tracked and evaluated, it will remain an unimplemented concept.

It is important to recognise that a performance management system is an important element for quality management. Supervising authorities need to direct the efforts of an organisation or a group

[465] For details, see Section IV.3.d of Chapter IV of the TARC report.

to know how, when, and where to institute a wide range of changes. These changes cannot be implemented sensibly unless they are based on appropriate information; hence, research is essential for continuous improvement in PMS.

***xiv)* Technology scan in SPV/DG (Systems)**

Recognising that technology is constantly changing and is an important mechanism to improve tax administration, the TARC in Chapter VII of its first report stated that the way to bridge the gap between present reality and a potential future state cannot be simply to "catch up" through an incremental approach; rather, there is need to have a strategic approach aimed at being ahead in the field of technology.[466] The TARC further stated that ICT, instead of being seen as merely a support function, must be seen as a key lever that enables the achievement of the tax administration's goal to maximise compliance and minimise the tax gap in a fair, transparent and customer focused way. It is, therefore, imperative to take note of the digital technology universe and its evolution to exploit the potential of ICT to transform business processes. The adoption of new technologies will need to be carried out in an organised manner after due impact analyses and with adequate time allowed for systems to be changed before they are given effect. This will require a clear road-map with well delineated milestones as well as regular and periodic research on technology so that a reliable, comprehensive ICT system, which covers all processes and is user friendly (both to the staff and to taxpayers) can be built.

***xv)* Fairness in tax disputes**

One of the key elements in improving the tax climate is to improve the dispute resolution mechanism. A certain level of tax disputes is a normal part of a system of taxation based on the rule of law. Tax disputes are not unique to India, but a large number and frequent disputes, may be on the same issue, have an adverse impact on how taxpayers see the Indian tax administration.[467] This needs quick correction. The TARC had made a number of recommendations in Chapter V of its report. These recommendations and observations need to be analysed, and can prepare the background to developing a range of approaches to reduce the number of cases that go to court, based on background research and analysis. The new mechanisms for dispute resolution will have to take into account the needs of the tax administration, its current practices, and the legal framework for taxation. Hasty and undiscussed mechanisms may prove to be an impediment rather than a help. This may require detailed discussion with taxpayers, lawyers, chartered accountants and other stakeholders. All these require detailed research inputs so that the discussions are meaningful and informed.

[466]See page 350 of the TARC report.

[467] In assessing whether there is a problem with excessive disputes, attention should also be paid to whether particular subsets of disputed cases are characterised by a common contentious issue.

xvi) Corruption in tax administration

Corruption in tax administration has attracted considerable attention from both the Boards. A number of measures have also been undertaken to reduce it, notable among these being the introduction of ICT in different processes.[468] The TARC also gave its full attention and focus on this malaise, given its wide societal and government public management implications, and made important recommendations covering the taxpayer, staff, structure and processes dimensions in Chapters II, IV, VI, VIII and XII of its report to address several aspects of the issue at all levels of the tax administration. But given the dynamic nature of the problem, the endeavour cannot be considered to be only this and needs to be carried out with a longer time horizon. The tax administration must continually undergo change to accommodate the dynamic nature of the problem. It can be highly dangerous if the public comes to doubt the integrity of tax personnel, as this directly influences the taxpayer's willingness to pay taxes.

Corruption affects and distorts what should be arms' length, or objective and unbiased, relationships between government officials and private sector individuals. Some individuals succeed in getting favourable treatment in their economic activities from public officials by paying bribes. Such treatment often reduces the cost of economic activities in which these individuals are engaged; it also creates new opportunities for them that are not available to others. Acts of corruption may be initiated either by private individuals or by government officials. When it is the latter, they may offer favourable treatment in exchange for a bribe or some other favour. When this happens, corruption disrupts the competitive situation that exists in the market and may give a competitive advantage to favoured individuals or enterprises.

In many situations, new rules and regulations may complicate the existing one without significant advantage to the tax administration, and become instruments used to extract bribes from individuals. The asymmetry of information between officials and taxpayers, and the multiplicity of rules and their frequent changes often make it impossible or very costly for taxpayers to ascertain whether the regulation makes management better for both the taxpayer and for the tax administration. It may also be the case that new tax procedures are used as a tool to harass and extract bribes because taxpayers are not adequately informed about them.

It, therefore, is imperative that regular and periodic research be carried out to identify procedures that give rise to corruption. There is also need for concurrent taxpayer surveys to identify which rule or regulation gives rise to corrupt practices. Surveys could also include a measure for customer satisfaction and user feedback. Clear procedures setting out the circumstances under which suspicious payments may be reported to other law enforcement agencies or authorities should be provided to the taxpayers. Measuring staff satisfaction will be part of the same exercise as dissatisfied staff is likely to be more corrupt than happy staff.

---

[468] World Bank Policy Research Series paper 6988 also suggests that tax simplification lowers corruption.

XV.4.b External research collaboration

Most advanced tax administrations access and engage with, collaboratively or as partners, a broad base of external research organisations for short-term or long-term research projects. This is in addition to in-house research on topics that need to be worked on only within the tax administration. The value of using external research institutions comes from the possibility of accessing superior professional knowledge and research techniques. This, when combined with the tax officer's familiarity with data and its interpretation, results in quality tax analysis and research for the tax administration. However, for such collaborative research to be successful, a well-defined system to manage such alliances needs to be put in place. Hence, an arrangement based on long-term collaboration and partnership will have to be put in place by the CBDT and CBEC for research and analysis on topics requiring external assistance. This will also require making data available to external agencies. Encrypting and anonymising taxpayer data and external data or third-party data, and establishing a data bank would help. The collaboration will need to also include a framework for dissemination of research results, as external researchers and academics will be interested in sharing the research results and getting them published in refereed accredited journals. A framework, which includes data sharing, examination of research results, publishing the research and other rights and duties, will have to be clear spelt out and be made part of the collaboration and partnership agreement for external research. Such a framework will not only be useful to the tax administration but will also be of significant advantage to external research organisations as they can leverage their research investment from seed-funding that this engagement will bring to more substantial opportunities.

One dimension that will need special attention in collaborations with external research institutions is the high expectation that the tax administration may have about the nature and quality of research outputs or deliverables. The expectation of the tax administration will be that the deliverables to be provided and communicated be such that they create value for their work and can be used as it is, without being reworked. This may need to be guarded against. For an effective collaboration and to understand each other's expectations, it will be important for research or academic institutions to clearly document the proposed activities and the expected outcomes and discuss them with tax administration so that there is no gap between what was expected and what was delivered. If there are any risks that could affect project outcomes, these should be discussed in detail between the two parties to enable them to develop a sustained partnership. Often, lack of clear understanding on the time-frame for research proves to be a major impediment in developing sustainable partnerships.

It will also be important to point out that research collaboration should be demand-driven and there should be clear organisational intent on the part of the tax administration to requisition the research input with full ownership of the demand. It should not be the case that the research is given out on the basis of one person's need. In such cases, the tax administration does not know how to utilise the research input once it has been received, often leading to its being junked without explanation. Organisational ownership will also mean that research projects are well monitored and are funded in a timely manner. Besides, the tax administration can analyse the results and provide its inputs after the preliminary results are available to make the research output meaningful and usable.

It will also be important that research collaborations are made not only with reputed national organisations like the NIPFP, NCAER, NIFM, ICRIER, etc., but are more broad-based so that regional institutes are also included. This will widen the pool of research institutes available for collaboration. Although geographical distribution cannot be the sole basis for enlarging the base, it needs to be taken into account to ensure that the capacity of these institutes also gets upgraded over a period of time. In this context, it may be mentioned that the TARC had recommended in Chapter III that research should be carried out not only at the central or apex level but also at the operational levels. Each vertical such as taxpayer services, dispute management, people management, etc., were recommended to carry out their own research – whether conducted in-house or through external entities – to improve delivery on assigned tasks. It was also recommended that the Knowledge, Analysis and Intelligence (KAI) centre and Tax Policy and Analysis (TPA) unit and the Independent Evaluation Office (IEO) conduct the more strategic research, having pan-organisational implications and requiring legislative or procedural change. A vast pool of research collaborators is in line with these recommendations.

Regular external research is normally in areas requiring taxpayer surveys, an evaluation of the effectiveness of the tax administration's policies, ways to improve the policy and its administration, and identification of the lessons learnt – areas in which the tax administration does not have any expertise. These survey-based researches require large manpower and are time-consuming. Often, they require engaging psychologists to frame the right questions so that the taxpayer does not feel that they are being probed. Besides, it is also possible that questions posed by the tax department may elicit a response different from the one a taxpayer would give to an external agency; it may also not always be the correct response. An example of customer surveys can be compliance perception surveys of taxpayers to measure perceptions of tax compliance among individuals and small and medium enterprises over time. In this, data on taxpayers' perceptions of the prevalence and acceptability of evasion, fairness of treatment, and the likelihood and consequences of being caught evading taxes can be collected. Based on this, the compliance gap can be measured. This type of survey will also help in building a deeper understanding of the drivers of taxpayer behaviour and how these affect the tax gap. Other connected areas of work by external research agencies could be measuring compliance costs,[469] change in compliance cost in response to the tax administration's actions, developing knowledge on the taxpayers' attitudes towards the tax administration, and exploring and measuring taxpayer perception and experience with the tax administration's services. An evaluation of survey results can help the tax administration effect operational changes and issue guidance notes to tax officers and staff. If such research is carried out on a regular basis, say with an interval of three to four years, it could help measure the change in compliance cost over time.

[469] The CBDT has commissioned a report from ICRIER on compliance cost.

XV.4.c Dissemination of research

Dissemination of research and associated knowledge management can play an important role in building the intellectual capital to improve the effectiveness of the tax administration. It will be important that the research output is suitably informed, so that officers and staff have access to the knowledge and information needed to perform their tasks with greater efficiency and effectiveness. The availability of knowledge and information on one designated portal, which is a norm in most advanced tax administrations, will result in better work performance by officers and an improvement in the quality of taxpayer services.

As part of research dissemination, the TPA (Tax Policy and Analysis) unit can bring out annual or biannual reports on revenue trends and forecasts and monthly supplements to such forecasts; the KAI centre could do so on the operational needs of the tax departments. These research outputs, on different aspects of tax policy and governance, should be brought out as a periodical – quarterly or half-yearly or annual – in the form of a compendium of research articles, studies and surveys. Publishing taxation research in a periodical and making them available to the public will have the benefits of publicising the work, crediting the officer who did the research, and promoting quality research. It is essential that the process of including a research article in the journal is transparent and merit-based. The NADT and NACEN, training institutions within the tax administration, can host the activities and publish research journals.

The TPA unit and KAI centre, along with NADT and NACEN, can also conduct seminars and conferences and bring out reports on the deliberations in those seminars. These conferences and seminars will provide officers an opportunity to interact with tax experts from the private sector and academia. This will provide a forum for two-way analysis on topics of tax policy, administration and law.

XV.4.d Human resources for research

The TARC in Chapter III of its report has discussed the staff requirement for the KAI centre, the TPA unit and in field formations for carrying out research work. The TARC recommended that to begin with, 60 statisticians and operational research specialists should be placed in the TPL and in the TRU when the two Boards are separate. The numbers can be augmented to a consolidated staff strength of 400 when the two Boards converge to have a single TPA unit. It was also recommended that the TPA unit could engage experts from outside, from universities and other research institutions for specific studies at hand. Besides, the TARC had recommended in Chapter III that the KAI centre (a shared service between the two Boards) should be the hub of analytical activity, intended to release the huge potential for exploiting the value lying in the rich data that the Boards hold. It was recommended that KAI staff should develop data-analytics capabilities so that the tax administration can increasingly undertake detailed analytics to develop breakthrough insights into the tax administration's operations, which will be helpful in improving processes and changing laws for delivery gain to the tax administration.

Since both the TPA and KAI centre will undertake high level research, people posted in these units need to have strong data handling skills and skills needed to carry out detailed analysis. Such research needs to adopt a multi-disciplinary approach on topics such as tax compliance and tax evasion. Since public finance and economic theories drive much of the work in the field of tax, other than tax laws, those disciplines naturally influence the work and research of the tax administration. People working in the TPA and KAI and involved in research, therefore, should have university-level education in economics, psychology, statistics, management, or law, with adequate experience in public policy formulation. But this in no way suggests that those with other educational backgrounds cannot produce useful research work. In fact, their inputs will also be of immense value given the multi-dimensional nature of work in the tax administration. It is also the case that officers often acquire the requisite knowledge while on the job; they can also be engaged in research work if they have proven ability to do so. Thus, the induction of officers and staff to undertake research work should be flexible, but due care needs to be taken regarding the capability of the officers and staff to handle research work or research projects. They need to be cognizant of developments in other tax administrations, so that these can be examined, and if found suitable, adapted and used. People involved in research will also need to have access to reputed journals. Regular scanning of journals and other such material can be useful to improve research inputs for improving tax governance.

In this regard, the Chief Economist in both the TPL and TRU will have to play an important role in identifying, guiding, monitoring, and evaluating research topics. He will have to provide quality assurance of the tax administration's analytical work. He should be responsible for presenting the analysis, examining evidence, and using analytical input to support the formulation of tax policy change and for internal administrative decisions. Importantly, he should be signing off on these inputs so that quality assurance is maintained.[470]

Research work requires certain critical skills. These skills, as already stated, should be in the fields of data analysis, policy analysis, and tax laws analysis. While some officers acquire the needed skills on the job, there is no focused training to develop these skills at present. Such training can be delivered either in-house at NADT or NACEN or by reputed research institutions. NADT and NACEN can also call for faculty from reputed institutions to train officers on research methodologies, and give practical training on data analysis. Participation in technical seminars and events in specialised areas will also add to the knowledge base of officers. They can be deputed to attend research seminars or courses, whether in India or abroad, for which adequate funds should be made available. Officers attending research seminars or courses should be required to present their learning to others so that knowledge acquired in these seminars or courses are disseminated to a larger base of officers. It will also be important to provide suitable training to staff, as they provide the backbone to this knowledge creation. A proper training policy should be drawn up to

[470]The Chief Economist in the HMRC signs off on all regulatory impact assessment statements used for budgetary work before these are presented for law making.

do so. Regular and continuous interaction with research institutions should also be encouraged for officers involved in research work, as that is another sure way to enhance the skills of the officers.

All IRS officers must publish at least one peer-reviewed research paper on topics of tax administration or tax policy before promotion to the grade of Commissioner. This will help develop research capability among officers, and prepare them for jobs requiring more analytical capabilities. These papers can then be made available with their name on the knowledge portal. A mechanism to incentivise quality research should also be put in place. This could be in the form of providing incentives such as a one-time award or an additional increment for research papers published in research journals of national or international repute.

XV.4.e Allocation of funds for research

To give adequate emphasis and focus to research in tax administration, sufficient funds should be provided in the budget under a separate head. To sustain the focus and to establish continuity, no re-appropriation of funds should be allowed from this head of research.

Participation in technical seminars and events in specialised areas will also add to the knowledge base of officers. Officers can be considered to be deputed to attend research seminars or courses, whether in India or abroad. Adequate funds will need to be allocated for this purpose to ensure that it is done on a regular basis.

XV.5 Summing up

The importance and challenges of networking and knowledge sharing to develop research inputs to improve tax governance is not an easy task. Many with a traditional mindset in the tax department may not consider developing research as being part of the core work of the tax administration. Traditional mindsets, however, need to be transformed to develop processes, taking into account human, organisational and institutional considerations, which encourage research to provide the inputs needed to improve tax administration. The effort should be supplemented, in a constructive manner, by accessing information and expertise existing elsewhere with external research institutions, to support research work. To encourage research, apex level supervisory officers need to understand organisational, sociological and economic needs and to recognise that investment in developing research skills is a fundamental management requirement. This will require that apex level supervisory officers recognise organisational, sociological, and economic needs and to recognise that investment in developing the requisite research skills is a fundamental management need if they are to reap the benefits that meaningful and detailed research would eventually offer. This will help the tax administration improve its ability to react to uncertainty and complexity, and build a resource for professional and organisational innovation.

Research, if carried out jointly by the CBDT and CBEC, will help the tax administration define and solve problems together. This will improve the joint working, so intensely desired, and help them co-ordinate their taxpayer programmes, policies, and services, besides improving both ICT infrastructure and information content.

XV.6 Recommendations

The TARC recommends the following:

i) Role of research in improving tax governance

a) The requirements of the tax administration are not static; they have a dynamic character, requiring constant evaluation and assessment to enable the tax administration to seamlessly modernise itself and look into its future needs. These demands require continuous, on-going research in tax governance so that there is sufficient and modern thinking available to improve processes, structures, and people functions in the tax administration, leading to better tax governance. (Section XV.1)

b) Many with a traditional mindset in the tax department may not consider developing research as being part of the core work of the tax administration. Traditional mindsets, however, need to be transformed to develop processes, taking into account human, organisational and institutional considerations, which encourage research to provide the inputs needed to improve tax administration. (Section XV.5)

c) Research in tax administration needs to include international comparisons by identifying good practices adopted by different tax administrations, understanding them and drawing lessons from them to raise the standards of tax governance. (Section XV.1)

d) The Indian tax administration is attempting to enhance delivery. Research in tax governance, with a multi-disciplinary approach involving disciplines such as economics, accounting, finance, law, management, behavioural science, ICT, and statistics, can help review current practices and design new approaches to expand the tax base and increase tax revenue, while remaining cognizant of the government's aim to promote economic development and tax justice. (Section XV.1)

e) Research needs to be evidence-based and needs to be meticulously designed, implemented, and executed so that the output is relevant to and usable to the tax administration. This will require an organisational framework, finance, staff, the development of linkages for collaboration – internal-governmental mechanism or with outside agencies – and the building of skills needed to evaluate outputs to assess their practical utility. (Section XV.1)

f) Research cannot be a one-off exercise; it has to be embedded in the tax administration so that there is two-way movement, top-down as well as bottom-up, within the various tax departments to build an ecosystem to undertake meaningful research. (Section XV.1)

ii) Areas of research

g) Important areas of research, only indicative by enumeration and by no means exhaustive, can be as follows:

- Compliance tracking
- Identifying the rich and wealthy
- Measuring voluntary compliance
- Identifying audit/scrutiny risk elements
- Impact analysis
- Taxpayer surveys
- Trade and customs
- Taxing hard-to-tax sectors
- Tax effects on investment
- Tax incentives for R & D
- Taxing natural resources
- Taxation of the insurance sector
- Performance management
- Technology scan in SPV/DG (Systems)
- Fairness in tax disputes
- Corruption in tax administration (Section XV.4.a)

h) Some of the research work can be carried out jointly by the CBDT and CBEC. (Section XV.4.a)

iii) External research collaboration

i) Research work, on short-term or long-term basis, can be carried out by external research organisations. This will be in addition to in-house research on topics that need to be worked on only within the tax administration. (Section XV.4.b)

j) The CBDT and CBEC will have to put in place an arrangement for partnering with external research organisations on a long-term basis for research and analysis on topics requiring external assistance. (Section XV.4.b)

k) This will also require making available data. Encrypting and anonymising taxpayer data and external or third-party data, and establishing a data bank will help make data available to external researchers. (Section XV.4.b)

l) Collaboration with external research organisations will need to include a framework for dissemination of research results, as external researchers and academics will be interested in sharing research results and getting them published in refereed accredited journals. (Section XV.4.b)

m) Partnership and collaboration with external research organisation will not only be useful to the tax administration, but will also be of significant advantage to external research organisations as they can leverage their research investment from seed-funding that this engagement will bring to more substantial opportunities. (Section XV.4.b)

n) The expectations of the tax administration on the deliverables by external research organisations will have to be clearly spelt out and communicated so that they are not misunderstood. (Section XV.4.b)

o) Research collaboration should be demand-driven and there should be clear organisational intent on the part of the tax administration to requisition the research input, with full ownership of the demand. (Section XV.4.b)

p) It is also important that research collaborations are made not only with reputed national organisations, like NIPFP, NCAER, NIFM, ICRIER, etc but are made broad-based to include other regional institutes so that there is a larger pool of research institutes available for collaboration. Although geographical distribution cannot be the sole basis for enlarging the base, it needs to be taken into account to ensure that the capacity of these institutes also get upgraded over a period of time. (Section XV.4.b)

q) The tax administration should consider engaging external research organisations for taxpayer surveys, evaluation of the effectiveness of the tax administration's policies, suggestions on ways to improve policy and administration and identifying lessons learnt**.** Apart from the fact that such research requires large manpower and are time-consuming, these are also areas in which the tax administration does not have any expertise. (Section XV.4.b)

iv) Dissemination of research

r) Dissemination of research and the associated knowledge management can play an important role in building intellectual capital to improve the effectiveness of the tax administration. (Section XV.4.c)

s) As part of research dissemination, the TPA (Tax Policy and Analysis) unit can bring out annual or biannual reports on revenue trends and forecasts and monthly supplements to such forecasts; the KAI centre could do the same on the operational needs of the tax departments. (Section XV.4.c)

t) The TPA unit and KAI centre, along with NADT and NACEN, can also conduct seminars and conferences and bring out reports on the deliberations in those seminars. These conferences and seminars will provide officers an opportunity to interact with tax experts from the private sector and academia. (Section XV.4.c) *v) Human resources for research*

u) In line with the recommendations in Chapter III of the TARC report, both TPA and KAI centre should undertake high level research. People posted in these units need to have highly developed skills in data handling and carrying out detailed analysis. Research work undertaken needs to adopt a multi-disciplinary approach. (Section XV.4.d)

v) People working and involved in research in the TPA and KAI centre should have university-level education in the field of economics, psychology, statistics, management, or law, with adequate experience in public policy formulation. But this in no way suggests that those with other educational backgrounds cannot work in the TPA and KAI. (Section XV.4.d)

w) The induction of officers and staff to undertake research work should be flexible, but due care needs to be taken to ensure that the officers and staff selected have the capability to handle research work or research projects. (Section XV.4.d)

x) The Chief Economist in both the TPL and TRU should play an important role in identifying, guiding, monitoring and evaluating research topics. He will have to provide quality assurance of the tax administration's analytical work. He should be responsible for presenting the analysis, examining the evidence, and using analytical input to support the formulation of tax policy change and for internal administrative decisions. (Section XV.4.d)

y) There should be intense training, either in-house at NADT or NACEN or at reputed research institutions, of officers involved in research. They should be trained on research methodologies, and be given practical training on data analysis. (Section XV.4.d)

z) All IRS officers must publish at least one peer-reviewed research paper on topics of tax administration or tax policy before promotion to the grade of Commissioner. (Section XV.4.d)

aa) If the research paper is published in research journals of national or international repute, the officer should be incentivised, either through a one-time award or by allowing one extra increment, so as to encourage more people to do the same. (Section XV.4.d)

bb) The attention of apex level supervisory officers is the key to developing human resources for research input. (Section XV.5)

vi) Allocation of funds for research

cc) To give adequate emphasis and focus to research in tax administration, sufficient funds should be provided in the budget under a separate head. To sustain the focus and to establish continuity no re-appropriation of funds should be allowed from this head of research. (Section XV.4.e)

dd) Participation in technical seminars and events in specialised areas will also add to the knowledge base of officers. Officers can be considered to be deputed to attend research seminars or courses, whether in India or abroad. Adequate funds will need to be allocated for this purpose to ensure that it is done on a regular basis. (Section XV.4.e)

APPENDICES

Chapter XIII

Revenue Forecasting

Appendix XIII.1

Calculating tax elasticity

The proportional adjustment method is used to make adjusted revenue series (*AT*) that reflect the revenue collection if the last previous fiscal year's tax structure was applied to earlier years. A formal mathematical presentation for calculating the adjusted revenue series is as follows:

$$AT_t \quad = \quad T_t \qquad \text{for } t = n$$

$$AT_t \quad = \quad T_t \times \prod_{i=t}^{n} \frac{T_i}{(T_i - DC_i)} \qquad \text{for } t = 1, \ldots, n\text{–}1$$

where T_t : Unadjusted (observed) tax revenue at current year t

AT_t : Adjusted tax revenue at current year t

DC_i : Revenue impact of discretionary changes at year i

t : Current year

n :The final year

The term $T_i/(T_i - DC_i)$ is called the coefficient of change. The symbol $\prod_{i=t}^{n}$ indicates cumulative multiplication of the coefficient of change for the current year t to the last year for which the data is available – the year preceding the forecast year. Assuming the final fiscal year is 2014, the application of above relationships can be shown as follows:

$$AT_{2014} = T_{2014}$$

$$AT_{2013} = T_{2013} \times \frac{T_{2014}}{T_{2014} - DC_{2014}}$$

$$AT_{2012} = T_{2012} \times \frac{T_{2013}}{T_{2013} - DC_{2013}} \times \frac{T_{2014}}{T_{2014} - DC_{2014}}$$

and so on ……

The following steps can be followed to estimate the tax elasticity and project tax revenue based on the estimated tax elasticity:

Step 1: Calculate the real values of GDP, tax revenue, and revenue impact of discretionary changes.

$$RealValue = NominalValue \times \frac{100}{Deflator}$$

Normally, the GDP deflator is used as the *deflator*. In other cases, the appropriate deflator, if available, needs to be used; for example, the import price index can be used to calculate real imports and real revenues from import duties.

Step 2: Calculate the coefficient of change for each fiscal year.

Step 3: Calculate the cumulative coefficient of change for each fiscal year.

Step 4: Calculate the adjusted tax revenue series.

Step 5: Calculate the log of real GDP and adjusted tax revenue. Log transformation is needed to estimate the tax elasticity using regression.

Step 6: Regression analysis is used with log real revenue as the dependent variable, and log real GDP as the independent variable. The coefficient of real GDP is the estimated tax elasticity, using the following equation:

$$Log(AdjRealRev) = \varepsilon \cdot Log(\mathrm{RealGDP}) + \mathrm{Constant}$$

From the definition of tax elasticity, the percentage change in tax revenue can be calculated by:

$$\%\Delta AT = \varepsilon \times \%\Delta Y$$

Thus, the tax revenue in year t+1can be forecast based on the revenue in year t by:

$$TaxRev_{t+1} = TaxRev_t \times (1 + \%\Delta AT)$$

$$TaxRev_{t+1} = \mathrm{TaxRev}_t \times (1 + \varepsilon \times \%\Delta Y_{t+1})$$

The above relationship gives the projected revenue in year t+1 at the price level of year t. To give the projected nominal revenue at the price level of year t+1, the above equation will have to be adjusted with the expected inflation in year t+1:

$$TaxRev_{t+1} = TaxRev_t \times (1 + E \times \%\Delta Y_{t+1})(1 + InflationRate_{t+1})$$

Use of Dummy variables

A dummy variable indicates a proxy variable to show the absence or presence of particular effects that may be expected to shift the outcome. It takes the binary value – 0 or 1; 0 for the absence of the effect and 1 for the presence of the effect. Suppose tax revenue over a number of years follows a straight line,[1] and makes a shift due to certain policy changes, such as the computerisation of tax returns or a tax amnesty scheme, etc. This can be shown by formulating a common-slope model given as

$$TR_i = \alpha + \beta\ GDP_i + \gamma\ D_i + \varepsilon$$

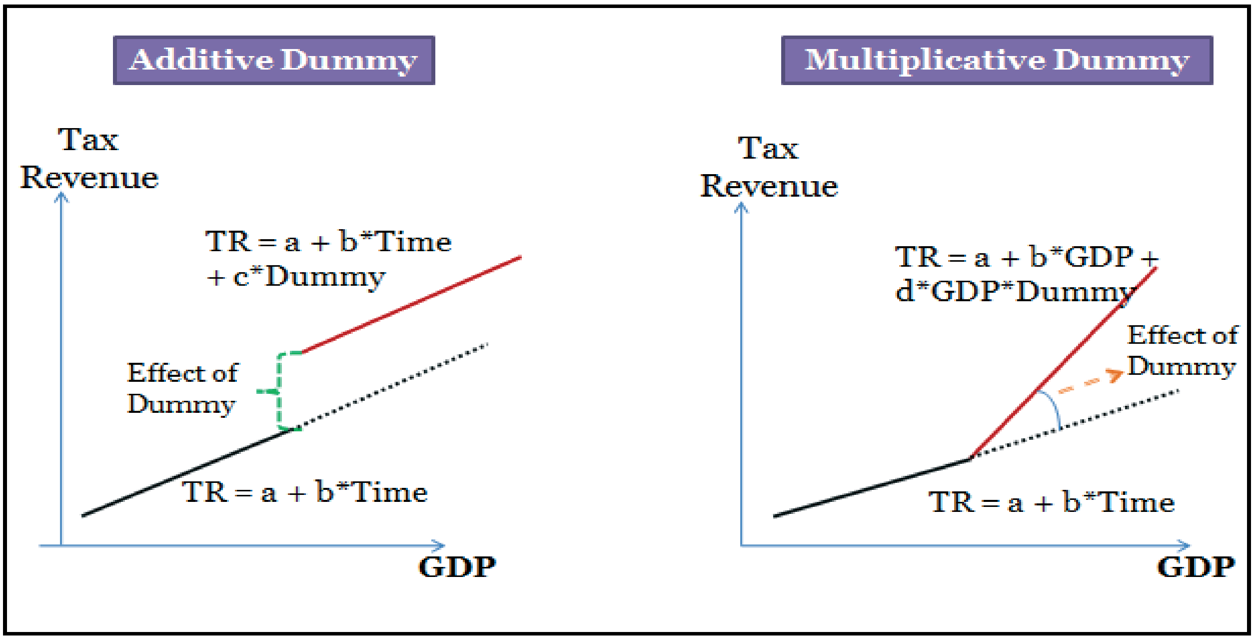
where D is called a dummy or indicator variable, and is 0 before the event and 1 after the event. If expanded, the two periods will be shown as

$$TR_i = \alpha + \beta\ GDP_i + \varepsilon, \qquad D = 0$$

$$TR_i = \alpha + \beta\ GDP_i + \gamma + \varepsilon, \quad D = 1$$

Such a relationship can be depicted in two forms – either tax revenues takes a quantum jump but there is no change in buoyancy or elasticity or there is a change in buoyancy or elasticity. The first one is called an additive effect; the dummy is used in an additive manner and in the second one, the use of dummy is called multiplicative dummy. These relationships are shown below in Diagram 13A.1.

Diagram 13A.1: Graphical depiction of dummy variables



[1] If that relationship is not a straight line, it can be assumed to be a straight line in first approximation.

The above two situations can be shown through the following two equations:

Additive Dummy: TR = a + b*GDP + c*Dummy

Multiplicative Dummy, TR = a+ b*GDP + d*GDP*Dummy

So, when the dummy = 0, then TR = a + b*GDP, for both the above equations, which is the case before the change.

When dummy = 1, then for the additive dummy, the equation becomes, TR = (a + c) + b*GDP. This is depicted in the first part of the Diagram 13A.1.

For multiplicative dummy, the equation becomes, TR = a + (b + d)*GDP. This is depicted in the second part of the Diagram 13A.1.

Appendix XIII.2

Communications from the CBDT and CBEC on tax forecasting methods adopted by them

a) TPL, CBDT

No.134/03/2015 TPL
Government of India
Ministry of Finance
Department of Revenue
Central Board of Direct Taxes
(TPL Division)

New Delhi, January 2015

Office Memorandum

The Budget Estimates (BE) for revenue collections for any financial year are basically made in the following manner:-

There are two factors which form the basis for estimation of BE. First is the revenue collection during the past three years. Second is the corresponding GDP growth during these three years. Further, the GDP growth for the relevant year, i.e., the year for which the Budget Estimates are being made is also considered. For the purpose revenue estimation, the rate of growth of direct tax, separately for Corporate Income Tax (CIT) and Personal Income Tax (PIT) is taken for the past three years. Buoyancy is worked out for each of these three years separately for CIT and PIT, which is defined to be:

$$\text{Buoyancy} = \text{Tax growth rate} / \text{GDP growth rate}$$

2. The forecast of GDP growth rate for the relevant year for which the estimates are being made is provided by the Department of Economic Affairs (DEA). The average buoyancy for PIT and CIT is multiplied by GDP growth rate to arrive at the tax growth rate for the purpose of projecting revenue estimates. This is basically done for gross collections of tax. The refund growth rate is also factored in and net revenue estimates is (*sic*) worked out separately for Corporate Income Tax and Personal Income Tax. In addition to this, the revenue raising measures or the giveaways in the Finance Bill are reduced or added to the estimates arrived from average buoyancy figures. An estimate based on the previous year's collections, which are major head and minor head wise, is also taken into account before arriving at the final estimate. The estimates made by

the TPL division are subject to the final approvals of the same, taken by Budget division of the DEA.

3. Therefore, the above method for revenue estimation is based on historical data and uses the GDP growth rate of the year in reference to arrive at the Budget Estimates. Over the past five years, the method has been tested against actual collection and in all years, the difference between actual collection and revised estimates has not been more than *three per cent.*

4. Further, with regard to manpower available for revenue forecasting in this division, it is submitted that presently two joint secretaries and one director are principally involved in the preparation of budget estimates. A larger team for the TPL division would be beneficial in better analysis (*sic*) of the revenue impact of the legislative provisions. It would also facilitate a better comparative study of global best practices with reference to revenue forecasting in this regard.

This issues (*sic*) with the approval of the Chairperson CBDT

(Raman Chopra)
Director (TPL-II)

Secretary,
Tax Administrative Reforms Commission,
Ministry of Finance, Govt of India

a) TRU, CBEC

F.No.333/13/2014-TRU
Government of India
Ministry of Finance
Department of Revenue
(Tax Research Unit)
***

146, North Block, New Delhi,
Dated: 7th January, 2015.

Office Memorandum

Subject: Input for 4th report of TARC-Regarding.

The undersigned is directed to refer to D.O letter No. TARC/Meeting/1/2013-14 dated 15th December, 2014 on the above subject and to enclose herewith the requisite materials/inputs for further action please.

Encl: As above

This issues (*sic*) with the approval of Chairman (CBEC)

(N. K. Vidyarthi)
Deputy Secretary
Tel.-23093361

To,
Sh. Sanjay Kumar
Secretary, TARC
Govt. of India, Ministry of Finance
Tax Administrative Reform Commission (TARC)
NBCC Plaza, 3rd Floor, Pushp Vihar
Saket, New Delhi-110017

Copy to:
Sh. Surendra Singh, US (CX-9), CBEC with compliance to OM NO. 296/253/2014-CX-9, dated 17th December, 2014(point no. 1).

Brief note on forecasting, analysing and monitoring of revenue targets

❖ Indirect tax revenue estimation/forecasting

At present, the work relating to indirect tax revenue estimation/forecasting at aggregate level is carried out by the Tax Research Unit (TRU) for preparation of the budget estimates (BE) for indirect taxes. This includes preparation of mid-term revised estimates (RE) of indirect taxes the need for which arises on account of mid-year policy level changes or changes in the assumption criterion/changes in international trade pattern and economic criterion, etc.

Before enumerating the exact methodology used for indirect tax revenue estimation it is necessary to understand the complexities involved in such estimation in the prevalent tax regime in India and why the standard text book economics and statistical methodology, based on pure economic criterion and a set of standardized assumption parameters, cannot solely be used for such estimation. Even though tax estimates based on buoyancy factor may turn out to be more realistic in the case of direct taxes than indirect taxes (*sic*). This is because direct taxes are progressive in nature in comparison to indirect taxes.

Indirect tax revenue at central level accrues from three taxes, namely, Customs, Central Excise and Service Tax, each of which is influenced by different factors which are described below:

a) **Customs revenue** in a particular year is a function of import and export volumes and the policy changes made in the fiscal policy. However, prediction of import and export volumes at the estimation stage is highly complex as these are a function of international prices of imported goods, monetary exchange rates, world economic and political scenario, FTAs entered into by various economic groups during the course of the year, trade restrictions imposed by importing/exporting countries, country specific tariff barriers imposed in terms of antidumping and safeguard duties and mid-year policy changes that may be implemented to address various concerns such as that of current account deficit. In recent times there have been huge volatility in the international prices of key commodities and also in exchange rates.

b) **Central Excise** duty is levied on manufacture. As of now about 60% of the total Central Excise revenue comes from specific rated commodities (like petroleum & products, tobacco product, sugar etc). Rest of the revenue comes from ad-valorem rates. A tax on manufacture cannot have a predictable co-relation to GDP numbers and tax buoyancy in the previous few years for the following reasons:

   - Manufacture does not have a predictable co-relation with growth in GDP. With increased FTA, there is likelihood of import substituting for manufacture. This is unlike consumption of goods and services, where there is predictability.

- Excise duty often used (*sic*) as a fiscal tool used for giving stimulus packages for ailing industries and the estimation of such fiscal stimulus is difficult to predict with any kind of accuracy.
- While a majority of excise revenue is a function of quantity of production, GDP numbers are in value terms. Therefore, revenue from these commodities cannot be based on GDP estimation. Further, production numbers of specific rated commodities cannot be predicted for the reason that this may vary for various reasons like ban imposed by the states on tobacco products or changes in consumption pattern.
- While taxes imposed on income side have a predictable co-relation with the GDP, inflation and other economic indicators, the taxes on expenditure side are dependent on the consumption pattern, which in itself is a function of price elasticity of a commodity, the saving rate, etc.
- Index of industrial production is estimated at the post production stages. This number cannot be estimated in advance and hence of not much help (*sic*). Further, it is based on value, whereas a major part of excise revenue is dependent on quantity.
- A large number of exemptions ensure selective taxation of manufacturing sector. However, such de-segregated sectoral data of manufacturing sector is not generally available.
- It is difficult to estimate the revenue augmentation by increased tax efficiency even though the increased level of automation, streamlining of business process and better tax administration contribute to the increase in revenue.

c) **Service Tax** is a tax on consumption of services. Services constitute major portion of the GDP and consumption of services has a co-relation with GDP. Therefore fair estimation of service tax revenue is feasible by economic forecasting technique based on tax buoyancy/tax elasticity method with certain underlying assumptions. However, this estimation could be much more accurate if service sector is comprehensively taxed. Even with the introduction of the Negative List based tax regime a significant portion of the service sector (illustratively, a significant portion of public services/government services, health, education, public transport by road and rail, non-commercial construction, sports, entertainment and cultural services, financial services, agricultural sector related services and transport of goods) is outside the tax net. Therefore, as the annual budget exercise entails several policy level changes, including review of exemptions, rates, changes in rules and regulations aimed at achieving the objective such as tax neutrality and other policy changes directed towards increasing the tax efficiency by increased compliance and effective tax administration, a high level of accuracy of estimation based on the broad numbers may not be feasible

The indirect tax revenue forecasting/estimation is done within the above stated constraints using the top down model. Broadly, the steps involved are,-

(i) The sectoral analysis of revenue estimates taking into account the past trends, growth prospect, the impact of policy level changes, sector analysis available in credible publication and the information provided by the respective ministries e.g., the growth projection in petroleum sector is provided by the Petroleum Planning and Analysis Cell (PPAC), Ministry of Petroleum and Natural Gas.

(ii) The import/export volume growth is estimated based on trends.

(iii) Feedback and revenue projection is taken into accounts from all zonal offices (CCs) in the 4th quarter.

(iv) Sectoral analysis of top 10 commodities/sectors w.r.t Customs, central excise and service tax.

(v) Based on the above inputs the numbers are consolidated and validated with nominal/Real GDP growth as reported by CSO, Ministry of Statistics, taking into accounts *(sic)* the likely outgo in refunds/Drawback, etc.

(vi) The estimates are finalized after due deliberation with the Budget and Economic Affairs Division taking into account the overall economic and fiscal scenario.

Basis for indirect tax revenue estimate for BE

- Advance estimate of nominal GDP growth provided by Economic Division of DEA.
- Change in tax policy, ARM, stimulus package/exemptions relief (if any), etc. during the Union Budget.
- Sectoral analysis by TRU.
- Feedback and projection from the field formations.
- Inputs from concerned Ministries, e.g. POL sector growth as provided by PPAC, Ministry of Petroleum, industry associations and sectoral analysis available in credible publications.
- Import and export sector analysis based on trends.
- Estimation of refund and duty drawback outgo.
- Inputs from Budget and Economic Affair Division.
- Tax buoyancy factor (multiplier) usually provided by budget division after assessing the overall revenue potential and fiscal target constraints.

Analysing and monitoring of revenue targets:

- After finalization of All India indirect tax head-wise targets, it is further allocated to the respective formations/zones of Customs, Central Excise and Service tax across the Country.

- Moving targets, based on average of previous few years *(sic)* month-wise revenue collection trends, is communicated to all zonal offices at the beginning of the commencement of the financial year for monitoring the month wise revenue collections vis-à-vis moving targets.
- At an all India level indirect tax revenue is monitored by TRU with regard to economic parameters viz., Quarterly release of GDP and monthly release of IIP, Imports, WPI inflation, etc.
- All zonal heads do a detailed analysis of revenue viz. top assessees wise, sector-wise, additional revenue measures, etc. and report to the Board on a monthly basis. Further, conferences including video conferencing and meetings are held regularly at the level of the Board and above with field offices to ascertain the reasons for variations in revenue trends with regard to all India growth, on a regular basis.
- The estimates of revenue are done regularly throughout the year. However, in the process of Budget Formulation, this is disclosed only once annually as BE/RE.
- The Directorate of Data Management (DODM) and DG (System) are entrusted with the work to monitor and maintain the on line data warehouse (EDW/ACES/ICEGATE) on various technical and revenue matters that are being utilized to fulfil (*sic*) the need of Department/Ministry for analytical reports.

Gap (Resources and Skill Set)

Staff: At present, TRU has two Joint Secretaries, five Director/DS level officers and 13 Under Secretaries/Technical officers. Out of these, the statistical section consists of one Deputy Secretary and three officers from the Indian Statistical Services (ISS). The statistical section of TRU mainly looks after the work relating to forecasting, analysing and monitoring of revenue targets. Sectoral analysis and impact of policy changes is dealt with by respective commodity officers (Under Secretary/Technical Officers). While officers from the ISS provide technical inputs on the methods used for revenue forecasting, this cannot be delinked from the primary function of this unit and the requirement of additional staff would have to be considered in a holistic manner.

In order to have statistical software package (viz. SAS, EViews, STATA, SPSS etc) based data analysis including data mining and forecasting (based on time-series, cross-section, and panel data) at large scale (*sic*), services of professionals like economist, econometrician, may be considered separately.

Appendix XIII.3

Sources of information in the CBEC and CBDT

A large volume of information relating to economic transactions is collected by the tax administration in the course of enforcement. This is separately collected by the CBEC and CBDT. As the two Departments increasingly move to a non-intrusive method of enforcement, there is considerable scope for duplication. This will impose unwarranted compliance cost on entities liable to provide the requisite information.

i) Collection of information by CBEC

The CBEC has access to a large volume of data as there is an obligation on tax payers to furnish a number of returns (monthly, quarterly and annual). It starts with the application for registration. The application for registration needs various data like PAN, constitution of the business, address, investment in land, plant and machinery, details of bank accounts, business transaction numbers obtained from other government agencies/departments like VAT registration, details of the directors and also the major excisable goods manufactured. This is just the beginning. Once one starts producing, there are several periodic returns. Some of the major ones are:

Description	**Short Title**
Monthly return for production and removal of goods and other relevant particulars and CENVAT Credit	**ER-1**

This has many details:

a) Sub-heading wise information of monthly production (both value and quantity) and duty payable
b) Receipt and clearance details of inter unit transfer of intermediate goods
c) Details of duty paid through cash and credit
d) Break up of credit from input goods, capital goods and services and goods obtained through dealers
e) Details of payment of arrears, interest and other dues

Monthly return for hundred per cent export-oriented undertakings in respect of goods manufactured, goods cleared and receipt of inputs and capital goods. **ER-2**

This has details like

a) Sub-heading wise information of monthly production (both value and quantity) and duty payable
b) Details of clearance for exports, deemed exports and DTA clearances and duty payable

c) Details of duty-free (domestic and imported) inputs received
d) Duty-free capital goods received
e) CENVAT availed and utilised

<u>Quarterly Return for Clearance of Goods and CENVAT Credit (For SSI Units)</u> **ER-3**

This is only for units availing of clearance-based exemption.

The return is similar to the information contained in ER 1.

<u>Annual Financial Information Statement/Form (Only for units paying duty of Rs.1 crore or more)</u> **ER-4**

This is a detailed return requiring extremely comprehensive details like

a) Expenditure details – value of inputs going into dutiable and exempt goods
b) Value of raw materials consumed as per the profit and loss account
c) Value and quantity of each major raw material consumed in manufacture of goods
d) Details of other expenditure like freight, advertising, R & D expenditure, sales promotion, commission paid, wages, power and fuel
e) Details of goods manufactured through job workers
f) Total sales value as per profit and loss account
g) Value and quantity of each major finished goods sold
h) Details of all trading activities
i) Sale value of exempted goods
j) Details of goods exported
k) Total VAT paid
l) Details of all other income like warranty charges, advertising charges recovered, forwarding charges, erection and commissioning charges, consultancy charges, CENVAT credit on inputs, capital goods and services

<u>CENVAT – Annual return of information relating to principal inputs</u> **ER-5**

This basically contains information relating to principal inputs and outputs, including their description and central excise headings and the input-output ratio for all products

<u>CENVAT – Monthly return of information relating to principal inputs</u> **ER-6**

This seeks details of receipt and consumption of principal inputs and finished excisable goods as well as waste and scrap arising during manufacture and cleared/destroyed.

Annual Installed Capacity Statement **ER-7**

This contains details like

a) Description of each class of goods manufactured and annual production capacity
b) Description and technical specification of main plant and machinery installed as well as the year of installation
c) Details of electricity connection like supplier company, electricity meter numbers, electricity consumer number
d) Sanctioned electricity load
e) Details of captive power plant

Quarterly return for goods charged to 1 per cent duty **ER-8**

This seeks heading-wise details of all production, clearance and duty excise.

Quarterly Return under Rule 7 of CENVAT Credit Rules, 2002 for Registered Dealers. **A**

This is for first and second stage dealers whose invoices can be used for taking CENVAT credit.

They have to give details of all duty paid goods received and how they have been distributed to various buyers.

Half-yearly service tax return u/s 70 of the Finance Act 1994 **ST 3**

a) This is for service providers only. This contains details like:
b) Value of taxable services
c) Exemption notification number, if availed
d) Value of taxable service, service tax payable and gross amount charged
e) Details of exempt services goods
f) Details of CENVAT credit availed and utilised.

Quarterly return to be filed by manufacturers in Uttarakhand/ Himachal Pradesh. **Form A**

These units are exempted. The return seeks data on their clearance and production data.

The source for Central Excise and Service Tax data is the ACES (Automation of Central Excise and Service Tax) system, which has the modules for filing returns online by an assessee. All modules of ACES (including the Audit module) and all modules are reportedly functional. Old customs returns data (from SERMON) has been migrated to some extent into ACES and is thus available. Service tax returns from the SERMON system have not been migrated and are not available. ACES has now been rolled out to all locations and the number of ST returns is expected to grow substantially as e-filing of returns has been made mandatory.

As against excise and service tax, there are no returns prescribed in customs. The entire data is culled from the bills of entry and the shipping bills. Since it is obtained electronically and processed, the data base is really impressive. It is possible to get data country wise, importer wise, as well as 8 digit customs-heading wise. To make the data base really useful, particularly for direct taxes, it will be useful if the bill of entry and the shipping bill also contains the PAN-based registration number. Most importers are anyway registered with CBEC, being manufacturers or service providers. Other commercial importers may be registered like central excise and if the registered number is mentioned in the bills of entry or shipping bills, it will help to correlate all the business activities of an importer.

ii) Collection of information by CBDT

The Income Tax Act, 1961, and Wealth Tax Act, 1957, obligate tax payers to furnish returns of their income activities in different formats as prescribed. These returns need to be filed annually or quarterly. This provides a large amount of data relating to economic transactions for use by the Income tax Department for designing its enforcement strategy. The data is partly linked to the PAN of the taxpayer. The information in these returns include all financial transaction details, wealth statement (if the taxpayer falls within the wealth tax bracket), tax audit report (if liable to audit), house property details (if there is income from house property), and any transaction of a capital nature.

1. Income Tax Return Classification

There are eight types of Income Tax returns for various kinds of assessees. The categories are given below.

a) ITR 1 – This income tax return is also called SAHAJ. This return is only for individuals having income from salary and interest. This ITR contains the particulars of salary and interest details of the individuals. This ITR also contains the details of TDS and taxes paid by the individual assessee.

b) ITR 2 - This return is filed by either individuals or Hindu Undivided Family (HUF) who do not have income from business or profession. The ITR contains details of salary, income from house property, details of capital gains and income from other sources. It also has details of various taxes paid by the assessee as well as TDS details.

c) ITR 3 - This ITR is for individuals/HUFs being partner in firms but not carrying out business or profession under any proprietorship. This ITR contains the details of salary, income from house property, profits and gains from business or profession and income from other sources. It also has details of losses, if any, incurred either in the current year or in an earlier year, which has been brought forward to the current year. It also has an asset-liability schedule where total income exceeds Rs.25 lakh.

d) ITR 4 – This ITR is for individuals or HUFs having income from proprietary business or profession. It has details of the business performed by the assessee along with details of alternate minimum tax and unabsorbed depreciation.

e) ITR 4S – This ITR is also known as SUGAM. It is for individuals/HUFs having income from presumptive business. This ITR gives details of profits and gains of business of plying, hiring, or leasing goods carriages under Section 44AE of the I-T Act.

f) ITR 5 – This ITR is for firms/Association of persons/body of individuals and limited liability partners. It contains details of income from house property, profit and gains from business and profession, capital gains and income from other sources. It also has a tax relief summary for taxes paid outside India and details of foreign assets.

g) ITR 6 – This ITR is for companies other than companies claiming exemption u/s 11 of the IT Act. It contains details of the nature of business, balance sheet and P&L a/c. It has details of minimum alternate tax, dividend distribution tax, details of income outside India along with a tax relief summary for taxes paid outside India and details of foreign assets.

h) ITR 7 – This ITR is for persons, including companies, who are required to furnish return under section 139 (4A) or section 139 (4B) or section 139 (4C) or section 139 (4D). It includes income derived from property held under trust or other legal obligations wholly for charitable or religious purpose, details of income of a political party, income of a research association, news agency, association or institution, funds and other association and income of a university, college or institution.

2. TDS/TCS Returns

TDS/TCS is one of the modes of collection of taxes, in which a person making specific kinds of payments to another deducts a certain amount, which is then remitted to the government account. The concept of TDS/TCS envisages the principle of pay as you earn. It facilitates the sharing of responsibility of tax collection between the deductor and the tax administration and ensures a regular inflow of cash resources to the government. It acts as a powerful instrument to prevent tax evasion and to expand the tax net.

The TDS/TCS return statement is a summary of all transactions made during a quarter, filed in form 26Q (non-salaried), 24Q (salaried), 27Q (non-resident) and 27EQ (TCS) as described under the Income Tax Act.

The form contains the following details.

i) TAN of the deductor

ii) PAN of the deductee

iii) Amount of tax deducted

iv) Amount which has been paid

v) Under which section tax is deducted

vi) Period for which tax is deducted

vii) Date when paid to government's account

There are penal rates of TDS/TCS for not providing PAN during a transaction. It ensures that there is no revenue loss to the government. The payment received by the department is mapped to the Online Tax Accounting System (OLTAS) for reconciliation and to give the tax credit to the deductees. There are provisions in the Act for lower or nil deduction of tax at source. This is done by filing Form 15 H/15G (applicable for senior citizens).

3. Online Tax Accounting System

OLTAS is a system through which tax payments deposited by taxpayers in banks has been linked to the AST modules of the Income Tax Department. Uploading tax deposit details from a bank to the TIN central system database having tax payment information facilitates verification of tax deposits as claimed in TDS statements and income tax returns. It also provides taxpayers the facility to view their payment information online. National Securities Depositories Limited (NSDL) is the nodal agency collecting and maintaining this database and providing access to the department and taxpayers. It is also to be mentioned that it is mandatory for companies and audited individuals to deposit their tax only through online payments.

4. Data collection through form 15G/15H

According to section 197A (1a) of the Income tax Act, 1961, a declaration has to be made by an individual (not being a company or firm) claiming certain receipts without deduction of tax in the form 15G. Besides, section 197A (1c) of the I-T Act, 1961, stipulates that a declaration has to be made by an individual who is of the age of sixty years or above claiming certain receipts without deduction of tax in form 15H. These forms are to be submitted by these individuals to the respective deductors, who deposit them with the jurisdictional CIT charge. These forms contain the following information:

i) PAN of the declarant

ii) Jurisdictional AO

iii) Details of Investment

iv) Estimated Income

5. Wealth Tax Return

Wealth Tax is applicable on individuals, HUFs and companies whose net wealth exceeds Rs.30 lakh. Wealth tax is charged @ 1 per cent of net wealth exceeding Rs.30 lakh. Net wealth means assets minus debt incurred for such asset. From assessment year 2014-2015 onwards, a company or an assessee being an individual or HUF liable to audit u/s 44AB is required to furnish Form BB (Return of Net Wealth) electronically under a digital signature. Other assesses can file their wealth tax return manually. Wealth tax return form BB (e-filing) is applicable from the assessment year 2014-15; for earlier years, Form BA will continue. The following information is collected in wealth tax returns.

i) Motor Cars (other than those used by the assessee in the business of running them on hire or used by the assessee as stock-in-trade)

ii) Yachts, boats and aircrafts (other than those used by the assessee for commercial purposes)

iii) Jewellery, bullion, furniture, utensils or any other article made wholly or partly of gold, silver, platinum or any other precious metal. (other than those used by the assessee as stock-in-trade)

iv) Any building or land (with some exceptions as given below):

   a) One residential home is exempt from wealth tax on urban land measuring 500 square metres or less

   b) Any residential property, which has been let out for a minimum period of 300 days in the previous year

   c) Any house occupied by the assessee for the purpose of any business or profession carried on by him

   d) Commercial establishments or complexes

v) Cash in excess of Rs.50000 in the case of individuals or HUFs

vi) Deemed assets, i.e., assets transferred without consideration to family, etc.

vii) Assets of minor child barring some exceptions

viii) Value of assets in a partnership firm to be clubbed with the assets of partner

6. Annual Information Return (AIR) Information

Local authorities, registrar of companies, registrar for properties registration, stock exchange, Reserve Bank of India (RBI) and various depositories are obligated under Section 285BA of the I-T Act, 1961, to furnish AIR to the I-T Department on specified financial transactions. Under this legal obligation, these agencies are required to submit information in respect of specified financial transactions within a stipulated time, in a manner prescribed by Rules 114B to 114D of the I-T Rules. The CBDT has notified the NSDL as the prescribed agency to receive AIR from the specified persons. At present, seven categories of persons are required to compulsorily file AIR, namely

i) Banks accepting cash deposit of Rs.10 lakh or more in a year from any person

ii) Bank or company issuing credit cards where payment against bills exceeds Rs.2 lakh in a year for any person

iii) Mutual funds collecting Rs.2 lakh or more for sale of units by any person

iv) A company receiving Rs.5 lakh or more against issue of shares

v) a company receiving Rs.5 lakh or more against issue of bonds/debentures, registrar/sub-registrars in respect of sale/purchase of immovable property exceeding Rs.30 lakh

vi) RBI for issue of bonds exceeding Rs.5 lakh

The filing of the AIR is monitored through the ITD system.

7. Central Information Branch (CIB)

Third party information from internal as well as external sources, in addition to the AIR, is collected through the CIB by the I-T Department. CIB collects information from 40 internal and external source codes of which 12 are compulsory. This information is on financial transactions like investment, expenses, payment of taxes, etc., and details of persons involved in these transactions. CBDT has made PAN quotation compulsory under Rule 114B of the I-T Rules for these transactions. CIB has both PAN as well as non-PAN data from its sources and PAN is populated in these transaction through the ITD to create the individual transaction statement of the assessee. The following are the steps in uploading CIB information in the ITD module.

a) In compulsory codes, the filing agency information (Part A) is first uploaded through an excel sheet

b) Transaction information (Part B) is uploaded through an excel sheet. The Part B file is uploaded through a night scheduler.

AIR and CIB are used to create an individual transaction statement (ITS), which also includes transactions coming through AIR, TDS returns, CIB and tax payment through *challan* (OLTAS), ITR, FIU and OECD. The individual transaction statement (ITS) report gives a comprehensive financial profile of the tax payer to the AO.

These are the primary sources of information to the Income Tax Department. Other information that is specific in nature is regularly accessed by the investigation wing through manual intelligence. All these sources provide a 360-degree profile of an assessee to the department, which is used to combat tax evasion.

Appendix XIII.4

Error in estimation

Revenue forecasts provide crucial information to support government decisions on sustainable policy settings and new policy initiatives. Revenues are inherently more variable and less within the government's control. Even relatively small percentage errors in the revenue forecasts can have a large impact on budget surpluses or deficits. Accordingly, forecasting methods and processes need to be accompanied with estimates of errors or variability.[2] Some of these errors can be different from forecasting errors. Hence, the errors can be categorised on the basis of technical issues, such as data accuracy, forecasting methodology, process and agency structures, and other issues such as the effects of fiscal objectives and the economic cycle. Hence, forecasting agencies generally review and improve data and models on an ongoing basis. These errors may not mean that the forecasts are effectively 'wrong' most of the time; the error should be statistical.

There are several statistical methods available to evaluate forecast performance. Table 13A.1 lists the commonly used methods/measure to estimate error.

Table 13A.1: Statistical techniques for error measurement

Technique	Measures
Mean Squared Error	The average of squared errors over the sample period
Mean Error	The average dollar amount or percentage points by which forecasts differ from outcomes
Mean Percentage Error	The average of percentage errors by which forecasts differ from outcomes
Mean Absolute Error	The average of absolute dollar amount or percentage points by which a forecast differs from an outcome
Mean Absolute Percentage Error	The average of absolute percentage[3] by which forecasts differ from outcomes

All these measures are subject to interpretation. For example, a simple value of mean or mean squared error provides useful information for a particular variable; mean percentage error means the relative errors can be compared across a number of variables. Ignoring the sign of the error

[2] Some of these errors may be intentionally biased – more likely to be off in one direction than another – because forecasters believe there is an "asymmetric risk" and that the most accurate estimate is not always the best estimate. This type of error has not been discussed in this Appendix.

[3] Absolute vale is the value being taken as only positive, even the negative value.

term by adopting absolute changes, one can get an idea of the magnitude of the errors generated by the forecasting technique. While these measures provide useful information on the errors in forecasts, they are not exhaustive. Of them, the mean squared error is the most widely used measure to estimate error due to its statistical properties and simplicity.

Mean squared error

The mean squared error (MSE) of an estimator measures the average of the squares of the "errors", that is, the difference between the estimator and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. The difference occurs because of randomness or because the estimator does not account for information that could produce a more accurate estimate. For the unbiased estimator, the MSE is the variance of the estimator. The MSE thus assesses the quality of an estimator or set of predictions in terms of its variation and degree of bias.

An MSE of zero, meaning that the estimator predicts observations of the parameter with perfect accuracy, is the ideal, but is practically never possible. The values of the MSE may be used for comparative purposes. Two or more statistical models may be compared using their MSEs to measure how well they explain a given set of observations. An unbiased estimator with the smallest variance among all unbiased estimators is the best prediction in the sense that it minimises the variance and is called the best unbiased estimator or MVUE (Minimum Variance Unbiased Estimator).

Root-mean-square error

The root-mean-square error (RMSE) is a frequently used measure of the differences between the value predicted by a model or an estimator and the values actually observed. Basically, the RMSE represents the sample standard deviation of the differences between predicted values and observed values. These individual differences are called residuals when calculations are performed over the data sample that was used for estimation, and are called prediction errors when computed out-of-sample. The RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure of predictive power. RMSE is a good measure of accuracy, but only to compare forecasting errors of different models for a particular variable and not between variables.

Mean percentage error

Mean percentage error (MPE) is the average of the percentage differences between budget and actual outcomes. Mean percentage error distorts the forecasting performance in the presence of negative numbers (when errors are not random). The MPE is the computed average of percentage errors by which the forecasts of a model differ from actual values of the quantity being forecast. The MPE is calculated as:

$$MPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{f_t - a_t}{a_t}$$

where, a_t is the actual value of the quantity being forecast, f_t is the forecast value, and n is the number of different times for which the variable is forecast. Because actual rather than absolute values of the forecast errors are used in the formula, positive and negative forecast errors can offset each other; as a result, the formula can be used as a measure of the bias in the forecasts. A disadvantage of this measure is that it is undefined whenever a single actual value is zero.

Mean absolute error

The mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. The MAE is a linear score, which means that all individual differences are weighted equally in the average.

Mean absolute percentage error

The mean absolute percentage error (MAPE) is a measure of the accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation and it usually expresses accuracy as a percentage. The mean absolute percentage error (MAPE) is the mean or average of the absolute percentage errors of forecasts. Error is defined as actual or observed value minus the forecast value. Percentage errors are summed without regard to sign to compute MAPE. This measure is easy to understand because it provides the error in terms of percentages. Further, because absolute percentage errors are used, the problem of positive and negative errors cancelling each other out is avoided. A smaller MAPE implies a better forecast.

Learning from forecast and target errors

Forecasting and targeting are both dynamic exercises. Forecasts are bound to differ in reality from the corresponding actuals and targets are bound to differ both from the forecasts and the actuals. If the forecasts are based largely on rules of thumb or continuation of linear trends (trend growth rate, pas period buoyancy), forecasts and targets derived from such forecasts are bound to fail. Such failures are likely to be typically large in periods where there are cyclical movements, particularly closer to cyclical troughs and peaks.

These exercises can be improved by understanding and identifying all causes of systematic errors. These errors may be due to biases and mis-prediction of variance in the relevant tax bases. The latter often arises when cyclical changes and structural breaks occur, which may have roots in the external as well as the domestic economy. We need institutional mechanisms to regularly analyse the extent of such errors and any lessons learnt should be utilised in undertaking forecasting and targeting exercises in the forecasts and targets of the next period.

The main summary statistics that we have used for this purpose are (a) mean square error and (b) Theil inequality statistics (TIL). We also use a decomposition of the mean square error to examine whether any systematic errors are identifiable.

The summary statistics and the decomposition of the mean square error are given below. The mean square error [Mp] and root mean square error (RMSQ) are defined below.

$$Mp = [1/n\ \Sigma\ (P_t - A_t)^2] \text{ and } RMSQ = \sqrt{Mp}$$

These have a minimum value of zero in the case of perfect forecasts. There is no upper limit. Their inadequacy lies in not having a proper unit of measurement. They give the same weight to a deviation whether a variable is measured in rupees or billion rupees or percentages. They however, have interesting mathematical and statistical properties and lend themselves to useful decompositions.

The Theil inequality coefficient with respect to change in variables is defined as follows:

$$TIL = [\Sigma\ (\Delta P_t - \Delta A_t)^2]^{1/2} / (\Sigma\ \Delta A_t^2)^{1/2}$$

The mean square error M_P can be decomposed as follows:

$$M_P = (\mu_P - \mu_A)^2 + (S_p - r\ S_A)^2 + (1 - r^2) S^2_A$$

where μ_P and μ_A are the sample means of predictions and realisations, S_P and S_A are their standard deviations and r is the correlation coefficient between them. The division of the terms on the right-hand side by the mean square error gives rise to the following quantities which are called 'inequality proportions':

$U^M = (P - A) / M_P$ mean proportion

$U^R = (S_P - r\ S_A)^2 / M_P$ slope proportion

$U^D = (1-r^2)\ S_A / M_P$ disturbance proportion

The terms thus provide information on the relative importance of one source of error vis-à-vis another. The mean proportion has a positive value if $\mu_P \neq \mu_A$. This, therefore, is due to 'bias'. The derivation of S_P for r S_A is due to a slope error, and the third term is a disturbance component.

In forecasting terminology, the cost of errors should eventually be estimated through a 'loss function'. The loss function should ideally be asymmetric, distinguishing between costs that arise when there is a shortfall relative to the forecast or target and when there is an overachievement relative to the forecast or target.

ANNEXURES

Annexure - I

TARC meetings with stakeholders

Date	Name of the Stakeholder
22.12.2014	Meetings with officers of TPL, CBDT
23.12.2014	Meetings with officers of TRU, CBEC
23.12.2014	Meetings with DGIT (Inv.), Delhi, DGIT (Inv.), Lucknow and DGIT (I&CI) from CBDT
23.12.2014	Meetings with DGRI and DGCEI from CBEC
08.01.2015	Meeting with SAS

Annexure -II

Composition of Focus groups

Sl. No.	Topic	Focus Group
1.	To review the existing mechanism and recommend appropriate means including staff resources for forecasting, analysing and monitoring of revenue targets	Prof. D. K. Srivasatava, ex-Director, Madras School of Economics Shri Gautam Ray, ex-CCE Shri Arbind Modi, I-T Prof. N R. Bhanumurthi, NIPFP Shri Siddhartha Roy, Tata Sons
2.	To review the existing mechanism and recommend measures to enhance predictive analysis to detect and prevent tax/economic offences	Shri Ronmoy Das, I-T Shri L. Satya Srinivas, CCE Shri Satpal Gulati, I-T Shri K. Balaji Majumdar, CCE Dr. Sanjay Kagwade, Terradata Prof. Manoj K Srivastava, MDI
3.	To review the existing policy and recommend measures for research inputs to tax governance	Shri Arbind Modi, I-T Shri Parneet S. Sachdeva, I-T Shri J. Albert, I-T Shri. Nagendra Kumar, CCE Shri S.P. Sahu, CCE

Note: I-T: Income Tax Department

CCE: Custom & Central Excise Department

Annexure - III

TARC meetings

Date of the meetings
16th December, 2014
17th December, 2014
27th December, 2014
29th December, 2014
6th January, 2015
7th January, 2015
8th January, 2015

Annexure - IV

Gazette Notification constituting TARC

MINISTRY OF FINANCE
(Department of Revenue)
NOTIFICATION

New Delhi, the 21st August, 2013

F.No.A.50050/47/2013-Ad.I. –The Government in its Budget, 2013-14, had, inter-alia, announced the setting up of a Tax Administration Reform Commission (TARC) with a view to reviewing the application of Tax Policies and Tax Laws in the context of global best practices and recommend measures for reforms required in Tax Administration to enhance its effectiveness and efficiency. Accordingly, it has been decided to constitute the Tax Administration Reform Commission with the following composition:

i)	Dr. Parthasarathi Shome	Chairman
ii)	Shri Y. G. Parande	Full-time Members
iii)	Ms. Sunita Kaila	
iv)	Shri M. K. Zutshi	Part-time Members
v)	Shri S.S.N. Moorthy	
vi)	Shri M.R. Diwakar	
vii)	Shri S. Mahalingam	

2. The Commission will have a fixed tenure of 18 months from the date of its constitution and work as an advisory body to the Ministry of Finance. The Commission will give its first set of recommendations with six months of its constitution and thereafter submit periodic reports after every three months.

3. The Terms of Reference of the Commission will be as follows:-

a) To review the existing mechanism and recommend appropriate organizational structure for tax governance with special reference to deployment of workforce commensurate with functional requirements, capacity building, vigilance administration, responsibility of human resources, key performance indicators, assessment, grading and promotion systems, and structures to promote quality decision making at the highest policy levels.

b) To review the existing business processes of tax governance including the use of information and communication technology and recommend measures tax governance best suited to Indian context.

c) To review the existing mechanism of dispute resolution, covering time and compliance cost and recommend measures for strengthening the same. This includes domestic and international taxation.

d) To review the existing mechanism and recommend capacity building measures for preparing impact assessment statements on taxpayers compliance cost of new policy and administrative measures of the tax Departments.

e) To review the existing mechanism and recommend measures for deepening and widening of tax base and taxpayer base.

f) To review the existing mechanism and recommend a system to enforce better tax compliance – by size, segment and nature of taxes and taxpayers, that should cover methods to encourage voluntary tax compliance.

g) To review existing mechanism and recommend measures for improved taxpayer services and taxpayers education programme. This includes mechanism for grievance redressal, simplified and timely disbursal of duty drawback, export incentives, rectification procedures and refunds etc.

h) To review the existing mechanism and recommend measures for "Capacity building" in emerging areas of Customs administration relating to Border Control, National Security, International Data Exchange and securing of supply chains.

i) To review the existing mechanism and recommend measures for strengthening of Database and inter-agency information sharing, not only between Central Board of Direct Taxes (CBDT) and Central Board of Excise and Customs (CBEC) but also with the banking and financial sector, Central Economic Intelligence Bureau (CEIB), Financial Intelligence Unit (FIU), Enforcement Directorate etc. and use of tools for utilization of such information to ensure compliance.

j) To review the existing mechanism and recommend appropriate means including staff resources for forecasting, analyzing and monitoring of revenue targets.

k) To review the existing policy and recommend measures for research inputs to tax governance.

l) To review the existing mechanism and recommend measures to enhance predictive analysis to detect and prevent tax/economic offences.

m) Any other issue which the government may specify during the tenure of the Commission.

4. The Commission will be supported by a Secretariat consisting of a Secretary at the level of Joint Secretary to the Government of India and other officials and support staff. They will be appointed on deputation/contract basis.

5. The Commission will be provided information and quantitative data of Central Board of Direct Taxes/Central Board of Excise and Customs to enable it to do statistical analysis for making recommendations.

6. The Headquarters of the Commission will be in Delhi.

M. L. MEENA
Joint Secretary

Tax Administration Reform Commission
Government of India, Ministry of Finance
3rd Floor, NBCC Plaza, Pushp Vihar, Saket
New Delhi – 110017
Phone:- 011- 29565091, Fax:- 011- 29565092
Website: www.finmin.nic.in

Designed & Printed by India Offset Press, New Delhi • www.indiaoffsetpress.com